

Bayesian spatio-temporal modelling of the relationship between mortality and malaria transmission in rural western Kenya

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät der
Universität Basel

von
Nyaguara Ombek Amek
aus Asego-Kanyada, Kenya

Basel, 2013

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von
Prof. Dr. Marcel Tanner, PD Dr. Penelope Vounatsou und Prof. Dr. Willem Takken

Basel, den 21. Februar 2012

Prof. Dr. Martin Spiess

Dekan

Dedicated to my parents Raphael, Caren and Mary

and

my brother Patrick Onduru

Table of content

LIST OF FIGURES.....	I
LIST OF TABLES.....	III
SUMMARY	V
ZUSAMMENFASSUNG	IX
ACKNOWLEDGEMENTS	XV
CHAPTER 1: INTRODUCTION.....	1
1.1 MORTALITY AND MALARIA BURDEN.....	1
1.2 MALARIA DISEASE AND TRANSMISSION	2
1.2.1 <i>The malaria parasite</i>	2
1.2.2 <i>The Anopheles vectors of human malaria</i>	4
1.2.3 <i>Malaria transmission</i>	5
1.2.4 <i>Malaria control policies</i>	7
1.3 MALARIA RELATED MORTALITY AND TRANSMISSION	9
1.4 AN OVERVIEW OF HEALTH AND DEMOGRAPHIC SURVEILLANCE SYSTEMS (HDSS).....	10
1.5 MALARIA TRANSMISSION INTENSITY AND MORTALITY BURDEN ACROSS AFRICA INITIATIVE	11
1.6 CHARACTERISTICS OF THE MTIMBA DATA	12
1.6.1 <i>Geostatistical data</i>	12
1.6.2 <i>Spatial misaligned data</i>	13
1.6.3 <i>Seasonality and temporal data</i>	13
1.6.4 <i>Zero inflated entomological data</i>	14
1.7 OBJECTIVES OF THE THESIS	15
CHAPTER 2: SPATIO-TEMPORAL MODELLING OF SPARSE GEOSTATISTICAL MALARIA SPOROZOITE RATE DATA USING A ZERO INFLATED BINOMIAL MODEL.....	17
2.1 INTRODUCTION	19
2.2 THE MOTIVATING EXAMPLE/DATA	21
2.3 MODEL FORMULATION	23
2.3.1 <i>Geostatistical-temporal binomial and ZIB models</i>	24
2.3.2 <i>Model fit</i>	25
2.3.3 <i>Model validation</i>	25
2.4 APPLICATION.....	26
2.5 CONCLUSIONS	31
CHAPTER 3: SPATIAL AND TEMPORAL DYNAMICS OF MALARIA TRANSMISSION IN RURAL WESTERN KENYA.....	35
3.1 INTRODUCTION	37
3.2 METHODS	39
3.2.1 <i>Study site</i>	39
3.2.2 <i>Entomological data</i>	40
3.2.3 <i>Entomological inoculation rate (EIR)</i>	40
3.2.4 <i>Climatic and Environmental data</i>	41
3.2.5 <i>Statistical analysis</i>	41
3.3 RESULTS	43
3.3.1 <i>Abundance/density of vector species</i>	43
3.3.2 <i>Entomological inoculation rate</i>	47
3.4 DISCUSSION	50
3.5 APPENDIX.....	53
3.5.1 <i>Negative binomial and Zero-Inflated negative binomial models</i>	53
3.5.2 <i>Geostatistical zero inflated negative binomial model</i>	53

CHAPTER 4: USING HEALTH AND DEMOGRAPHIC SURVEILLANCE SYSTEM (HDSS) DATA TO ANALYSE SPATIO-TEMPORAL PATTERNS OF SOCIO-ECONOMIC STATUS; AN EXPERIENCE FROM KEMRI/CDC HDSS.....	57
4.1 INTRODUCTION	59
4.2 MATERIAL AND METHODS	61
4.2.1 <i>Study Site and Population</i>	61
4.2.2 <i>Data collection</i>	61
4.2.3 <i>Data management and analysis</i>	62
4.2.4 <i>Ethical consideration</i>	63
4.3 RESULTS	63
4.4 DISCUSSION	71
CHAPTER 5: INFANT AND CHILD MORTALITY IN RELATION TO MALARIA TRANSMISSION IN RURAL WESTERN KENYA	73
5.1 INTRODUCTION	75
5.2 METHODS	77
5.2.1 <i>Study area and population</i>	77
5.2.2 <i>Cause-specific mortality</i>	78
5.2.3 <i>Socioeconomic status</i>	78
5.2.4 <i>Entomological Inoculation rate (EIR) and ITN</i>	78
5.3 STATISTICAL ANALYSIS	79
5.4 RESULTS	80
5.4.1 <i>Descriptive statistics</i>	80
5.4.2 <i>Model-based results</i>	82
5.5 DISCUSSION AND CONCLUSION	85
5.6 APPENDIX	88
CHAPTER 6: MORTALITY IN RELATION TO MALARIA TRANSMISSION: A COMPARISON ACROSS AGE GROUPS IN RURAL WESTERN KENYA	91
6.1 INTRODUCTION	93
6.2 MATERIALS AND METHODS	95
6.2.1 <i>Study area and Population</i>	95
6.2.2 <i>Explanatory variables</i>	96
6.3 STATISTICAL ANALYSIS	97
6.3.1 <i>Excess mortality rate attributed to malaria exposure</i>	98
6.4 RESULTS	98
6.4.1 <i>Descriptive statistics</i>	98
6.4.2 <i>Model-based results</i>	100
6.5 DISCUSSION AND CONCLUSION	104
6.6 APPENDIX	107
CHAPTER 7: GENERAL DISCUSSION	109
7.1 STATISTICAL CONTRIBUTION	109
7.2 EPIDEMIOLOGICAL CONTRIBUTION	110
7.3 LIMITATIONS AND CHALLENGES	114
CHAPTER 8: EXTENSION AND FUTURE RESEARCH.....	117
CHAPTER 9: CONCLUSION	119
BIBLIOGRAPHY.....	121
CURRICULUM VITAE.....	135

List of Figures

Figure 1.1 Geographical distribution of malaria in 2008.....	2
Figure 1.2 Life cycle of malaria parasite	4
Figure 1.3 Countries with HDSS field sites and members of INDEPTH Network.....	11
Figure 2.1 Surveyed locations with infected and none-infected mosquitoes.....	27
Figure 2.2 Proportion of test locations with sporozoite rate falling in between 5% to 95%. Credible intervals of the posterior predictive distribution	28
Figure 2.3 Maps of predicted sporozoite rate during wet and dry seasons based on spatial zero inflated binomial model.	30
Figure 2.4 Map of prediction error of sporozoite rate during wet and dry seasons based on spatial zero inflated binomial model	30
Figure 3.1 Location of the KEMRI/CDC HDSS site.....	39
Figure 3.2 Monthly pattern of average number of Anopheles gambiae and funestus species.....	43
Figure 3.3 Monthly pattern of observed, fitted and predicted indoor residual densities of Anopheles gambiae mosquito	44
Figure 3.4 Proportion of test locations with none-zero mosquitoes falling in between 5% to 95% credible intervals of the posterior predictive distribution	45
Figure 3.5 Temporal pattern of observed and predicted entomological inoculation rate	48
Figure 3.6 Predicted EIR maps	50
Figure 4.1 Histogram of SES index obtained by Ordinary PCA	65
Figure 4.2 Histogram of SES index obtained by Polychoric PCA	66
Figure 4.3 Histogram of SES index obtained by MCA	66
Figure 4.4 Distribution of all household in each MCA quintile by year	69
Figure 4.5 Distribution of cohort household in each MCA quintile by year	70
Figure 4.6 Spatial and temporal distribution of SES quintiles.....	71
Figure 5.1 Main causes of death among infants	81
Figure 5.2 Main causes of death among 1-4years.....	82
Figure 5.3 Excess mortality for under-five age groups.....	85
Figure 6.1 Crude death rates per year	99
Figure 6.2 Main causes of death by age groups for older children and adults	100
Figure 6.3 Excess mortality for under-five age groups.....	103
Figure 6.4 Excess mortality for 5 years old above.....	103
Figure 6.5 Effect of malaria transmission in acquired immunity	106
Figure 7.1 Distribution of all-cause and malaria specific mortality and malaria transmission .	113

List of Tables

Table 2.1 Posterior estimates of geostatistical models	29
Table 3.1 Posterior estimates of zero inflated geostatistical density models.....	46
Table 3.2 Distribution of EIR by area in relation to wet and dry months during study period	47
Table 4.1 Variables included in and weights obtained from first component of different techniques	64
Table 4.2 Percentage of households owning assets by quintile of MCA-based SES index	68
Table 5.1 Childhood mortality rates per 1000 live births	81
Table 5.2 Hazard ratio (HR) estimates of predictors of all-cause and malaria specific mortality for under-five age categories from spatiotemporal models	83
Table 5.3 Hazard ratio (HR) estimates of predictors of all-cause and malaria specific mortality for under-five age categories spatiotemporal models (EIR and ITN).....	84
Table 6.1 Hazard ratio (HR) posterior estimates of all-cause mortality by age groups (EIR only)	101
Table 6.2 Hazard ratio (HR) posterior estimates of all-cause mortality by age groups (EIR and ITN).....	102

Summary

Sub-Saharan Africa (SSA) still bears the highest burden of the global mortality despite recent dramatic decreases. The majority of these deaths occur in children younger than 5 years and malaria infection is thought to be a leading cause of these deaths. Because of this belief, many studies have documented the effects of malaria transmission on childhood, but everyone living in malaria endemic areas is exposed to malaria parasites and is at risk of dying of malaria or malaria related causes. Besides the immediate threat to human survival, consequences of repeated clinical malaria infection places enormous economic and emotional impact on the households and systems.

Over a century, a number of malaria control strategies have been implemented to reduce or eradicate the malaria burden. However, some of these interventions were never successful in SSA due to weak health systems, political goodwill and anti-malarial drug resistance among other factors. A global health initiative to roll back malaria (RBM) was initiated in 1998 aiming to halve the malaria-related mortality by year 2010 and to eliminate the disease by 2030 through evidence-based malaria control approaches. However, monitoring of the progress and achieving the above objective requires (i) reliable all-cause and malaria specific mortality which is often lacking in most of this region, and (ii) precise knowledge on the nature of the relationship between mortality and transmission which remains unclear. INDEPTH, a network of health and demographic surveillance systems (HDSS), initiated the malaria transmission intensity and mortality burden across Africa (MTIMBA) project in the year 2002 with the aim to improve our understanding of this relationship in its malaria endemic member sites. The HDSS exist in various parts of the low and middle-income countries where routine vital registration systems are weak or nonexistent, and routinely monitor demographic and health events at household level in a geographically defined area. It also collects information on causes of death, entomological data in randomly selected houses (locations) among others.

The MTIMBA data are characterized by the presence of spatio-temporal correlation and the sparsity of the entomological data. Spatial correlation arises because locations in close proximity have similar risks due to common exposures. Sparse data occurs when large number of survey

Summary

locations has zero (no) mosquitoes or proportions of infected mosquitoes. Standard statistical models are not appropriate to analyze these data because they assume independence between locations, leading to incorrect parameter estimates. In addition, excess zeros introduce overdispersion. Ignoring the extra zeros result in poor fit. Geostatistical temporal models adjust for spatial and temporal correlation by introducing location and time specific random effects respectively. Zero-inflated analogues of these models assume that a proportion of zeros arise from a count distribution and the remaining ones are observed with probability one. Spatio-temporal models have large number of parameters. Bayesian methods can fit highly parameterized models by employing Markov chain Monte Carlo (MCMC) simulation algorithms, hence overcome the computational problems of the likelihood-based methods.

The objectives of this thesis was (i) to develop data driven Bayesian geostatistical models to assess the relationship between mortality and malaria transmission and (ii) apply these models to analyze the MTIMBA data extracted from KEMRI/CDC HDSS database with the aim to (a) estimate transmission heterogeneity and produce smooth maps of transmission intensity of the study area (b) assess the spatio-temporal changes and obtain smooth surfaces of socioeconomic status and (c) assess the relationship between malaria transmission and mortality across ages taking into account intervention efforts, socioeconomic status and demographic factors.

In chapter 2, Bayesian zero inflated binomial (ZIB) geostatistical models were developed and compared with standard binomial analogues to analyze sparse sporozoite rate (SR) data adjusting for environmental/climatic factors and seasonality. The models also included spatial and temporal correlation. Smooth maps of SR during wet and dry season were produced. The results showed that ZIB models fit the data better and estimate predictors with lower uncertainty compared to standard binomial models. The analysis also revealed spatial and seasonal heterogeneity in SR. SR was high during the wet season and in most parts of the northern and in a few locations in the southern part of the study area. Rainfall and altitude (distance above sea level) were the main drivers of SR in this area.

The method used to obtain high resolution entomological inoculation rate (EIR) surfaces is discussed in chapter 3. EIR is a product of sporozoite rate (binomial data) and mosquito density (count data). Therefore we developed Bayesian zero inflated negative (ZINB) geostatistical

Summary

models to analyze sparse mosquito density data. The ZINB models included predictors similar to the ZIB model in chapter 2. Model based predicted estimates of SR and mosquito densities were multiplied to obtain EIR estimates. High resolution (250 m by 250 m) temporal (monthly) EIR maps were produced for the study area. The results showed that distance to water bodies and vegetation are the main factors influencing the mosquito density in the study area. In addition, there was strong evidence of spatial and temporal patterns in mosquito density and EIR in the study area.

In chapter 4, we used the household assets and characteristics data routinely collected in the KEMRI/CDC HDSS to compare different methods used to calculate household socioeconomic index based on assets as a proxy to household socioeconomic status. We ranked households into quintiles using generated household index and assessed changes in household quintiles over time. The results reveal that multiple correspondence analysis (MCA) explains our data better than ordinary and polychoric principal component analysis. The gap between the poorest and the least poor households increased in the ratio of 1:6 at the end of the study period. Spatial analysis also showed a gradual increase in least poor households in the southern part of the study area as the year progresses.

High resolution EIR monthly estimates obtained in chapter 3 were linked to locations of mortality outcome in the study area. The relationship between malaria transmission intensity measured by EIR, all-cause and malaria specific mortality was assessed using Bayesian spatio-temporal geostatistical Cox proportion hazard models. The models included EIR estimates (with their uncertainty), age, household socioeconomic quintiles, ITN use and parameters describing space-time correlation. EIR was included in the model as an errors-in-variable covariate to take into account the prediction uncertainty. The study population was categorized into the following age groups neonates (0-28 days), post-neonates (1-11 months), child (1-4 years), 5-14, 15-29, 30-59 and ≥ 60 years. Analysis was carried out in each age group and results were discussed in chapters 5 and 6.

Summary

The results of these analyses suggest that the effect of malaria transmission on all-cause and malaria-specific mortality is age-dependent. Under-five year old children have the highest risk of dying from malaria or any disease with increase in transmission intensity. No trend is observed in older children and adults ≤ 59 years old. This re-enforces the need for malaria interventions to selectively target the affected age groups thus making control effective. However, the effects of malaria transmission on all-cause and malaria specific mortality in under-five age groups were similar when compared. This could be attributed to poor specificity of verbal autopsy in identifying malaria deaths in malaria endemic areas.

Higher transmission intensity appeared to have a protective effect to elderly population. These suggest gene selection and acquisition of immunity due to long exposure of malaria infection from childhood. Use of ITN has shown a reduction in all-cause mortality in almost all age groups except in child (1-4 year), but the effect is only strong in post-neonates and adults aged 30-59 years olds. Higher household socioeconomic status was also associated with lower all-cause mortality, but surprisingly was not associated with malaria specific mortality in the study area.

The results of this work improve our understanding of the relation between malaria transmissions, all-cause and malaria specific mortality in the KEMRI/CDC HDSS. Results from Rufiji DSS suggest also similar trends implying that our results may be used to generalize the transmission-pattern. However, we still need to compare our results with those from other HDSS sites before making a general conclusion. Similarly, these results are important in developing effective malaria control interventions. Another contribution of this work is the development of spatio-temporal models for sparse entomological data which can be used to fit other epidemiological datasets and the estimation of high resolution EIR surfaces.

Zusammenfassung

Subsahara-Afrika (SSA) leidet noch immer an der höchsten globalen Mortalität trotz der enormen jüngsten Rückgänge. Die meisten dieser Todesfälle treten bei Kindern unter 5 Jahren auf, wobei die Hauptursache dieser Todesfälle vermutlich Malaria-Infektionen sind. Aufgrund dieser Vermutung haben viele Studien die Auswirkungen von Malaria auf Kinder untersucht. Allerdings sind alle Menschen die in Malaria-endemischen Gebieten leben den Malaria-Parasiten ausgesetzt und laufen damit Gefahr an Malaria oder Malaria-bezogenen Ursachen zu sterben. Neben der unmittelbaren Bedrohung für das Überleben der Menschen haben wiederholte klinischen Malaria-Infektionen enorme wirtschaftliche und emotionale Konsequenzen auf die Haushalte und Systeme.

Seit mehr als ein Jahrhundert gibt es eine Reihe von Malaria-Kontroll-Strategien zur Verringerung oder Beseitigung der Malaria. Allerdings waren einige dieser Interventionen nie wirklich erfolgreich in SSA aufgrund der unzureichenden Gesundheitssysteme, politischem Willen und Arzneimittel Resistenzen neben anderen Faktoren. Eine globale Gesundheitsinitiative zur Verdrängung der Malaria (RBM) wurde im Jahr 1998 initiiert mit dem Ziel die Malaria-Mortalität bis 2010 zu halbieren und die Krankheit mittels evidenzbasierter Ansätze bis zum Jahr 2030 auszurotten. Die Überwachung der Fortschritte und der Erreichung der oben genannten Ziele erfordert (i) zuverlässige Daten über die Gesamt- und Malaria-bezogene Mortalität, welche oft in diesen Regionen fehlen, und (ii) genaue Kenntnisse über die Beziehung zwischen Übertragung und Mortalität, welche noch unklar sind. INDEPTH, ein Netzwerk gesundheitlicher und demographischer Überwachungssysteme (HDSS), initiierte im Jahr 2002 ein Projekt über die Malaria-Übertragungsintensität und -mortalität in Afrika (MTIMBA) mit dem Ziel unser Verständnis dieser Beziehung in endemischen Ortschaften die Teil des Projektes sind zu verbessern. HDSS existiert in verschiedenen Teilen der Länder mit niedrigem und mittlerem Einkommen, in denen Melderegister nicht richtig funktionieren oder nicht vorhanden sind. Dieses System überwacht routinemäßig demographische and gesundheitliche Ereignisse auf Haushaltsebene in einem räumlich begrenztem Gebiet. Zusätzlich sammelt es auch Informationen zu Todesursachen und entomologische Daten in zufällig ausgewählten Häusern.

Zusammenfassung

Die MTIMBA Daten sind durch räumlich-zeitliche Korrelationen und der Seltenheit der entomologischen Daten gekennzeichnet. Räumliche Korrelation entsteht, weil benachbarte Standorte ähnlichen Risiken aufgrund gemeinsamer Einflüsse aufweisen. Die Seltenheit der Daten entsteht, wenn an einer große Anzahl von Standorten keine Mücken bzw. keine infizierten Mücken gefunden wurden. Die standard-statistischen Modelle sind nicht geeignet um diese Daten zu analysieren, da sie von der Unabhängigkeit der einzelnen Standorte ausgeht, was zu falschen Parameterschätzungen führen kann. Darüber hinaus führt die große Anzahl an Nullen zu Überdispersion, was bei Nichtbeachtung zu einer schlechten Modellanpassung führt. Geostatistische zeitliche Modelle erfassen die räumliche und zeitliche Korrelation durch orts- und zeitspezifische zufällige Effekte. Die Moskitoanzahl wird durch eine diskrete Verteilung angenähert deren zusätzlicher Anteil an Nullen mit einer Wahrscheinlichkeit von eins angenommen wird. Räumlich-zeitliche Modelle haben große Anzahl von Parametern. Bayes'sche Methoden können diese hoch parametrisierten Modelle durch den Einsatz von Markov-Ketten-Monte-Carlo (MCMC) Simulationsalgorithmen ermitteln und überwinden damit die rechnerischen Probleme der Wahrscheinlichkeits-basierten Methoden.

Ziel dieser Arbeit war es (i) die Beziehung zwischen Mortalität und Übertragung von Malaria mittels Bayes'scher geostatistischer Modelle zu entwickeln und (ii) diese Modelle zu verwenden, um die MTIMBA Daten von KEMRI/CDC HDSS zu analysieren, und um (a) die Heterogenität der Übertragung zu ermitteln und Karten der Intensität im Untersuchungsgebietes zu erstellen, (b) räumlich-zeitliche Veränderungen zu ermitteln und den sozioökonomischen Status, (c) die Beziehung von Malaria-Übertragung und Mortalität zum Alter zu untersuchen unter Berücksichtigung von Interventionen, sozioökonomischen Status und demographischen Faktoren.

In Kapitel 2 werden so genannte Bayes'sche Zero-inflated (ZIB) geostatistische Modelle mit binomial-verteilten Analoga verglichen um die Sporozoiten-Rate (SR) unter Einbeziehung von Umwelt- und Klimafaktoren und Saisonalität zu analysieren. Diese Modelle beachtetten ebenso räumliche und zeitliche Korrelationen in den Daten. Mittels der Modelle wurden Karten der SR für die Regen- und Trockenzeit produziert. Die Ergebnisse zeigten, dass ZIB Modelle die Daten besser erklären und die Variablen mit geringerer Unsicherheit im Vergleich zu den binomial-

verteilten Modelle abgeschätzt werden konnten. Die Analyse fand darüber hinaus räumliche und saisonale Heterogenität in der SR. Besonders hoch war die SR während der Regenzeit und in den meisten nördlichen Teilen des Untersuchungsgebietes als auch einige wenige Ortschaften im Süden. Niederschlag und Höhenlage (Abstand zum Meeresspiegel) waren die Haupteinflussfaktoren der SR.

Die Methodik um hochauflösende Karten der entomologische Inokulationsrate (EIR) zu erhalten wird in Kapitel 3 besprochen. EIR ist das Produkt der SR (binomial-verteilt) und der Moskito-Dichte (Zählwerte). Wir entwickelten daher Bayes'sche Zero-inflated negativ-binomial-verteilte (ZINB) geostatistische Modelle um die Moskito-Daten zu analysieren. Die ZINB Modelle enthalten Prädiktoren ähnlich denen des ZIB-Modells aus Kapitel 2. Modellbasierte Schätzungen der SR und Moskito-Dichte wurden multipliziert, um EIR Schätzungen zu erhalten. Monatliche EIR-Karten mit einer räumlichen Auflösung von 250m mal 250m wurden für das Untersuchungsgebiet erstellt. Die Ergebnisse zeigten, dass die Entfernung zu Gewässern und die Vegetation die wichtigsten Einflussfaktoren auf die Moskito-Dichte im Untersuchungsgebiet sind. Darüber hinaus gab es starke Hinweise auf räumliche und zeitliche Muster in der Mückendichte und dem EIR.

In Kapitel 4 haben wir die routinemäßig erfassten Daten des KEMRI/CDC HDSS auf Haushaltsebene zu den Vermögenswerten und weiteren Charakteristika genutzt, um verschiedene Methoden verwendet, um den sozioökonomischen Index der Haushalte mittels verschiedener Methoden zu berechnen und vergleichen. Wir unterteilten den generierten Haushalt-Index in Quintile mit und erfassten die Veränderungen der Quintile über die Zeit. Die Ergebnisse zeigen, dass multiple Korrespondenzanalyse (MCA) unsere Daten besser als normale und polychoric Hauptkomponentenanalyse erklärt. Die Kluft zwischen den ärmsten und den am wenigsten verarmten Haushalten erhöhte sich auf das 6-fache bis zum Ende des Untersuchungszeitraumes. Die räumliche Analyse zeigte zudem eine schrittweise Erhöhung der Anzahl der am wenigsten verarmten Haushalte im südlichen Teil des Untersuchungsgebietes im weiteren Jahresverlauf.

Die hochauflösenden monatlichen EIR Schätzungen aus Kapitel 3 wurden mit der Sterblichkeit an verschiedenen Ortschaften des Untersuchungsgebietes verbunden. Die Beziehung zwischen Malaria-Übertragung, ermittelt durch den EIR, und der gesamten und Malaria-spezifischen

Mortalität wurde mittels Bayes'scher räumlich-zeitlicher geostatistischer Cox proportionaler Hazard-Modelle untersucht. Die Modelle enthielten EIR Schätzungen (mit Unsicherheit), Alter, haushaltsspezifische sozioökonomische Quintile, ITN Verwendung und Parameter zur Beschreibung der Raum-Zeit-Korrelation. Der EIR wurde im Modell als ein Fehler-in-Variable Einflussfaktor berücksichtigt um die Unsicherheiten in der Schätzung zu behalten. Die Population der Studie wurde in die folgenden Altersgruppen unterteilt: Neugeborene (0-28 Tage), post-Neugeborene (1-11 Monate), Kinder (1-4 Jahre), 5-14, 15-29, 30-59 und ≥ 60 Jahre. Die Analyse wurde für alle Altersklassen durchgeführt und die Ergebnisse werden in Kapitel 5 und 6 diskutiert.

Die Ergebnisse dieser Analysen lassen vermuten, dass die Wirkung der Malaria-Übertragung auf die Gesamt- und die Malaria-spezifische Mortalität altersabhängig ist. Kinder unter fünf Jahren haben das höchste Risiko an Malaria oder anderen Krankheiten zu sterben mit zunehmender Intensität der Übertragung. Dagegen ist kein Trend bei älteren Kindern und Erwachsenen unter 60 Jahren zu beobachten. Dies belegt erneut die Notwendigkeit von Malaria-Interventionen die selektiv auf die betroffenen Altersgruppen eingehen um eine wirksame Kontrolle zu erzielen. Allerdings waren die Auswirkungen der Malaria-Übertragung auf die Gesamt- und Malaria-spezifische Mortalität bei allen Kindern unter fünf Jahren ähnlich. Dies könnte auf eine geringe Spezifität der verbalen Autopsie bei der Identifizierung der Malaria-bedingten Todesfälle in endemischen Gebieten zurückzuführen sein.

Eine erhöhte Übertragungsintensität scheint einen schützenden Effekt auf die ältere Bevölkerung zu haben, was auf genetische Selektion und Erwerb von Immunität aufgrund der langjährigen Malaria-Gefährdung hindeutet. Das Verwenden von ITNs zeigt eine Verringerung der Gesamtmortalität in fast allen Altersgruppen, außer bei Kindern (1-4 Jahre), aber dieser Effekt ist lediglich bei post-Neugeborenen und Erwachsenen im Alter von 30 bis 59 Jahren signifikant. Ein besserer sozioökonomischer Status ist ebenfalls mit einer niedrigeren Gesamtmortalität assoziiert, aber überraschender Weise nicht mit der Malaria-spezifischen Mortalität in der untersuchten Region.

Zusammenfassung

Die Ergebnisse dieser Arbeit verbessern unser Verständnis der Beziehung zwischen Malaria-Übertragungen und der Gesamt- und malaria-spezifischen Mortalität im KEMRI/CDC HDSS. Die Ergebnisse aus dem Rufiji DSS zeigen ähnliche Trends auf und legen nahe, dass unsere Ergebnisse dazu verwendet werden können die Übertragungs-Muster zu verallgemeinern. Allerdings müssen wir noch unsere Ergebnisse mit denen von anderen HDSS Standorten vergleichen, bevor wir eine allgemeine Schlussfolgerung ziehen können. Ebenso wichtig sind die Ergebnisse bei der Entwicklung wirksamer Interventionen zur Malaria-Kontrolle. Ein weiterer Beitrag dieser Arbeit ist die Entwicklung von räumlich-zeitlichen Modellen für seltene entomologische Daten, die für andere epidemiologischen Datensätze verwendet werden können und zur Ermittlung hoch auflösender EIR Karten.

Acknowledgements

As I finalize writing this thesis, I wish to acknowledge a number of people who have contributed to its success either directly or indirectly. To begin, I'm very grateful to my supervisor PD. Dr. Penelope Vounatsou for the mentorship and with the patience for me as I went through the ropes of Bayesian modeling. Thanks for sharing your immense experience with me.

I extend special thanks to Prof. Dr. Tom Smith for considering me for the MTIMBA PhD program. I appreciate your constructive comments, edits in my work and willingness to respond to my questions even when I asked them in wrong places. I remember your kindness in welcoming me with a lunch and gift of a warm jacket to a freezing Basel - am thankful for that show of hospitality. Special thanks also go to Prof. Dr. Willem Takken, who was willing to and acted as the co-referee for this thesis.

I would like to thank the Swiss TPH family for the friendly environment and support during my stay in Basel. Many thanks go to Christine Mensch, Margrit Slaoui, Zsuzsanna Gyorffy, Maya Zwygart, Dagmar Batra, Christine Walliser, and Beatrice Wackerlin for administrative support accorded to me. I, also, appreciate the assistance I got from the IT and library sections. Thanks Mike Schur for rescuing my laptop. To Dr. Konstantina Boutsika, thanks for your generosity! It was nice enjoying Greek food at your place. Many thanks also go to Prof. Dr. Mitchell Weiss and Prof. Dr. Nino Kunzli for creating a good working environment at the EPH department. Last, but not least, I would like to thank Prof Dr. Marcel Tanner for a warm welcome to the institute and for always asking me “kila kitu iko sawa” whenever we bumped into each other.

Many thanks go to the Biostatistics unit. In particular I would like to thank Dr. Nadine Schur, Dr. Ronaldo Scholte, Dr. Laura Goşoniu, Dominic Goşoniu, Federica Giardina, Verena Jürgens, Frédérique Chammartin for your professional friendship and support. I do appreciate your kindness by allowing me to run my models that could even take more than two weeks, not to mention the “memory issue” in your computers. You did not turn me away even when you had deadlines to meet. I'm grateful for the “birthdays”. To Dr. Nadine Schur, many more thanks for translating the summary of this thesis to German. Special thanks go to the MTIMBA team: Simon Kasasa “the big man”, Eric Diboulo and “madada zangu” Dr. Susan Rumisha and

Acknowledgements

Bernadette Huho. The ideas, challenges and support shared across as we cracked the data were very supportive. I do appreciate it.

I enjoyed the friendship of many fellow students and other people I met in Basel. To mention a few, I would like to thank the Tanzanian colleagues. Dr. Henry Mwanyika, Dr. Boniphace Idindili, Angel Dillip Dr. Susan Rumisha, Felister Mwingira, Judith Kahama, Angelina Lutambi, Bernadette Huho, Dr. Mwifadhi Mrisho, Dr. Fatuma Manzi, Dr. Pax Masimba, Simon Vendeline, Dr. Rashid Khatibu and any whom I may have forgotten. It was kind of you for organizing the social gatherings to make us feel at home. I really enjoyed the discussions in these occasions. To Suleiman Jembe, I do appreciate your friendship and the many dinners you offered particularly in the last month of this work. Thanks for coming around almost every weekend just to encourage me. To Lydia Mwangi, I'm very grateful for the Christmas dinners at your place.

Back in Kenya, I'm deeply indebted to Kisumu KEMRI/CDC collaboration under the leadership of Dr. Kayla F. Laserson, the current HDSS chief Dr. Frank Odhiambo and the former "my senior brother" the late Dr. Adazu Kubaje. I don't know the right words to use to express my gratitude to all of you for all the support you accorded me throughout this work. To my "senior brother" I know you would be very happy and laughing wherever you are as I submit this thesis and shout "bingo" as the university confer the PhD title. I also wish to thank Dr. Nabie Bayoh, Dr. Mary Hamel, Dr. Kim A. Lindblade, Dr. John Gimnig, Dr. Laurance Slutsker and Dr. Kayla F. Laserson for their constructive comments in this work. To my fellow colleagues in the office, many thanks for your prayers, words of encouragement and the great work you do to improve the health status of the residents of the study area through research. I am very happy and pleased to see how a number of you took on additional responsibilities to allow me concentrate in my school work. It is my belief that the experience acquired during this work will go a long way to fill the gap much needed to contribute to realizing the public health research agenda.

I thank the MTIMBA principal investigators (PIs) for conceptualization of the project, and all of whom I met at Bagamoyo while putting the database together: Dr. Salim Abdullah, Dr. Ricardo Thompson, Dr. Diadier Diallo, Dr. Seth Owusu-Agyei and Dr. Honorati Masanja. My apologies if I missed out someone. Without this noble idea, I would not have written this thesis in relation

Acknowledgements

to that idea. Thanks Dr. Honorati Masanja for encouraging me to apply for this MTIMBA PhD position.

I greatly appreciate the prayers and words of encouragement on phone, by text messages, e-mails and word of mouth from my friends back at home. In particular, I want to acknowledge Paul Okayo, Dickson Odhiambo, Evans Illah, George Olilo, Sammy Khagayi, Obor David, Maquins Sewe, James Ojwang, Peter Nyawach and Rev. Fr. Oscar. Thanks guys, those words made me stronger to continue with this work.

My heartfelt thanks go to Amek's family for the unconditional love, encouragement, inspiration and emotional support. Special thanks to Patrick, Tom, Odero, Akinyi, Anyangi and your families. Many thanks also go Lwanga, Odhiambo, Milka, Rev. Fr. Adede, Rev. Sr. Susan, Dancun, Mourice and Abonyo. I do appreciate your perpetual encouragement. To my best friends Dorcas and mum, thanks for being there to offer emotional support. I can't thank you enough.

This thesis is dedicated to my deceased parents Raphael, Caren and Mary for the sacrifices they made to make sure I was in school and to my brother Patrick who took over the responsibility when they were gone. I thank God for His love, guidance, good health throughout my life and this work. May You continue to bless me abundantly.

The analysis and write up of this thesis was partly supported by the Swiss National Science Foundation (Project Nr. 325200_118379) and a Swiss-South African Joint Research Programme (Project Nr. JRP IZLSZ3_122926).

Chapter 1: Introduction

1.1 Mortality and malaria burden

Globally, over 50 million people are estimated to die each year and majority of these deaths occur in Africa mainly in children less than five years old (Murray and Lopez, 1997; UN report, 2009). Malaria infection is implicated as the leading cause of global illness (Snow et al., 2005; WHO malaria report, 2009) and mortality (Snow et al., 1999; Rowe et al., 2006; WHO malaria report, 2009). About half of the world population is at risk of being infected. In 2008, there were 109 malaria endemic countries (Figure 1.1) and 3.3 billion people were at the risk of malaria. About 243 million people become ill and almost a million die of malaria each year. The majority of malaria illnesses (85%) and deaths (89%) occur in Africa (WHO malaria report, 2009) affecting mainly children under five years of age and pregnant women in endemic areas. The magnitude of malaria burden does not only extend beyond immediate threats to survival, but also the consequences of repeated clinical infection place a burden on households, systems and thus impact the country's economy (Gallup and Sachs, 2001). Although recent studies (Bhattarai et al., 2007; Okiro et al., 2007; Murray et al., 2012) reported dramatic decreases in the malaria-specific deaths and/or hospital admissions due to scale up of combined malaria control strategies, malaria infection still remains a major public health and development problem in Sub-Saharan Africa. Moreover, the relationship between malaria transmission intensity and mortality remains unclear (Snow et al., 1997; Smith. et al., 2001; Gemperli et al., 2004). INDEPTH, a network of demographic surveillance systems in developing countries initiated malaria transmission intensity and mortality burden across its member sites in Africa (MTIMBA) project in the year 2002 to improve our knowledge on this relationship. Clear understanding of the transmission-mortality relation in is important for planning effective interventions thus reduces high mortality.

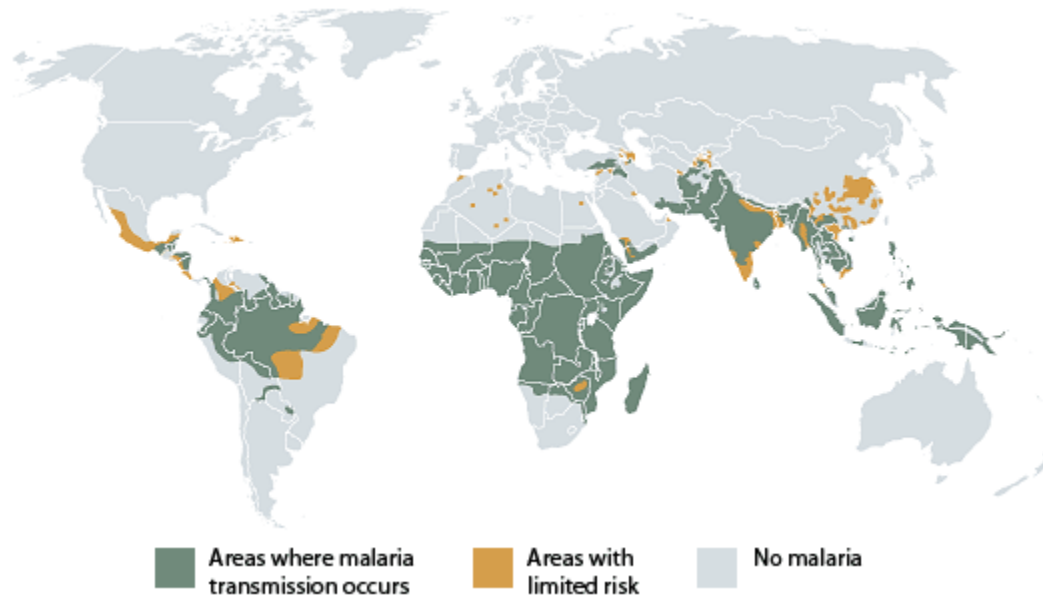


Figure 1.1 Geographical distribution of malaria in 2008 (source: World Health Organization: www.who.int/ith/en)

1.2 Malaria disease and transmission

1.2.1 The malaria parasite

Malaria is a vector-borne disease caused by protozoan parasites of the genus *Plasmodium*. There are four malaria parasite species which infect humans, namely *Plasmodium falciparum*, *P. vivax*, *P. malariae* and *P. ovale*, with *P. falciparum* being the predominant (in Africa) and most likely to cause severe complications and death among the four (Greenwood et al., 2005). *P. vivax* is less virulent, though malaria due to *P. vivax* has been implicated for a huge health burden, life expectancy and productivity of the general population (Mendis et al., 2001; Genton et al., 2008). It is mostly common in Middle East, Asia and the Western pacific (Mendis et al., 2001). *P. malariae* and *P. ovale* are thought to have symptoms that are usually less severe (Mueller et al., 2007) with the former found in tropical and sub-tropical regions across the globe and the latter primarily in Sub-Saharan Africa. However, *P. ovale* infections have also been found in various parts of Southeast Asia (Baird et al., 1990; Kawamoto et al., 1999; Win et al., 2002), the Middle East (Al-Maktari et al., 2003) and the Indiana (Jambulingam et al., 1989). Recently, a fifth species *P. knowlesi* previously thought to mainly infect only monkeys (long-tailed and pig-tailed), was found widespread among humans in Malaysia (Singh et al., 2004).

Chapter 1: Introduction and objectives

The malaria parasite is transmitted from human to human via the bite of infected female mosquitoes of the genus *Anopheles* that most often bite at night. Figure 1.2 shows the life cycle of malaria parasite. The first stage of malaria infection begins when a female mosquito carrying malaria-causing parasites injects the parasite (sporozoites) into a human blood stream and then rapidly invades liver cells. In the liver, sporozoites multiply and transform into merozoites. The duration of this process depends on the species of *Plasmodium*. However, some malaria parasite species such as *P. vivax* remains dormant for a longer period in the liver, causing relapses even several years after the first attack (Krotoski WA., 1989). The merozoites are then released from the liver and invade red blood cells within the blood circulation. In the blood circulation, merozoites either develop into (female/male) gametocytes or undergo repeated cycles of replication. The merozoites that continue to multiply may result in many sporozoite-infected in the host blood stream, leading to illness and complication of malaria if not treated.

When a female mosquito bites an infected human, it ingests the gametocytes. Within the mosquito, gametocytes mature into gametes and sexual replication takes place producing zygote. The zygote further develops into mature oocyst that burst, releasing sporozoite which migrate to salivary glands. These sporozoites are injected when the mosquito bites another human and a new life cycle begins (Garnham, 1988). The duration of the mosquito cycle depends on the species of *Plasmodium*. For instance, during the exoerythrocytic cycle it takes 43 to 48, 50 and 72 hours for *P. falciparum* and *P. vivax*, *P. ovale*, and *P. malariae* infections respectively (Garnham, 1988).

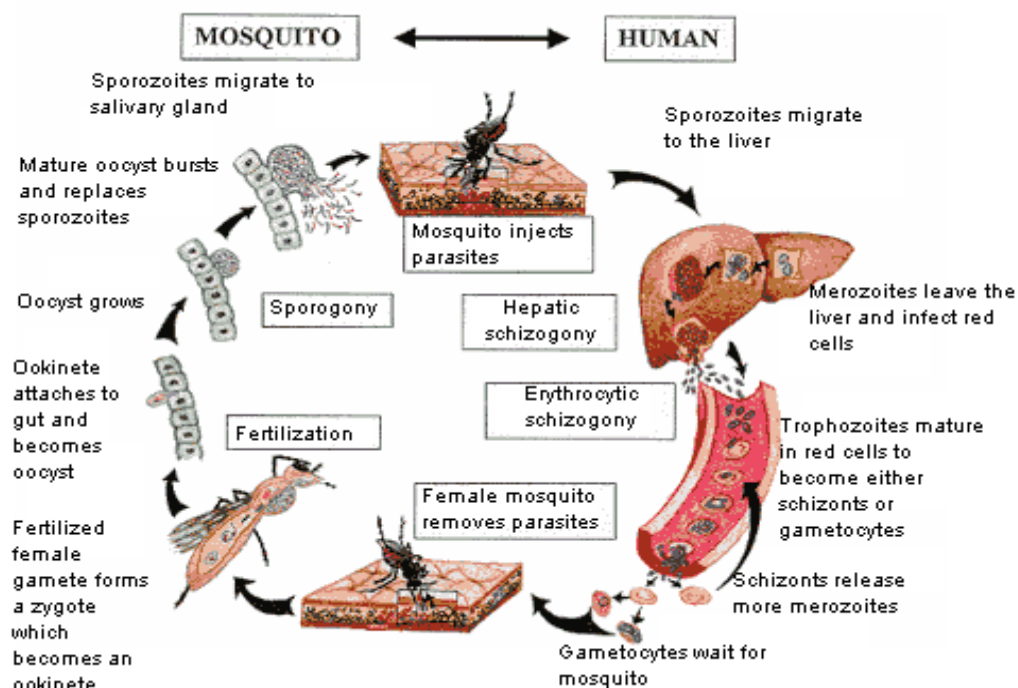


Figure 1.2 Life cycle of malaria parasite (Source: <http://www.emro.who.int/rbm/images/malarialifecycle-1.gif>)

1.2.2 The Anopheles vectors of human malaria

Four hundred and sixty-two (462) species of *Anopheles* have been named (Harbach, 1994, 2004), at least 70 of which are capable of transmitting human malaria (Service and Townson, 2002) and from these, *Anopheles gambiae* complex (*An. gambiae sensu stricto*, *An. arabiensis*, *An. quadriannulatus*, *An. bwambae*, *An. merus* and *An. melas*) (White, 1974; Coetzee et al., 2000) and *Anopheles funestus* are the main vectors associated with malaria transmission in Africa which bears the highest burden of malaria infection. These vectors have different behavioral characteristics and feeding preferences. For instance, *An. gambiae sensu stricto* prefers humid areas, rests indoors (endophilic), feeds indoors (endophagic) mainly on humans and breeds in fresh water such as rice fields, sunlit pools. *An. arabiensis* are common in arid areas, rest outdoors (exophilic), prefer feeding on domestic animals and breed mostly in pools produced by rain water. However, these two sub-species of *An. gambiae* are broadly sympatric (White, 1974; Lindsay et al., 1998; Coetzee et al., 2000). *Anopheles funestus* breed mostly in permanent water bodies especially with vegetation, is highly anthropophilic and is both endophagic and exophilic.

1.2.3 Malaria transmission

Measures of malaria intensity and transmission

Malaria transmission intensity is an important determinant of the burden of the disease. Measures of malaria transmission are broadly categorized into those associated with infection in humans (parasite prevalence, incidence rate, and seroprevalence rate), measures related to mosquito-human interaction (entomological inoculation rate) and measures associated with vector (mosquito density, infectious reservoir).

Malaria prevalence is the most commonly used measure of intensity. It is the number of individuals found to be infected with malaria parasites in their blood smears out of the total screened. This information is obtained from cross-sectional community surveys or through longitudinal investigations of a specific through longitudinal investigations of study population. Prevalence data is often restricted to children in high endemic areas, but surveys should be extended to include all age groups in areas of low endemicity. Survey data across locations are often heterogeneous in the age groups involved and the seasons that the surveys were carried out.

Malaria incidence is also a measure of the amount of malaria transmission. It is the number of new cases of malaria diagnosed in a population per unit time thus a direct measure of the amount of malaria transmission. Incidence data is mostly obtained from health facilities. These estimates are often unreliable especially in Africa where several factors such as cultural/systems beliefs, socioeconomic status, distance influence the decision of whether to or not to visit a health facility (Obrist et al., 2007). Furthermore, most of the health facilities in the malaria endemic areas lack equipment (such microscopy) for parasite confirmation thus incidence of fever is used as a proxy for malaria incidence. In areas where microscopy is available, skills required to detect parasites and accuracy are also weak if not lacking, but rapid diagnostic test are increasingly widespread.

The entomological inoculation rate (EIR) is a more direct measure of malaria intensity. It estimates the number of infectious mosquito bites an individual is exposed to per unit time (i.e. day, month or year) (MacDonald G, 1957). EIR is expressed by the product of the “human biting rate” and the sporozoite rate (SR) which is the fraction of mosquitoes with sporozoites in their salivary glands (Birley and Charlewood, 1987; Molineaux et al., 1988). The “human biting rate” refers to the number of mosquito bites a person receives per unit time. It is obtained by mosquito

collection techniques such as pyrethrum sprays collection, light trap and human landing catch. Although, EIR estimates may not be accurate as mosquito density is markedly heterogeneous and sporozoite rates are usually very low even in high endemic areas. It is still a better estimate of transmission intensity particularly when efforts are geared to reduce human-vector contact. In the recent past it has been suggested that seroprevalence rates could be used as a proxy measure of transmission intensity (Drakeley et al., 2005).

Climatic determinants of malaria transmission

Malaria transmission is influenced by a number of factors such as climatic and environmental conditions, socioeconomic status, population movement, access to and quality of health care (treatment of parasite), interventions targeting reduction of vectors, human activities (irrigation, dam construction) and drug resistance. These factors affect the distribution and abundance of mosquito population. However, climatic factors are the main drivers of malaria transmission by influencing both the parasite and vector directly besides vector control and treatment of parasites. For instance, temperature influences both the distribution of mosquito vector and its life cycle. Optimum temperature for mosquito parasite development (sporogonic) cycle ranges from 27° to 31°C, but as the temperature decreases, the number of days required to complete the cycle increases. Parasite development ceases at 16°C (for *P. falciparum*). Increase in temperature speeds up the mosquito development and reduces the intervals between blood-meals, leading to higher transmission. Temperatures above 32°C causes high vector turnover that are weak thus high vector mortality (Craig et al., 1999).

Rainfall is an important factor influencing malaria transmission. It provides breeding sites for mosquitoes to lay eggs thus increasing mosquito population and it also regulates temperature and humidity, two important factors for mosquito survival. However, unusual heavy rainfall can flood away breeding larvae or eggs and/or lower temperature, reducing malaria transmission intensity in highland areas. Conversely in rare cases, lack of rainfall can turn rivers/streams into pools which act as mosquito breeding mosquito sites leading to malaria epidemic. The duration of rainfall season is important in areas where rainfall is seasonal and high temperature, this is because transmission peak shortly after the onset of rains due to short sporogonic cycles (Craig

et al., 1999). Vegetation index has been suggested to be a good proxy of humidity and rainfall (Thomson et al., 1997).

Altitude also plays a role in malaria transmission. Anopheline are generally found in altitudes below 2500 meters above the sea level (Service and Townson, 2002) and as altitude increases, the duration necessary to complete the sporogonic cycle increases, so transmission becomes unlikely because few mosquitoes survive long enough to transmit. Non-malaria transmitting anophelines may still be common at higher altitudes (highland areas).

1.2.4 Malaria control policies

A number of malaria control strategies have been implemented over the last century, aiming to reduce or eliminate the burden of malaria. Larval control and source reduction were the main malaria control methods before war world II. Even though these methods significantly helped to eliminate malaria in many countries, including Italy, Israel and United States (Kitron and Spielman, 1989), malaria continued to be a serious problem worldwide.

In 1939, Paul Muller discovered the highly effective insecticide dichlorodiphenyltrichloroethane (DDT) and the focus shifted to adult mosquito control. The effectiveness of DDT gave hope that malaria could be eradicated through a combination of indoor residual spraying (IRS) and disease surveillance to detect and treat any remaining infections. In line with this notion, the World Health Organization conducted a global eradication program in the 1950s and 1960s. Although it was initially successful in many countries such as India, Sri Lanka and the former Soviet Union, the success was never sustained due to high cost of the programme, unwillingness of many communities to repeatedly spray their houses, and the emergence of DDT resistance. Apart from a few pilots, in highly endemic areas and in particular Sub-Saharan Africa the eradication program was never implemented (Greenwood and Mutabingwa, 2002; Muturi et al., 2008).

Despite challenges associated with the global eradication program, over 30 counties were freed of malaria reinforcing the belief that global eradication is achievable (Greenwood, 2009). However, the overall goal to globally eradicate malaria was never realized and thus was considered widely as a failure (Greenwood et al., 2005). This failure led to neglect of malaria control and research for a period of about 25 years (early 1970s to late 1990s). As an example,

Chapter 1: Introduction and objectives

during the period 1975-1996 only 3 out of 1223 new drugs developed were antimalarial (Trouiller and Olliaro, 1998).

Following the failure to achieve global eradication, the malaria eradication policy was abandoned and recommendations were made at the 22nd World Health Assembly targeting alternative approaches for malaria control in areas where elimination was not feasible. Six years later (1978) a resolution was adopted that malaria control strategies be based on local epidemiological and socioeconomic factors and integrated into primary health care systems. However, this strategy was never implemented in most countries due to lack of funds, political goodwill, and emergence of chloroquine resistance *Plasmodium falciparum* in Africa and rapid deterioration of primary health care. This led to an unprecedented increase in malaria morbidity and mortality in the 1980s (Greenwood and Mutabingwa, 2002; Breman et al., 2004; Greenwood et al., 2005; Muturi et al., 2008).

The continued increase of burden of malaria into early 1990s gave birth to the Global Malaria Control Strategy in 1992. In the ministerial conference in Amsterdam, local analysis of malaria effect as a way of assessing sustainability and cost effectiveness was adopted (Muturi et al., 2008). That is, case management through early detection and prompt treatment. However, since the meeting, no significant progress in terms of implementation has been achieved due to lack of funding in endemic countries.

In 1998, the Roll Back Malaria (RBM) partnership was initiated with an aim to develop global and local partnership to halve the deaths due to malaria by 2010 (Yamey, 2000) and ensure that malaria ceases to be a public health problem by 2030, through the implementation of four key technical strategies: insecticide-treated nets, improved case management, control of malaria in pregnancy, and early warning and containment of epidemic (WHO, 1998). The RBM global partners include WHO, the United Nations Children Fund (UNICEF), the World Bank and the United Nations Development Program (UNDP).

For four decades from 1969, malaria control strategies dominated the control efforts to reduce malaria burden, and eradication of malaria was never been suggested within malaria control community due to perceived failure of the global eradication program in the 1950s and 60s.

However in October 2007, Melinda and Bill Gates renewed the campaign to eradicate malaria which was supported by WHO and RBM (Roberts and Enserink, 2007; Greenwood, 2009). However, at the moment malaria experts doubt its practical feasibility with the current tools, weak health systems and poor understanding of malaria transmission heterogeneity (Greenwood et al., 2005; Tanner and Savigny, 2008; Greenwood, 2009).

1.3 Malaria related mortality and transmission

A large proportion of childhood deaths in high malaria transmission areas are attributed to malaria infection (Velema et al.; Alonso et al., 1993; Salum et al., 1994; Smith. et al., 2001). Recent studies (Bhattarai et al., 2007; Rajaratnam et al., 2010; Hamel et al., 2011; Murray et al., 2012) have reported a dramatic decline in all-cause and/or malaria specific mortality in African children apparently due to high coverage of combined malaria control strategies which aim to lower transmission intensity, but mortality burden remains intolerable. Furthermore, there are concerns that any intervention targeting reduction of malaria exposure might delay the acquisition of immunity, shifting the burden of disease to an older age group (Snow et al., 1997; O'Meara et al., 2008). There are also suggestions that long term transmission control in high endemic areas might delay severe infection or death and perhaps even increase all-cause mortality (Snow and Marsh, 1995; Trape and Rogier, 1996; Snow et al., 1997). A better understanding of the relationship between mortality and malaria transmission intensity is needed for proper planning, effective implementation and evaluation of interventions to reduce high childhood mortality thus achieving the Millennium Development goal for child survival. However, this relationship remains unclear (Snow et al., 1997; Smith. et al., 2001; Gemperli et al., 2004). More on this topic is discussed in chapter 5 and 6.

The majority of the current estimates of malaria as a cause of death in Africa are derived from verbal autopsy techniques (VA) (Snow et al., 1999). VA is based on information obtained from a caregiver on the deceased's terminal illness. Questions asked include clinical signs, symptoms and duration during the fatal illness. The information collected is reviewed independently by a panel of two or three physicians to assign probable causes of death. In situations where the information is reviewed by more than two reviewers, concurrence by atleast two is required to assign cause of death (Snow et al., 1992; Adazu et al., 2005). The process of assigning cause of

death has a number of challenges not limited to systematic differences in interpreting VA information (reliability), expensive (time consuming) and lack physicians. Furthermore, the poor sensitivity and specificity of the tool in measuring malaria mortality has been highlighted (Snow et al., 1992; Todd et al., 1994; Anker et al., 1999). Currently VA is the only approach in obtaining cause of death at community level in Africa. Most recently, alternative (computer) methods for coding VA data, based on expert algorithm and data driven of assigning causes of death have emerged. These methods are currently in being validated (Reeves and Quigley, 1997; Freeman et al., 2005; Oti and Kyobutungi, 2010).

1.4 An overview of health and demographic surveillance systems (HDSS)

Reliable estimates on all-cause and malaria specific mortality are difficult to obtain, if not completely lacking in poor developing countries, particularly in Sub-Saharan Africa the home of the highest burden of mortality and malaria transmission. This is mostly due to weak health systems, poor attendance of health facilities, weak civil registration and death certification systems, as most of the deaths including those caused by malaria occur at home without any contact with the health system (Korenromp et al., 2003; Mathers et al., 2005). To overcome some of these weaknesses, the Health and Demographic Surveillance Systems (HDSS) were established in various parts of the developing countries to supplement where routine vital registration systems are poorly developed or nonexistent. They track a limited and common set of key variables determining population dynamics and demographic trends, in a well defined population through routine collection and processing information on births, deaths and migrations. In addition, the HDSS collect information on health outcomes (such as causes of death using verbal autopsy, incidence and prevalence of particular diseases of public health importance), carry out routine surveillance of the entomological correlates of malaria in randomly selected households, and conduct education and socio-economic surveys (Adazu et al., 2005).

In 1998, a global network (International Network for the Demographic Evaluation of Populations and Their Health in Developing Countries-INDEPTH) of all HDSS was established with an aim to strengthen capacity of its members and inform policy in low- and middle-income countries

though multi-site research. In November 2006, INDEPTH had a total of 37 member sites in 19 countries (Abdullah et al., 2007). Figure 1.3 shows the distribution of member countries.



Figure 1.3 Countries with HDSS field sites and members of INDEPTH Network in November 2006
(Source http://library.wur.nl/frontis/environmental_change/09_sankoh.pdf)

1.5 Malaria transmission intensity and mortality burden across Africa initiative

Monitoring the progress of the roll back malaria (RBM) goals of halving malaria mortality and morbidity rates by year 2010 requires reliable baseline all-cause and malaria specific mortality data which is often lacking in malaria endemic areas. Moreover, RBM advocate for vector control mainly through wide scale use of insecticide treated nets as one of the main strategies for achieving the objective which has been documented to have remarkable reduction on malaria transmission, malaria morbidity and all-cause childhood mortality (Binka et al., 1996; Nevill et al., 1996; Diallo et al., 2004; Lengeler, 2004; Lindblade et al., 2004). However, the relationship between malaria mortality and transmission intensity remains unclear (Snow et al., 1997; Smith et al., 2001; Gemperli et al., 2004).

In 2002, INDEPTH network initiated the MTIMBA (Malaria Transmission Intensity and Mortality Burden across Africa) project with the financial support from MIM/TDR and RBM in a number of its malaria endemic member sites across Africa to generate reliable information that will guide malaria control policies in Africa. The main objective of MTIMBA is to improve our

understanding of the relationship between mortality and malaria transmission intensity in malaria endemic areas. Other objectives include collaboration between RBM malaria initiative at different levels for monitoring and evaluation, and build capacity in the field of malaria research and control in Africa.

MTIMBA is a multi-centre study, involving 18 malaria-endemic sites in Africa, but only eight of these sites from West (Navrongo, Nuona, Kourweogo and Oubritenga), East (KEMRI/CDC, Ifakara and Rufiji) and Southern (Manhica) Africa provided comprehensive disaggregated data at household or individual level on all-cause mortality, entomology and malaria control activities that can be used to study the relationship between mortality and malaria transmission intensity (Kasasa et al.; Abdullah et al., 2007). To our knowledge, the MTIMBA database is the only current entomological and mortality database that can be used to assess space-time variation in malaria transmission and efficacy on interventions in multiple sites.

1.6 Characteristics of the MTIMBA data

1.6.1 Geostatistical data

HDSS data are collected at fixed geographical locations. This type of data is known as geostatistical data. Observations collected at locations close to each other in space are correlated because locations in close proximity are characterized by similar risks due to common exposures. Standard statistical models assume independence of observations. Therefore analyzing these data without taking into account the spatial correlation could result in incorrect model estimates (Cressie, 1993; Thomson et al., 1999).

Spatial models take into account the spatial correlation according to the way the geographical information is available. For instance, in geostatistical data, spatial models introduce an extra parameter (random effect) at each location. These parameters are considered as latent observations of a spatial process and are modeled via a multivariate distribution which incorporates spatial correlation in the covariance matrix, typically assuming that the covariance between any pair of locations is a function of distance between the locations. The number of parameters increases with the number of locations surveyed. Hence these models are highly parameterized when large number of locations is involved (as in the case of MTIMBA data) and

can not be estimated by most commonly used maximum likelihood methods. Bayesian computational methods are suitable in fitting highly parameterized models by employing Markov chain Monte Carlo (MCMC) simulation algorithms (Gelfand and Smith, 1990). Diggle et al (1998) formulated geostatistical models using the Bayesian framework of inference. These models have been applied and further developed for mapping malaria transmission (Kleinschmidt et al., 2000; Diggle et al., 2002; Gemperli, Sogoba, et al., 2006; Gemperli, Vounatsou, et al., 2006; Gosoni et al., 2006, 2009; Kazembe et al., 2006; Sogoba et al., 2007; Hay et al., 2009; Riedel et al., 2010; Gething et al., 2011) and mortality (Gemperli et al., 2004; Kazembe et al., 2007; Sartorius et al., 2011).

For large number of locations (e.g. over 1000) computations involving the covariance matrix of the spatial process during model fit are not feasible. Recent developments in geostatistical modeling estimate the spatial process from a subset of locations and use approximations to obtain the random effects at the observed locations (Banerjee et al., 2008). These methods have been used in analyzing MTIMBA Rufiji data in Tanzania (Rumisha et al., 2012) and mortality data from the Agincourt DSS in South Africa (Gosoni et al., 2012)

1.6.2 Spatial misaligned data

Entomological data was collected in randomly selected houses (locations), while mortality outcome status was obtained from all locations within the study area. The locations of the two datasets do not necessarily match and thus the datasets are spatially misaligned (Banerjee and Gelfand, 2002). In 2003, Gamperli and colleagues analyzed misaligned malaria survey and mortality data extracted from independent databases: the demographic and health surveys (DHS) and the mapping malaria risk in Africa (MARA) database, respectively. They linked the data by developing geostatistical models to predict malaria prevalence at the mortality locations. Subsequently survival models with errors-in-covariates were fitted to take into account the prediction error of the malaria covariate.

1.6.3 Seasonality and temporal data

Malaria transmission is driven by environmental factors such as rainfall and temperature. Therefore transmission intensity and vector population fluctuate over time in areas where environmental factors are seasonal. In addition entomological data was collected biweekly

introducing temporal correlation in the data. Ignoring seasonal and/or temporal correlation when analyzing these data may lead to incorrect model estimates. Studies have adjusted for seasonality within a modeling framework by introducing a binary covariate indicating wet (transmission) and dry (no transmission) seasons (Abeku et al., 2002; Gemperli, Sogoba, et al., 2006). (Zhang et al., 2007; Briët et al., 2008) have used seasonal autoregressive integrated moving averages (SARIMA) models to assess seasonality and take into account temporal correlation. Furthermore, harmonic functions have been employed to model seasonal trends in time-series data (Stolwijk et al., 1999; Griffin et al., 2010). However, in malaria epidemiology literature is sparse in model-based approach in estimating seasonal trends in non-Gaussian data. In a Bayesian formulation temporal correlation can be modeled by introducing into the model random parameters at each time point (e.g. month) modeled via autoregressive process of various orders. The Deviance Information Criterion (Spiegelhalter et al., 2002) is used to identify the best fitting order.

1.6.4 Zero inflated entomological data

Mosquito entomological data are usually collected at fix locations over time. Therefore besides being correlated in space and time, they are also characterized by large number of locations with either no mosquitoes or proportion with parasites (sporozoites in glands). The occurrence of large number of “zero” (no) mosquitoes or infected ones could be due to (i) seasonality: population of mosquitoes is high in wet season as it favors their development and survival as opposed to dry season which is characterized by unsuitable weather conditions resulting to high mortality of mosquito and (ii) interventions targeting survival of mosquitoes prevents their development and kills older ones that are likely to be infectious. These lead to no mosquitoes or proportion of infected mosquitoes.

The presence of large number of zero mosquitoes or proportion infected results in over-dispersion in the data and this is popularly known as “zero-inflation”. Zero inflated data contain extra zeros relative to underlying distribution. Standard models only estimate a certain frequency of zeros in the data. Therefore are not appropriate for analyzing zero inflated data because they predict fewer zeros than the number that are observed leading to poor fit. Zero inflated analogues of the standard models are appropriate to fit sparse data (Lambert, 1992; Hall, 2000). A zero inflated model is a mixture model having two components; one arising from a parent

distribution and the other corresponds to the excessive zeros that can not be accounted for by the distribution. The zeros from the parent distribution can be assumed to be random and driven by frequency determined by the parent distribution. The remaining excess zeros are assumed to be “structural” that may arise from unmeasured predictors of the outcome and/or seasonality. Zero inflated models have been applied in number epidemiological studies (Nobre et al., 2005; Ramis-Prieto et al., 2007; Barnes et al., 2008; Fernandes et al., 2009; Vounatsou et al., 2009; Berrang-Ford et al., 2010; Manh et al., 2011), but limited in malaria transmission adjusting for temporal correlation.

1.7 Objectives of the thesis

The main objectives of the thesis were: (i) to develop data-driven Bayesian geostatistical and temporal models to assess the relationship between mortality and malaria transmission and (ii) to implement the above models for the analysis of the MTIMBA data extracted from KEMRI/CDC HDSS database using MTIMBA protocol to (a) estimate transmission heterogeneity and produce smooth maps of transmission intensity in the study area, (b) assess spatio temporal changes and obtain smooth surfaces of socioeconomic status and (c) assess the relationship between malaria transmission and mortality across ages after adjusting for intervention effects, socioeconomic status and demographic factors.

The specific objectives

- Develop geostatistical zero inflated spatio-temporal binomial models to analyze sparse malaria sporozoite rate data (chapter 2)
- Estimate monthly EIR surfaces at high spatial resolution in the study area, during the MTIMBA project period (chapter 3)
- Assess spatio-temporal changes of socioeconomic status (SES) in KEMRI/CDC HDSS and estimate SES surfaces (chapter 4)
- Assess the relationship between (all-cause and malaria specific) childhood mortality and malaria transmission adjusted for socioeconomic status, effect of intervention and demographic factors in KEMRI/CDC HDSS (Chapter 5)
- Extend the analysis of the malaria transmission mortality relation across all age-groups in the study population (Chapter 6)

Chapter 2: Spatio-temporal modelling of sparse geostatistical malaria sporozoite rate data using a zero inflated Binomial model

Nyaguara Amek^{1,2,3}, Nabie Bayoh¹, Mary Hamel^{1,4}, Kim A Lindblade⁴, John Gimnig⁴, Kayla F. Laserson¹, Laurence Slutsker⁴, Thomas Smith^{2,3}, Penelope Vounatsou^{2,3*}

1. Kenya Medical Research Institute/Centers for Disease Control and Prevention (CDC), Research and Public Health Collaboration, P.O. Box 1578 Kisumu, Kenya
2. Swiss Tropical and Public Health Institute, Socinstr. 57, P.O. Box, 4002 Basel, Switzerland
3. University of Basel, Petersplatz 1 P.O. Box 4003 Basel, Switzerland
4. Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta GA 30301, Georgia, USA

*Corresponding Author

This paper has been published in Spatial and Spatio-temporal Epidemiology Journal 2 (2011) 283-290

Abstract

The proportion of malaria vectors harboring the infectious stage of the parasite (the sporozoite rates) is an important component of measures of malaria transmission. Variation in time and/or space in sporozoite rates contribute substantially to spatio-temporal variation in transmission. However, because most vectors test negative for sporozoites, sporozoite rate data are sparse with large number of observed zeros across locations or over time in the case of longitudinal data. Rarely are appropriate methods and models used in analyzing such data. In this study, Bayesian zero inflated binomial (ZIB) geostatistical models were developed and compared with standard binomial analogues to analyze sporozoite data obtained from the KEMRI/CDC health and demographic surveillance system (HDSS) site in rural Western Kenya during 2002-2004. ZIB models showed a better predictive ability, identified more significant covariates and obtained narrower credible intervals for all parameters compared to standard geostatistical binomial model.

Key words: Bayesian inference; Gaussian spatial process; Health Demographic Surveillance Systems; malaria entomological data; sparse data.

2.1 Introduction

Malaria infection is still one of the main public health problems in Africa despite continued attention and funding of control efforts. Approximately, 85% (208 millions) of global episodes of malaria occurred in Africa. About 1 million deaths are attributable to malaria yearly with 89% of these deaths occurring in Africa, mostly in children under 5 years old and pregnant women (WHO malaria report, 2009). Malaria parasites are transmitted from human to human via the bite of an infective anopheline mosquito. The mosquito larval development and survival strongly depend on prevailing climatic and environmental factors (Yé et al., 2008) which in turn influence malaria transmission intensity. These factors include land use and water management, flow accumulation, housing conditions, rainfall, humidity and temperature.

The most common measure of mosquito to human transmission is the entomological inoculation rate (EIR), which is the number of infective mosquito bites a person receives per unit time (usually a year) (Onori and Grab, 1980). The EIR is the product of anopheline mosquito density, the average number of mosquitoes biting each person per unit time (e.g. day) and the proportion of infective mosquitoes known as sporozoite rate (SR). Changes in SR have substantial impact on the EIR.

Malaria transmission and mosquito population fluctuate over the years and seasons (Billingsley et al., 2005) especially in areas where rainfall is highly seasonal. However, the proportion of infective mosquitoes is usually low, notwithstanding the fluctuations in the mosquito population. Most studies have reported very low sporozoite rates during periods of high mosquito density in both low and high transmission areas (Burkot et al., 1988; Beier et al., 1990). For instance, Burkot et al. (1988) reported SR lower than 1% in Madang Province in Papua New Guinea an area with perennial malaria transmission. Beier et al. (1990) also reported SR as low as 2% and 9.9% in Kisian and Saradidi areas respectively, in Western Kenya where transmission is high throughout the year. Therefore, it is common to observe zero (none) infective mosquitoes in a large number of locations or over time in the case of longitudinal data.

Transmission patterns also tend to vary in space and even within a high transmission area some locations will have more malaria compared to others (Smith et al., 1995). Transmission

heterogeneity is an important determinant of the likely impact of interventions. If foci of high transmission could be selectively targeted, then malaria control could be made easier, but usually we cannot locate the high transmission areas and in such situations, transmission heterogeneity makes control less effective than in a homogeneous population (Woolhouse et al., 1997; Filion et al., 2006). Quantification of heterogeneity in malaria transmission is challenging because the data are often sparse and standard statistical methods commonly used do not separate out sampling error from true spatial variation in transmission.

To assess spatial variation and seasonality in malaria transmission, mosquito data need to be collected at fixed geographical locations repeatedly over time. Observations collected at locations close to each other in space or time, are correlated because of similar environments and clustering of transmission. Analyzing these data without taking into account the spatial-temporal correlation will result in underestimation of the standard error of the covariate parameters (Cressie, 1993; Thomson et al., 1999). Spatial statistical methods take into account the spatial correlation according to the way the geographical information is available. In geostatistical data, spatial models introduce an extra parameter (random effect) at each location. These parameters are considered as latent observations of a spatial process and are modeled via a multivariate distribution which incorporates spatial correlation in the covariance matrix, typically assuming that the covariance between any pair of locations is a function of distance between the locations.

Levels of infectiousness of mosquitoes are typically reported as numbers of positive and total female mosquitoes screened for *Plasmodium falciparum* sporozoite antigens, and can be modeled by a binomial distribution. Logistic regression models can be used to link the logit of the binomial probability (sporozoite rate) with the factors driving malaria transmission (e.g. environmental, climatic) as well as with the spatial random effects. To our knowledge these models have not been used in real applications. Most of the analyses on sporozoite rate data are rather descriptive. Furthermore, it is common to observe many more locations with zero infected mosquitoes, than predicted by the standard binomial models. The extra zeros introduce overdispersion. Logistic regression models which ignore overdispersion have poor fit because they predict fewer zeros than the number that are observed.

In the recent past, there has been an increasing interest in models for count data that allow large number of zeros in the data. These models have been applied in a number of epidemiological applications of space-time data such as malaria incidence (Nobre et al., 2005), schistosomiasis intensity (Vounatsou et al., 2009), dengue fever (Fernandes et al., 2009), sleeping sickness incidence (Berrang-Ford et al., 2010) and cancer mortality data (Ramis-Prieto et al., 2007).

Lambert (1992) described zero-inflated Poisson (ZIP) regression, a class of models for count data with many zeros. Hall (2000) extended Lambert's ZIP model by introducing random effects to take into account correlation in repeated measurements from horticultural data. Using score test statistics, Ridout et al. (2001) showed that the zero-inflated negative binomial (ZINB) model can provide a very good fit if the nonzero counts are overdispersed. Vieira et al. (2000) proposed an overdispersed truncated binomial model for zero-inflated proportions when focus is only in the non-zero data. The zero-inflated binomial (ZIB) distribution was first introduced by (Kemp and Kemp, 1988). Subsequently, the model was reviewed by Hall (2000), who further developed it by including random effects to analyze proportion of life adult insects from horticultural data. To our knowledge there are limited applications of ZIB models to overdispersed geostatistical data.

This paper extends the ZIB model to fit geostatistical temporal mosquito sporozoite rate data having large number of locations with zero infective mosquitoes. In Section 2.2, we describe the data which motivated this application. Section 2.3 presents a Bayesian formulation of the geostatistical-temporal ZIB model. We report on the performance of this model and its ability to predict sporozoite data (Section 2.4), and discuss the results and model extension in Section 2.5.

2.2 The motivating example/Data

The entomological data which motivated this work were obtained from the KEMRI/CDC health and demographic surveillance system (HDSS) site in rural Western Kenya. The data were extracted using the Malaria Transmission Intensity and Mortality Burden across Africa (MTIMBA) protocol. MTIMBA was a multi-centre project during 2002-2004 with the aim to examine the relationship between mortality and malaria transmission intensity in HDSS malaria endemic sites in Africa.

Chapter 2: Modelling of sparse geostatistical malaria sporozoite rate data

The KEMRI/CDC HDSS area is characterized by gentle hills/slopes (elevation =1147-1388m) that are drained by several small streams and one permanent river. Rainfall occurs year-round with heavy rains falling from March through May and from November to December (wet season). The remaining months of the year receiving light showers (dry season). At the time the data were collected, the HDSS covered approximately 500 km² with a population of 135,000 residing in 33,990 households within 21,477 compounds in Asembo and Gem, two defined regions in the HDSS area (Adazu et al., 2005; Ombok et al., 2010). Malaria transmission fluctuates over the wet and dry seasons during the year with an overall entomological rate of seven infection bites per person per year (Adazu et al., 2005). Despite a high coverage of insecticide treated nets (ITN) malaria prevalence is still high and is the main cause of mortality (48 deaths per 1000 person years) in children under 5 years old.

Mosquitoes were collected monthly using the Centers for Disease Control and Prevention (CDC) light traps from 10 randomly selected houses each month from HDSS database and four additional houses neighboring each index house. In each house, a light trap was placed next to the sleeping place of an index person and mosquitoes were collected for two consecutive nights. The sleeping place was covered with an insecticide treated net to protect the index person from mosquito bites. All mosquitoes captured in the light trap were identified to morphological species level and members of *Anopheles gambiae* complex were further identified to sibling species level using a modified polymerase chain reaction (PCR) method (Scott et al., 1993). The female mosquitoes were tested for the presence of *P. falciparum* sporozoite antigens using enzyme linked immunosorbent assays methods (Wirtz et al., 1987). During the study period, 3850 houses had sporozoite data.

The sporozoite data were linked to the geographic information system (GIS) database to extract the coordinates of each house (location) where mosquito data were collected, resulting in 267 unique locations. A database was built linking these entomological data with environmental predictors including rainfall, season, elevation, distance to the nearest water source (the lake, streams and river), vegetation (normalized difference vegetation index) and land surface temperature.

Normalized difference vegetation index (NDVI) and land surface temperature (LST) were extracted at 250 m by 250 m and 1 km by 1 km spatial resolution, respectively from Moderate Resolution Imaging Spectroradiometer (MODIS). Elevation (distance above the sea level), was extracted at 1 km resolution from Digital Elevation Model (DEM), all these were obtained from US Geological Survey (USGS) EROS Data Center. Rainfall estimate (RFE) data with an 8 km by 8 km spatial resolution from Meteosat 7 satellite were also obtained from the Africa Data Dissemination Service (ADDS).

For each sampled location, we calculated monthly estimates of environmental factors over the years of the study period taking into account the possible lag time in the effect of environmental factors on sporozoite rate, arising from the latent periods in the life cycles of the mosquito and of the parasite. For each environmental factor, lag time ranged from the current to 3 month prior to mosquito collection.

2.3 Model formulation

The data are typically binomial as they represent the number of female *Anopheles* mosquitoes screened for the presence of *P. falciparum* sporozoite antigens and of the number found positive. Eighty-seven percent of our data locations have no (zero) sporozoite rate. Below we describe the standard binomial model and the ZIB analogue which takes into account the “excess zeros” observed with frequency higher than the one explained by the binomial distribution. The standard binomial model assumes that the number of positive mosquitoes Y_{it} at location i and month t arises from a binomial distribution: $Y_{it} \sim \text{Bin}(N_{it}, p_{it})$ where p_{it} corresponds to the sporozoite rate and N_{it} is the number of tested mosquitoes. Covariates X_{it} are modeled on the logit scale of p_{it} , that is $\text{logit}(p_{it}) = X_{it}^T \beta$ where, β is the vector of regression coefficients.

To model the high frequency of zero sporozoite rates which can not be estimated by the binomial distribution determined by the covariates in p_{it} , a ZIB specification has been developed. ZIB assumes that each zero has probability $(1-\pi_{it})$ to arise from a binomial distribution and probability π_{it} to be a structural zero observed due to unmeasured covariates such as proximity to a breeding site with young mosquitoes without sporozoite antigens in their glands. Therefore the “zero” is

modeled as a mixture of a binomial distribution and a degenerated distribution at mass zero with mixing proportion π_{it} (Lambert, 1992; Hall, 2000). The ZIB model is formally described as follows:

$$\Pr(Y_{it} = y | p_{it}, \pi_{it}, N_{it}) = \begin{cases} \pi_{it} + (1 - \pi_{it}) \text{Bin}(0 | p_{it}, N_{it}) & \text{for } y = 0 \\ (1 - \pi_{it}) \text{Bin}(y | p_{it}, N_{it}) & \text{for } y > 0 \end{cases}$$

where π_{it} is the mixing proportion. The term π_{it} can be assumed constant or dependent on covariates Z_{it} which could be similar to, or a subset of X_{it} that is $\text{logit}(\pi_{it}) = Z_{it}^T a$. In our application, the mixing proportion was assumed to be constant based on an exploratory analysis which showed that the proportion of zeros in our data was neither associated with seasonality nor the time of the mosquito collection.

2.3.1 Geostatistical-temporal binomial and ZIB models

We introduce spatial and temporal correlation by adding location and month-specific random effects ϕ_i and ε_t , respectively, on the logit scale of the sporozoite rate, that is $\text{logit}(p_{it}) = X_{it}^T \underline{\beta} + \phi_i + \varepsilon_t$. We assume that the ϕ_i 's are latent observations from a Gaussian spatial process and therefore arise from a multivariate normal distribution, $\underline{\phi} = (\phi_1, \dots, \phi_n)^T \sim N(0, \sigma_1^2 R)$ with σ_1^2 measuring spatial variation and R modeling the correlation between any pair of locations i and j . We adopt an exponential function for R : $R_{ij} = \exp(-\rho d_{ij})$, where d_{ij} is the distance between i and j and ρ is a smoothing parameter corresponding to the rate of correlation decay with increasing distance. The value $3/\rho$ estimates the minimum distance at which the spatial correlation is significant at 5% (Ecker and Gelfand, 1997). Temporal correlation was introduced by monthly random effects and modeled by autoregressive (AR) process of various orders. The deviance information criterion (DIC) (Spiegelhalter et al., 2002) was used to identify the best fitting order of the process, which was found to be 1, and therefore we consider that $\varepsilon_t \sim N(\gamma \varepsilon_{t-1}, \sigma_2^2)$ and $\varepsilon_1 \sim N(0, \sigma_2^2 / (1 - \gamma))$. The terms σ_2^2 and γ are the temporal variance and autocorrelation parameters respectively with $\gamma \in (-1, 1)$. Agarwal et al. (2002) proposed introducing spatial correlation on the mixing proportion, however such a formulation has not been used in our application.

2.3.2 Model fit

To complete the model formulation above, we specify prior distributions for the model parameters. For regression coefficients ($\underline{\beta}$), we choose a non-informative normal prior distribution with large variance, inverse gamma priors for σ_1^2 and σ_2^2 , a gamma prior for ρ and a Uniform prior for γ , that is $\sigma_1^2, \sigma_2^2 \sim IG(2.01, 1.01)$, $\gamma \sim U(-1, 1)$ and $\rho \sim G(0.1, 0.1)$. We further consider a constant zero-inflated mixing proportion across the area and time with a Uniform prior distribution $\pi \sim U(0, 1)$. The Markov chain Monte Carlo (MCMC) simulation algorithm was used to estimate model parameters (Gelfand and Smith, 1990). Starting with some initial values about the parameters, we ran two chains sampler discarding the first 5000 iterations. Convergence was assessed by Gelman-Rubin diagnostic (Gelman, 1992).

Bayesian kriging (Diggle et al., 1998) was used to predict the sporozoite rate at locations where the sporozoite data were not collected. In particular, for each sample of the parameters from the posterior distribution, a random effect is simulated from the Gaussian spatial process conditional on the random effects estimated at the observed locations. This is added on the regression term relating the covariates at the new location with the regression coefficients estimated during the model fit. The resulting equation estimates the sporozoite rate on the logit scale at the new location as a sample from the posterior predictive distribution. A grid of 7726 pixels with 250 m by 250 m spatial resolution covering the entire study area was used to predict sporozoite rate. The analysis was carried out in STATA 10 (Stata corporation, College Station, Texas, USA), WinBUGS version 1.4 (Imperial College and Medical Research Council, London, UK) and in specialized Bayesian geostatistical codes written by the authors in Fortran 95 (Digital Equipment Corporation) using standard numerical libraries (Numerical Algorithms Group Ltd.).

2.3.3 Model validation

We evaluated the predictive ability of the model by fitting the data to a training set of 230 (85%) randomly selected locations, and compared the model-based predictions for the remaining 41 (15%) locations (test set) with the observed data (Gosoni et al., 2009). The predictive performance of the models was assessed by calculating (1) credible intervals for each model and test location with varying probability coverage ranging from 5% to 95% of the posterior predictive distribution and (2) the mean square error between the observed and predicted data.

The model with the best predictive ability was the one with the highest percentage of locations within the interval of smallest coverage as well as the model with the smallest mean squared error.

2.4 Application

Preliminary analyses using a standard zero-inflated logistic model (without spatial and/or temporal effects) was carried out to assess the best combination of lag environmental variables (rainfall, vegetation and temperature) associated with sporozoite rate. All models fitted included a seasonality term, distance to water bodies and altitude, in addition to the environmental factors. Seasonality was modeled by (1) trigonometric functions with a cycle of 12 months corresponding to two transmission seasons (wet vs. dry) and (2) a binary variable (wet vs. dry). The wet (rainy) and dry seasons were defined based on the rainfall data, with wet season being from March through May and November to December, and the remaining months grouped in the dry season. Based on the Akaike's information criterion (Akaike, 1974), a model including wet and dry season, rainfall, average temperature and NDVI over the month preceding the mosquito collection was used in the spatial analysis.

An. gambiae s.l. was the predominant vector species representing 96.5% (n = 1608) of the tested mosquitoes. The remaining 3.5% belonged to the *An. funestus* species. Seventy-two percent of specimens identified by PCR were *An. gambiae* s.s., the remainder were *An. arabiensis*. Point estimates of the overall sporozoite rate (SR) were 3% and 0% for *An. gambiae* s.l. and *An. funestus* respectively during the study period. The distribution of surveyed compounds (locations) with infected and none (zero) infected mosquitoes is shown in Figure 2.1. About 87% of these locations had no infected mosquitoes and the SR for locations with infected mosquitoes was 19%. Overall, the sporozoite rate was significantly higher in Gem, which is the northern part of the study area compared to Asembo, which is the southern part during the wet season while this trend was reversed in the dry season with the sporozoite rate significantly higher in Asembo than Gem.

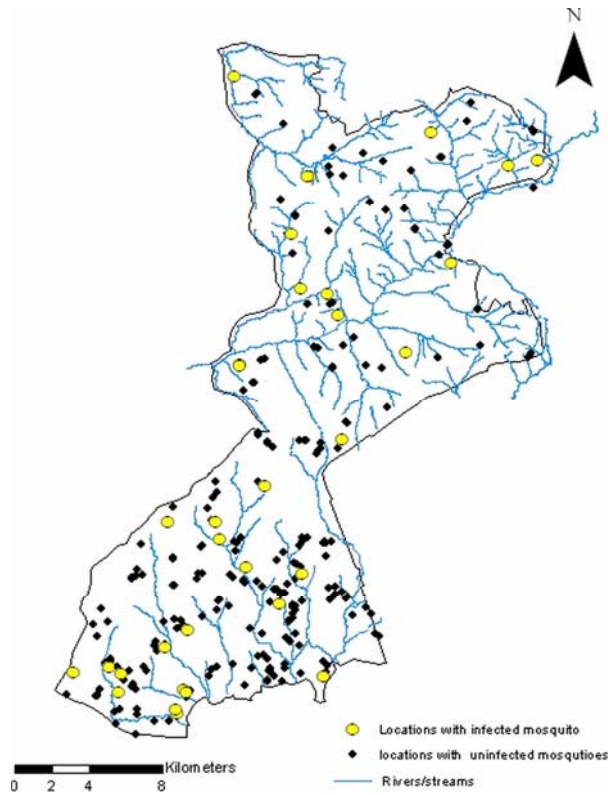


Figure 2.1 Surveyed locations with infected and non-infected mosquitoes

Model validation results in Figure 2.2 present the proportion of test locations predicted well by the different models within Bayesian credible intervals with probability cover ranging from 5% to 95%. Within an 80% credible interval, the ZIB spatial and spatio-temporal models included 80% and 68% of the test locations respectively as compared to 21% and 58% of the locations predicted well by their standard binomial analogues. Similar results were obtained using the mean square error measure (data not shown). Based on the above results we employed further the ZIB spatial model to produce a high geographical resolution SR map of the HDSS.

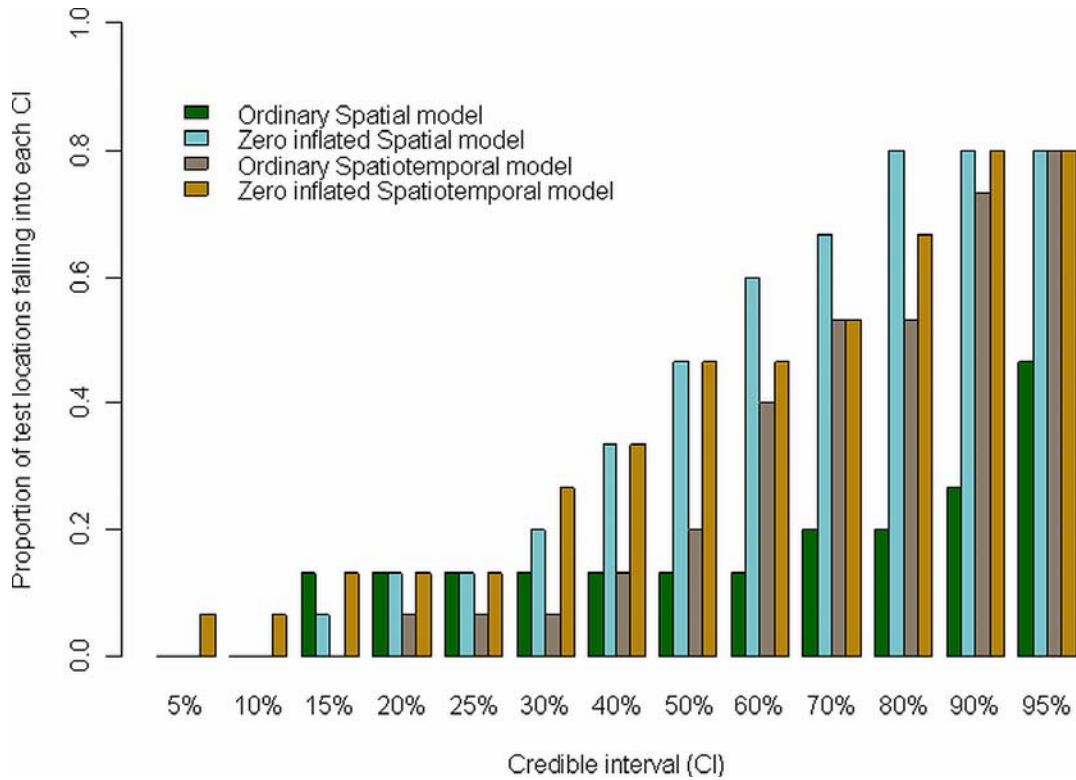


Figure 2.2 Proportion of test locations with sporozoite rate falling in between 5% to 95%. Credible intervals of the posterior predictive distribution

Model-based estimates of the parameters of the ZIB spatial model together with those from the standard binomial are shown in Table 2.1. The covariates significantly associated with SR are rainfall, altitude and seasonality. In particular, rainfall of the month preceding the mosquito collection is negatively associated with the sporozoite rate, implying that the lower the rainfall, the higher the sporozoite rate. A negative association was also found between seasonality (wet vs. dry) and sporozoite rate, implying high sporozoite rate during the wet season. Furthermore, the model suggests that the sporozoite rate increases with altitude within the study area. It is worth noting that seasonality and altitude are significant in the ZIB model, but not, in the standard binomial model. Distance to water body, NDVI and average temperature of the previous month of mosquito collection were not significantly associated with sporozoite rate. Overall, confidence intervals in the ZIB model are narrow as compared to the binomial model. The zero-inflated mixing proportion was estimated to be 0.62 (95% credible interval: 0.40-0.77) in comparison to the 87% of locations with zero sporozoite rate. The posterior median of the spatial decay parameter ρ was equal to 13.04 (95% credible interval: 8.21-34.47), meaning that the

Chapter 2: Modelling of sparse geostatistical malaria sporozoite rate data

minimum distance at which the spatial correlation is less than 5% is equal to 25.53 km (95% credible interval: 9.66, 40.63).

Table 2.1 Posterior estimates of geostatistical models

Covariates	Binomial model		Zero-Inflated binomial model	
	Median	95% CI	Median	95% CI
Intercept	-9.939	(-23.220,3.663)	-8.378	(-15.538,1.089)
Seasonality (wet vs. dry)	-1.125	(-2.227,0.252)	-1.296	(-2.458,-0.034)
Distance to water bodies	0.001	(-0.001,0.002)	0.001	(-0.001,0.003)
Altitude	0.002	(-0.010,0.010)	0.005	(0.002,0.008)
Rain	-0.014	(-0.021,-0.007)	-0.012	(-0.022,-0.004)
NDVI	6.659	(-0.567,14.640)	0.198	(-8.752,7.195)
LSTAV	0.083	(-0.177,0.325)	-0.176	(-0.446,0.081)
Spatial Variation	1.294	(0.130,6.922)	0.021	(0.007,0.321)
Range($3/\rho$) ^a	23.743	(7.198,40.282)	25.530	(9.657,40.626)
Zero-Inflated proportion	-	-	0.624	(0.397,0.771)

a : minimum distance in kilometers at which spatial correlation is significant at 5%

Figure 2.3 display the smooth maps of sporozoite rate generated from the ZIB spatial model for dry and wet seasons. Both maps predict high sporozoite rates in most locations in the northern and a few locations in the southern part of the study area, with sporozoite rate being high during the wet season. The prediction errors for both seasons are shown in Figure 2.4. The uncertainty in the predicted SR is highest in a few areas at the southern part of the study area.

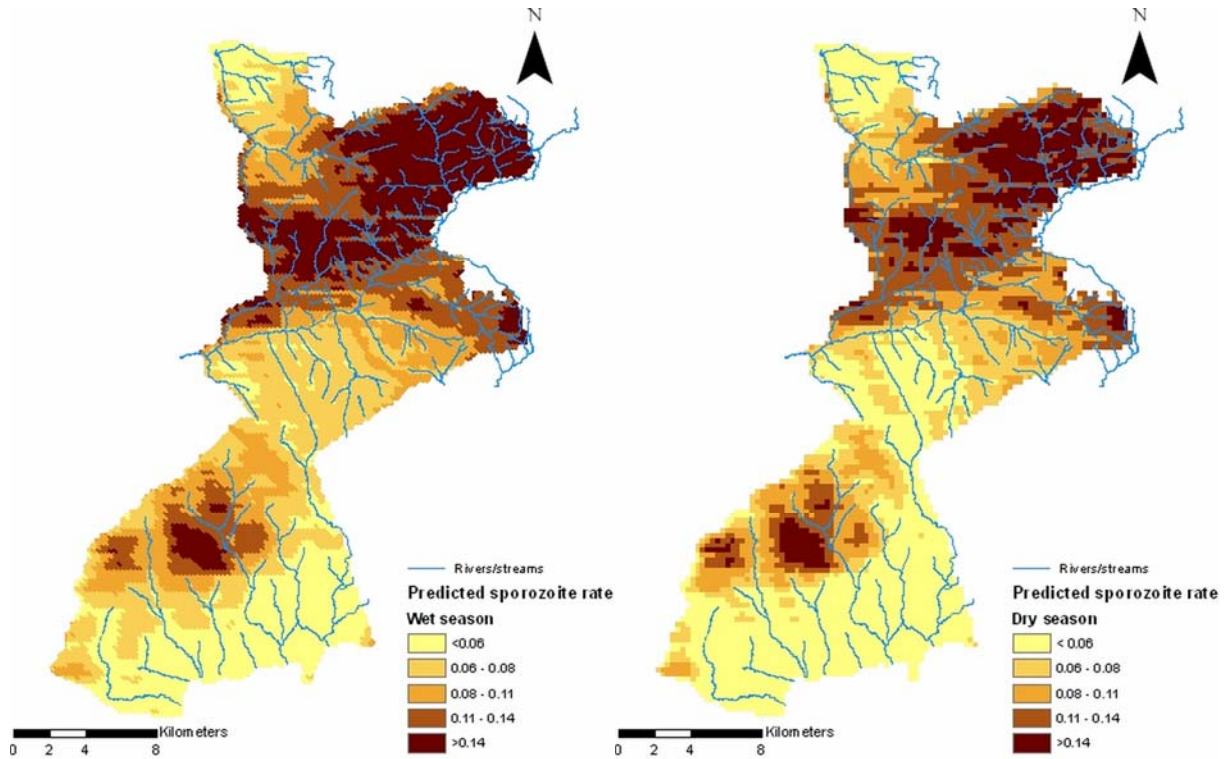


Figure 2.3 Maps of predicted sporozoite rate during wet season (March through May and November to December) and dry season (January to February and June through October) based on spatial zero inflated binomial model.

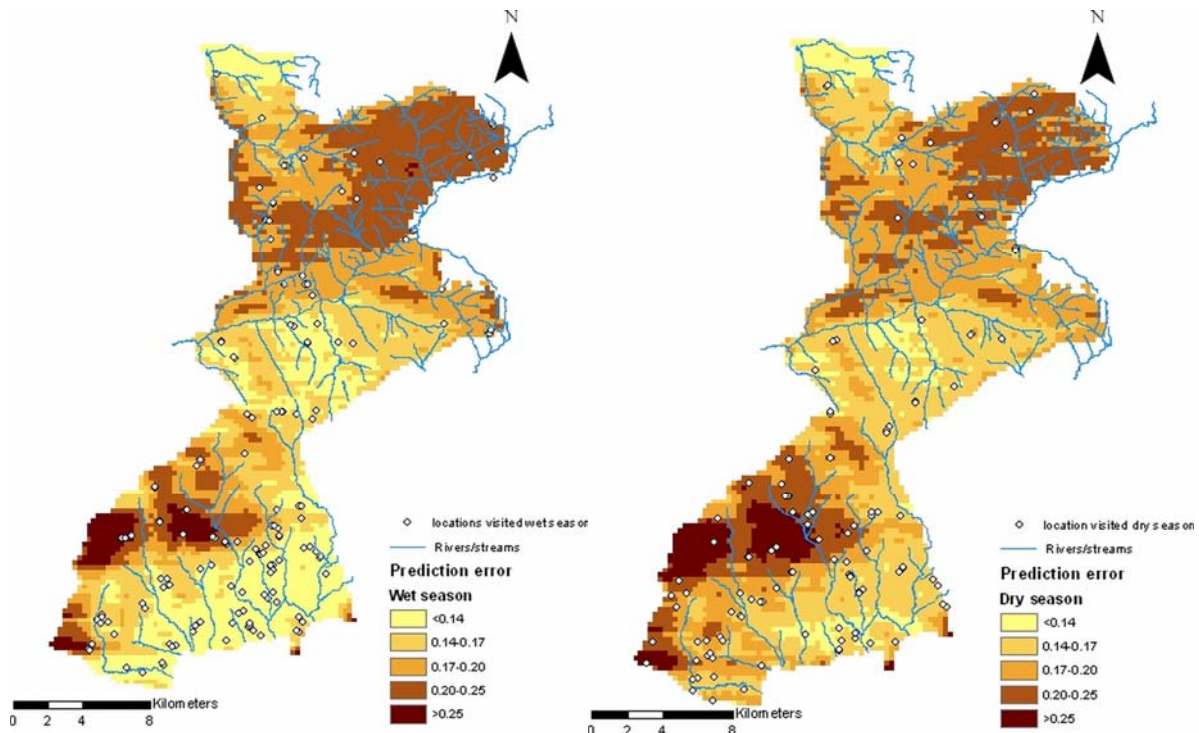


Figure 2.4 Map of prediction error of sporozoite rate during wet season (March through May and November to December) and dry season (January to February and June through October) based on spatial zero inflated binomial model

2.5 Conclusions

In malaria epidemiology, it is common to observe spatially correlated data with large number of locations having “zero” observations, but appropriate methods and models are rarely used to analyze this type of data. In this study we showed that Bayesian geostatistical zero inflated binomial models for sporozoite rate data observed with “many zeros” across locations have better predictive performance than standard binomial analogues. Incorporating temporal correlation in the standard binomial model improved its predictive ability, however the spatial ZIB resulted in a more parsimonious specification with the highest predictive performance. In addition the ZIB model identified more significant covariates and gave narrower credible intervals for all parameters.

Rainfall, seasonality and altitude are significantly related to sporozoite rate, with a negative association between rainfall in the previous month and sporozoite rate. This can be explained by the long duration of sporozoite development in relation to mosquito survival, implying that only older mosquitoes carry sporozoites. Rainfall leads to emergence of many young mosquitoes (after a delay due to larval development), which reduce the proportion of older ones in the population. The positive association of altitude with sporozoite rate can also be explained by the concentration of newly emerged mosquitoes near their larval habitats, which tend to be concentrated at lower elevations where water may collect (Mutuku et al., 2006). The absence of a significant relation with the distance to water bodies was perhaps surprising, but *An.gambiae* s.l. breeds in very small water bodies (Gimnig et al., 2003), that are not necessarily included in the geographical information system (GIS) layer, so the locations of streams, which in any case are rather widely dispersed in the study area, is not important for transmission. A study of the geographic distribution of adult mosquitoes in the same area also found no relationship with water bodies identified in a GIS database, particularly during the rainy season when small temporary habitats were abundant (Hightower et al., 1998).

NDVI, representing vegetation in the month prior to mosquito collection, was not significantly associated with sporozoite rate. While NDVI can be a reasonable indicator of suitability for mosquitoes at a continental or national scale small area variations in vegetation cover evidently do not have a simple relationship with breeding site locations.

Chapter 2: Modelling of sparse geostatistical malaria sporozoite rate data

The value of the zero inflated mixing proportion in our data shows that out of the 87% of the locations with zero infected mosquitoes, 62% could not be accounted for by the binomial distribution. Observed and predicted sporozoite rate revealed seasonal and geographical variation over the study area. Spatial correlation in sporozoite rate was significant in distances up to 26 km (95% credible interval: 9.66, 40.63). This could reflect the nature of mosquitoes moving longer distances and unmeasured local factors such as the mosquito breeding sites and socioeconomic status among others that are spatially correlated.

The ZIB specification applied in this study allows a proportion of the zeros to be “random” derived with frequency determined by the binomial distribution, and the remaining zeros (excess zeros) to be structural (Lambert, 1992; Hall, 2000). Structural zeros may arise due to unmeasured factors such as interventions targeting mosquito survival and therefore killing older mosquitoes that are likely to be infective. Dry months, unsuitable for malaria transmission may contribute with an increase frequency of non-infected mosquitoes, however malaria seasonality is taken into account as a covariate on the logit scale of sporozoite rate and therefore the binomial distribution should be able to explain those observed zeros. Vieira et al. (2000) proposed a truncated distribution to model the non-zeros assuming that all zeros are structured. Lam et al. (2006) applied truncated Poisson models to explain zero inflated count data. A model formulation assuming a truncated binomial distribution is less appropriate when modeling sporozoite rate data because observed factors such as seasonality will contribute to non-structural zeros.

The maps of sporozoite rates obtained in this study will be of value in constructing models of space-time patterns in the EIR that we can use to analyze relationships between malaria exposure and mortality for individual households in a number of HDSS sites across Africa. At the same time, our models take us nearer to understanding the extent of heterogeneity in exposure that sustains malaria transmission in Africa.

Acknowledgments

The authors are grateful to the KEMRI/CDC HDSS for providing the data and to the principal investigators of the MTIMBA project for initiating the project. The analysis of the data was partly supported by the Swiss National Science Foundation (Project Nr.325200_118379) and a Swiss-South African Joint Research Programme (Project Nr. JRP IZLSZ3_122926). The

Chapter 2: Modelling of sparse geostatistical malaria sporozoite rate data

KEMRI/CDC HDSS is a member of the INDEPTH network. The findings and conclusions in this study are those of the authors and do not necessarily represent the views of the centers for Disease Control and Prevention. This paper is published with permission from the Director KEMRI

Chapter 3: Spatial and temporal dynamics of malaria transmission in rural Western Kenya

Nyaguara Amek^{1,2,3}, Nabie Bayoh¹, Mary Hamel^{1,4}, Kim A Lindblade⁴, John Gimnig⁴, Frank Odhiambo¹, Kayla F. Laserson¹, Laurence Slutsker⁴, Thomas Smith^{2,3}, Penelope Vounatsou^{2,3*}

1. Kenya Medical Research Institute/Centers for Disease Control and Prevention (CDC)
Research and Public Health Collaboration, P.O. Box 1578 Kisumu, Kenya
2. Swiss Tropical and Public Health Institute, Socinstr. 57, P.O. Box, 4002 Basel, Switzerland
3. University of Basel, Petersplatz 1 P.O. Box 4003 Basel, Switzerland
4. Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta GA 30301, Georgia, USA

*Corresponding Author

This paper has been published in *Parasites & Vectors* journal 2012, 5:86

Abstract

Understanding the relationship between *Plasmodium falciparum* malaria transmission and health outcomes requires accurate estimates of exposure to infectious mosquitoes. However, measures of exposure such as mosquito density and entomological inoculation rate (EIR) are generally aggregated over large areas and time periods, biasing the outcome-exposure relationship. There are few studies examining the extent and drivers of local variation in malaria exposure in endemic areas. We describe the spatio-temporal dynamics of malaria transmission intensity measured by mosquito density and EIR in the KEMRI/CDC health and demographic surveillance system using entomological data collected during 2002-2004. Geostatistical zero inflated binomial and negative binomial models were applied to obtain location specific (house) estimates of sporozoite rates and mosquito densities respectively. Model-based predictions were multiplied to estimate the spatial pattern of annual entomological inoculation rate, a measure of the number of infective bites a person receive per unit of time. The models included environmental and climatic predictors extracted from satellite data, harmonic seasonal trends and parameters describing space-time correlation. *Anopheles gambiae* s.l was the main vector species accounting for 86% (n=2309) of the total collected mosquitoes with the remainder being *Anopheles funestus*. Sixty eight percent (757/1110) of the surveyed houses had no mosquitoes. Distance to water bodies, vegetation and day temperature were significantly associated with mosquito density. Overall annual point estimates of EIR were 6.7, 9.3 and 9.6 infectious bites per annum for 2002, 2003 and 2004 respectively. Monthly mosquito density and EIR varied over the study period peaking in May during the wet season each year. The predicted and observed densities and EIR showed a strong seasonal and spatial pattern over the study area. Spatio-temporal maps of malaria transmission intensity obtained in this study are not only useful in understanding variability in malaria epidemiology over small areas but also provides a high resolution exposure surface that can be used to analyse the impact of transmission on malaria related and all-cause morbidity and mortality.

3.1 Introduction

Malaria parasites are transmitted from human to human via the bite of an infected female anopheline mosquito. The life cycle of the mosquito and parasite within it are strongly influenced by climatic factors, primarily rainfall, temperature and humidity. Suitable rainfall provides mosquito breeding sites and temperature influences both vector and parasite development. By understanding the relations between environmental/climatic factors and malaria transmission in space and time, transmission intensity can be estimated in areas where data are otherwise lacking and high risk areas can be identified. Understanding spatial and temporal variation in vector density and transmission intensity is useful in planning effective malaria control programs and determining the optimal allocation of limited resources.

Malaria transmission intensity is often assessed by the entomological inoculation rate (EIR) which is the product of the vector biting rate and the sporozoite rate (SR) which is the proportion of mosquitoes with sporozoites in their salivary glands (Molineaux et al., 1988). EIR estimates the number of infective bites a person receives per unit time and thus the level of exposure of an individual to malaria parasites. Studies have shown strong correlation between EIR and malaria prevalence (Beier et al., 1994, 1999; Charlwood et al., 1998). Furthermore, EIR is the most accurate measure of transmission intensity (Burkot and Graves, 1995) particularly when efforts are made towards reducing human-vector contact.

Mosquito population size and sporozoite rates fluctuate between seasons and over years (Charlwood et al., 1995). Shortly after the onset of rainfall, mosquito populations increase to a peak. As the dry season sets in, mosquito populations decline in numbers since no new recruits are added to the population (Billingsley et al., 2005). A change in either mosquito density or sporozoite rate or both affects the EIR. Similarly, mosquito population distribution is heterogeneous (Smith et al., 1995; Hightower et al., 1998; Drakeley et al., 2003) and even within a defined geographical area mosquito densities vary widely in space and time.

In the KEMRI/CDC Health and demographic surveillance systems (HDSS), entomological data are collected from randomly selected locations (houses) as part of routine surveillance to assess the effects of interventions aimed at reducing malaria transmission intensity. The main

characteristics of the data are the presence of spatio-temporal correlation and the large number of locations without mosquitoes (zeros). Spatial correlation arises because neighbouring locations are influenced by similar exposures such as climate and environment due to close proximity of locations. Analyzing these data without taking into account these specific characteristics result in overestimation or underestimation of the statistical significance of the covariates (Cressie, 1993) and poor model fit respectively.

Several studies have reported large spatio-temporal variations in mosquito density, SR, and EIR (Smith et al., 1995; Lindsay et al., 1998; Drakeley et al., 2003; Dery et al., 2010). For instance Dery et al. (2010) reported sporozoite rate of 1.5 % and 4.7 % for *An. funestus* and *An.gambiae* respectively and annual EIR estimates of 267 and 231 infectious bites per person per year (ibpy) for first year and second year respectively in the forest-savannah transitional belt of Ghana. Drakeley et al. (2003) also reported SRs of less than 1% with EIRs ranging from 4 to 108 in the cool and wet seasons respectively in Ifakara, a semi-urban area in Tanzania. In the same study, EIR of 54 ibpy was reported in the eastern part of Ifakara town compared to only 15 ibpy at the center of the town. Smith et al. (1995) mapped mosquito (*An. funestus* and *An.gambiae*) densities in Namawala, a single village in Morogoro region of south eastern of Tanzania. Overall spatial pattern of mean log densities of both species were similar with higher density of *An. funestus* species in the southern edge of the village adjacent to rice growing fields. In the above studies, large number of locations had zero mosquitoes. However, appropriate statistical methods taking into account zero inflation were not used to assess variation in space and time. In addition, sporozoite rate are binomial data, whereas mosquito densities are count data which requires different modeling approaches to obtain EIR.

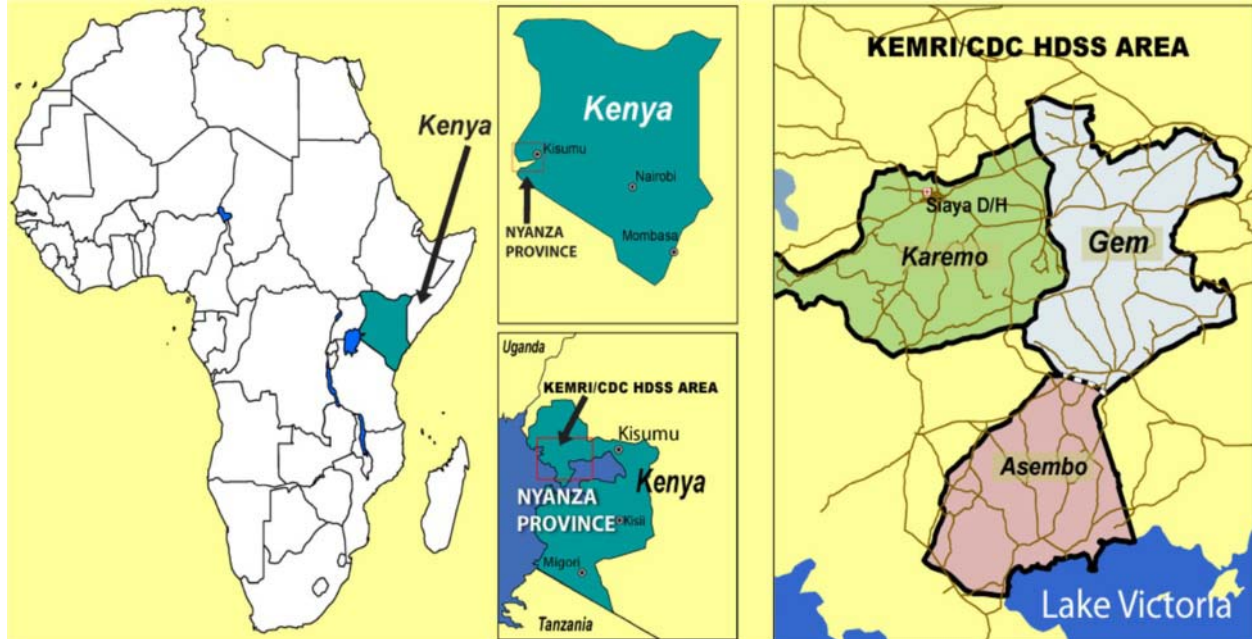
In our previous work (Amek et al., 2011), we developed spatio-temporal zero inflated models to analyze sparse sporozoite rate data. These models have been used to obtain spatially explicit estimates and maps of sporozoite rates in the KEMRI/CDC HDSS. In this study we extend our previous work by analyzing zero inflated mosquito density data. Spatio-temporal model based-estimates of mosquito density are combined with sporozoite rate model based-estimates obtained by Amek et al., 2011 to estimate the space-time pattern of EIRs.

3.2 Methods

3.2.1 Study site

This study was carried out in the KEMRI/CDC HDSS site located in Asembo (Rarieda Division, Bondo District), Gem (Yala and Wagai Divisions, Siaya District) and Karemo (Karemo Division, Siaya District) areas situated in Nyanza Province, rural western Kenya (Figure 3.1).

Figure 3.1 Location of the KEMRI/CDC HDSS site



During the study period, the KEMRI/CDC HDSS was only operating in Asembo, bordering Lake Victoria and Gem, adjacent to and North of Asembo. The HDSS has been described elsewhere in detail (Adazu et al., 2005). In brief, KEMRI/CDC HDSS area is characterized by gentle hills/slopes (elevation =1,147-1,388meters) that are drained by several small streams and one permanent river in Gem. Rainfall occurs year-round with heavy rains falling from March through May and from November to December (wet season). The remaining months of the year receive only light showers (dry season). Most inhabitants reside in traditional houses with mud walls and thatched roofs clustered into compounds. The compounds consist of clusters of one or more houses separated from other such clusters by the surrounding agricultural fields. At the time the data were collected, the study area covered approximately 500 km² with a population of 135,000 living in 33,990 households within 21,477 compounds.

Malaria is holoendemic in the KEMRI/CDC HDSS area where it is transmitted by *An.gambiae* s.l. and *An. funestus* (Phillips-Howard et al., 2003b; Adazu et al., 2005). A trial of insecticide-treated mosquito nets (ITNs) trial conducted from 1996 to 2002 reduced malaria transmission by 90% (Gimnig et al., 2003; Lindblade et al., 2004). However, despite the continued use of ITNs and a relatively low EIR of about seven ibpy (Adazu et al., 2005), malaria prevalence remains high and is still the main cause of child mortality (Kayla et al., 2009).

3.2.2 Entomological data

The entomological data (2002-2004) used in this study has been described elsewhere in detail (Adazu et al., 2005). In brief, *Anopheles* mosquitoes were collected monthly using Centers for Disease Control (CDC) light traps from 10 randomly selected houses (locations) each month from HDSS database along with four additional houses neighboring each index house. In each house, a light trap was placed next to the sleeping place of an individual who was randomly chosen from the list of household members and mosquitoes were collected for two consecutive nights. The sleeping place of the selected individual was covered with an insecticide treated net to protect the person from mosquito bites. Captured mosquitoes were initially identified morphologically while members of *Anopheles gambiae* complex were further identified to species using polymerase chain reaction (PCR) (Scott et al., 1993). Female *Anopheles* mosquitoes were tested for the presence of circumsporozoite antigens using an enzyme linked immunosorbent assay method (Wirtz et al., 1987).

3.2.3 Entomological inoculation rate (EIR)

The entomological inoculation rate (EIR), was calculated as the product of light trap densities and the proportion of infected mosquitoes (sporozoite rate). Mosquito density in the light traps was calculated by dividing the number of mosquitoes caught by the CDC light traps by the number of trap-nights. This estimate was then adjusted by multiplying by 1.605 as described by Lines and colleagues to calibrate the light trap estimates to those of human landing catch (Lines et al., 1991). The conversion factor adjusts for vector collection bias between human bait catch technique which is directly associated with mosquito feeding on humans and light trap collection which tends to underestimate the densities observed in human landing catches (Le Goff et al., 1997). High EIR resolution was obtained as a product of predicted mosquito SR and density at

locations where mosquitoes were not collected. The former was extracted from analysis in (Amek et al., 2011).

3.2.4 Climatic and Environmental data

Climatic and environmental predictors used in this study are similar to the ones used by Amek et al. (2011). Land surface temperature, normalized difference vegetation index, rainfall, and elevation were extracted from remote sensing data. Distance to the nearest water source (the lake, streams and river) was obtained from the KEMRI/CDC HDSS global positioning system (GPS) database. Land surface temperature for day and night (LST) and Normalized Difference Vegetation Index (NDVI) were extracted at 0.25 km by 0.25 km and 1 km by 1 km spatial resolution respectively from Moderate Resolution Imaging Spectroradiometer (MODIS). NDVI is a proxy measure of vegetation cover ranging from 1 to -1. Positive values indicate presence of vegetation and negative values and values close to zero represent barren or water surfaces. Elevation (distance above the sea level) data were extracted at 1 km resolution from Digital Elevation Model (DEM). MODIS and DEM were obtained from U.S Geological Survey (USGS) EROS Data Center. Rainfall estimate (RFE) data with an 8 km by 8 km spatial resolution from Meteosat 7 satellite were also obtained from the Africa Data Dissemination Service (ADDS).

All environmental factors were extracted for each location and lags up to 3 months were created to account for possible elapsing (lag) time, between the predictive variables (rainfall, LST and NDVI) and the outcome variable (mosquito density).

3.2.5 Statistical analysis

The lag time analysis was carried out in STATA (version 9.0) to determine the best combination of lags that estimated the mosquito density taking into account the seasonality, distance to water bodies and elevation. Seasonality was modeled by (i) trigonometric functions with a cycle of 12 months (Stolwijk et al., 1999) corresponding to two transmission seasons (wet vs. dry) and (ii) a binary variable indicating wet or dry season. The wet and dry seasons were defined based on rainfall data, with wet season being March through May and November to December classified as wet season, and the remaining months classified as the dry season. Trigonometric functions estimate the magnitude and the exact peak point (e.g. month, week or day) of the seasonal variation using the amplitude and the phase parameters respectively.

Chapter 3: Dynamics of malaria transmission

The Akaike's information criterion (Akaike, 1974) was used to select the best fitting model combining seasonality and environmental factors. The best model included seasonality with a cycle of 12 months, average NDVI and night temperature (LSTN) during the month of mosquito collection, average day temperature (LSTD) during the current and previous month of mosquito collection and total rainfall during the current and two previous months of mosquito collection. A Bayesian geostatistical version of the above model using a zero-inflation formulation was further fitted to assess space time variation. The model included year effect and an autoregressive term to take into account temporal correlation. Bayesian Kriging method, similar to the one used in our previous work (Amek et al., 2011) was used to predict mosquito density at locations (houses) where mosquitoes were not collected. Location specific predictions of sporozoite rate obtained by Amek et al. (2011) and density were multiplied to obtain the EIR estimates.

The assessment of model predictive ability was similar to the one carried out by Amek et al (2011). We assessed model predictive ability by fitting the models on a training set of 85% (943) of the randomly selected locations and compared the model-based predictions with the observed data at the remaining 15% (167) test locations (Gosoni et al., 2009). The best model was one with the highest percentage of test locations falling within the Bayesian credible interval of smallest coverage as well as the model with the smallest mean square error.

The Bayesian model was fitted in OpenBUGS version 3.1.2 (Imperial College and Medical Research Council, London, UK) and Kriging was carried out in a code written by the authors in Fortran 95 (Digital Equipment Corporation) using standard numerical libraries (Numerical Algorithms Group Ltd). A description of the Bayesian geostatistical formulation model fitted to mosquito count data is given in the appendix.

3.3 Results

3.3.1 Abundance/density of vector species

A total of 2309 anopheline mosquitoes were collected from 3850 catches in 1110 unique locations during the study period. About 68% of these locations had no mosquitoes. *An. gambiae* s.l. mosquito was the predominant vector species accounting for 86 % of the total *Anopheles* mosquitoes collected. The remaining 14 % were *An. funestus*. Average monthly abundance of *Anopheles* mosquitoes varied over the study period. Each year, the peak collecting period for *An. gambiae* was May, during the rainy season (Figure 3.2). *An. funestus* was very low throughout the study period except in the month of April and December in the year 2004. PCR tests on the *An. gambiae* s.l. samples indicated that the majority (72%) were *An. gambiae* s.s with *An. arabiensis* accounting for the rest of the tested mosquitoes.

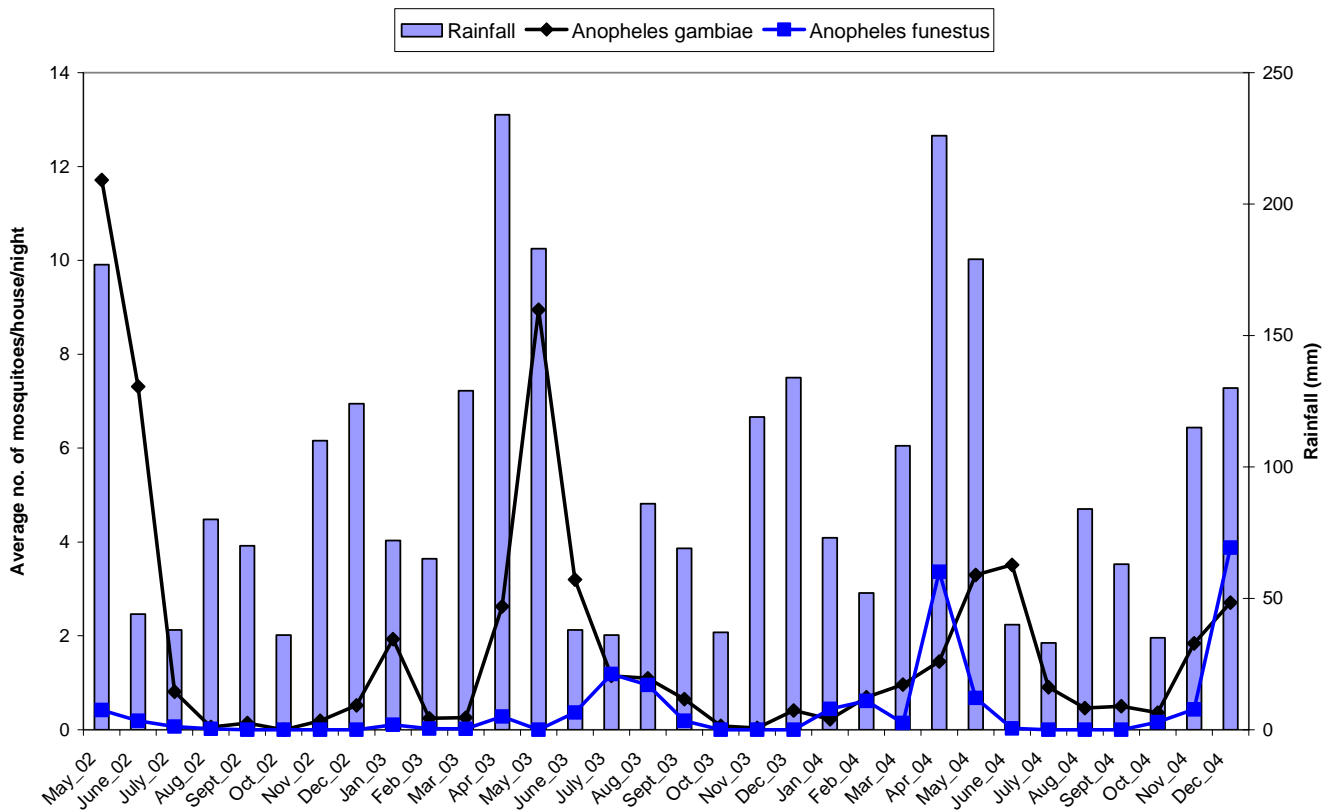


Figure 3.2 Monthly pattern of average number of *Anopheles gambiae* and *funestus* species in relation to Rainfall

Figure 3.3 shows the monthly pattern of observed, fitted and location-specific predicted density of *An. gambiae*. It should be noted that the observed density has a similar pattern to the location-specific predicted and fitted densities throughout the study period. *An. gambiae* density varied over the months with peaks May of each year. However, the absolute density during the peak month (May) significantly decreased over the 3 years of the study. Comparison between wet and dry months indicated that density was higher in wet months.

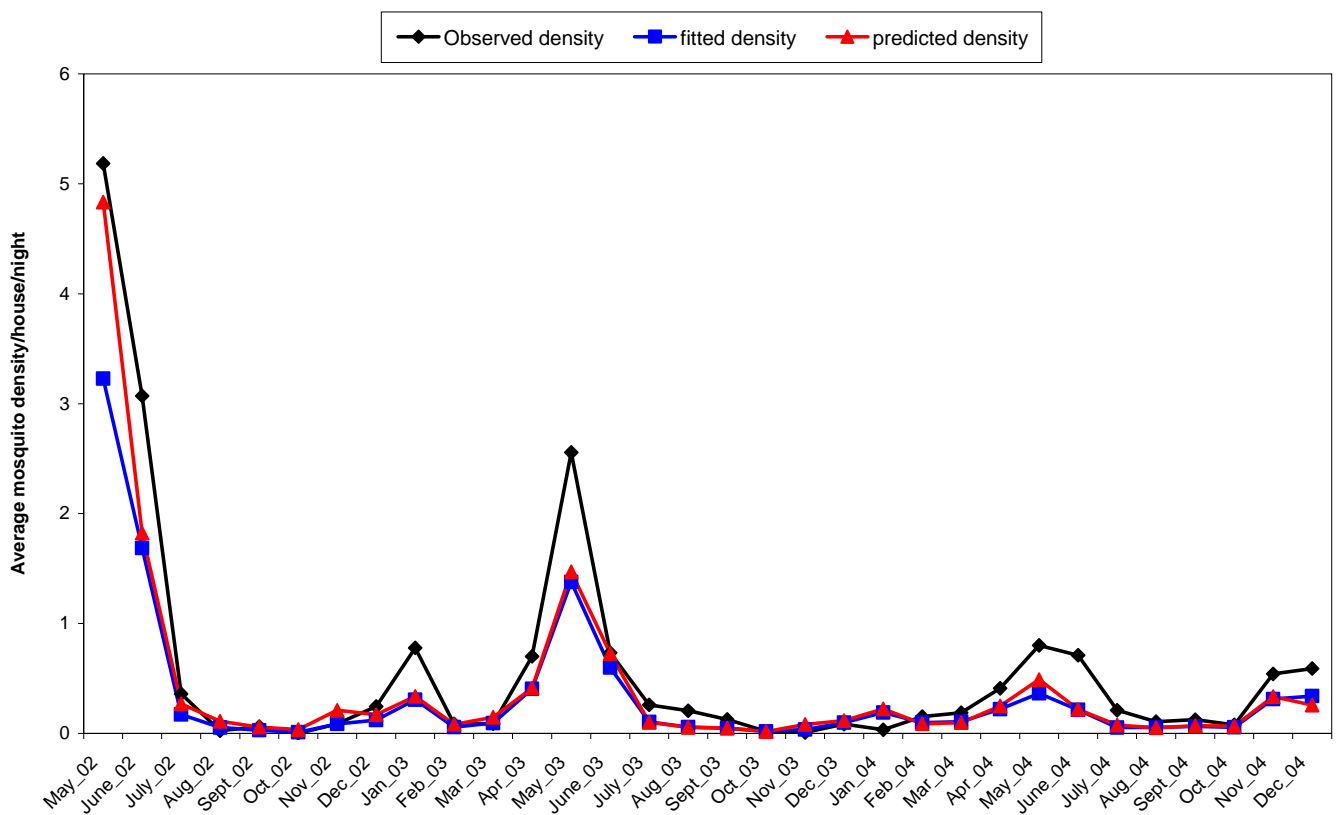


Figure 3.3 Monthly pattern of observed, fitted and predicted indoor residual densities of *Anopheles gambiae* mosquito

Model validation showed that 83% and 66% of the test locations had mosquito densities which were within the 95% credible intervals estimated from the zero inflated spatio-temporal negative binomial model and zero inflated spatial negative binomial model respectively. Furthermore, the zero inflated spatio-temporal negative binomial model consistently included the highest proportion of test locations in all the credible intervals compared to spatial negative binomial

model (Figure 3.4). Similar results were obtained using the mean square error measure (data not shown).

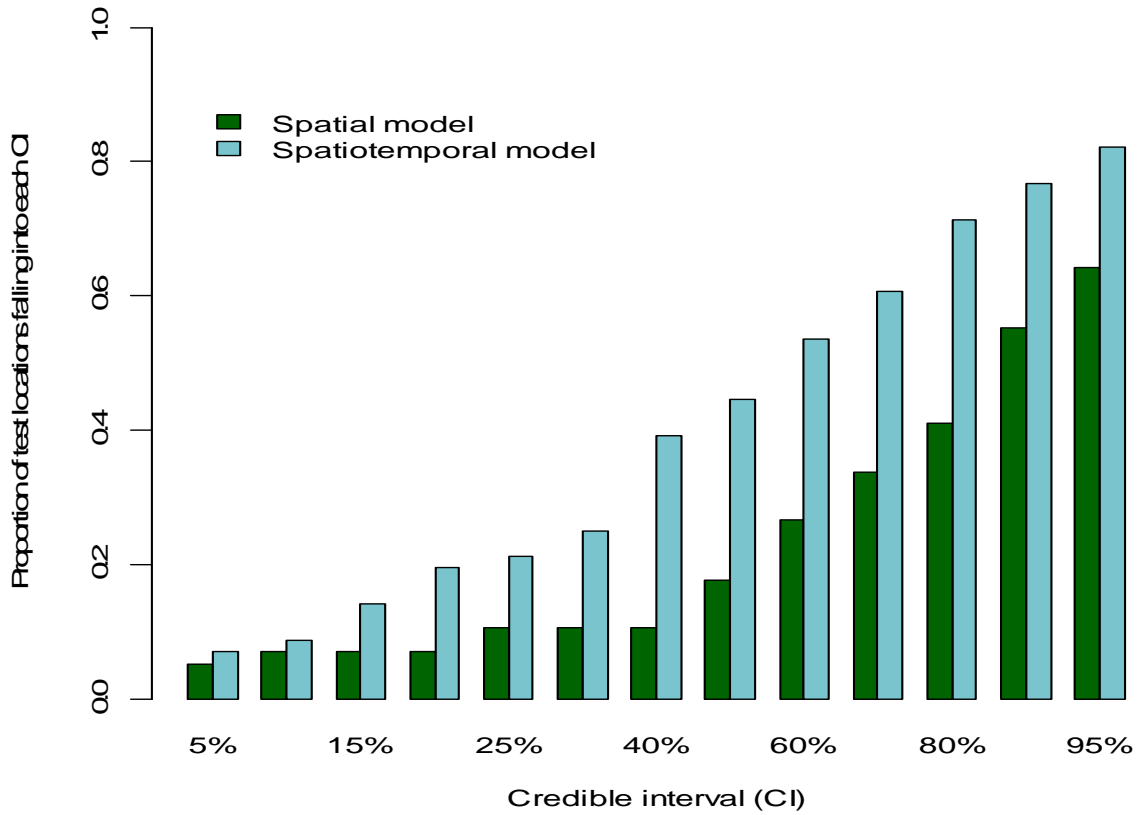


Figure 3.4 Proportion of test locations with non-zero mosquitoes falling in between 5% to 95% credible intervals of the posterior predictive distribution

The best fitting zero-inflated spatiotemporal model included the following parameters: distance to water bodies, elevation, average value of NDVI and LSTN during the month of mosquito collection, average LSTD during the current and the previous month of mosquito collection, total rainfall during the current and the two previous months of mosquito collection, year trend, trigonometric seasonality, spatial and temporal variations. The results of bivariate non-spatial and spatio-temporal zero-inflated negative binomial models are shown in Table 3.1 below.

Table 3.1 Posterior estimates of zero inflated geostatistical density models

Covariates	Bivariate non-spatial	Spatiotemporal model
	mean(95% CI)	Median (95% CI)
Intercept	-	4.634(0.005,7.098)
Distance to water body	-0.003(-0.006,0.001)	-0.007(-0.013,-0.002)
Altitude	0.002(-0.001,0.003)	-0.008(-0.041,0.020)
Rainfall ^{***}	0.006(0.005,0.008)	0.040(-0.041,0.113)
NDVI [*]	4.837(3.589,6.086)	4.170(1.308,6.725)
LSTD ^{**}	-0.139(-0.182,-0.096)	-0.246(-0.3752,-0.153)
LSTN [*]	-0.010(-0.065,0.085)	0.124(-0.031,0.234)
Year2	-0.276(-0.538,-0.013)	0.242(-0.356,0.852)
Year3	-0.404(-0.673,-0.135)	0.441(-0.244,1.122)
Cosine	0.642 (0.477,0.807)	1.75(0.570,2.913)
Sine	0.533 (0.364,0.701)	0.522(-0.597,1.590)
Amplitude	-	1.922(0.941,3.016)
Shift/phase	-	0.280(-0.291,1.033)
Over dispersion value	-	0.705(0.502,1.135)
Spatial Variation	-	0.874(0.516,1.417)
Temporal variation	-	0.322(0.140,0.898)
Range (3/ρ) ^a	-	3.039(1.337,6.482)
Zero-Inflated proportion	-	0.074(0.004,0.200)

a: minimum distance in kilometers at which spatial correlation is significant at 5%, CI= credible interval, *: environmental average value of the current month of mosquito collection, **: environmental average value of the current and previous month of mosquito collection, ***: environmental total value of the current and two previous months of mosquito collection.

Distance to water bodies, mean value of NDVI during the month of collection and average day temperature during the current and the previous month of collection were associated with mosquito density. In particular, distance to water bodies and average day temperature (LSTD) during the current and the previous month of mosquito collection were negatively related with mosquito density. Mean value of NDVI during the month of collection was positively associated with mosquito density. The average of the total rainfall during the current and two previous months of mosquito collection, mean night temperature (LSTN) during the month of collection and elevation were not associated with mosquito density. The minimum distance at which the spatial correlation was significant at 5% was 3.0 km (95% credible interval: 1.337, 6.482).

The 95% credible interval of the amplitude parameter revealed a strong monthly variation of mosquito density. The phase of 0.28 radians indicated that the maximum density occurred in the

months of May and the minimum in November. However, the average density during the second and third year was not strongly different than that of the first year.

3.3.2 Entomological inoculation rate

The overall point estimates of annual EIR were 6.7, 9.3 and 9.6 ibpy for the years 2002, 2003 and 2004 respectively. The estimates of EIR for this study were obtained exclusively from the *An. gambiae* mosquitoes because none of the *An. funestus* mosquitoes tested positive for the presence of *Plasmodium falciparum* sporozoite antigens. The estimates of EIR for 2002 are based on data from Asembo only since mosquito collection started in Gem in 2003. Gem had high EIR in both wet and dry seasons throughout the study period (Table 3.2 below).

Table 3.2 Distribution of EIR by area in relation to wet and dry months during study period

Area	2002		2003		2004	
	Wet	dry	Wet	Dry	Wet	Dry
Asembo	4.9	1.8	4.3	2.8	4.9	1.2
Gem	-	-	6.6	4.9	8.5	4.6

Figure 3.5 depicts the temporal pattern of observed and predicted EIR in relation to total monthly rainfall. Overall observed and predicted EIR display a similar trend during the study period. However, our model over-predicted EIR in May in 2002, May 2003 and June 2003. Monthly point estimates of EIR varied over the study period with highest inoculation rate of 4 ibpm (infectious bites per person month) occurring in May 2004 and lowest in October. Comparison between wet and dry months indicated that both the predicted and observed EIR are higher during the wet months (data not shown).

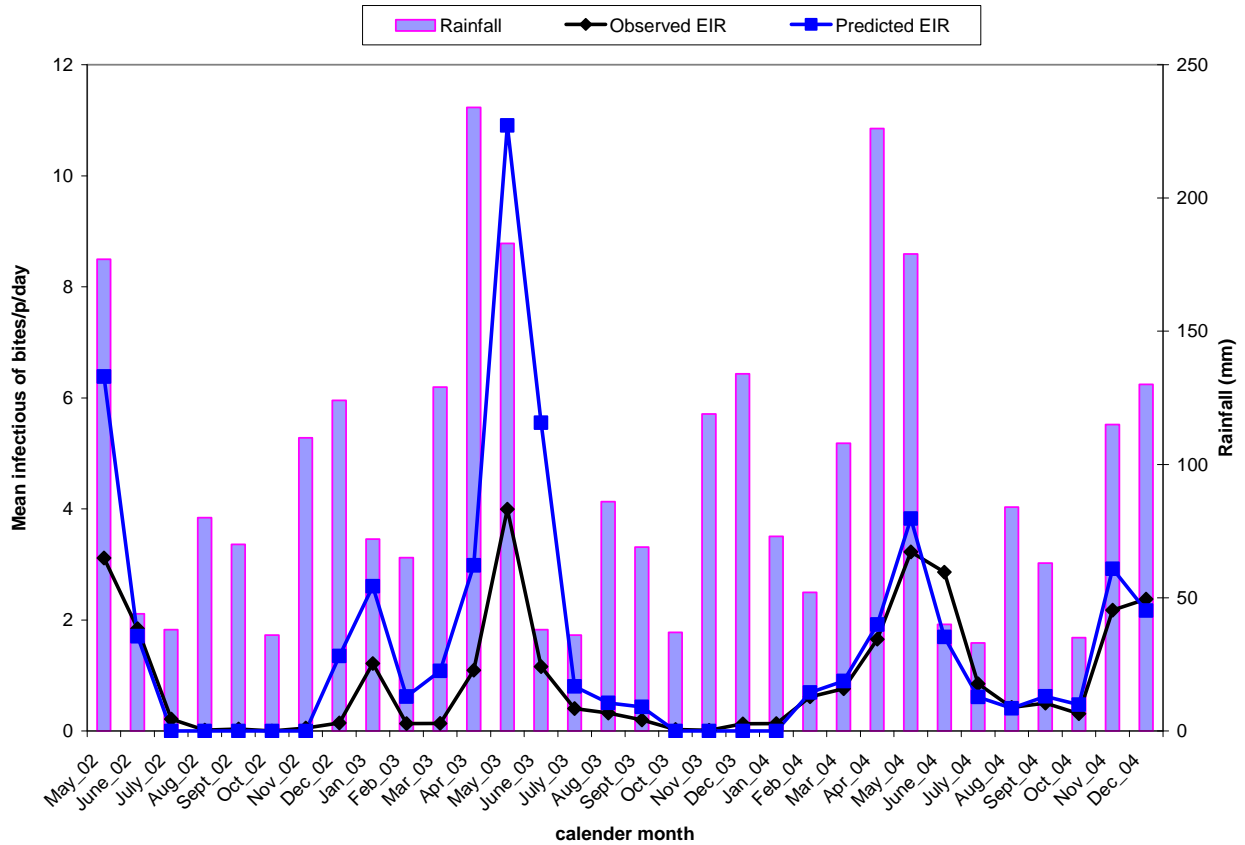
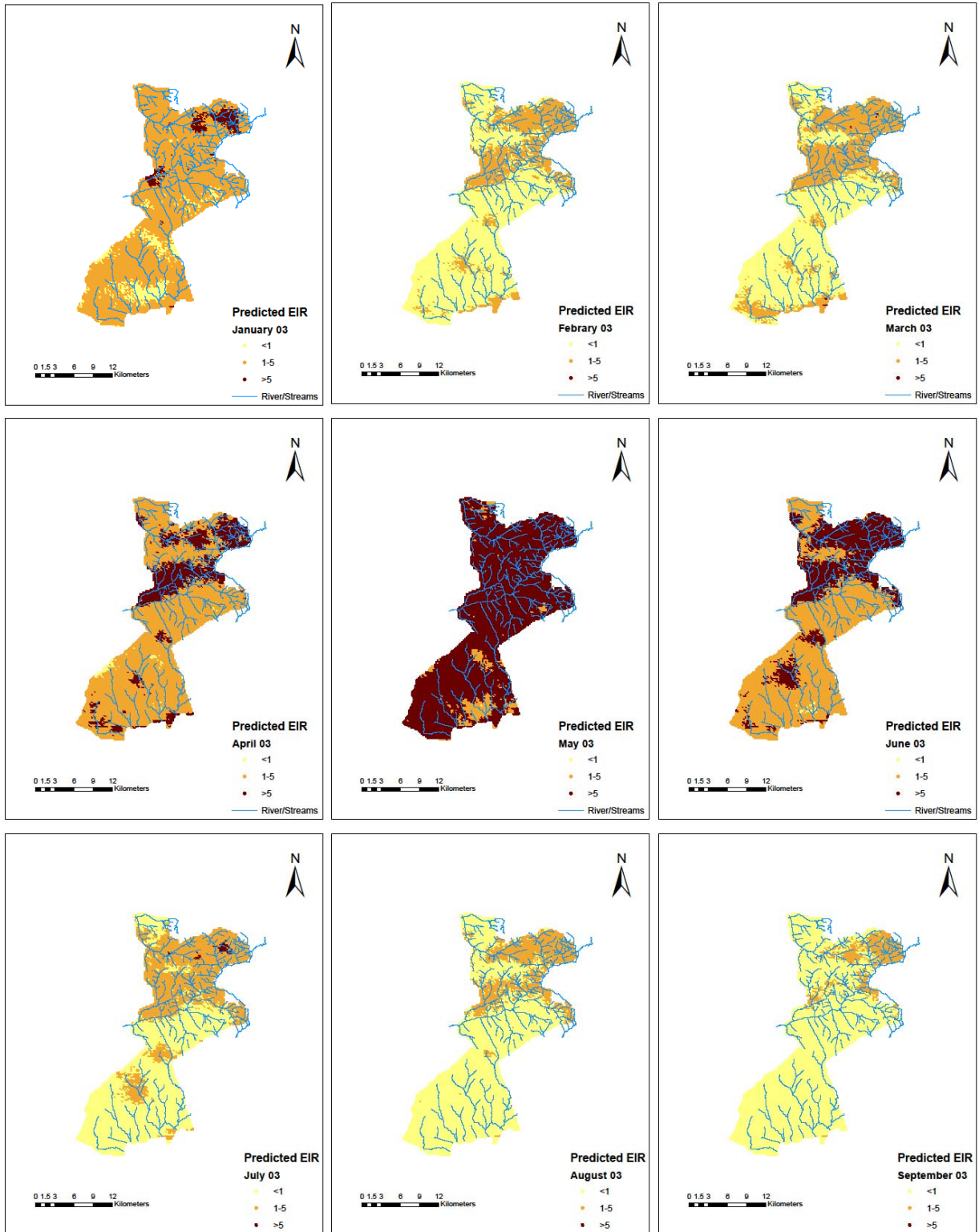


Figure 3.5 Temporal pattern of observed and predicted entomological inoculation rate in relation to rainfall during the study period

Smooth maps of monthly predicted malaria transmission 2003 are shown in figures 3.6 below. These predicted maps depict spatial variation within and between the months with areas of high predicted EIRs occurring in the northern part of the study area with a few locations with high predicted EIRs occurring during wet months in the Southern part of the study area. Similar trends were also observed in the maps for 2002 and 2004. Prediction error maps (not shown) were also produced.

Chapter 3: Dynamics of malaria transmission



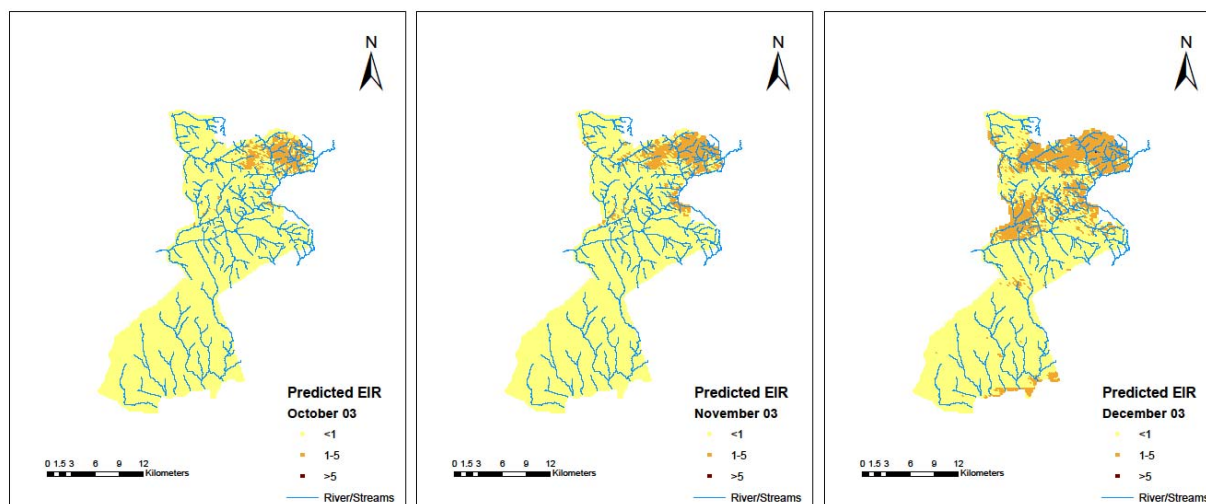


Figure 3.6 Predicted EIR maps

3.4 Discussion

In this study, we described and estimated malaria transmission patterns in the KEMRI/CDC HDSS site using mosquito density and entomological inoculation rate as a measure of malaria transmission intensity. Malaria transmission fluctuated over the months (see figures 3.2 & 3.5) in the HDSS with highest mosquito density/abundance and EIR occurring in May each year. Transmission intensity measured by EIR showed that residents in the HDSS were exposed to a range of 0-4 infective bites per month (see figure 3.5). The results also showed that *An. gambiae* is the main species driving transmission because the density and infectivity of *An. funestus* were very low. *An. gambiae* prefers temporary breeding sites which are common in the study area during the rainy seasons as opposed to *An. funestus* which are mainly found in permanent water bodies. A study (Gimnig et al., 2003) on the impact of ITNs on entomological indices in the same area also found a similar result (high density of *An. gambiae* compared to *An. funestus*).

The negative association between distance to water bodies and mosquito density in our results implies that many mosquitoes tend to be found close to the water bodies that act as the breeding sites. This probably applies to both newly emerged mosquitoes and adult mosquitoes that have limited dispersal ability. A study of geographic distribution of adult mosquitoes in the same area also found a significant relationship with water bodies identified in a GIS database during the dry season (Hightower et al., 1998). Although elevation was negatively associated with mosquito density, the relationship was not strong.

Chapter 3: Dynamics of malaria transmission

Temperature is an important factor related to mosquito development and survival and to the duration of the sporogonic cycle of the parasite (Onori and Grab, 1980). Temperatures above 22^o C are suitable for stable malaria transmission (Craig et al., 1999) and in our study, the average daily maximum temperature is about 29^o C. In our study, the average day temperature during the current and the previous month of mosquito collection had a strong negative effect on mosquito density.

NDVI, a proxy measure of vegetation was positively associated with mosquito density. The higher the NDVI value the greener the vegetation which is suitable for mosquito development.

The spatial correlation in mosquito density was strong at distances up to about 3 km (95% credible interval: 1.337, 6.482). However, a study by Midega and colleagues found a maximum distance of mosquito dispersal of only 0.7 km using a capture-recapture technique at the Kenyan coastal region (Midega et al., 2007). Mosquito dispersal is unlikely to explain this rather long distance correlation, which is probably due to unobserved/unmeasured spatially-correlated factors such as the spatial pattern of breeding sites and socioeconomic status.

The smooth maps generated in this study show that malaria transmission intensity in the HDSS varies over space and time, with high transmission occurring in a few pockets (hot spots). EIR peaks shortly after the onset of long rains in May of each year. Comparison between the study regions shows that EIR is consistently higher in Gem than Asembo which may be attributable to the occurrence of more rivers and streams in Gem that contribute to the creation of large numbers of mosquito breeding sites. Similarly substantial differences in the overall EIR between 2002 and 2003 could be due to earlier interventions in some parts of Asembo area (Gimnig et al., 2003).

Most analyses of mosquito sporozoite rate, density and EIR in relation to environmental/climatic factors and/or malaria incidence have been based on the assumption of independence between observations. However, mosquito's data are usually collected repeatedly over time at fixed geographical locations thus are spatially correlated due to common exposures. Similarly, mosquito density data are count data which are commonly analysed using the Poisson distribution. However, the Poisson distribution assumption implies that the mosquito average

equals the variance which is not always the case since entomological data usually have large number of locations with zero mosquitoes even in areas of high transmission. Our proposed Bayesian geostatistical zero-inflation model for assessing the relationship between mosquito density and environmental/climatic factors takes into account the underlying spatial processes and overdispersion associated with observed “excess zeros”. The model has a large number of parameters. However, simulation-based Bayesian computation allows simultaneous estimation of all parameters including the error of the location-specific predictions, a feature missing in the maximum likelihood based framework.

Our work used variable selection method based on standard models. Geostatistical variable selection has been applied in malaria epidemiology (Giardina et al 2012). However, this method could not be employed in our data which was collected over large locations. We are currently developing methodology to address this problem.

The maps of EIR produced in this study provide a high resolution exposure surface which is useful in analyzing the impact of transmission on malaria related and all-cause morbidity and mortality. At the same time, these maps help us understand the variability in malaria epidemiology over small areas.

Acknowledgments

We are grateful to the KEMRI/CDC HDSS for providing the entomological data and to the MTIMBA principal investigators for conceiving the project. The analysis of the data was supported by Swiss National Science Foundation (Project Nr. 325200_118379). The KEMRI/CDC HDSS is a member of the INDEPTH network.

The findings and conclusions in this study are those of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention. This paper is published with the permission of the Director of the Kenya Medical Research Institute.

3.5 Appendix

3.5.1 Negative binomial and Zero-Inflated negative binomial models

Mosquito density data are typical overdispersed count data, thus modeled using the negative binomial model: Let Y_{it} be the number of mosquitoes at location i at time t , arising from a negative binomial distribution; $Y_{it} \sim NB(\mu_{it}, r)$ with mean μ_{it} and parameter r measuring the extra variation (overdispersion) in our data. To capture the excess zero values that cannot be accounted for by the overdispersion parameter r , we used the zero-inflated negative binomial (ZINB) model which is a mixture model with two components: one arising from the negative binomial distribution and the other corresponds to the excess zeros. That is

$$f(Y_{it} = y_{it}) \sim \begin{cases} 0 & \text{with probability } p_{it} \\ NB(\mu_{it}, r) & \text{with probability } 1 - p_{it} \end{cases} .$$

The ZINB density is given by

$$f(y_{it}) = (1 - p_{it}) \frac{(y_{it} + r - 1)!}{y_{it}!(r - 1)!} \left(\frac{r}{r + \mu_{it}} \right)^r \left(\frac{\mu_{it}}{r + \mu_{it}} \right)^{y_{it}}, \quad r > 0, \text{ and } y_{it} > 0$$

with the mean equal to $(1 - p_{it})\mu_{it}$ and the variance given by $\text{var}(Y_{it}) = (1 - p_{it}) \left(1 + \frac{\mu_{it}}{r} + p_{it}\mu_{it} \right) \mu_{it}$. The term p_{it} is the mixing proportion. The above model reduces to zero inflated Poisson distribution as $r \rightarrow \infty$. The environmental and seasonality factors, \underline{X}_{it} were modeled on the $\log(\mu_{it})$ scale of the mean of the outcome, that is $\log(\mu_{it}) = \underline{X}_{it}^T \underline{\beta}$, where $\underline{\beta}$ is the vector of regression coefficients.

3.5.2 Geostatistical zero inflated negative binomial model

Mosquito data used in our analysis are collected at fixed geographical locations, sharing common exposures such as environmental and climatic factors thus correlated in space. To take into account the spatial correlation, we introduce spatial correlation parameter by adding location-

specific random effect φ_i no the mean structure of the above model, that is $\log(\mu_{it}) = \underline{X}_{it}^T \underline{\beta} + \varphi_i$.

We further assume that random effects are parameters from a latent Gaussian spatial process with covariance matrix and model spatial correlation between any pair of locations as a function of their distance irrespective of the direction. We used an exponential correlation function, that is $\underline{\phi} = (\phi_1, \dots, \phi_n)^T \sim N(0, \Sigma)$, $\Sigma_{ij} = \sigma_1^2 \text{corr}(d_{ij}, \rho)$ where σ_1^2 is the spatial variation d_{ij} is the distance between location i and j , and ρ is the smoothing parameter, measuring the rate of correlation decay with increasing distance. The value $3/\rho$ estimates the minimum distance at which spatial correlation is less than 5% (Ecker and Gelfand, 1997).

In addition to the above spatial correlation, mosquitoes were collected monthly in different locations during the study period and thus correlated in time too. We model temporal correlation by introducing monthly random effects (ε_t) to the above model: $\log(\mu_{it}) = \underline{X}_{it}^T \underline{\beta} + \varphi_i + \varepsilon_{it}$ and modeled by autoregressive (AR) process of various orders. The deviance information criterion (DIC) (Spiegelhalter et al., 2002) was used to identify the best fitting order of the process, which was found to be one. Thus we considered that $\varepsilon_t \sim N(\gamma\varepsilon_{t-1}, \sigma_2^2)$ and $\varepsilon_1 \sim N(0, \sigma_2^2 / (1 - \gamma))$. The terms σ_2^2 and γ are the temporal variance and autocorrelation parameters respectively with $\gamma \in (-1, 1)$.

Model fit

Prior distributions of the above model parameters were adopted to complete the Bayesian model specification above. In particular, we choose non-informative Normal prior distribution for the $\underline{\beta}$ parameters with large variance, an inverse gamma priors for σ_1^2 and σ_2^2 , a gamma prior for ρ and a Uniform prior for γ , that is $\sigma_1^2, \sigma_2^2 \sim IG(2.01, 1.01)$, $\gamma \sim U(-1, 1)$ and $\rho \sim G(0.1, 0.1)$. We further consider a constant zero-inflated mixing proportion across the area and time with a Uniform prior distribution $\pi \sim U(0, 1)$. To estimate the model parameters, we employed Markov Chain Monte Carlo (MCMC) simulation algorithm (Gelfand and Smith, 1990) and starting with some initials values about the parameters, we run two chains sampler discarding the first 5000 iterations. Convergence was assessed by Gelman-Rubin diagnostic (Gelman, 1992). Using

Chapter 3: Dynamics of malaria transmission

Bayesian kriging (Diggle et al., 1998) method that is for each sample of the parameters from the posterior distribution, a random effect is simulated from the Gaussian spatial process conditional on the random effects estimated at the observed locations. This is added on the regression term relating the covariates at the new location with the regression coefficients estimated during the model fit. The resulting equation estimates the mosquito density on the logit scale at the new location as a sample from the posterior predictive distribution. A grid of 7726 pixels with 250 meters by 250 meters spatial resolution covering the entire study area was used to predict density.

The Bayesian model was fitted in OpenBUGS version 3.1.2 (Imperial College and Medical Research Council, London, UK). Bayesian Kriging was carried out in a code written by the authors in Fortran 95 (Digital Equipment Corporation) using standard numerical libraries (Numerical Algorithms Group Ltd).

Chapter 4: Using health and demographic surveillance system (HDSS) data to analyse spatio-temporal patterns of socio-economic status; an experience from KEMRI/CDC HDSS

Nyaguara Amek^{1,2,3}, Penelope Vounatsou^{2,3*}, Benson Obonyo^{1,2,3}, Mary Hamel^{1,4}, Frank Odhiambo¹, Laurence Slutsker⁴, Kayla Laserson¹

1. Kenya Medical Research Institute/Centers for Disease Control and Prevention (CDC)
Research and Public Health Collaboration, P.O. Box 1578 Kisumu, Kenya
2. Swiss Tropical and Public Health Institute, Socinstr. 57, P.O. Box, 4002 Basel, Switzerland
3. University of Basel, Petersplatz 1 P.O. Box 4003 Basel, Switzerland
4. Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta GA 30301, Georgia, USA

*Corresponding Author

This manuscript is prepared for submission to Global Health Action Journal

Abstract

Continuous monitoring in Health and Demographic Surveillance sites (HDSS) allows collection of longitudinal data demographic, health related and socio-economic indicators of the site population. We sought to use household survey data collected between 2002-2006 as part of the KEMRI/CDC HDSS site in Asembo and Gem, western Kenya to estimate socio-economic status (SES) in two areas and assess changes of SES over time and space. Data on household assets and characteristics, mainly source of drinking water, cooking fuel and occupation of household head was annually collected from 44,313 unique households during the study period. SES index was calculated as a weighted average of assets using weights generated via Principal Component Analysis (PCA), Polychoric PCA and Multiple Correspondence Analysis (MCA) methods applied to the pooled data. The index from the best method was used to rank households into SES quintiles and assess their transition over time across SES categories. Kriging was employed to produce SES maps at the start and the end of the study period. First component of PCA, Polychoric PCA and MCA accounted for 13.7%, 31.8% and 47.3% respectively of the total variance of all variables. The gap between the poorest and the least poor increased from 1% at the start to 6% at the end of the study period. Spatial analyze revealed that the increase in least poor households was centered in the lower part of study area (Asembo) over time. No major changes were observed in Gem. We conclude that DSS sites can provide a platform to assess spatial-temporal changes in the SES status of the population. Evidence on how SES varied over time and space within the same geographical area may provide a useful tool to design interventions in health and other areas that have close bearing with the SES of the population.

4.1 Introduction

Information on and measurements of households' standing and their transition across socioeconomic status (SES) is central to health research and policy interventions targeting the poorest in a given population. Of interest to health research and studies on welfare is the differential effects on health outcomes, and utilization of health care services based on SES.

Standard indicators of measuring SES are based on monetary values of household incomes and consumption spending. However, difficulties are encountered in collecting accurate income and expenditure information about households. Collecting accurate income data is particularly difficult in low income countries and in situations where households draw income from multiple sectors such as formal, informal and also through remittances. Collecting comprehensive and accurate information in those settings requires considerable resource and time (Montgomery et al., 2000). In addition, capturing income from economic activities such as in agriculture, which may be in-kind, may also lead to difficulties in attaching true value of a household's well-being (McKenzie, 2005). Consumption data may be easier to collect compared to income especially in rural areas. The time and resources required to collect complete and accurate consumption data can be extensive and costly (Filmer and Pritchett, 2001).

Due to limitations of obtaining complete and accurate data on income and expenditure to measure SES, alternative methods for measuring SES have been developed. These methods allow the use of proxies of living standards such as household ownership of assets (TV, radio, car among others) as well as access to clean water, electricity, sanitation facilities, and household characteristics obtained from household surveys (Montgomery et al., 2000; Filmer and Pritchett, 2001; Houweling et al., 2003; Schellenberg et al., 2003; Kolenikov and Angeles, 2005; Vyas and Kumaranayake, 2006; Booysen et al., 2008).

Until recently, Principal Component Analysis (PCA) has been the standard technique that uses asset-based variables which are mostly qualitative ordinal indicators to measure SES (World Bank, 2001). However, PCA was designed for a set of quantitative variables. Unlike PCA, Multiple Correspondence Analysis (MCA) is applicable for both quantitative and qualitative

Chapter 4: Spatio-temporal patterns of socio-economic status

variables, and more recently Polychoric PCA which is more suited for discrete variables. Kolenikov and Angeles (2005) described polychoric PCA as an improvement to ordinary PCA and MCA. In general, PCA and MCA are usually applied to the indicator matrix, a matrix of zeros and ones representing assets ownership or household characteristics (0 is no and 1 is yes).

The asset indices are calculated as a weighted sum of the binary indicators with weights corresponding to the factor loadings of the first principal component. Unlike ordinary PCA and MCA that creates binary variables for each category, Polychoric PCA treats variables as ordinal for each category (Kolenikov et al., 2004). However, it is unclear whether assets weights obtained by statistical methods characterize the wellbeing of a household or individual in the local setting in which the data were obtained.

In addition to measuring SES in a static sense, analyzing its trends and space-time interaction is of interest. Understanding who moved and how they moved out of poverty is of special interest from an economic development point of view, especially as countries seek to evaluate progress towards achieving the MDG goal of eradicating poverty.

Conceptual problems on how to measure poverty (SES) trends in the absence of longitudinal data and difficulties of tracking and collecting data on household mobility, and how that affects whether they stay or move out of poverty have precluded robust analysis of poverty transition (Dercon and Shapiro, 2007). However, most health and demographic surveillance system (HDSS) sites collect longitudinal data on household assets from surveys carried out in all households within the sites. This is a wealth of information that can be used to assess the impact of health interventions not only on health but also on well-being of a household. This allows the construct socio-economic asset indices comparable in space and time.

Since 2002, Kenya medical research institute in collaboration with centers for disease control (KEMRI/CDC) has through its HDSS, collected longitudinal data on health, demographic events and socioeconomic indicators of the population in Asembo and Gem in western Kenya with an intention to explore the relationship between households' socioeconomic positions and health outcomes. These data provide a strong foundation to evaluate poverty transition in time as well

as space. However, no analysis has been done on space time dynamics of poverty (SES). Nor has a systematic approach of ranking the households based on their SES characteristics been applied.

We sought to determine and evaluate changes in SES over time and in space using socio-economic indicators collected from the KEMRI/CDC HDSS site from 2002 to 2006. Different statistical techniques, namely ordinary PCA, polychoric PCA and MCA methods have been applied to the SES data to compare their strength, and precision in estimating an SES index. Geographical information system (GIS) data were utilized to obtain maps of socio-economic status over time.

4.2 Material and Methods

4.2.1 Study Site and Population

The KEMRI/CDC HDSS site is located in Asembo (Rarieda Division, Bondo District), Gem (Yala and Wagai Division, Siaya District) and Karemo division (Siya District) areas situated in rural Nyanza Province in Kenya. During the study period, the HDSS was only operating in Asembo and Gem areas which is approximately 500 km² with a population of 135,000 residing in 33,990 households within 23,525 compounds and is one of the most impoverished areas of Kenya. The residents are predominantly Luos, a Nilotic ethnic group who earn their living through subsistence farming and fishing (Odhiambo et al, 2012).

4.2.2 Data collection

Households registered in the KEMRICDC HDSS area are visited every four months to update the statues of all household residents and to record pregnancies, births, deaths, in and out migration. Information on socioeconomic status was being collected once in a year during the study period and geographical positions of the compounds are updated yearly to account for addition of new compounds.

In this study, socioeconomic data collected in all existing households during 2002 to 2006 were used. Unfortunately, there were no available data for Gem in 2002 since socioeconomic data collection started in 2003 in that region of the study area. The variables used to generate

household SES index included occupation of household head, primary source of drinking water, use of cooking fuel, ownership of in-house assets (lantern lamp, sofa, radio, bicycles and television) and livestock possessions (poultry, pigs, donkey, cattle, sheep and goats). Household survey instrument and information collection recommended by World Bank and used by Demographic and Health Surveys (DHS) to examine socioeconomic differences in health and nutrition in population are similar to what we applied in our study (World Bank, 2001).

4.2.3 Data management and analysis

Data forms were scanned using Cardiff teleform and subjected to logical checks to ensure compatibility with the exiting information. Inconsistent data were returned for correction. Data analysis was done using STATA 10.0 (Stata Corporation, college station, TX, USA) and Arc View 3.2a

The asset index was constructed using PCA, polychoric PCA and MCA. The index with the highest total variation of the variables used and a more symmetric distribution was used to rank households into SES quintiles (categories) with the first quintile representing the most poor households, second more poor, third poor, fourth less poor and the fifth being the least poor households. This index was also used to assess their yearly transition across SES quintiles. To make sure that this index was comparable over time, it was constructed using pooled weighted estimates from all the five years data.

We linked the quintiles to geo-locations of households and displayed them in Arc View 3.2a. Using spatial statistics, smooth maps of SES for each year were generated by applying the Kriging method.

We obtained asset weights from the first component/dimension each of PCA, polychoric PCA and MCA as it explained the largest variation in the variables used (Houweling et al., 2003; Vyas and Kumaranayake, 2006). We measured how households moved from one quintile to another over time, the ratio between the most poor and the least poor at the beginning and end of the study period and geographical distribution of the quintiles.

4.2.4 Ethical consideration

Informed written consent was obtained from the compound heads for participation of their households. The HDSS activities of which our study is part of were reviewed and approved by the institutional review board of both CDC (Atlanta, GA) and KEMRI (Nairobi, Kenya).

4.3 Results

There were 44,313 unique households during the study period (2002-2006) of which 44,094 (99.9%) households had complete data and are included in the analysis. Table 4.1 shows the list of all variables used, with the grouped categories for each variable in the second column and the weights for each variable in the third, fourth, fifth and sixth column obtained from ordinary PCA, polychoric PCA and MCA, respectively.

Ordinary PCA assigned the highest weight to owning a bicycle followed by owning a lantern lamp with the lowest weight for not owning a bicycle. Polychoric PCA assigned the highest weight to owning more than four sheep, followed by owning a donkey with the lowest weight for not owning a radio. MCA gave the highest weight to owning a donkey followed by owing a TV with the lowest weight for households not owning a radio.

Generally, in each analysis, variables with higher weights are associated with higher SES, and conversely variables with low weights are associated with lower SES. Thus all weights in Table 4.1 are in the expected direction. The weights in each method are not directly comparable as they are on different scales. However, the relative magnitude of weights across methods can be compared. For instance, the ratio of the weight assigned to owning a radio to the weight assigned to owning a TV is 1.3 in the ordinary PCA index, 0.028 in the polychoric PCA index and 0.26 in the MCA index.

Table 4.1 Variables included in and weights obtained from first component of different techniques

Variable	Categories	Weights		
		Ordinary PCA	Polychoric PCA	MCA weights
Goats	Owns >= 4 goats	0.1588	0.4779	2.9045
	Owns < 4 goats	0.1032	0.2394	1.3741
Cattle	No goat	-0.1961	-0.1340	-0.7942
	Owns >= 5 cattle	0.1750	0.5871	3.4211
	Owns < 5 cattle	0.1409	0.2542	1.4677
Sheep	No cattle	-0.2364	-0.1943	-1.1316
	Owns >= 4 sheep	0.1465	0.6200	4.1411
	Owns < 4 sheep	0.1231	0.4000	2.4808
Poultry	No sheep	-0.1928	-0.0763	-0.4866
	Owns >= 7 poultry	0.1740	0.2771	1.6441
	Owns < 7 poultry	-0.0673	-0.0417	-0.4194
Donkey	No poultry	-0.1208	-0.3519	-1.6117
	Owns a donkey	0.0979	0.6145	4.4713
Pigs	No donkey	-0.0979	-0.1271	-0.0858
	Owns a pig	0.0469	0.3470	2.2691
Radio	No pig	-0.0469	-0.0065	-0.0388
	Owns a radio	0.2470	0.1696	1.0867
Bicycle	No radio	-0.2470	-0.3543	-2.2453
	Owns a bicycle	0.2818	0.3061	1.9359
Sofa	No bicycle	-0.2818	-0.2611	-1.6411
	Owns a sofa	0.2643	0.2923	1.8672
Lantern lamp	No sofa	-0.2643	-0.2337	-1.4963
	Owns a lantern	0.2779	0.2234	1.4287
TV	No lantern	-0.2779	-0.3400	-2.1616
	Owns a TV	0.1846	0.6067	4.2245
Cooking Fuel	No TV	-0.1846	-0.0475	-0.3227
	Charcoal/firewood	-0.0144	-0.0002	-0.0073
Drinking water source	Gas/ paraffin(kerosene)	0.0144	0.0248	1.1492
	Borehole/well	0.0319	0.0701	0.4587
	Lake/pond/river	-0.0780	-0.0713	-0.2702
	Rainfall	0.0834	0.0912	1.9795
Occupation (Household head)	Other	0.0069	0.0069	0.9282
	Doing business	0.0419	0.1387	0.7084
	Cmmercial farming	0.0168	0.1049	1.6506
	Housewife	-0.0104	-0.3318	-0.5016
	Salaried worker	0.0888	0.2737	2.4279
	Skilled labor	0.0158	0.1964	0.4966
	Unskilled labor	-0.0413	-0.2228	-0.1850
	Subsistence farming	-0.0449	-0.0197	-1.0474
Other	-0.0293	-0.1705	-1.0156	

Chapter 4: Spatio-temporal patterns of socio-economic status

Ordinary PCA, polychoric PCA and MCA yielded 40, 14 and 22 components/dimensions respectively. The first component of each analysis accounted for 13.7%, 31.8% and 47.3% , respectively of the variance. Histograms of SES indices obtained from ordinary PCA, polychoric PCA and MCA are shown in Figure 4.1, 4.2 and 4.3 respectively. None of the techniques revealed a symmetric distribution. However, Ordinary PCA and MCA indices are more skewed than the Polychoric PCA index.

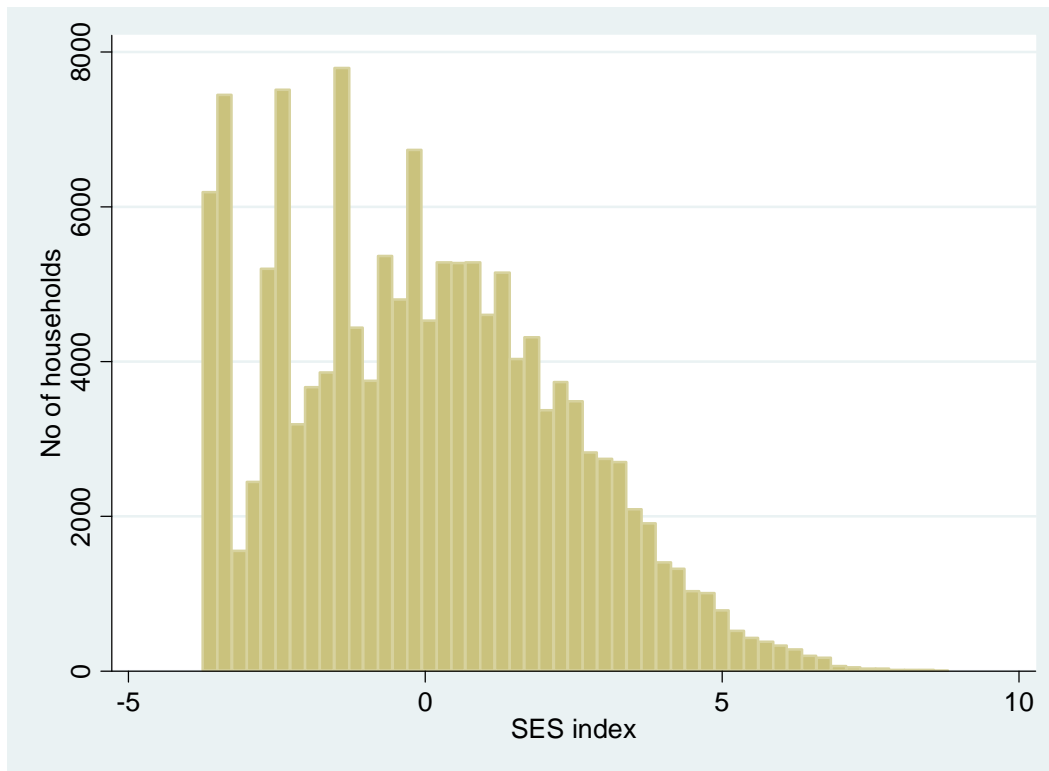


Figure 4.1 Histogram of SES index obtained by Ordinary PCA

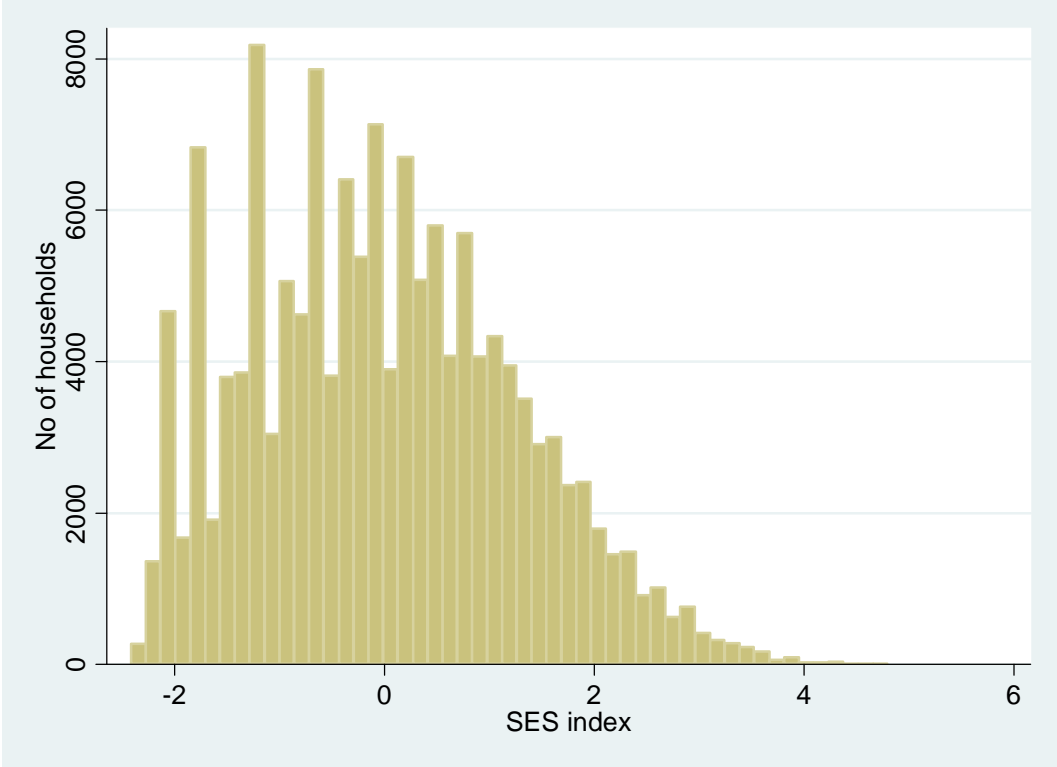


Figure 4.2 Histogram of SES index obtained by Polychoric PCA

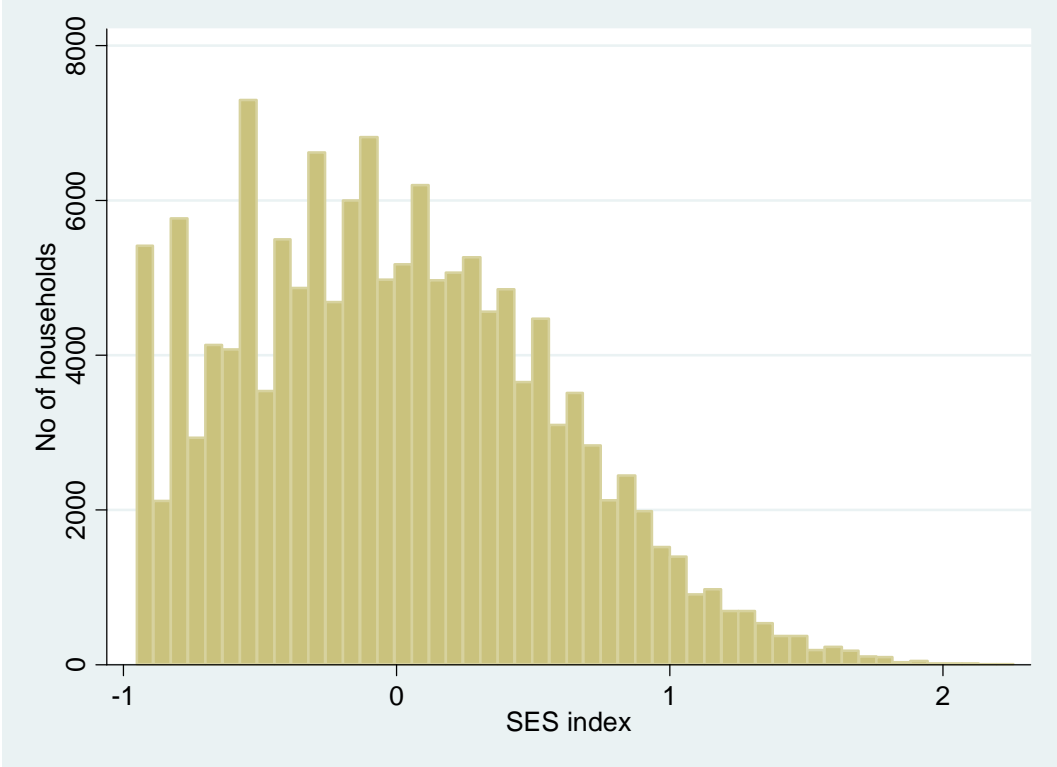


Figure 4.3 Histogram of SES index obtained by MCA

Chapter 4: Spatio-temporal patterns of socio-economic status

All indices (from the three techniques) show many households with same or similar weights at the left side of the histograms which is a common problem (clumping) with the SES indices due to a large proportion of households having similar assets, source of cooking fuel and drinking water resulting in small differences by which one can distinguish SES quintiles.

Ordinary PCA asset index was highly and statistically significantly correlated with the index based on MCA ($r = 0.997$, $p < 0.01$), and somewhat less with the index based on polychoric PCA ($r = 0.991$, $p < 0.01$). Similarly, Polychoric PCA was highly and statistically significantly correlated with the index based on MCA ($r = 0.995$, $p < 0.01$). Although, the three techniques did not place households in the same quintiles, comparisons showed quintile deviations restricted to one down or up of the total observed households. For instance, comparison between Ordinary PCA and MCA shows that 93% of households were placed in the same quintile by these methods, 87% of the households by Ordinary PCA and Polychoric, and 91% by Polychoric and MCA.

With all three techniques (results available on request), the proportion of assets owned by the least poor households are much higher than the ones owned by the poorest households. For example, in Ordinary PCA and MCA, 92% of those in the least poor quintile owned a bicycle compared to 3% of those in the poorest quintile. A similar trend was observed in MCA (Table 4.2 below). Over 80% of the poorest households tend to get their drinking water from the lakes or ponds or river as compared to least poor households who tend to get their water from rainfall (15%) as measured by all the techniques. A household headed by a salaried worker or business person tends to be in the least poor quintile than those headed by an unskilled laborer, housewife, or subsistence farmer. It is interesting to note that the presence of a salaried household head and possession of any asset are highly correlated ($p < 0.001$), with the highest correlation being with owning four goats or more.

Household quintiles were generated from MCA index given that the technique (MCA) gave the highest percentage (47.3%) of the total variation of the variables used and it is suitable for both quantitative and qualitative variables.

Table 4.2 Percentage of households owning assets by quintile of MCA-based SES index

Variable	Categories	% of Assets/characteristics in each quintiles				
		1st	2nd	3rd	4th	5th
Goats	Owens >= 4 goats	0.18	2.54	5.33	12.01	33.00
Cattle	Owens >= 5 cattles	0.08	1.35	2.79	8.46	34.31
Sheep	Owens >= 4 sheeps	0.01	0.30	0.91	2.94	19.44
Poultry	Owens >= 7 poultry	5.42	17.46	30.04	40.31	61.46
Donkey	Owens a donkey	0.01	0.26	0.56	1.20	7.33
Pigs	Owens a pig	0.17	0.83	1.34	2.12	3.90
Radio	Owens a radio	20.57	54.00	79.05	88.61	96.01
Bicycle	Owens a bicycle	2.84	19.67	44.03	71.72	91.68
Sofa	Owens a sofa	5.06	19.81	43.51	65.06	88.46
Lantern lamp	Owens a lantern	3.89	43.33	70.98	87.26	96.28
TV	Owens a TV	0.01	0.25	0.87	5.42	28.99
Cooking Fuel	Gas/ paraffin	0.19	0.45	0.80	0.93	0.81
Drinking water source	Borehole/well	11.49	16.52	15.01	19.66	18.35
	Lake/pond/river	86.79	79.61	79.79	73.06	65.88
	Rainfall	1.59	3.64	5.02	7.06	15.43
	Other	0.12	0.23	0.18	0.22	0.35
Occupation (Household head)	Doing business	5.49	12.53	11.82	16.50	15.39
	Commercial farming	0.07	0.31	0.34	0.44	0.89
	Housewife	2.36	1.86	1.81	1.49	1.08
	Salaried worker	0.52	1.78	3.34	5.96	13.92
	Skilled labor	1.83	4.15	4.07	5.56	4.05
	Unskilled labor	11.35	9.48	5.98	5.12	2.53
	Subsistence farming	72.15	66.68	70.37	62.68	60.17
Other	6.23	3.21	2.27	2.26	1.96	

Note: 1st=poorest, 5th=least poor

Figure 4.4 shows changes in SES quintiles over the years for all households. The number of the most poor (1st quintile) households decreases gradually over the study period while the number of the second most poor (2nd quintile) households increases in the first year of study and decreases steadily between third and fourth year of study period. The number of less poor (4th quintile) households increases throughout the study period while the number of least poor (5th quintile) households decreases in the first year and thereafter increases throughout the study period. In the year 2002, 22.45% of the total households were in the most poor quintile as compared to the least poor quintile (19.17%) in the same year. In the year 2003, 21.15% of the total households were most poor as compared to least poor (17.57%). However, at the end of study period, 18.88% of the total households were most poor as compared to least poor (22.25%).

Overall, there was a decrease by 4% of the number of most poor households and an increase by 3% of the number of least poor households during the study period.

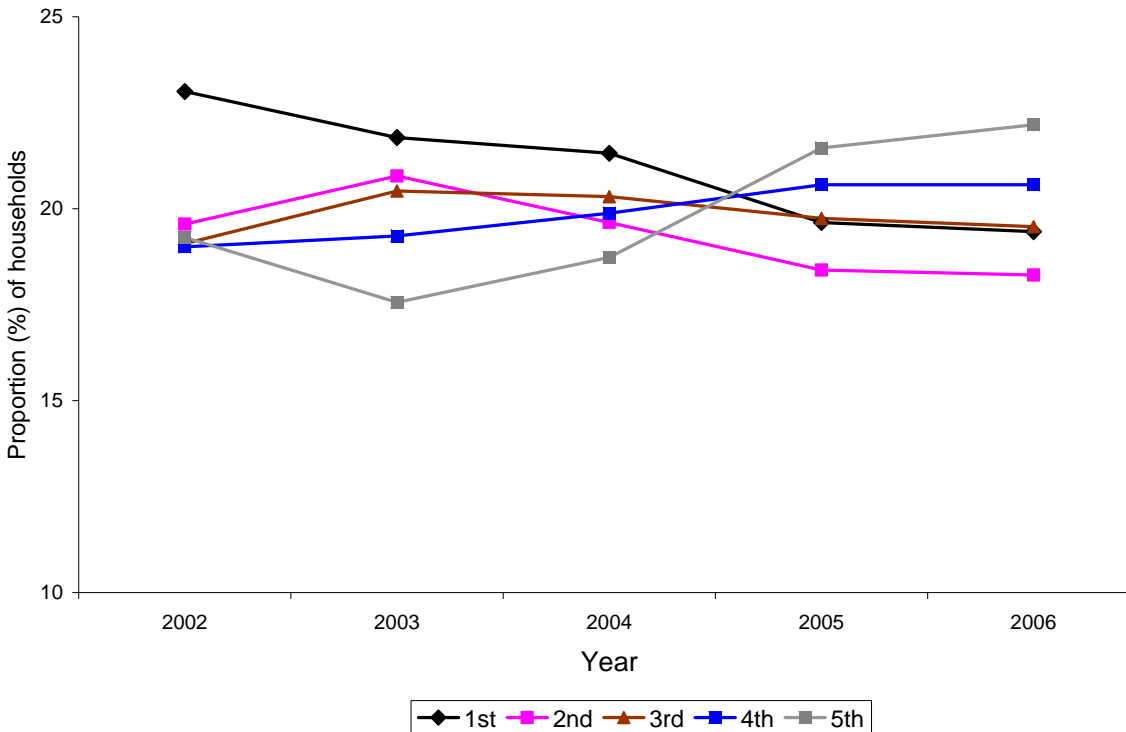


Figure 4.4 Distribution of all households in each MCA quintile by year

Cohort households

There were 42,998 unique households from 2003 to 2006 of which 46% (19,261) households were observed consecutively for four years (2003-2006). Few households from this cohort acquired assets during the period. For instance the number of households that had a radio, a bicycle, a TV and a sofa increased by 2%, 3%, 2% and 9% respectively during the four years. While goats (≥ 4), cattle (≥ 5) and sheep (≥ 4) ownership increased by 3%, 2% and 1% respectively for the first three years, households headed by business persons decreased by 5% over the four years.

Figure 4.5 show the change in cohort household quintiles over time. In 2003, 18.17% of the total households were most poor as compared 20.69% being least poor. However, at the end of study period, 18.33% of the total households were most poor as compared to 24.92% being least poor. Thus there was no significant change (0.2%) on the number of most poor households, but

Chapter 4: Spatio-temporal patterns of socio-economic status

households in more and poor categories decreased by 2% each while number of least poor households increased by 4% over the study period. These trends shows that some households in the lower quintiles acquired more assets thus moving into higher quintile as shown by the positive and negative change in the slopes. Further analysis on the individual households indicated that socioeconomic status of 31%, 26% and 43% households improved, declined and never changed during the period, respectively.

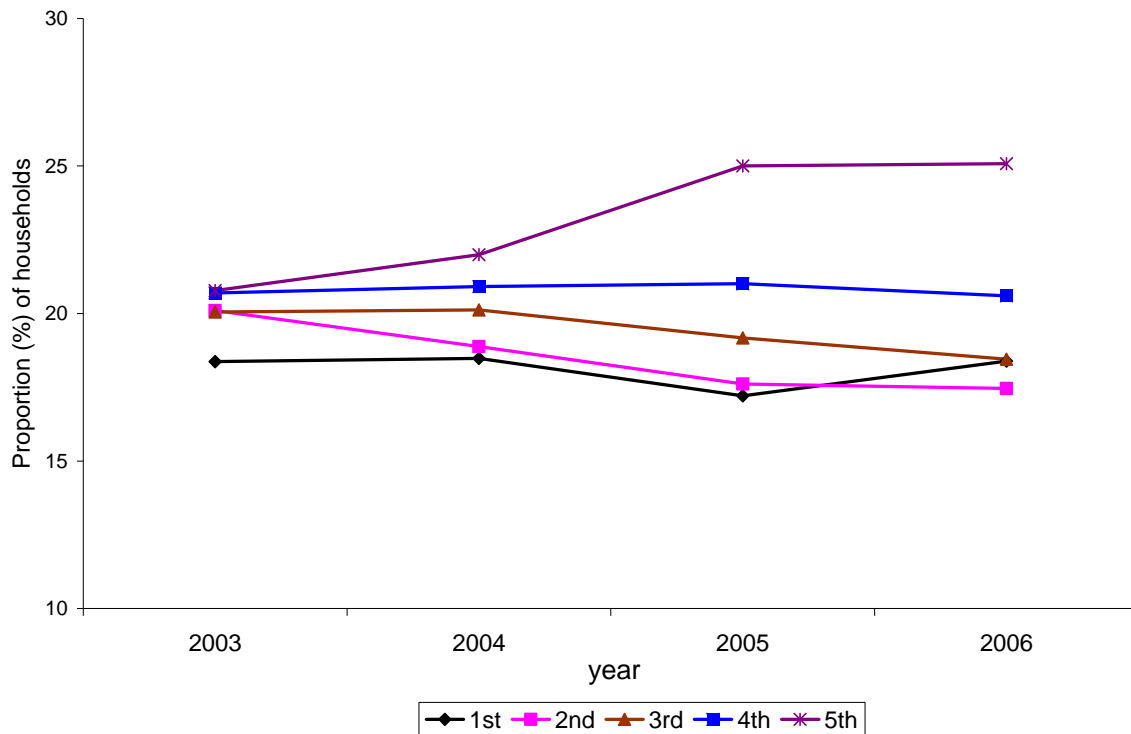


Figure 4.5 Distribution of cohort household in each MCA quintile by year

Figure 4.6 shows the spatial and temporal distribution of SES quintiles for the cohort households for each year of the study period. Overall, the lower part of the study area (Asembo) had a higher number of least poor households as compared to the upper part of the study area (Gem). There is a clear increase in the number of least poor household over the study period.

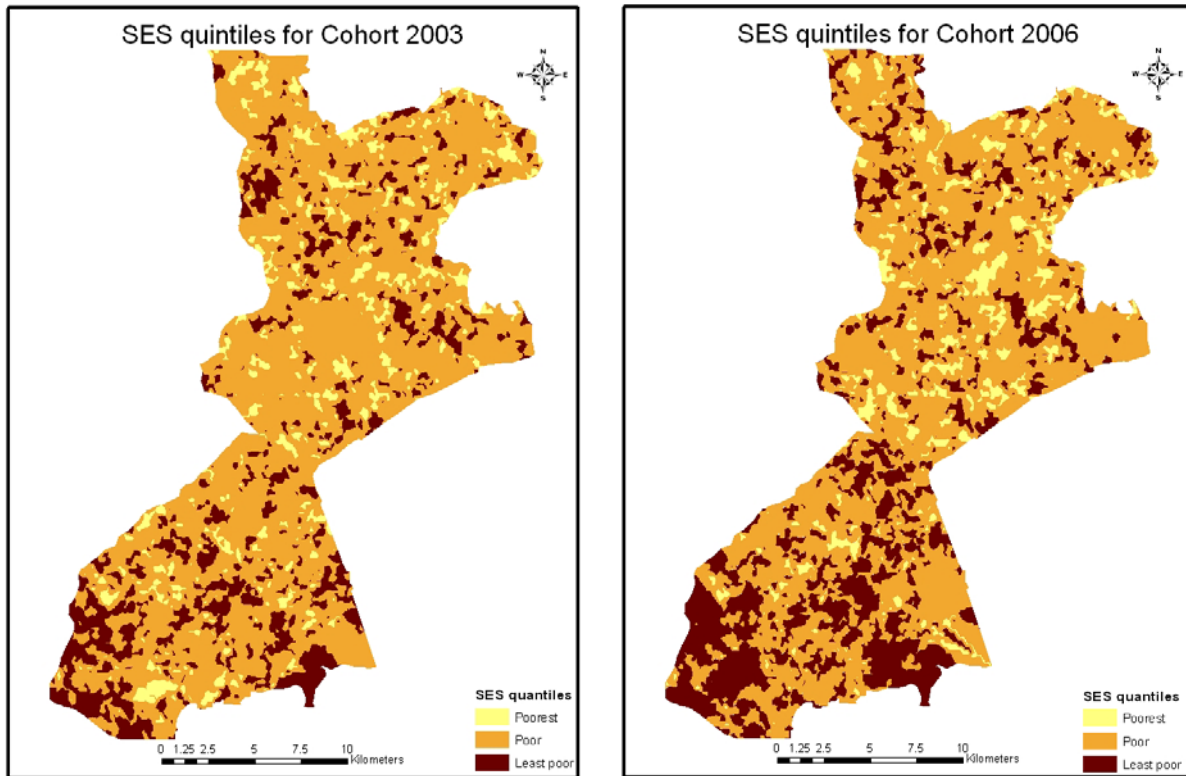


Figure 4.6 Spatial and temporal distribution of SES quintiles

4.4 Discussion

The objective of this analysis was to determine and evaluate changes in SES over time by applying the best method that provides reliable weights corresponding to the socio-economic values of assets in the local setting. By using the three methods (Ordinary PCA, Polychoric PCA and MCA) to obtain weights of index components, we have shown their relative strengths and differences in generating scores for ranking households into SES with MCA being a better method among the three methods.

The proportion of households owning assets grouped by quintiles generated by MCA reveals additional interesting results (table 4.2). In general the poorest households are below average in most assets to which the least poor have access. This is support the notion that well-off households tends to own more assets that tend to distinguish them from the poorest households.

Chapter 4: Spatio-temporal patterns of socio-economic status

The significant correlation between salaried household head with ownership of assets which is a proxy measure of household socio-economic status revealed that creating formal employment may reduce or alleviate poverty in the community.

The study results from any of the three methods also show a general increase in the number of households acquiring assets over time that reflected an increasing higher standard of living with households moving from lower quintiles to higher quintiles (figure 4.4 and 4.5). Overall, the gap between the poorest and the least poor is 3.80 % at the start of the study and declined to 2.79% at the end of the study period. By contrast, the analysis of the cohort households revealed that the gaps between the least poor and poorest are 1% and 6% at the start and end of the four years, respectively. The HDSSs continuously monitor health and demographic changes of a defined population over time through longitudinal data collection thus these platforms provide an avenue for detecting annual dynamics/changes in SES within a population.

The spatial distribution of SES (Figure 4.6) revealed a concentration of many least poor household residing in the lower part of the study area (Asembo). We could not establish the reasons for the spatial distribution of the SES. Though by speculation, in 2003 to early 2005 there was construction of a road through Asembo, as a result, this opened the area and residents may have been more likely to buy assets like TVs, bicycles and radios among others that could change their SES since most of them were employed by the road construction company. However, the exact cause of the difference in SES distribution is a matter for further investigation. Such analysis should also explore if there is any difference in health outcomes associated with SES in the study area. Currently we are in the process of modeling the relationship between mortality and malaria transmission in space and time to evaluate ongoing interventions in the study area of which SES is a risk factor.

Chapter 5: Infant and child mortality in relation to malaria transmission in rural western Kenya

Nyaguara Amek^{1,2,3}, Mary Hamel^{1,4}, Kim A Lindblade⁴, Kayla F. Laserson¹, Laurence Slutsker⁴, Thomas Smith^{2,3} and Penelope Vounatsou^{2,3*}

1. Kenya Medical Research Institute/Centers for Disease Control and Prevention (CDC)
Research and Public Health Collaboration, P.O. Box 1578 Kisumu, Kenya
2. Swiss Tropical and Public Health Institute, Socinstr. 57, P.O. Box, 4002 Basel, Switzerland
3. University of Basel, Petersplatz 1 P.O. Box 4003 Basel, Switzerland
4. Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta GA 30301, Georgia, USA

*Corresponding Author

This manuscript is prepared for submission to PLoS ONE

Abstract

Mortality in under-five year old children remains a major public health problem in Africa with malaria being one of the leading causes of death in this age group. Many approaches to malaria control involve reducing transmission intensity but little is known of how much mortality is reduced by specific reductions in transmission. We assess the relationship between under-five mortality and malaria transmission intensity measured by entomological inoculation rate (EIR). Entomological data were collected in the KEMRI/CDC health and demographic surveillance system during 2002-2004. Time to death for each child was treated at monthly interval and Bayesian geostatistical Cox proportion hazard models with time-dependent covariates were fitted by employing logistic regression. The models included all-cause and malaria specific mortality, EIR, insecticide treated nets (ITN), socio-economic status (SES) and parameters describing space-time correlation. The overall under-five mortality rate was 237 (95% CI: 219, 255) deaths per 1000 live births during the study period with the highest mortality rate in post neonates (102 per 1000 live births). Eighty-one percent of the total deaths were assigned causes of death, with malaria assigned as the main cause of death in infants and children (aged 1-4years), whereas neonatal sepsis was the main cause of deaths in neonates. All-cause mortality decreased with age. EIR was significantly associated with all-cause and malaria specific mortality. Similarly, the use of ITN is associated with reduction in all-cause mortality in postneonates (hazard ratio: 0.75, 95% CI: 0.52, 0.97). No trend was observed in malaria specific mortality. Our results indicate that successful malaria intervention measures targeting malaria transmission intensity will not only reduce significantly malaria-diagnosed mortality but also mortality assigned to other causes, in under-five year old children in endemic areas.

5.1 Introduction

Under-five mortality still remains a major public health problem in Sub-Saharan Africa (SSA) despite recent dramatic decreases. Of the 8.8 million global annual deaths, about 50% occur in SSA. In Kenya, one in twelve children (84 per 1000 live births) dies before their fifth birthday (UNICEF, 2008). On a global scale, most under-five (childhood) deaths have been attributed to pneumonia, diarrhea, malaria, neonatal sepsis, malnutrition, preterm delivery and asphyxia at birth (Bryce et al., 2005). All these conditions/diseases are either preventable or treatable with minimum interventions (Black et al., 2003).

In SSA, malaria/or malaria associated conditions are thought to be the leading causes of death and morbidity in children (WHO malaria report, 2009). Understanding the relationship between mortality and malaria transmission intensity is important in developing effective and evaluating efficacy of malaria control programs thus achieving the Millennium Development goal for child survival. A number of studies (Smith. et al., 2001; Snow and Marsh, 2002; Gemperli et al., 2004; Rumisha et al., 2011a, 2011b) have discussed the relationship between mortality and malaria transmission intensity, but this relationship remains unclear. Assessment of the efficacy of these interventions to reduce mortality requires reliable/accurate information on all-cause and cause-specific mortality which is often lacking in SSA due to poor vital registration system (Mathers et al., 2005) and the fact that most children die at home without any contact with the health system.

In an analysis of published and unpublished data on all-cause child mortality rates and EIR across Africa, Smith and colleagues found that all-cause mortality is significantly associated with entomological inoculation rate (EIR) (which is a measure of transmission intensity) in infants but no clear trend was observed in children (12-59 months) (Smith. et al., 2001). However, the study used convenience sampling and only a small number of sites could be compared. It did not take into account confounding variables which might independently affect both malaria transmission and mortality, thus introducing potential ecological confounding. Other studies have attempted to allow for this, for example (Gemperli et al., 2004) linked the Mapping Malaria Risk in Africa (MARA) and Demographic and Health Surveys (DHS) datasets to identify factors related with geographical differences in infant mortality risk in Mali. This study showed that geographical distribution of malaria is not a major determinant of the pattern of all cause infant

mortality. However, these two datasets were not comparable in space and time, since MARA data are historical data collected over long period of time while DHS are surveys carried out at different seasons and include different age groups at various locations, thus confounders associated with misalignment of the two datasets were introduced. Similarly, a study by Rumisha and colleagues found no significant trends in the relationship between all-cause mortality in children less than 5 years old and malaria transmission intensity measured by EIR and insecticide-treated net adjusted for space-time variation (Rumisha et al., 2011b) using longitudinal data from Rufiji demographic surveillance system (HDSS).

More recent studies (Bhattarai et al., 2007; Okiro et al., 2007; Steketee et al., 2008; Murray et al., 2012) argue that scaling up of malaria interventions such as use of insecticide treated nets (ITN) and intermittent preventive treatment (IPT), have led to dramatic declines in the number of deaths and/or hospital admission attributable to malaria in African children, but the relationship between mortality and reducing malaria transmission remains unclear, and inferences about mortality impact have largely been based on extrapolating from the results of randomized trials.

Current estimates on the causes of death in SSA are mainly obtained using Verbal autopsy (VA) techniques (Greenwood et al., 1987; Garenne and Fontaine, 2006; Abdullah et al., 2007; Murray et al., 2012). VA entails interviewing the caregiver concerning circumstances that led to the deceased's death and clinical signs and symptoms during the terminal illness. The information collected is reviewed independently by two or more physicians to assign most probable cause of death. Usually concurrence by at least two physicians is required to assign cause of death in situations where more than two reviewers are involved. However, sensitivity and specificity in measuring some of the main causes of death such as malaria is poor (Snow et al., 1992; Todd et al., 1994; Quigley et al., 1996; Korenromp et al., 2003). For instance (Snow et al., 1992) compared hospital-based causes of death in children with the ones assigned by VA in the Coastal region of Kenya. The study showed that VA correctly identified (sensitivity) less than 50% of the actual malaria deaths. In malaria endemic areas, over reporting of malaria deaths is common. For example febrile illness with no other confirmed etiology is usually recorded as malaria (Abdullah et al., 2007). Malaria also shares symptoms with other diseases such as acute respiratory-tract

infection, meningitis in children which are often assigned as malaria using VA (Snow et al., 1992).

In this study, we analyze all-cause and malaria specific child mortality data in relation to malaria transmission intensity from KEMRI/CDC health and demographic surveillance system (HDSS) site, a member of INDEPTH network which participated in the INDEPTH malaria transmission intensity and mortality burden across Africa (MTIMBA) project (Kasasa et al. in preparation; Amek et al., 2011; Rumisha et al., 2012). MTIMBA aims to improve our understanding of the relationship between mortality and malaria transmission intensity by analyzing both between-site and within-site variation in these quantities in HDSSs. HDSS routinely monitor population dynamics and demographic trends and also use verbal autopsies to assign cause of death (INDEPTH network, 2002; Adazu et al., 2005). For a period of three years, the sites participating in MTIMBA also measured malaria exposure in the form of the Entomological Inoculation Rate (EIR), which is the product of vector biting rate and the proportion of mosquitoes with sporozoite rates in their salivary glands per month (Molineaux et al., 1988).

5.2 Methods

5.2.1 Study area and population

KEMRI/CDC health and demographic surveillance system (HDSS) is located in three regions namely Asembo (Rarieda Division, Bondo District), Gem (Yala and Wagai Divisions, Siaya District) and Karemo (Karemo, Division, Siaya District) in Nyanza Province, rural Western Kenya. During the study period, the HDSS operated in Asembo and Gem, an area of approximately 500 km² with a population of 135,000 living in 33,990 households in 21,477 compounds in 217 villages. The residents of the study area are predominantly from the Luo ethnic group, and derive their livelihood mainly from subsistence farming. This area is one of the most deprived in Kenya with over 66% of the inhabitants living below the poverty level (Krishna et al., 2004). The study area has high under-five mortality (Adazu et al., 2005) and malaria infection is holoendemic, mainly transmitted by *An.gambiae* s.l. (Bayoh et al., 2010). An insecticide-treated mosquito nets (ITN) trial conducted from 1996 to 2002 in the area reduced malaria transmission by 90% (Gimnig et al., 2003; Lindblade et al., 2004). However, despite the continued high prevalence of ITN use and a relatively low EIR of about seven infectious bites

per year (Adazu et al., 2005), malaria prevalence is still high and is thought to be the main cause of child mortality (Phillips-Howard et al., 2003b; Adazu et al., 2005). Below, we describe methods used to obtain cause-specific mortality data and explanatory variables.

5.2.2 Cause-specific mortality

The Verbal autopsy (VA) technique was used to assign cause of death within the study area (Adazu et al., 2005; van Eijk et al., 2008). VA interviews were conducted by trained workers using VA questionnaires. They interviewed the main caregiver on the signs and symptoms of the child's terminal illness and care seeking behavior during the illness. Information from these forms was independently reviewed by a panel of at most three clinical officers and used to assign a most probable cause of death (Adazu et al., 2005).

5.2.3 Socioeconomic status

The socioeconomic indicators routinely collected in the HDSS (2002-2004) were used to generate household socioeconomic index employing multiple correspondence analysis (MCA) technique. The analysis of socioeconomic indicators has been described elsewhere in details (Amek et al. in preparation). In brief, the variables used to obtain the SES quintiles included occupation of household head, primary source of drinking water, use of cooking fuel, ownership of in-house assets (lantern lamp, sofa, bicycle, radio and television) and livestock possessions (poultry, pigs, donkey, cattle, sheep and goats). SES index was calculated as a weighted average of the above indicators divided by the number of household members. Household indices were then grouped into five quintiles with the first quintile representing the poorest households followed by very poor, poor, less poor and the last household being least poor.

5.2.4 Entomological Inoculation rate (EIR) and ITN

Bayesian geostatistical zero inflated binomial and negative binomial predictive models were fitted to entomological data to obtain high resolution estimates of sporozoite rate and mosquito densities respectively, at monthly intervals over the study area and thus align in space and time the mortality data. The models included environmental and climatic factors extracted from satellite data, harmonic seasonal trends and parameters describing space-time correlation. Model-based predictions allowed estimation of SR and mosquito densities at 250 m by 250 m spatial resolution. These estimates were multiplied and taking into account a conversion factor to adjust

for vector collection bias between human bait catch and light trap collection techniques as described by Lines and colleagues (Lines et al., 1991) to obtain smooth EIR surfaces. Modeling details are provided by (Amek et al., 2011, 2012). ITN use data included in this study was from a one time survey carried within the study area in 2002 to access the ITN coverage.

5.3 Statistical analysis

The analyses included all under-five year old children who were residents between May 2002 and December 2004 as defined by HDSS residency rule (Adazu et al., 2005). These children were grouped into five categories namely neonates, post neonates and child (1-4 years old). Time at risk for each child was defined as the number of months that child was a resident during the study period and aged below 5 years old. The under-five mortality rate was calculated as the total number of children dying before their fifth birthday divided by the number of live births. The child (1-4 years old) mortality rate derived as the number of deaths among children between 1-4 years divide by the number of live births. The post neonatal mortality rate was obtained as the number of deaths among ≥ 29 days old children who died before their first birth day divided by the total number of live births and the neonatal mortality rate was also calculated as the number of deaths in live-born children less than 29 days old divided by the number of live births. To account for non-linearity between EIR and mortality, different transformations of EIR such as logarithmic, categorization and fractional polynomial functions of different orders were assessed. The Akaike's information criterion (AIC) (Akaike, 1974) was used to select the best transformation, which was found to be logarithm of EIR estimate a month previous to the mortality status outcome.

For analyzing the effects of risk factors, time to death for each child was treated as discrete with monthly intervals. Cox proportional hazard models were fitted using binary logistic regression (Singer and Willett, 1993; Manda and Meyer, 2005). The models included EIR estimates, SES quintiles, ITN use or age. The prediction error of the EIR estimate was introduced into the model as a measurement error-in the covariate. Exploratory analysis was carried out in STATA 10 (Stata Corporation US) to assess the bivariate relations with either all-cause and malaria specific mortality. All covariates that were significant at 15% significant level were further included into a Bayesian geostatistical spatiotemporal logistic regression models. Spatial correlation was

modeled via village-specific random effects, which are considered as latent observations of a spatial Gaussian process. Correlations between any pairs of village locations were considered as an exponential function of their distance, irrespective of direction and modeled by the variance covariance matrix of the process (Diggle et al., 1998). Temporal correlation was modeled by introducing monthly random effects into the spatial model, modeled via autoregressive processes of various orders.

Bayesian models were fitted in OpenBugs version 3.1.2 (Imperial College and Medical Research Council London, UK). A description of the Bayesian geostatistical formulation model is given in the appendix

5.4 Results

5.4.1 Descriptive statistics

During the study period, 32709 (under 5 years old) children were observed contributing 47170 person-years of time. There were 3,107 deaths among these children with 670, 1234 and 1203 occurring in 2002, 2003 and 2004 respectively. It is worth mentioning that the 2002 data was only for 8 months period. The median age at death was 11 months and 53% of the total deaths occur in infants. Twenty percent of the deaths occurred in poorest households compared to 18% in least poor households. Table 5.1 shows all-cause mortality rates per 1000 live births in each year of the study period. For instance, under-five mortality decreased from 240 (95% CI: 222-258) deaths per 1000 live births in 2003 to 239 (95% CI: 220, 256) deaths per 1000 live births in 2004. Infant mortality rates also decreased from 127 (95% CI: 115-139) deaths per 1000 live births in 2002 to 121 (95% CI: 109, 133) deaths per 1000 live births. However, child (1-4 years) mortality rate increased in the same period from 109 (95% CI: 99, 119) deaths in 2003 to 118 (95% CI: 107, 129) deaths per 1000 live births, in 2004. Similar trends were observed in each study region (Asembo and Gem).

Table 5.1 Childhood mortality rates per 1000 live births

Age category	Calendar year		
	2002 (95% CI)	2003 (95% CI)	2004 (95% CI)
neonate	22 (17, 27)	19 (15, 24)	21 (19, 26)
Post-neonate	105 (92, 118)	109 (99, 119)	102 (93, 111)
infants	127 (115, 139)	126 (114, 138)	121 (109, 133)
Child (1-4 years)	102 (92, 112)	109 (99, 119)	118 (107, 129)
Under five	228 (212, 244)	240 (222, 258)	239 (220, 256)

Cause of death was assigned to 81% of the total deaths. Figures 5.1 and 5.2 depict the main causes of death for infants and child (1-4 years). Malaria was the main cause of death followed by anemia then pneumonia in the two age groups. In neonates, sepsis and then prematurity were the leading causes of death (data not shown). A tendency for mortality to decrease over time was observed for all the main causes of death except for malaria, HIV and diarrhea.

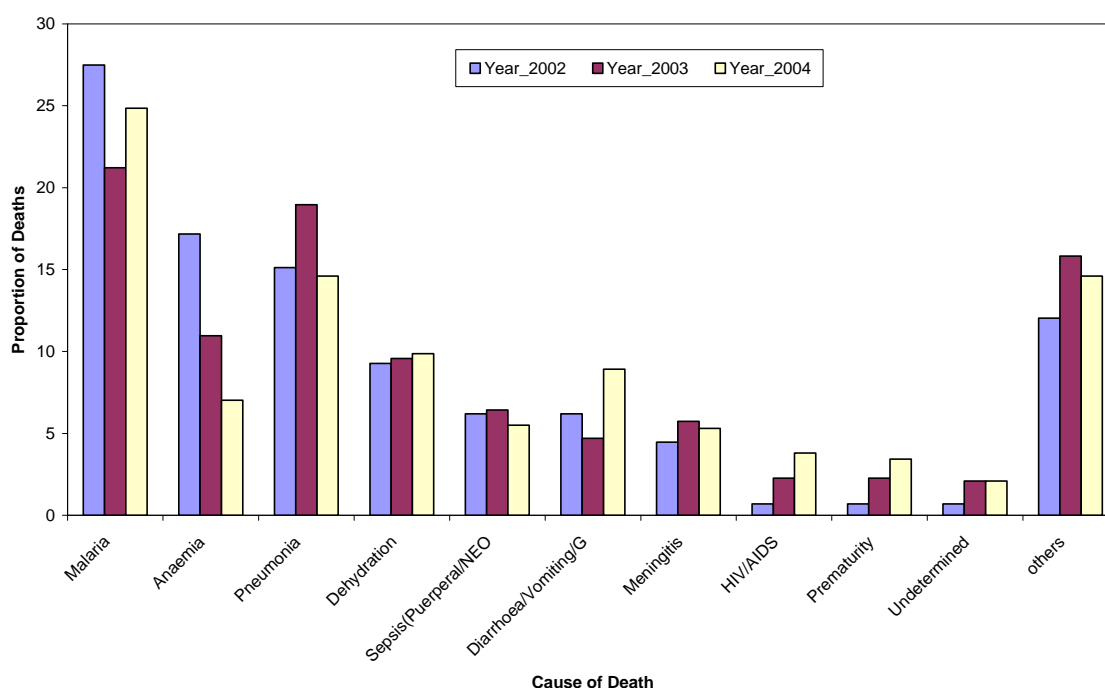


Figure 5.1 Main causes of death among infants

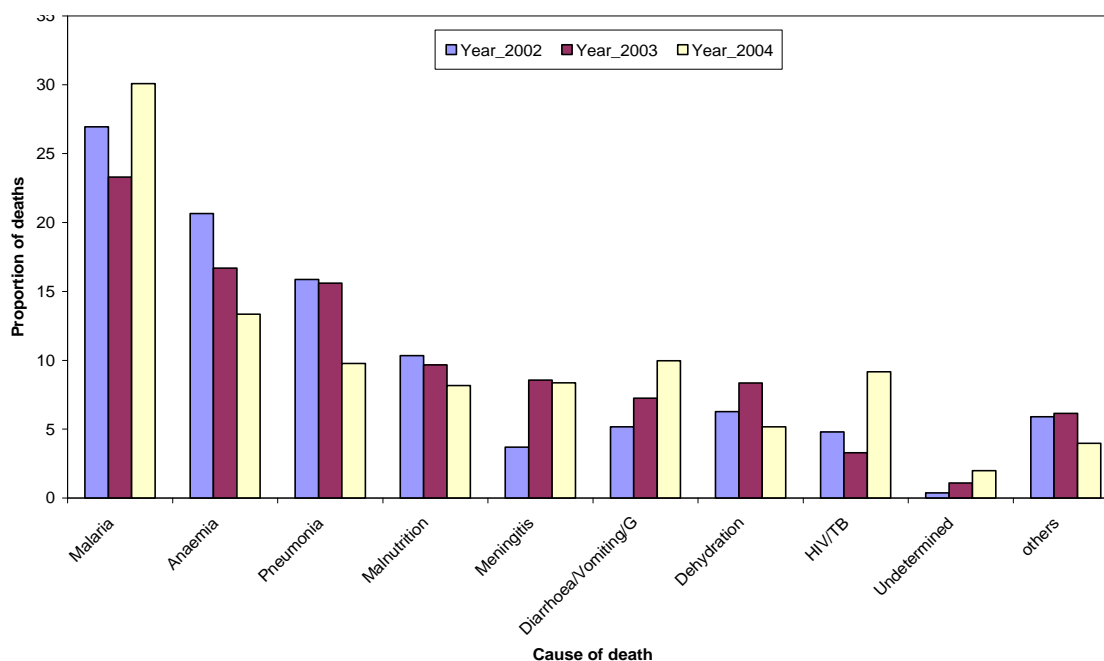


Figure 5.2 Main causes of death among 1-4years

5.4.2 Mode-based results

Table 5.2 presents the hazard ratio (HR) of predictors of all-cause and malaria specific mortality obtained from geostatistical spatiotemporal models adjusted for EIR and SES. The EIR was strongly associated with all-cause mortality in all age groups and the risk of dying from any illness was high in child (1-4 years) (HR= 1.58, 95% CI: 1.43, 1.74) compared to neonates (HR= 1.44, 1.30, 1.59, 95% CI: 1.27, 1.47) and post-neonates (HR= 1.34, 95% CI: 1.2). Older children had lower all-cause mortality in the first year of life, but there was no clear age trend after the first birthday

Results from malaria specific mortality models (Table 5.2) showed that malaria exposure is associated with malaria diagnosed mortality in all age groups, with post-neonates (1-11 months) experiencing the highest risk (HR= 1.60, 95% CI: 1.37, 1.82). Similarly age had a negative effect on malaria specific mortality in post-neonates, but not trend was observed in neonates and child (1-4 years). The estimated spatial correlation for both all-cause and malaria specific mortality was strong. Comparison between the all-cause and malaria specific models shows that spatial range from the latter model had lower spatial ranges with narrower confidence intervals. Similar trends were observed in models including EIR only (data not shown). Higher socioeconomic

Chapter 5: Infants and child mortality in relation to malaria transmission

quintiles were associated with reduction in all-cause mortality in all age groups but no significant effects were observed in relation to malaria specific mortality.

Table 5.2 Hazard ratio (HR) estimates of predictors of all-cause and malaria specific mortality for under-five age categories from spatiotemporal models

Covariate	All-cause mortality			Malaria specific mortality		
	Neonates HR (95% CI)	Post-neonate HR (95% CI)	Child(1-4yr HR (95% CI)	Neonates HR (95% CI)	Post-neonate HR (95% CI)	Child (1-4yr) HR (95% CI)
Constant	0.11 (0.07,0.18)	0.02 (0.01,0.04)	0.22 (0.10,0.65)	0.10 (0.01,0.40)	0.03 (0.01,0.10)	0.04 (0.01,0.09)
Age	0.96 (0.94,0.99)	0.94 (0.93,0.95)	0.99 (0.98,1.01)	1.01 (0.92,1.10)	0.94 (0.91,0.96)	0.99 (0.97,1.02)
Log eir	1.44 (1.30,1.59)	1.34 (1.25,1.43)	1.58 (1.43,1.74)	1.46 (1.08,1.80)	1.60 (1.37,1.82)	1.58 (1.33,1.86)
SES quintile						
1 st	1.00	1.00	1.00	1.00	1.00	1.00
2 nd	1.06 (0.89,1.24)	0.94 (0.81,1.08)	0.84 (0.62,1.11)	1.00 (0.68,1.49)	0.88 (0.66,1.23)	0.75 (0.44,1.34)
3 rd	0.95 (0.82,1.12)	0.88 (0.77,1.02)	0.67 (0.50,0.91)	1.01 (0.69,1.51)	0.85 (0.62,1.17)	0.55 (0.30,1.00)
4 th	0.65 (0.55,0.78)	0.65 (0.56,0.76)	0.65 (0.47,0.87)	0.90 (0.60,1.33)	0.74 (0.54,1.02)	0.97 (0.56,1.72)
5 th	0.73 (0.62,0.88)	0.71 (0.61,0.83)	0.64 (0.49,0.85)	0.98 (0.67,1.48)	0.85 (0.62,1.19)	0.71 (0.40,1.28)
Random Error	0.09 (0.06,0.13)	0.10 (0.07,0.14)	0.15 (0.09,0.23)	0.23 (0.12,0.41)	0.14 (0.08,0.23)	0.22 (0.11,0.45)
Spatial variation	0.15 (0.06,0.34)	0.61 (0.18,4.14)	0.23 (0.11,0.53)	0.30 (0.13,0.77)	0.22 (0.10,0.57)	0.33 (0.14,0.82)
Temporal variation	0.15 (0.08,0.47)	0.13 (0.08,0.24)	0.30 (0.13,2.78)	0.63 (0.21,7.14)	0.21 (0.12,0.44)	0.22 (0.11,0.48)
Range (3/ ρ) ^a	19.98 (1.11,41.07)	26.64 (9.99,41.07)	29.97 (6.88,41.07)	15.54 (3.33,39.96)	21.09 (3.33,39.96)	19.98 (4.44,39.96)

a : minimum distance in kilometers at which spatial correlation is significant at 5% , CI=credible intervals

The geostatistical spatiotemporal models including ITN data (Table 5.3) showed that EIR was still associated with all-cause and malaria specific mortality, though the effect was not strong except in all-cause child (1-4 years) mortality (HR= 1.97, 95% CI: 1.26, 2.89). Age was also associated with reduction in all-cause and malaria mortality in the first year of life (HR= 0.95, 95% CI: 0.83, 0.99), but no trend was observed thereafter. ITN use was associated with reduction

Chapter 5: Infants and child mortality in relation to malaria transmission

in all-cause and malaria specific mortality in post-neonates and child (1-4 years). For instance, post-neonates who slept under the ITN were less likely to die due to any illness compared with their counterparts who did not use an ITN (HR= 0.15, 95% CI: 0.02, 0.43). However, protective effect of ITN in mortality decreases with age (25% and 4% for all-cause mortality in post-neonates and child respectively).

Table 5.3 Hazard ratio (HR) estimates of predictors of all-cause and malaria specific mortality for under-five age categories spatiotemporal models (EIR and ITN)

Covariate	All- cause mortality		Malaria specific mortality	
	Post-neonates HR (95% CI)	Child (1-4yr) HR (95% CI)	Post-neonates HR (95% CI)	Child (1-4yr) HR (95% CI)
Constant	0.05 (0.01,0.80)	0.04 (0.01,0.5)	0.02 (0.00,0.03)	0.05 (0.01,0.37)
Age	0.95 (0.83,0.99)	1.01 (0.96,1.04)	0.75 (0.41,0.99)	0.96 (0.89,1.03)
Log eir	1.80 (0.89,2.55)	1.97 (1.26, 2.89)	1.27 (0.50,2.00)	1.67 (0.89,2.78)
ITN	0.75 (0.52,0.97)	0.96 (0.76, 1.16)	0.80 (0.57,1.10)	0.98 (0.64,1.25)
Random Error	0.56 (0.17,2.70)	0.37 (0.16,1.13)	0.62 (0.18,1.66)	0.52 (0.17,1.29)
Spatial Variation	0.86 (0.18,1.33)	0.52 (0.17,1.04)	0.62 (0.18,1.66)	0.60 (0.18,1.29)
Temporal Variation	0.72 (0.21,1.48)	0.48 (0.18,1.97)	0.65 (0.19,1.40)	0.52 (0.18,1.61)
Range (3/ ρ) ^a	11.10 (2.22,34.41)	17.76 (3.33,39.96)	26.64 (9.99,41.07)	27.75 (9.99,41.07)

a : minimum distance in kilometers at which spatial correlation is significant at 5%

Figure 5.3 depicts the excess all-cause and malaria specific mortality as a function of malaria exposure. Overall, excess mortality rates in all-cause and malaria specific mortality increases with malaria exposure in all age groups. Comparing excess all-cause mortality rates and that of malaria specific reveals that excess all-mortality is much higher than that of the malaria specific mortality. This implies that malaria indirect causes are an important factor to consider. The effect of indirect causes is much higher in younger children than older ones. However, the excess mortality shows a sign of leveling off after 59 infective bites per person per month in all age groups except in all-cause neonatal mortality.

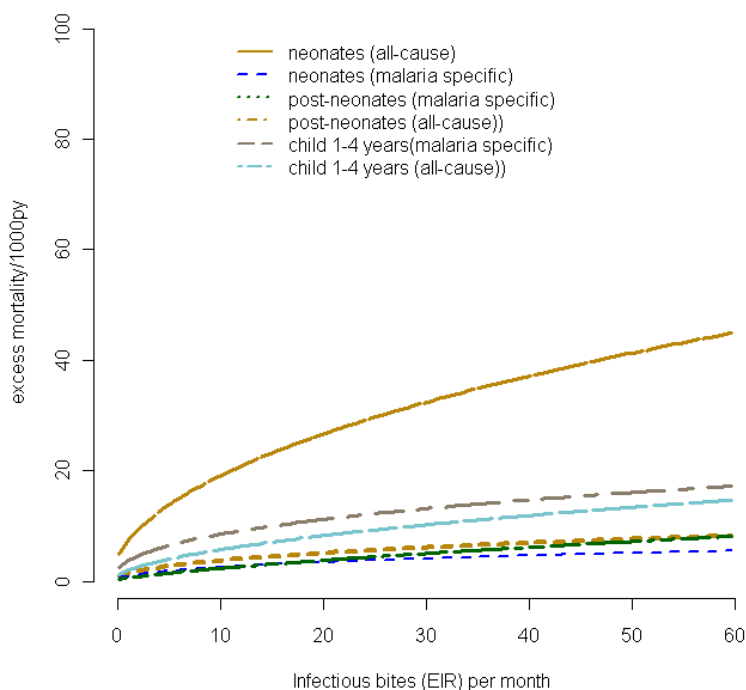


Figure 5.3 Excess mortality for under-five age groups

5.5 Discussion and conclusion

The present study assesses the effects of malaria exposure measured by EIR on all-cause and malaria specific mortality in under-five year old children in the KEMRI/CDC HDSS. Overall, under-five mortality declined during the study period as reported previously (Hamel et al. 2011; Abdullah et al. 2007). However, there was an age shift in the mortality with rates actually increasing in the older children (aged 1-4 years). Reductions in major causes of death occurred in all age groups during the period, except for malaria, HIV and diarrhea in 2004.

Results from bivariate (not shown) and Bayesian geostatistical spatio-temporal models indicated that all-cause and malaria specific mortality increases strongly with the malaria transmission intensity (EIR) of the previous month in a logarithm scale in neonates, infants and children (aged 1-4 year). The positive relationship in these age groups also means that, overall, childhood (<5 years old) all-cause and malaria specific mortality appears to significantly increase with EIR. This implies that decreasing transmission intensity will reduce under-five all-cause and malaria specific mortality in the study area and particularly in infants, who experienced highest mortality

Chapter 5: Infants and child mortality in relation to malaria transmission

rate (125 deaths per 1000 person-years). The positive association between malaria exposure and all-cause mortality is consistent with a recent study (Rumisha et al., 2011b) in a similar setting in which they found that the effect of EIR decreases with age. However (Smith. et al., 2001) showed that all-cause mortality increased strongly with malaria exposure in infants, but not trends in older children (1-4 years) and (Ross et al., 2006) also reported similar relationship with malaria specific mortality.

It is usually expected that the effect of malaria transmission intensity will be strongly associated with malaria specific deaths compared to all-cause mortality. However, a comparison on the effect of EIR on all-cause and malaria specific mortality (Table 5.2) was not significant. This implies that malaria exposure seems to have similar effect on both. This could be due to misclassification error of VA technique (Snow et al., 1992; Todd et al., 1994). Although it has been shown that VA have a low level of sensitivity and specificity in identifying malaria deaths in endemic areas, currently it's the only method available to ascertain cause of death at community level in developing countries (resource-poor settings), where deaths mostly occur at home without any contact with the health system (Snow et al., 1992; Adjuik et al., 2006; van Eijk et al., 2008). Recently, computer-based methods such as expert algorithms and data driven (statistical) methods have been proposed as better methods of obtaining cause of death. However, most of these methods are still in developing phase or need further validation (Reeves and Quigley, 1997; Freeman et al., 2005).

The estimated spatial correlations for both all-cause and malaria-specific in under-five children were very strong with lower spatial ranges and narrower confidence intervals being observed in malaria specific mortality age groups compared to all-cause mortality age groups which was also high. However, spatial correlation for malaria transmission intensity was significant at a lower distance of 4 kilometers (Amek et al., 2012). This rather higher distance could be due to unobserved spatially-correlated factors such as distance to the health facilities.

In addition to EIR, we also assessed the effects of SES and ITN use in all-cause and malaria specific mortality after adjusting for malaria transmission. Higher SES quintiles were protective against all-cause mortality in under-five age groups. This implies that children from well-off households were more likely to experience lower death rates than their counterparts in poorer

Chapter 5: Infants and child mortality in relation to malaria transmission

households. These results are consistent with many studies (Wagstaff, 2000; Sasiwongsoj, 2010; Po and Subramanian, 2011) which have also reported the effects of socioeconomic inequalities on child mortality. In particular, they found strong significant relationship between SES and all-cause mortality in under-five old children, with a high mortality rate occurring in the worse-off (lower) quintiles. However, it is surprising that SES was not strongly relationship with malaria specific mortality in childhood.

The protective effect of ITN use in childhood mortality confirms the results from previous studies that linked ITN use with reduction in all-cause mortality (Diallo et al., 2004; Lengeler, 2004; Lindblade et al., 2004). In particular ITN was associated with a 25% reduction in post-neonates, but no trend was observed in older children and in relation to malaria specific mortality. A previous study (Lindblade et al., 2004) in the same area also reported similar results with 22% reduction of all-cause mortality in post-neonates (1-11 months). A recent study in Tanzania (Rumisha et al., 2011b) reported a decreasing protective of ITN with age, which is consistent with our results in which ITN use was associated with 3% reduction older children (1-4 years) mortality. However, our analysis on the effect of ITN use on all-cause and malaria specific was carried out with data for 2002 only because the ITN data was from a one time survey carried out to assess the net usage in the area. Since then the government through the ministry of health and non-profit organizations have been distributing ITNs to under-five children and pregnant women in the area. Unfortunately the HDSS did not monitor the net use during the study period. A number of interventions targeting children are currently on going and the impact of some of these interventions have shown dramatic reduction in child mortality (Hamel et al., 2011). Similarly, dramatic decrease in deaths and/or hospital admissions attributed to malaria has been reported in malaria endemic areas with scale up of combined malaria control strategies (Bhattarai et al., 2007; Okiro et al., 2007). However, further analysis on the impact of these interventions taking into account spatiotemporal effects of transmission intensity are needed.

5.6 Appendix

Let Y_{ijt} be the mortality (all-cause or malaria specific) status of child i at village j and time interval t and \underline{X}_{ij} be the covariates associated with child i at location j . We assume that Y_{ijt} arises from a Bernoulli distribution. That is $Y_{ijt} \sim \text{Be}(p_{ijt})$ where, p_{ijt} is the probability of child i dying at time interval t . We modeled spatial correlation via village-specific random effects ϕ_j (which is considered as latent observations of a spatial Gaussian process) on the logit, as $\text{logit}(p_{ijt}) = \underline{X}_{ij}^T \underline{\beta} + \phi_j + \theta_j + \varepsilon_i$ where θ_j measures the remaining non-spatial variation at village level and $\underline{\beta}$ is the vector of regression coefficients. EIR is modeled on the log scale as a covariate with measurement errors (because it is estimated by separate Bayesian geostatistical models), that is $\text{LogEIR}_i \sim N(\log \tilde{\text{EIR}}_i, \sigma_{ei}^2)$ predicted at the household of child i where $\tilde{\text{EIR}}$ and σ_e^2 correspond to the mean and variance respectively obtained from posterior prediction distribution of EIR at household i . We assumed $\underline{\phi} \sim \text{MVN}(0, \Sigma)$, Σ is the covariance matrix with elements Σ_{kl} accounting for the covariance between any pair of villages k and l irrespective of the direction (isotropy). Using an exponential correlation function, the covariance matrix is defined by $\Sigma_{kl} = \sigma_1^2 \exp(-\rho d_{kl})$ where σ_1^2 is the spatial variation, d_{kl} is the distance between villages k and l , and ρ is the rate of correlation decay with increasing distance. The minimum distance at which the spatial correlation is significant at 5% is called range and can be obtained from the value $3/\rho$ (Ecker and Gelfand, 1997). We assume an exchangeable prior distribution for θ_j , that is $\theta_j \sim N(0, \sigma_2^2)$.

We modeled temporal correlation via the monthly random effects ε_t ($t=1, \dots, 32$) and assumed an autoregressive (AR) process. The deviance information criterion (Spiegelhalter et al., 2002) was used to identify the best fitting order of the process, which was found to be one. Thus we considered that $\varepsilon_t \sim N(\gamma \varepsilon_{t-1}, \sigma_3^2)$, and $\varepsilon_1 \sim N(0, \frac{\sigma_3^2}{1-\gamma})$. The terms σ_3^2 and γ are the temporal variance and autocorrelation parameters respectively with $\gamma \in (-1, 1)$.

Chapter 5: Infants and child mortality in relation to malaria transmission

A Bayesian model formulation requires the specification of prior distributions for all model parameters. In particular, we choose a non-informative normal prior distribution with mean zero and large variance for the $\underline{\beta}$ parameters, regression coefficients ($\underline{\beta}$), an inverse gamma priors for σ_e^2 , σ_1^2 , σ_2^2 and σ_3^2 . A gamma prior for ρ and a uniform prior for γ , that is $\sigma_e^2, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sim IG(2.01, 1.01)$, $\gamma \sim U(-1, 1)$ and $\rho \sim G(0.1, 0.1)$.

The model was fitted using Markov Chain Monte Carlo (MCMC) simulation algorithm in OpenBugs version 3.1.2 (Imperial College and Medical Council, London, UK) to estimate model parameters (Gelfand et al., 2000). Starting with some initial values about the parameters, we run two chains sampler discarding the first 5000 iterations. Convergence was assessed by Gelman-Rubin diagnostic (Gelman, 1992).

Chapter 6: Mortality in relation to malaria transmission: a comparison across age groups in rural western Kenya

Nyaguara Amek^{1,2,3}, Mary Hamel^{1,4}, Kim A Lindblade⁴, Kayla F. Laserson¹, Laurence Slutsker⁴, Thomas Smith^{2,3}, Penelope Vounatsou^{2,3*}

1. Kenya Medical Research Institute/Centers for Disease Control and Prevention (CDC) Research and Public Health Collaboration, P.O. Box 1578 Kisumu, Kenya
2. Swiss Tropical and Public Health Institute, Socinstr. 57, P.O. Box, 4002 Basel, Switzerland
3. University of Basel, Petersplatz 1 P.O. Box 4003 Basel, Switzerland
4. Centers for Disease Control and Prevention, 1600 Clifton Rd., Atlanta GA 30301, Georgia, USA

*Corresponding Author

This manuscript is prepared for submission to Malaria Journal

Abstract

Most studies of the effect of malaria transmission on mortality in endemic areas have focused on children, yet everyone in these areas is at risk of illness and/or dying due to malaria or other causes indirectly linked to malaria due to malaria infection. As transmission is reduced moreover, this relationship remains unclear. We assessed the relationship of malaria transmission intensity measured by entomological inoculation rate (EIR) with mortality in older children and adults in the KEMRI/CDC health and demographic surveillance system during 2002-2004. Time to death for each resident was analysed using monthly intervals and Bayesian spatio-temporal geostatistical Cox proportion hazard models were fitted via a logistic regression formulation. The models were adjusted for EIR and their measurement errors, age, insecticide treated nets (ITN), socioeconomic status (SES) and parameters describing space-time correlation. EIR exposure surfaces were obtained from a geostatistical model relating entomological data and climatic factors. EIR prediction error was incorporated in survival model. Model estimates were used to calculate excess mortality. Study population was 136801 of which 18.9% were under-five year old children and 5.8% were elderly people (≥ 60 years of age). Females constituted 53% of the population. Crude death rate was 24 and 18 per 1000 person-years for entire population and individuals ≥ 5 years of age respectively. An overall decline in mortality was observed during the period and malaria being the main cause of death for younger (<5 years) and older children (5-14 years), whereas HIV was implicated for people ≥ 15 years of age. Age was associated with all-cause mortality in all older children and adult age groups. A higher EIR was associated with lower mortality in the elderly mortality and ITN had a protective effect in adults aged 30-59 years old mortality. Excess mortality was higher in older children (5-14 years) compared to other age groups. Reduction in elderly mortality in higher EIR could be linked to selection phenomena, and to acquisition of immunity as a result of long exposure to malaria since childhood.

6.1 Introduction

Malaria is one of the main causes of morbidity and mortality in Africa. According to the World Health Organization, there are over 200 million cases of malaria every year, resulting in 2-3 million severe cases and about 800,000 deaths. Most of these deaths occur in Africa (WHO malaria report, 2009). Mortality as a result of malaria infection mainly occurs in children under-five years of age and pregnant women in endemic areas. As a result, most studies (Salum et al., 1994; Nevill et al., 1996; Smith. et al., 2001; Bhattarai et al., 2007) have focused on the effects of malaria transmission on childhood mortality, yet everyone in these areas is exposed to malaria parasites thus at risk of dying due to malaria or other causes indirectly related to malaria (Molineaux, 1997). The relationship between malaria transmission intensity and mortality or morbidity is complex. It has been postulated (Snow et al., 1994, 1997) that any intervention targeting reduction of malaria exposure might delay the acquisition of immunity and thus shift the burden of disease to older age group. There have also been suggestions that, other factors being constant, long-term transmission control in high endemic areas might delay severe clinical consequences of infection or death and perhaps even increase all-cause mortality (Snow and Marsh, 1995; Trape and Rogier, 1996). These considerations make it important to understand the relationship between malaria transmission intensity and mortality for all age groups at risk.

A study in Senegal (Trape and Rogier, 1996) suggested that a tenfold decrease or increase in malaria transmission is associated only with a twofold decrease or increase in malaria morbidity. In that study, three populations with different exposures were compared: Dakar (1 infective bite per person per year), Ndiop (20 infective bites per person per year) and Dielmo (200 infective bite per person per year). Despite these major variations in transmission intensity, the cumulative number of malaria attacks by the age of 60 years was similar (30, 62 and 43 respectively). The study also reports that the incidence of clinical attacks is probably directly proportional to the level of transmission in adults as in children in low transmission areas (e.g. ≤ 0.1 infective bites person per year). However, this does not hold at higher levels of transmission such as >10 infective bites per person per year, suggesting that malaria morbidity rates in endemic settings vary by a factor of 2-3 according to the level of transmission (McGuinness et al., 1998).

Chapter 6: Mortality in relation to malaria transmission across age groups

Other studies have also reported (Snow et al., 1997) high severe malaria hospitalization rate among children from low-moderate transmission levels and in particular, high frequency of severe disease being observed in older children (Snow et al., 1994) with higher fatality rate than infants.

A study of under-five year old children did not show clear increase in malaria specific mortality rates with transmission intensity (Snow and Marsh, 1995). Smith et al., (2001) found that all-cause infant mortality rate increases in relation to transmission intensity, but no trend was observed with older children (12-59 months). On the other hand, data from different ranges of malaria morbidity with varying parasite prevalence showed that malaria-specific and all-cause under-five mortality rates increased with the prevalence of *P. falciparum* infection (Korenromp et al., 2003).

The use of insecticide-treated nets have shown a reduction in malaria transmission by 90% (Gimnig et al., 2003) and all-cause mortality of infants (28days-11months olds) by 23% (Phillips-Howard et al., 2003) in an area of high transmission. Similar results were reported in older children (aged 1-4 years old) in areas with low or intense but highly seasonal transmission (Binka et al., 1996; Nevill et al., 1996). Furthermore, studies (Binka et al., 2002; Diallo et al., 2004; Lindblade et al., 2004) have found no evidence to suggest that insecticide-treated nets use from birth increases all-cause mortality in older children in both stable and unstable high endemic areas.

Recently, studies (Rumisha et al., 2011a: 2011b) in the Rufiji demographic surveillance system in Tanzania reported increase of all-cause mortality in children aged ≤ 14 years with malaria exposure measured by entomological rate, but a decrease in individuals aged ≥ 15 years. Furthermore, strong effect of EIR in mortality was observed in school aged children (5-14 years).

The conflicting results above and others, indicate that the precise nature of the relationship between malaria transmission intensity and mortality is not known, perhaps due to methodological inadequacies and the fact that spatial correction, which independently affects both malaria transmission and mortality, was not considered in these studies.. Furthermore, most of these studies have focused on child mortality and morbidity, yet everyone is at exposed to

malaria infection thus at risk of the disease. These uncertainties hamper effective planning, implementation and evaluation of malaria control strategies. It therefore becomes paramount to clearly understand how malaria transmission intensity relates to mortality in all age groups for proper planning, effective implementation and evaluation of interventions.

In this study, we assess the mortality burden attributed to malaria in older children and adults by analyzing all-cause and malaria specific mortality, and entomological inoculation rate data from the KEMRI/CDC health and demographic surveillance system (HDSS) site, a member of INDEPTH network. The site also participated in the INDEPTH malaria transmission intensity and mortality burden across Africa (MTIMBA) project (Kasasa et al.). The HDSS routinely monitors demographic (births, death and migration) and health events at household level in a geographically defined area. In addition, it collects information on causes of death using verbal autopsy, carry out routine surveillance of the entomological correlates of malaria in randomly selected houses and conduct education and socio-economic surveys (Adazu et al., 2005).

6.2 Materials and methods

6.2.1 Study area and Population

KEMRI/CDC HDSS was launched in September 2001 in Asembo (Rarieda Division, Bondo District) and Gem (Yala and Wagai Divisions, Siaya District) by U.S Centers for Disease Control and Prevention (CDC) in collaboration with the Kenya Medical Research Institute (KEMRI). In 2007, the HDSS expanded to Karemo Division, which is in the northwest, bordering Gem. During the study period, the HDSS was only operating in Asembo and Gem which is approximately 500 km² with a population of 135,000 residing in 33,990 households within 217 villages (75 in Asembo and Gem 142).

The HDSS was established to measure the burden of infectious diseases and evaluate public health interventions designed to reduce these diseases. After the baseline census, all registered including new households were visited after every 4 months to update the status of all individuals residing in the household and record pregnancies, births deaths and migration status that occurred since the previous visit. Data on education and socioeconomic indicators were collected annually. Verbal autopsy questionnaires were used to ascertain most probable cause of

death and entomological data were collected monthly from 10 randomly selected houses and 4 additional houses neighbouring each selected houses (Adazu et al., 2005).

Malaria infection is holoendemic and mainly transmitted by *An. gambiae* s.l (Bayoh et al: Amek et al 2011). Malaria transmission is intense and perennial, though insecticide treated mosquito nets (ITN) trial conducted from 1996 to 2002 in the area reduced it by 90% (Gimnig et al., 2003; Lindblade et al., 2004). Malaria prevalence is still high and is the main cause of child mortality (Adazu et al 2005; unpublished HDSS annual reports)

The population of the study area are culturally homogenous (over 95% are of the Luo ethnic tribe) who depend mainly from substance farming for their livelihood (Adazu et al 2005). The study population consisted of individuals who were resident between May 2002 and December 2004 as defined by HDSS residency rule (Adazu et al., 2005)

6.2.2 Explanatory variables

Household socioeconomic status (SES), age of the individual, entomological inoculation (EIR) rate and insecticide treated nets (ITN) were included as covariates. Household socioeconomic index was calculated as a weighted average of socioeconomic indicators using weights generated via multiple correspondence analysis (MCA). The variables included were occupation of household head, primary source of drinking water, use of cooking fuel, ownership of in-house assets (lantern lamp, sofa, bicycle, radio and television) and livestock possessions (poultry, pigs, donkey, cattle, sheep and goats) . The indices were then grouped into five quintiles with the first quintile representing the poorest households, poorer, poor, less poor and the fifth being the least poor households (Amek et al. in preparation).

EIR monthly estimates at high geographical resolution were obtained over the study area from Bayesian geostatistical zero inflated models, which provided EIR estimates aligned in space and time with the mortality data. These models included environmental and climatic factors extracted from remote sensing data, harmonic seasonal trends and parameters describing space-time correlation. Detailed descriptions of the methodology and analysis of the EIR are provided elsewhere (Amek et al., 2011, 2012). In brief, the EIR exposure surfaces were obtained by the product of geostatistical model based estimates of sporozoite rate (SR) and densities. A

multiplication factor of 1.605 was included to adjust the mosquito densities for the trapping efficiency of light traps (Lines et al., 1991). The ITN-use data included in this study was from a one time survey carried within the study area in 2002 to access the ITN coverage.

6.3 Statistical analysis

The study population was grouped into seven categories namely 0-28 days, 1-11 months, 1-4, 5-14, 15-29, 30-59 and ≥ 60 years old and analysis carried out separately for each age group. Time to death for each individual in each age group was rounded to the nearest month and Cox proportional hazard models were fitted using binary logistic regression formulation (Singer and Willett, 1993; Manda and Meyer, 2005). Predictors included in all models were estimated EIR (on a logarithm scale, base 10) of a month prior to mortality outcome status, household SES quintiles, ITN use and age. The prediction error of the EIR estimate was included in the model as a measurement error in the covariate (Bernadinelli et al., 1997). Exploratory analyses were carried out in STATA 10 (Stata Corporation US) to assess the bivariate relations between all-cause and malaria specific mortality.

In a second phase of analysis, Bayesian geostatistical spatiotemporal logistic regression models were used to allow for spatial and/or temporal correlations in the data (which otherwise results in biased estimates of the predictors) (Cressie, 1993). Both the malaria exposure data and the mortality data were geo-located at the level of the household, which was a fixed geographical locations over a continuous period of the study thus correlated in space and time. Spatial correlation arises because locations in close proximity have similar risks due to shared exposures. Standard statistical methods are not appropriate for analyzing spatially correlated data because they assume independence between observations. Spatial correlation was modeled via village-specific random effects assuming that they (random effects) follow a multivariate normal distribution with variance covariance matrix related to an exponential correlation function between any pair of villages irrespective of direction. Temporal correlation was modeled by introducing monthly random effects and modeled by autoregressive (AR) processes of various orders. The models also included exchangeable (clustering) random effects at the level of the village. All covariates that were significant at a 15% significant level in the exploratory analysis were included in the Bayesian models, which were fitted in OpenBugs version 3.1.2 (Imperial

College and Medical Research Council London, UK) .A detailed description of the Bayesian geostatistical formulation model is given in the appendix

6.3.1 Excess mortality rate attributed to malaria exposure

The excess mortality rate (EMR) was calculated as the difference between the mortality rates (MR) at which the values of EIR is greater than zero and at zero: $EMR = MR(EIR > 0) - MR(EIR = 0)$. In particular, we computed the probability (P_{kj}) of death for each age group j from the logistic regression model over a range of EIR between 0.1 and 60 infectious bites per person per month covered by intervals of 0.1 using model coefficients, monthly EIR and midpoint for each age group. The probability at zero level of EIR was obtained using a Taylor series approximation, that is $\log(EIR) \sim (EIR - 1) - \frac{1}{2}(EIR - 1)^2 + \frac{1}{3}(EIR - 1)^3 - \dots$ where $EIR = 0$. The probability of death for each age group j was converted to a rate and expressed per 1000 person years, that is rate = $[-\ln(1 - P_{kj})]/t$ per 1000py, where k =EIR interval, j =age group and t =one month

6.4 Results

6.4.1 Descriptive statistics

During the study period, the area had a population of 136801 residing in 217 villages. Children less than 5 years of age constituted 18.9% of the population and adults ≥ 60 years of age made up of 5.8 %. Overall, females were slightly higher than males accounting for approximately 53% of the total population. There were 5192 deaths for residents aged ≥ 5 years of age resulting to a crude death rate of 18 deaths per 1000 person-years. Fifty two percent of the total deaths were in females. Results of under-five age groups analysis is presented by Amek et al. (in preparation). Figure 6.1 below show the crude death rates per year. Mortality decreased annually during the period except in 2003 where there was an increase in mortality for under-five and above 60 years of age.

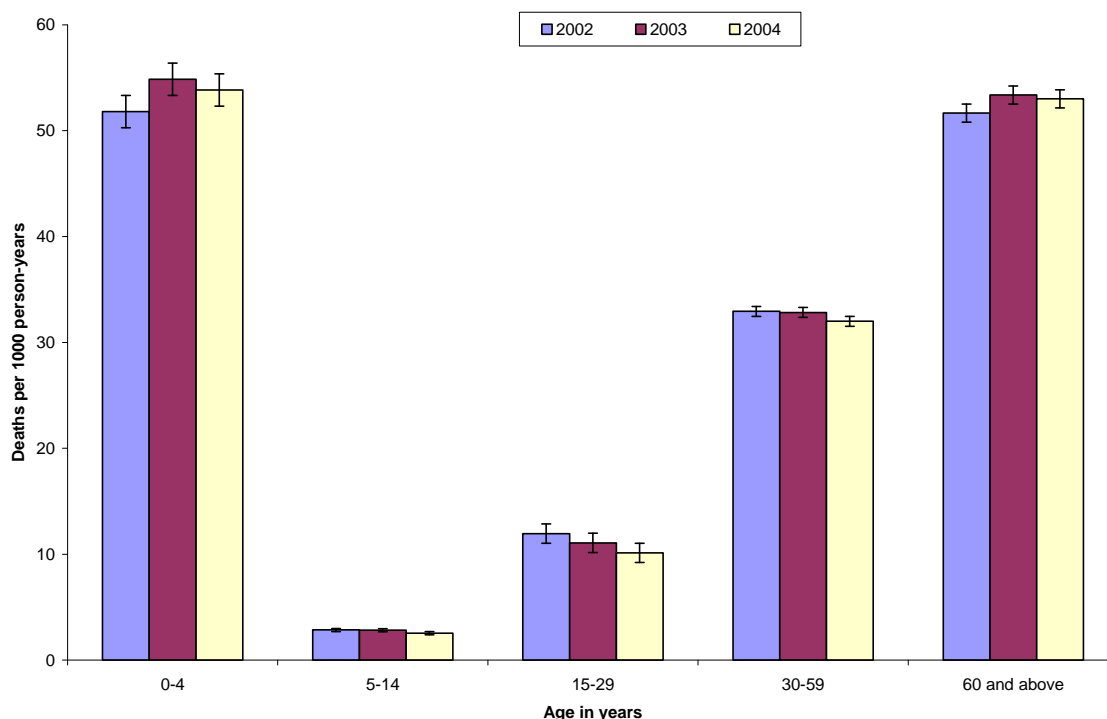


Figure 6.1 Crude death rates per year

Of the total deaths reported, 6136 (74%) were assigned causes by VA. The low percentage is due to the fact that VA for adolescents and adults (>11 years) started in 2003 (van Eijk et al., 2008) or lack of respondent (30 cases) due to loss to follow up. In all, HIV/TB related deaths (23.12%) were the main cause of death followed by malaria (15.17%) and pneumonia (10.28%) of the total reviewed deaths. In under-five year old children, malaria was the main cause of death followed by anemia then pneumonia (Amek et al., in preparation). Whereas in children ≥ 5 years old of age HIV/TB was the main cause of death followed by malaria then pneumonia. Results of under-five age groups analysis are presented by Amek et al., (in preparation). Figure 6.2 shows the five main causes of death by age groups. Malaria was the main cause of death for individuals aged 5-14 years while HIV was implicated for deaths in adults ≥ 15 years old. However, malaria still remains as one of the main causes of death in the remaining age groups.

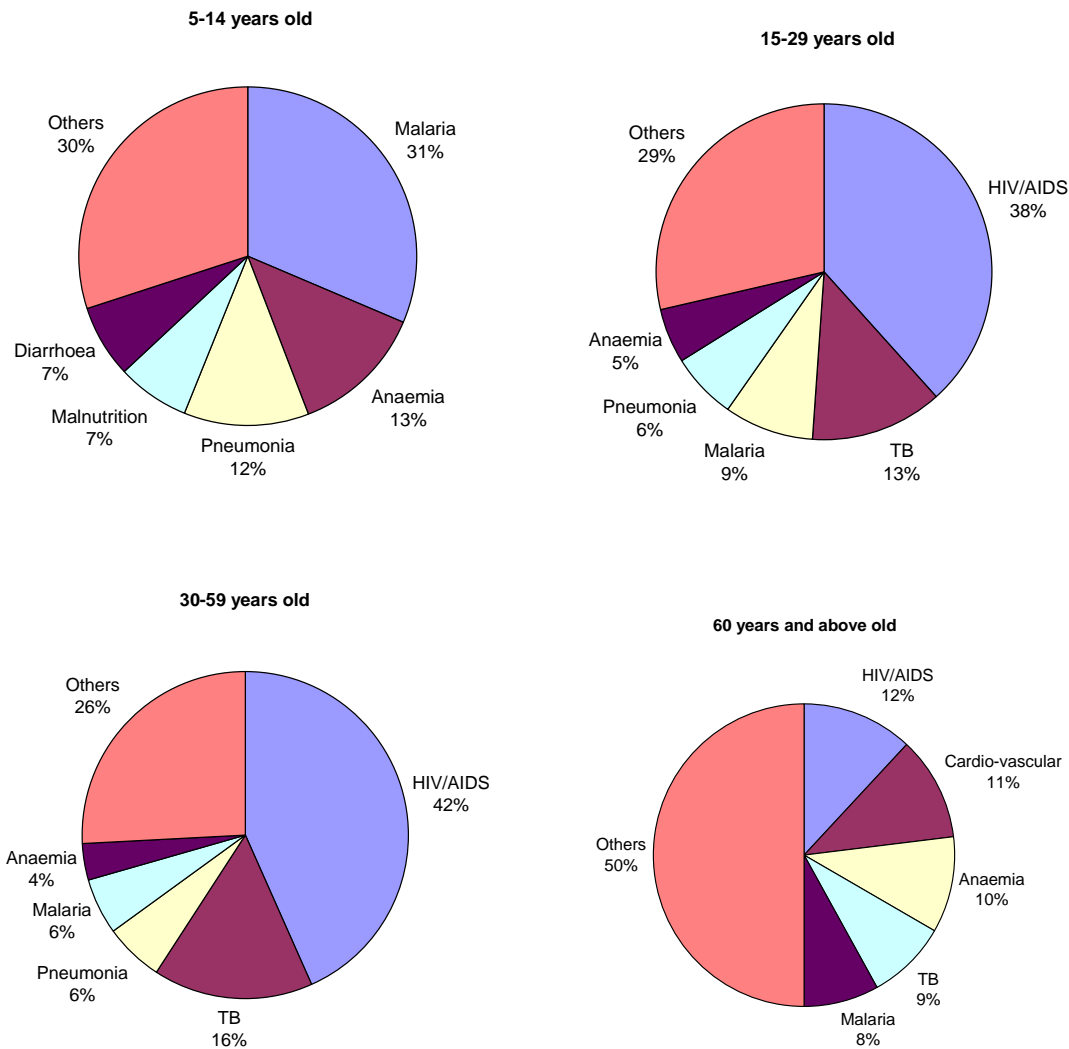


Figure 6.2 Main causes of death by age groups for older children and adults

The bivariate analyses revealed that EIR, ITN use, age and higher socioeconomic status are associated with all-cause mortality. However, there was no relationship between malaria specific mortality and these covariates in older children (age ≥ 5 years old) and adults.

6.4.2 Model-based results

Table 6.1 shows the results obtained from Bayesian geostatistical binary logistic models excluding SES and ITN covariates. All-cause mortality decreased strongly with increase in EIR in elderly individuals (aged ≥ 60 years old). No trend was observed in older children (≥ 5 years old) and adults aged ≤ 59 years old. Age was positively associated with mortality in all age

Chapter 6: Mortality in relation to malaria transmission across age groups

groups, though the association was not very strong: 5-14 years (HR=1.01, 95% CI: 1.00-1.02), 15-29 years (HR=1.04, 95% CI: 1.03-1.05), 30-59 years (HR=1.02, 95% CI: 1.01-1.03) and 60years and above (HR=1.05, 95% CI: 1.04-1.06).

The estimated spatial correlation was associated with all-cause mortality in all age-groups and the minimum distance at which it is significant at 5% (spatial range) was 6.11 km (95% CI: 1.11-21.09 km), 5.55 km (95% CI: 0.67-24.42 km), 33.30 km (95% CI: 14.43-41.07 km) and 26.64 km (95% CI: 11.10-39.96 km) for aged 5-14, 15-29, 30-59 and ≥ 60 years old respectively. Temporal variation was related to mortality in adults aged between 15-29 and 30-59 years old.

Table 6.1 Hazard ratio (HR) posterior estimates of all-cause mortality by age groups (EIR only)

Covariate	Neonates* HR (95% CI)	Infants* HR (95% CI)	1-4yrs* HR (95% CI)	5-14 HR (95% CI)	15-29 HR (95% CI)	30-59 HR (95% CI)	60 and above HR (95% CI)
Constant	0.09 (0.06, 0.14)	0.12 (0.04, 0.51)	0.19 (0.06,0.72)	0.09 (0.02,0.20)	0.07 (0.01,0.30)	0.02 (0.01,0.04)	0.02 (0.01,0.03)
Age	0.96 (0.94, 0.98)	0.94 (0.92, 0.95)	0.99 (0.98,1.00)	1.01 (1.00,1.02)	1.04 (1.03,1.05)	1.02 (1.01,1.03)	1.05 (1.04,1.06)
Log EIR	1.45 (1.30, 1.60)	1.37 (1.27,1.47)	1.58 (1.42,1.74)	0.74 (0.40,1.20)	0.99 (0.47,1.29)	0.96 (0.63,1.12)	0.79 (0.48,0.97)
Random Error	0.09 (0.06, 0.13)	0.08 (0.05,0.11)	0.15 (0.09,0.25)	0.30 (0.14,0.71)	0.29 (0.14,0.58)	0.14 (0.10,0.23)	0.19 (0.12,0.26)
Spatial Variation	0.16 (0.07, 0.35)	0.15 (0.08, 0.28)	0.22 (0.10,0.52)	0.32 (0.13,0.92)	0.29 (0.12,0.81)	0.23 (0.11,0.58)	0.09 (0.05,0.16)
Temporal Variation	0.15 (0.08, 0.49)	0.37 (0.12, 4.48)	0.31 (0.13,3.79)	0.49 0.14,1.75	0.37 (0.14,0.98)	0.23 (0.11,0.51)	0.60 (0.18,4.51)
Range (3/ ρ) ^a	23.31 (1.11,41.07)	34.41 (12.21,41.07)	28.86 (6.66,41.07)	6.11 (1.11,21.09)	5.55 (0.67,24.42)	33.30 (14.43,41.07)	26.64 (11.10,39.96)

a : minimum distance in kilometers at which spatial correlation is significant at 5% , *= results from Amek et al. (in preparation), CI=credible intervals

A model including SES and EIR (data not shown) depicts similar relationship between malaria exposure and all-cause mortality in each age group as reported in table 6.1. Higher SES quintiles were associated with reduction in all-cause mortality in all age groups. This implies that individuals belonging to households with fewer assets (low quintile) are more likely to experience high mortality compared to their counterparts from households with more assets.

The model including the ITN use (Table 6.2) showed that ITN was strongly associated with reduction in all-cause mortality in adults aged 30-59 years (HR= 0.74, 95% CI: 0.60, 0.91). A decrease in the protective effect of ITN with age was also observed in individuals aged ≥ 15

Chapter 6: Mortality in relation to malaria transmission across age groups

years old (44%, 26% and 19% for age groups 15-29, 30-59 and 60 plus respectively). However, after adjusting for ITN use, protective effect of malaria exposure in elderly mortality was not strong. Age was positively associated with mortality in adults aged 15-59 years. No trend was observed in older children (5-14 years) and elderly population (≥ 60 years).

Table 6.2 Hazard ratio (HR) posterior estimates of all-cause mortality by age groups (EIR and ITN)

Covariate	Post-neonate* HR (95% CI)	1-4yrs* HR (95% CI)	5-14 HR (95% CI)	15-29 HR (95% CI)	30-59 HR (95% CI)	60 and above HR (95% CI)
Constant	0.05 (0.01,0.80)	0.04 (0.01,0.5)	0.23 (0.08,0.71)	0.02 (0.01,0.18)	0.05 (0.02,0.12)	0.09 (0.03,0.20)
Age	0.95 (0.83,0.99)	1.01 (0.96,1.04)	0.99 (0.98,1.01)	1.03 (1.02,1.04)	1.04 (1.03,1.06)	1.00 (0.97,1.04)
Log EIR	1.80 (0.89,2.55)	1.97 (1.26, 2.89)	0.78 (0.30,1.22)	1.02 (0.45,1.54)	1.02 (0.72,1.30)	0.92 (0.48,1.35)
ITN	0.75 (0.52,0.97)	0.96 (0.76, 1.16)	1.09 (0.69,1.45)	0.66 (0.40,1.10)	0.74 (0.60,0.91)	0.81 (0.36,1.68)
Random error	0.56 (0.17,2.70)	0.37 (0.16,1.13)	0.25 (0.13,0.53)	0.32 (0.15,0.72)	0.13 (0.08,0.23)	0.34 (0.14,1.05)
Spatial Variation	0.86 (0.18,1.33)	0.52 (0.17,1.04)	0.37 (0.12,1.56)	0.41 (0.14,1.40)	0.23 (0.10,0.57)	0.47 (0.16,2.09)
Temporal Variation	0.72 (0.21,1.48)	0.48 (0.18,1.97)	0.27 (0.11,0.71)	0.27 (0.13,0.66)	0.15 (0.10,0.28)	0.36 (0.14,1.16)
Range (3/ ρ) ^a	11.10 (2.22,34.41)	17.76 (3.33,39.96)	8.88 (1.11,29.97)	14.76 (3.33,39.96)	28.86 (7.77,41.07)	19.98 (4.44,39.96)

a : minimum distance in kilometers at which spatial correlation is significant at 5% , *= results from Amek et al. (in preparation), CI=credible intervals

Figure 6.3 and 6.4 shows the excess mortality as a function of malaria exposure. In children <15 years old of age, there was a clear increase in mortality rate with EIR. The highest burden of mortality rate is observed in neonates followed by infants and the least being in children aged 5-14 years old. Interestingly, a protective effect was noted in individuals of aged ≥ 15 years old with very low excess mortality in elderly population (aged >60 years old) due to malaria exposure

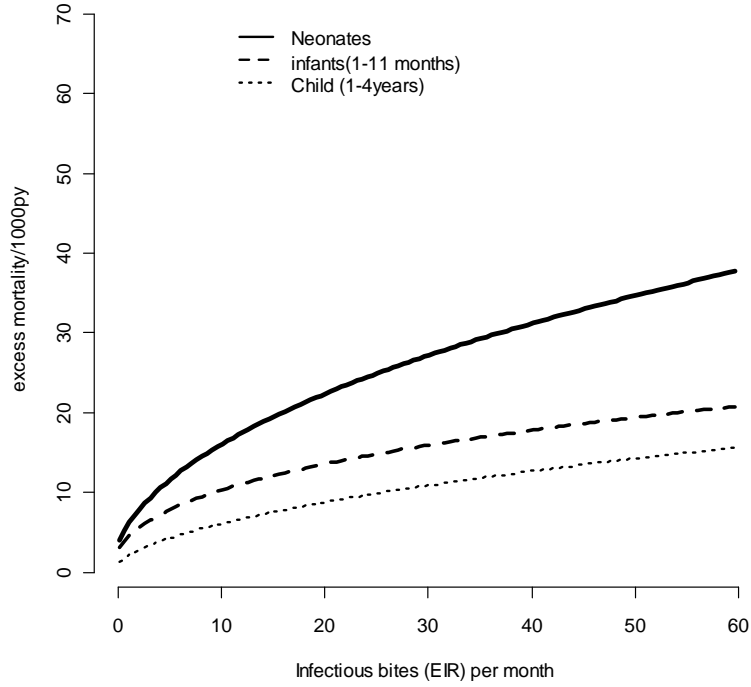


Figure 6.3 Excess mortality for under-five age groups

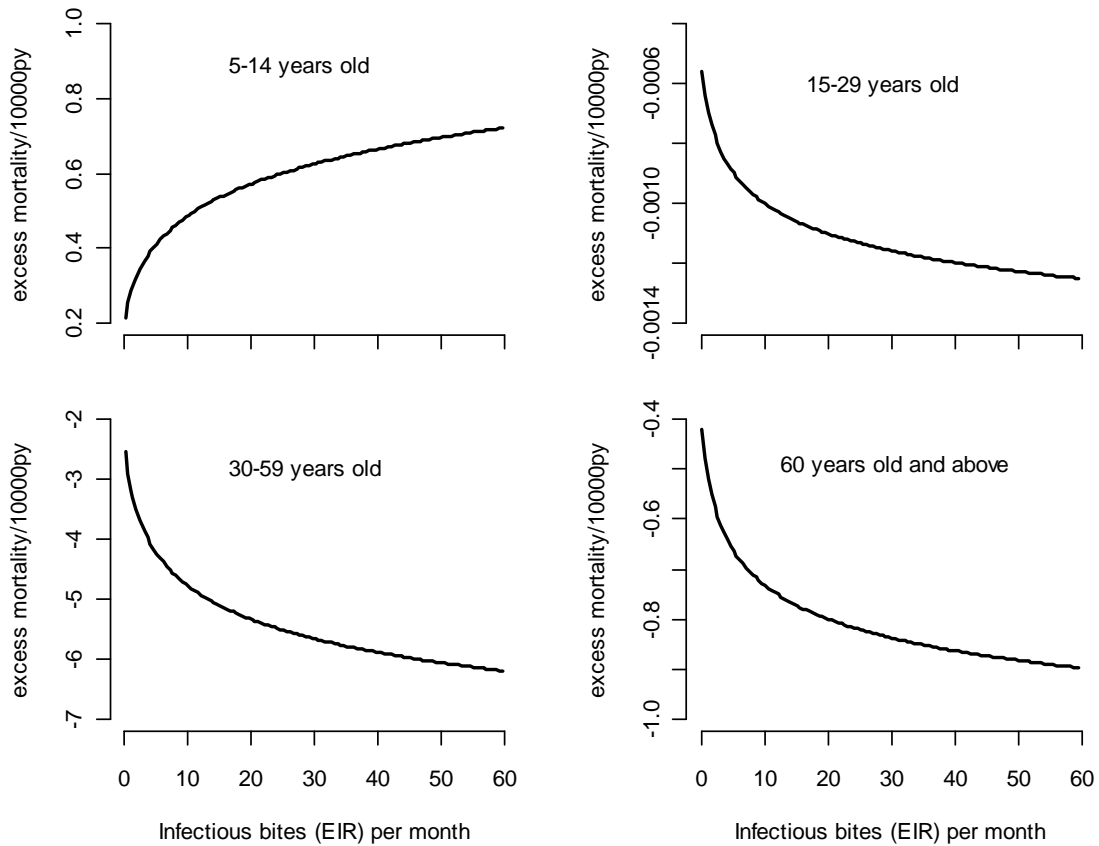


Figure 6.4 Excess mortality for 5 years old above

6.5 Discussion and conclusion

In this study we applied Bayesian geostatistical spatio-temporal binary logistic models to assess the age patterns in the relationship between all-cause and malaria specific mortality and malaria transmission intensity measured by EIR in the KEMRICDC HDSS. The models were adjusted for village clustering, spatial and temporal variations. High resolution estimates of predicted EIR at household level were taken from Bayesian spatio-temporal predictive models and, because these were estimates and not precise values, we modeled EIR as a covariate with measurement errors. Mortality outcome status was linked with the EIR estimates of the previous month of that mortality status to allow for the temporal effect of malaria exposure on mortality outcome.

An overall decline in mortality was observed during the study period. However, mortality increased in the elderly (≥ 60 years) population in 2004. A similar trend was also observed in children aged 1-4 years (Amek et al., in preparation). Compared with the rest of Kenya, the study area experienced significantly higher mortality during the study period (Central Bureau of Statistics et al., 2004). The Kenyan national crude death rate was 15 per 1000 population in 2002. Children who survive childhood experience lower mortality until age 15 years when mortality begins to increase in the study area (see Figure 6.1). However, overall age specific mortality trends were similar to those reported in Rufiji DSS and other studies (Adjuik et al., 2006; Abdullah et al., 2007).

Although our VA data suggested that malaria is the main cause of death in individuals aged 5-14 years old, it was not significantly related malaria transmission intensity in bivariate analysis. This could be due to small number of deaths that occurred in the age group. Our data also showed that HIV was the main cause of death in individuals aged ≥ 15 years old with the highest burden occurring in reproductive age (see figure 6.2).

Although malaria transmission was reduced by 90% in the study area as a result of an insecticide-treated net trial conducted in the study area between 1996 to 2002 (Lindblade et al., 2004) and the EIR remains low (8.5 infective bites per person per year) (Amek et al., 2012), malaria is still the main cause of childhood and individuals aged 5-14 years old mortality. This confirms the results by other studies (Beier et al., 1999) that showed *P. falciparum* prevalence

Chapter 6: Mortality in relation to malaria transmission across age groups

exceeding 40% in areas with EIR < 5 infective bites per person per year. Beier and colleagues further suggest that significant malaria morbidity and mortality may continue to occur until transmission is lowered to less than one bite per person per year.

In contrast to strong effect of EIR in all-cause mortality in under-five age groups (Amek et al., in preparation), no trend was observed in older children (aged ≥ 5 years old) and adults ≤ 59 years old. The highest increase in mortality risk with malaria exposure was observed in children aged 1-4 years (58%) and the lowest in infants (37%). The highest rate of excess mortality was observed in neonates followed by infants (Figure 6.3). A study in Rufiji DSS (Rumisha et al., 2011a) during the same period reported similar trend of on the effect of EIR in all-cause mortality in under-five age groups. However, the EIR was negatively correlated with all-cause mortality for elderly population (aged ≥ 60 years olds). This could be because the older people at risk in very high exposure areas are a selected group who successfully acquired natural immunity protecting them against severe malaria. Those at risk of malaria mortality in these areas died during childhood. We illustrate this in figure 6.5 by comparing the effect of malaria transmission intensity on mortality in low and high transmission areas. The top row illustrates the situation in high transmission areas, where children are exposed at early stage of life to intense mosquito parasites leading to severe risk of the disease and high rates of death. Thereafter, the risk falls rapidly as immunity is acquired in the survivors. For instance, by 5 years of age the acquired immunity is able to protect the child from severe disease. In contrast (second scenario), their counterparts born in low transmission settings experience low exposures resulting in longer period of risk which spreads the risk of severe disease across the entire childhood period. However, in both scenarios similar numbers of deaths are observed over the whole age-range. A study (Rumisha, Smith, et al., 2012) in Rufiji DSS also found that higher malaria exposure reduces mortality in elderly population.

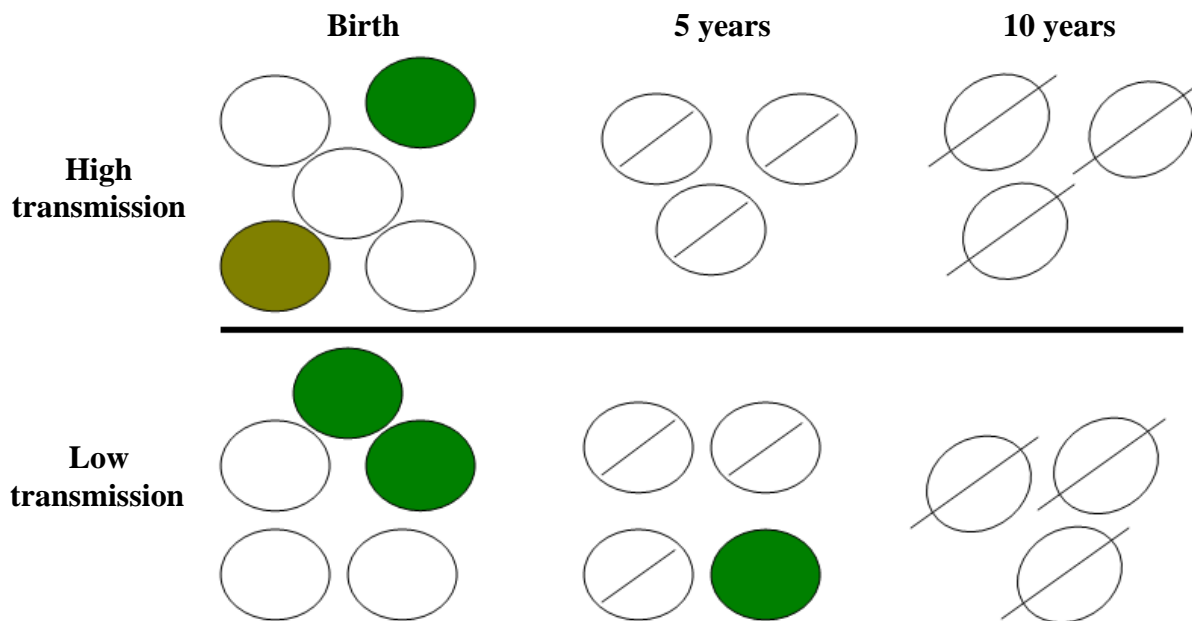


Figure 6.5 Effect of malaria transmission in acquired immunity

ITN use was significantly protective against all-cause mortality in infants and adults aged 30-59 years old. This confirms the results of a study on the sustainability of reductions of malaria transmission and mortality with use of ITN carried out in the same area (Lindblade et al., 2004) that showed a significant decrease in all-cause mortality in infants but no trend in children aged 1-4 years old. However, it should be noted that even though we used ITN to measure exposure, they only measured recent exposure since ITN data was collected in the same with mortality data. A randomized control trial on the effect of ITNs on mortality in an area of intense malaria transmission did not find association between ITN use and adult mortality (Binka et al., 2002). But a study in moderate transmission area showed that mosquito net use is associated with reduction in all-cause mortality in adult aged ≥ 40 years old (Smith et al., 2002). Similarly, another study in India (Dhingra et al., 2010) using VA showed an increase in malaria specific mortality with age from young adults. A review on malaria mortality for the last 30 years (1980-2010) revealed higher estimates than reported previously (Murray et al., 2012). These suggest that malaria could be a major cause of death in adults.

In conclusion, our study found a strong negative relationship between mortality and malaria exposure in the elderly (aged ≥ 60 years old). Insecticide-treated nets use also showed a reduction in mortality in adults aged 30-59 years old. These results complement earlier analyses

(Amek et al., in preparation) in children under-five years of age in the same area showing that malaria transmission intensity is associated with all-cause and malaria specific mortality in neonates, infants and child (1-4 years).

6.6 Appendix

Let Y_{ijt} be the mortality status of individual i in village j at time interval t . We assume that Y_{ijt} arises from a Bernoulli distribution. That is $Y_{ijt} \sim \text{Be}(p_{ijt})$ where, p_{ijt} is the probability of individual i dying at time interval t . Let \underline{X}_{ij} denote the covariates associated with individual i in village j . These covariates include age, ITN use, SES and EIR. Also let ϕ_j and θ_j be village specific spatial and non spatial random effects respectively, and ε_t be the temporal (monthly) random effect. We model the association between the above covariates and mortality status of individual i at village j and at time interval t adjusting for random effects above via logistic regression: $\text{logit}(p_{ijt}) = \underline{X}_i^T \underline{\beta} + \phi_j + \theta_j + \varepsilon_t$, where $\underline{\beta}$ is the vector of regression coefficients. EIR was estimated by a separate Bayesian geostatistical models, therefore was modeled as a covariate with measurement error, in particular on a logarithm scale. That is $\text{LogEIR}_i \sim N(\log \tilde{\text{EIR}}_i, \sigma_{ei}^2)$ predicted at the household of child i where $\tilde{\text{EIR}}$ and σ_e^2 correspond to the mean and variance respectively obtained from posterior prediction distribution of EIR at household. We assumed that ϕ_j are parameters from latent observations of a spatial Gaussian process with covariance matrix and model spatial correlation between any pair of villages as a function of their distance irrespective of the direction using an exponential correlation function, that is $\underline{\phi} \sim \text{MVN}(0, \Sigma)$, $\Sigma_{kl} = \sigma_1^2 \exp(-\rho d_{kl})$ where σ_1^2 is the spatial variation, d_{kl} is the distance between villages k and l , and ρ is the rate of correlation decay with increasing distance. The minimum distance at which the spatial correlation is significant at 5% is called range and can be obtained from the value $3/\rho$ (Ecker and Gelfand, 1997). We also assume an exchangeable prior distribution for θ_j , that is $\underline{\theta} \sim N(0, \sigma_2^2)$. The temporal effect (ε_t) was modeled by autoregressive (AR) process of various orders. The deviance information criterion (Spiegelhalter et al., 2002) was used to identify the best fitting order of the process, which was found to be one: $\varepsilon_t \sim N(\gamma \varepsilon_{t-1}, \sigma_3^2)$,

and $\varepsilon_1 \sim N(0, \frac{\sigma_3^2}{1-\gamma})$. The terms σ_3^2 and γ are the temporal variance and autocorrelation parameters respectively with $\gamma \in (-1,1)$.

To complete the Bayesian model formulation, we specify prior distributions for all model parameters. In particular, we choose a non-informative normal prior distribution with mean zero and large variance for the $\underline{\beta}$ parameters, regression coefficients ($\underline{\beta}$), an inverse gamma priors for σ_e^2 , σ_1^2 , σ_2^2 and σ_3^2 . A gamma prior for ρ and a uniform prior for γ , that is $\sigma_e^2, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sim IG(2.01, 1.01)$, $\gamma \sim U(-1, 1)$ and $\rho \sim G(0.1, 0.1)$.

The model was fitted using Morkov Chain Monte Carlo (MCMC) simulation algorithm in OpenBugs version 3.1.2 (Imperial College and Medical Council, London, UK) to estimate model parameters (Gelfand and Smith, 1990). Starting with some initials values about the parameters, we run two chains sampler discarding the first 5000 iterations. Convergence was assessed by Gelman-Rubin diagnostic (Gelman, 1992).

Chapter 7: General discussion

This thesis contributes to malaria epidemiology by analyzing the KEMRI/CDC HDSS MTIMBA database to (i) assess the geographical and seasonal distribution of infective mosquitoes (sporozoite rate), (ii) map space-time variation in entomological inoculation rates (EIR), (iii) assess space-time changes in socioeconomic status and (iv) assess the relationship between all-cause and malaria specific mortality in relation to malaria transmission. For the above applications, data-driven geostatistical models and novel statistical methodology for sparse binomial data were developed. In this thesis, maps of the malaria transmission (exposures) have been produced adjusting for the large number of locations with either no mosquitoes or zero proportion of infected mosquitoes. Furthermore, high geographical resolution (250 m by 250 m) EIR surfaces were estimated over the study area as a product of geostatistical model based estimates of mosquito sporozoite rates and density, allowing alignment in space and time of the malaria transmission and mortality data. The uncertainty in the predicted EIR estimates was further included in assessing the relation between mortality and malaria transmission intensity. We have also produced maps of household socioeconomic quintiles as predictors of mortality.

The results of this thesis contribute to a better understanding of the relationship between malaria transmission and mortality in malaria endemic areas. This work yielded six manuscripts which form the main chapters of this thesis. One manuscript (chapter 2) was already published in *Spatial and Spatio-temporal Epidemiology* journal (Amek et al., 2011), another one (chapter 3) was submitted in *Parasite & Vectors* and it has been accepted conditional on minor revisions (Amek et al., 2012), the remaining three (chapter 4, 5 and 6) have been prepared for submission in *Global Health Action*, *PLoS ONE* and *Malaria* journal, respectively. In each chapter, detailed methodology, discussions of the results and conclusions is provided. In the next sections, we report a summary of the main contribution and findings (grouped into statistical and epidemiological), their implication to malaria control, limitations and recommendations for further research

7.1 Statistical contribution

In chapters 2 and 3, we extended zero inflated models to fit overdispersed geostatistical spatio-temporal entomological data. Sparse entomological data arises because in most surveyed

locations either no mosquito was caught or most of them tested negative for the presence of sporozoites in their salivary glands. Strategies for handling zero inflated count data are well documented (Lambert, 1992; Hall, 2000; Vieira et al., 2000; Agarwal et al., 2002; Yau et al., 2003; Martin et al., 2005, 2005; Vounatsou et al., 2009). To our knowledge, no analysis had yet applied to zero-inflated, binomial, geostatistical data with seasonal and temporal characteristics. The zero inflated geostatistical models developed and described in this thesis are not only limited to entomological data but can also be applied to other epidemiological datasets. A number of free software such as *R* (<http://www.r-project.org/>) and *BayesX* (<http://www.stat.uni-muenchen.de/~bayesx/bayesx.html>) have in-built zero inflation algorithms. However, these algorithms do not adjust for temporal correlation, limiting their application to our entomological data.

The approach used to obtain EIR estimates described in chapter 3 is novel statistical method in malaria epidemiology. In practical applications, EIR is assumed to be a continuous variable and is fitted using linear regression and transformation techniques, such as logarithm, if the normality assumption is violated (Githeko et al., 1993; Gemperli, Sogoba, et al., 2006). It is difficult to achieve normality in data with large number of location having zero mosquitoes or zero proportion of infected mosquitoes because no approach will distribute out the zeros values. Furthermore, EIR is a product of mosquito sporozoite rate and density which are binomial and count data types respectively. These two data types require different modeling techniques. In particular sporozoite rate and mosquito density follow binomial and negative binomial distributions respectively.

7.2 Epidemiological contribution

Disease risk maps are useful tools for spatially targeted proper planning, effective implementation and evaluation of interventions. In addition, they are helpful to identify higher risk areas, so called “hot spots”. In chapters 2 and 3 model-based high resolution maps of malaria transmission were produced for the KEMRI/CDC HDSS area: Maps of predicted sporozoite rate in wet and dry (Figure 2.3, chapter 2) seasons, and monthly EIR (Figure 3.5 chapter 3). These maps were generated at 250 m by 250 m spatial resolution using transmission models, adjusted for environmental/climatic predictors which are extracted from remote sensing data and

parameters describing space-time correlation. Different sets of environmental and climatic predictors were evaluated and the best combination that favored vector population and infectivity was included in the geostatistical models. These maps were shared with area entomologist who confirmed that they reflect known patterns of transmission in the area revealing high predictive ability of the models (Figure 2.4 chapter 2). An important characteristic of these maps is the ability to show small scale and space-time transmission heterogeneity. During the study period, the HDSS covered an area of approximately 500 km² and entomological data was collected at 1110 unique locations. However, existing transmission maps cover wide geographical areas ranging from country (Omumbo et al., 1998) and regional (Gemperli, Sogoba, et al., 2006) levels to the entire world (Gething et al., 2011) based on few survey locations in relation to the area covered. Furthermore, temporal variation was not taken into account in the models that produced these maps. The established relation between environmental factors and malaria transmission could predict changes in malaria transmission in the presence of environmental changes.

It has been documented that household socioeconomic status (SES) has an effect on mortality (Wagstaff, 2000; Savigny et al., 2002; Sasiwongsoj, 2010; Po and Subramanian, 2011) and malaria infection (Abdulla et al., 2001). However, most of these studies have employed principal component analysis (PCA) technique to analyze qualitative ordinal asset-based variables yet PCA was designed for a set of quantitative variables. In chapter 4, we calculated household SES and assessed changes in household socioeconomic positions in relation to other households over time and in space by comparing three different techniques namely PCA, multiple correspondence analysis (MCA) and polychoric PCA. The best method (MCA) was further applied to obtain households SES used as a covariate in assess the relationship between mortality and malaria transmission (Chapter 5 and 6)

A clear understanding of the relationship between mortality and malaria transmission is important in developing effective and monitoring the impact of malaria interventions. In chapters 5 and 6, we assessed the relationship between malaria transmission all-cause and malaria mortality in children and adults. A number of studies have discussed this relationship before mainly in children (Snow et al., 1997; Smith. et al., 2001; Gemperli et al., 2004). In addition,

effects of insecticide treated nets on child mortality have been documented (Binka et al., 1996; Schellenberg et al., 2001; Lengeler, 2004; Lindblade et al., 2004; Lim et al., 2011). However, these studies and others have reported contradicting results perhaps due to limitation in methodology used and the fact that spatial correlation, which independently affects both malaria transmissions and mortality, was ignored. This made it challenging to understand the general pattern between mortality and malaria transmission. In this thesis, the models developed, data used and the rigorous analyses employed to assess the amount of mortality attributed to malaria transmission are plausible. Therefore the results obtained may reflect the nature of mortality-transmission pattern.

The majority of malaria deaths are known to occur in childhood and pregnant women in malaria endemic areas therefore many studies on malaria focus on these age groups yet everyone in these areas is at risk of malaria infection and may result in death due to or contributed to malaria. In chapter 6, we report the relationship between age-specific mortality and malaria transmission intensity which is useful to selectively target high risk age group thus making control effective. However, it would be interesting to evaluate cumulative transmission exposure and how it is associated with acquired immunity and assess effects of incidence-transmission relation on mortality. The decreasing risk in elderly mortality in the study area suggests there may be selective effect as a consequence of malaria immunity. We propose further research to determine the duration of immune memory and level required to avert severe disease (Filipe et al., 2007).

Figure 7.2 combines results from chapters 2, 3 and 4 to determine seasonality and relationship between malaria transmission and, all-cause and malaria specific mortality and in the KEMRI/CDC HDSS. The figure depicts seasonal and temporal patterns in mortality and transmission intensity in the area with the highest peak occurring in the months of May and June for malaria transmission and mortality respectively. In particular all-cause and malaria specific mortality peaks a month after transmission peak. This pattern reinforces our choice of linking EIR estimates of the previous month with the mortality outcome status of the current month, incorporating the relationship between both all-cause and malaria specific mortality and transmission. This pattern is consistent with a similar study carried out in Rufiji DSS in the same

period (Rumisha et al., 2011b) in which they found that EIR is positively associated with all-cause mortality in children younger than 5 years.

The area receives rainfall throughout the year with heaviest rains falling in the months of April. Surprisingly we observe a decrease in malaria deaths in that month. This could be attributed to washing away of the mosquito larvae as a result of heavy downpours. All-cause mortality peaks two month after the heaviest rainfall or a month after the end of the heavy rainy season which is May. Further research is needed to determine the determinants of this higher mortality other than malaria transmission. Malaria mortality occurs throughout the year in the study area despite very low transmission observed in the months of August to September. This point to challenges in using EIR as a measure of transmission intensity and uncertainty in determining to what level should it should be reduced in order to significantly impact on all-cause and malaria specific mortality and morbidity. Significant mortality and morbidity has been suggested to occur at EIR as low as 1 infectious bite per person per year (Beier et al., 1999)

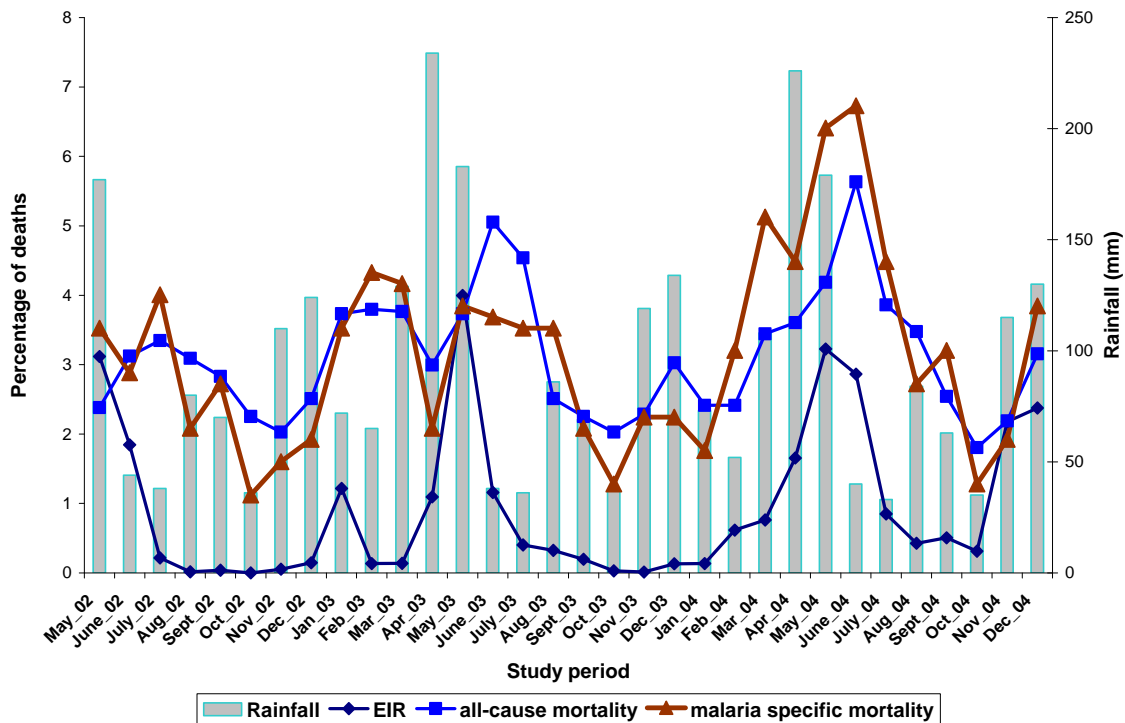


Figure 7.1 Distribution of all-cause and malaria specific mortality and malaria transmission during the study period

7.3 Limitations and challenges

The MTIMBA initiative was designed on a HDSS platform utilizing its infrastructure to collect or extract existing comprehensive entomological, demographic and intervention data. However, not all of these data were collected. For example, the KEMRI/CDC HDSS did not routinely monitor ITN use during the MTIMBA period (2002-2004). Analysis including ITN use in this thesis was only possible for 2002 data, which was obtained from a one time survey in 2002 in the area on net coverage. We could not assume the same status throughout the study period because there after, Kenyan government and other organizations through the Kenya ministry of health distribute ITNs to children under five years of age and pregnant women in the area. Unfortunately this was not routinely tracked by the HDSS.

Verbal autopsies (VA) among adolescents and adults (>11 years old) was introduced in 2003 in the study area (van Eijk et al., 2008). Therefore, the effect of malaria transmission on cause-specific mortality for this age group in 2002 adjusting for ITN use could not be assessed. . In general, VA technique tends to over estimate malaria deaths in malaria endemic areas, febrile illness with no other confirmed etiology are usually coded as malaria death (Abdullah et al., 2007). Malaria often shares symptoms with a number of diseases (e.g. anaemia, meningitis, acute respiratory infection) (Snow et al., 1992; Müller et al., 2003) which may also be assigned as malaria death resulting to over diagnosis of malaria deaths. Using validated VA malaria specific mortality based on hospital malaria deaths may give a better estimate of malaria community deaths and a more accurate pattern of malaria mortality-transmission relation in the study area.

Environmental and climatic predictors used in this work were extracted from satellite data. These data are generated in continuous scale at different spatial and temporal resolutions. However, entomological data was collected biweekly thus discrete. Lag analysis was used to align the two datasets. In particular, lag analysis refers to a technique of summarizing covariate values over a specific period (i.e. one, two and three months) prior to survey dates. The summarized values are linked to the study outcome to assess the time intervals that best explains the data. In our work, both datasets were aggregated on a monthly time series scale. This assumes that data collected at the start of the month e.g. 3rd have similar temporal values with the ones collected at the end of that month e.g. 28th. When this data is aggregated by obtaining the mean value for the month, it

mask the between point estimates variation which may lead to incorrect parameter estimates. Similarly, manual processing of satellite data for large spatial and temporal (e.g. HDSS) data is a challenge.

Selecting a set of predictors (for the optimum model) that explains best the observed data is an important process in statistical analysis. In our work, environmental and climatic covariates were selected by fitting non-spatial models using different combinations (Sogoba et al., 2007). The best combination was discussed with local entomologists and confirmed in the literature. However this procedure ignored the spatial and temporal correlation in the data. In Bayesian statistics, several variable selection methods have been developed including Kuo & Mallick (Kuo and Mallick, 1998), Gibbs variable selection (George and McCulloch, 1993) and Stochastic search variable selection (Dellaportas et al., 2002). But the implementation of these methods for geostatistical data is limited. Most recently, Gosoniu et al (2011) suggested that Gibbs variable selection has a better predictive ability in geostatistical modeling by analyzing malaria indicator survey data from Liberia. The use of Bayesian variable selection for entomological data may lead to a more parsimonious model. Most of these approaches can easily be formulated and implemented in OpenBugs, but can be computationally expensive for large datasets.

Mostly, the relationship between the predictor and outcome is not linear. Common approaches to adjust for non-linearity include logarithmic transformation, categorization and use of polynomial functions (Royston et al., 1999; Gemperli, Sogoba, et al., 2006; Gosoniu et al., 2008; Schur et al., 2011). In this thesis, we employed logarithmic transformation in assessing the relationship between transmission and mortality. Other approaches use penalized spline functions (Gosoniu et al., 2009; Magalhães et al., 2011), but the obtained coefficient estimates are difficult to interpret and cannot be applied easily.

Chapter 8: Extension and future research

This thesis has used EIR as a measure of malaria transmission. EIR is the most preferred measure (tool) to quantify the impact of control interventions such as ITNs, indoor residual spraying (IRS) on malaria transmission (Shaukat et al., 2010). However, EIR is influenced by a number of environmental factors such as rainfall (water for mosquito to lay eggs), temperature (affect sporogonic cycle), land use (water bodies-breeding sites) and control interventions. A change in any of these factors may affect the development and survival of mosquito at different stages leading to a decrease in EIR. Therefore it would be important to take into account the age of mosquito and mosquito survival probability when assessing variation in transmission intensity using EIR (Billingsley et al., 2005; Cook et al., 2008).

The malaria specific mortality data used in this work was obtained from VA. It would be interesting to compare pattern of transmission-malaria specific mortality relation using adjusted malaria specific mortality based on hospital malaria deaths with our results.

Chapter 9: Conclusion

One of the objectives of the MTIMBA initiative was to compare malaria transmission-mortality relation across its member sites. The results obtained in this thesis are from only one out of eight participating sites with some of these sites experiencing different transmission levels. Our results on the transmission-mortality relation are consistent with the ones obtained in Rufiji DSS which is also a member site (Rumisha et al., 2011a, 2011b). However, it would be necessary to compare these results with other sites before generalizing the conclusions outside these two settings. The findings reported in these two settings act as a baseline insight on the relationship between malaria transmission and mortality which could be used to monitor the progress of ongoing interventions. Recent studies in some of these settings have shown dramatic reduction in child mortality (Bhattarai et al., 2007; Hamel et al., 2011) and hospital admission attributed to malaria (Okiro et al., 2007) as a result of combined malaria control strategies. However, further analysis taking into account spatial and temporal effects are needed.

Bibliography

- Abdulla, S., Kikumbih, N., Masanja, H., Mshinda, H., Nathan, R., Savigny, D. de, 2001. Mosquito nets, poverty and equity in rural southern Tanzania.
- Abdullah, S., Adazu, K., Masanja, H., Diallo, D., Hodgson, A., Ilboudo-Sanogo, E., Nhalo, A., Owusu-Agyei, S., Thompson, R., Smith, T., Binka, F.N., 2007. Patterns of Age-Specific Mortality in Children in Endemic Areas of Sub-Saharan Africa. *The American Journal of Tropical Medicine and Hygiene* 77, 99 -105.
- Abeku, T.A., de Vlas, S.J., Borsboom, G., Teklehaimanot, A., Kebede, A., Olana, D., van Oortmarssen, G.J., Habbema, J.D.F., 2002. Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best. *Trop. Med. Int. Health* 7, 851-857.
- Adazu, K., Lindblade, K.A., Rosen, D.H., Odhiambo, F., Ofware, P., Kwach, J., Vaneijk, A.M., Decock, K.M., Amornkul, P., Karanja, D., Vulule, J.M., Slutsker, L., 2005. Health and demographic surveillance in rural western Kenya: a platform for evaluating interventions to reduce morbidity and mortality from infectious diseases. *Am J Trop Med Hyg* 73, 1151-1158.
- Adjuik, M., Smith, T., Clark, S., Todd, J., Garrib, A., Kinfu, Y., Kahn, K., Mola, M., Ashraf, A., Masanja, H., Adazu, K., Adazu, U., Sacarlal, J., Alam, N., Marra, A., Gbangou, A., Mwangi, E., Binka, F., 2006. Cause-specific mortality rates in sub-Saharan Africa and Bangladesh. *Bull. World Health Organ.* 84, 181-188.
- Agarwal, D., Gelfand, A., Pousty, C., 2002. Zero-inflated models with application to spatial count data. *Environmental and ecological statistics* 9, 341-355.
- Akaike, H., 1974. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* 19, 716-723.
- Al-Maktari, M.T., Bassiouny, H.K., Al-Hamd, Z.S., Assabri, A.M., El-Massry, A.G., Shatat, H.Z., 2003. Malaria status in Al-Hodeidah Governorate, Yemen: malariometric parasitic survey & chloroquine resistance *P. falciparum* local strain. *J Egypt Soc Parasitol* 33, 361-372.
- Alonso, P.L., Lindsay, S.W., Schellenberg, J.R.M.A., Gomez, P., Hill, A.G., David, P.H., Fegan, G., Cham, K., Greenwood, B.M., 1993. A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of The Gambia, West Africa: 2. Mortality and morbidity from malaria in the study area. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 87, Supplement 2, 13-17.
- Amek, N., Bayoh, N., Hamel, M., Lindblade, K.A., Gimnig, J., Laserson, K.F., Slutsker, L., Smith, T., Vounatsou, P., 2011. Spatio-temporal modeling of sparse geostatistical malaria sporozoite rate data using a zero inflated binomial model. *Spatial and Spatio-temporal Epidemiology* 2, 283-290.
- Amek, N.O., Bayoh, N., Hamel, M., Lindblade, K.A., Gimnig, J., Odhiambo, F., Laserson, K.F., Slutsker, L., Smith, T., Vounatsou, P., 2012. Spatial and temporal dynamics of malaria transmission in rural Western Kenya. *Parasites & Vectors* under review.
- Amek, N.O., Vounatsou, P., Obonyo, B., Hamel, M., Slutsker, L., Laserson, K.F., Using health and demographic surveillance system (HDSS) data to analyse spatio-temporal patterns of socio-economic status; an experience from KEMRI/CDC HDSS. in preparation.

Bibliography

- Anker, M., Back, R.E., Coldham, C., Kalter, H.D., Ross, D., Snow, R.W., 1999. A standard verbal autopsy method for investigating causes of death in infants and children [WWW Document]. WHO. URL <http://www.who.int/whosis/mort/verbalautopsystandards/en/>
- Baird, J.K., Purnomo, Masbar, S., 1990. Plasmodium ovale in Indonesia. *Southeast Asian J. Trop. Med. Public Health* 21, 541-544.
- Banerjee, S., Gelfand, A.E., 2002. Prediction, Interpolation and Regression for Spatially Misaligned Data. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 64, 227-245.
- Banerjee, S., Gelfand, A.E., Finley, A.O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. *J R Stat Soc Series B Stat Methodol* 70, 825-848.
- Barnes, K.I., Little, F., Mabuza, A., Mngomezulu, N., Govere, J., Durrheim, D., Roper, C., Watkins, B., White, N.J., 2008. Increased Gametocytemia after Treatment: An Early Parasitological Indicator of Emerging Sulfadoxine-Pyrimethamine Resistance in *Falciparum* Malaria. *Journal of Infectious Diseases* 197, 1605 -1613.
- Bayoh, M.N., Mathias, D.K., Odiere, M.R., Mutuku, F.M., Kamau, L., Gimnig, J.E., Vulule, J.M., Hawley, W.A., Hamel, M.J., Walker, E.D., 2010. Anopheles gambiae: historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province, Kenya. *Malar. J* 9, 62.
- Beier, J.C., Killeen, G.F., Githure, J.I., 1999. Short report: entomologic inoculation rates and Plasmodium falciparum malaria prevalence in Africa. *Am. J. Trop. Med. Hyg* 61, 109-113.
- Beier, J.C., Oster, C.N., Onyango, F.K., Bales, J.D., Sherwood, J.A., Perkins, P.V., Chumo, D.K., Koech, D.V., Whitmire, R.E., Roberts, C.R., Diggs, C.L., Hoffman, S.L., 1994. Plasmodium falciparum Incidence Relative to Entomologic Inoculation Rates at a Site Proposed for Testing Malaria Vaccines in Western Kenya. *Am J Trop Med Hyg* 50, 529-536.
- Beier, J.C., Perkins, P.V., Onyango, F.K., Gargan, T.P., Oster, C.N., Whitmire, R.E., Koech, D.K., Roberts, C.R., 1990. Characterization of Malaria Transmission by Anopheles (Diptera: Culicidae) in Western Kenya in Preparation for Malaria Vaccine Trials [WWW Document]. *Journal of Medical Entomology*. URL <http://www.ingentaconnect.com/content/esa/jme/1990/00000027/00000004/art00024>
- Bernadinelli, L., Pascutto, C., Best, N.G., Gilks, W.R., 1997. Disease Mapping with Errors in Covariates. *Statistics in Medicine* 16, 741-752.
- Berrang-Ford, L., Berke, O., Sweeney, S., Abdelrahman, L., 2010. Sleeping sickness in southeastern Uganda: a spatio-temporal analysis of disease risk, 1970-2003. *Vector Borne Zoonotic Dis* 10, 977-988.
- Bhattarai, A., Ali, A.S., Kachur, S.P., Mårtensson, A., Abbas, A.K., Khatib, R., Al-mafazy, A., Ramsan, M., Rotllant, G., Gerstenmaier, J.F., Molteni, F., Abdulla, S., Montgomery, S.M., Kaneko, A., Björkman, A., 2007. Impact of Artemisinin-Based Combination Therapy and Insecticide-Treated Nets on Malaria Burden in Zanzibar. *PLoS Med* 4, e309.
- Billingsley, P., Charlwood, J.D., Knols, B.G., 2005. Rapid assessment of malaria risk using entomological techniques: taking an epidemiological snapshot [WWW Document]. URL <http://library.wur.nl/ojs/index.php/frontis/article/viewArticle/960>
- Binka, F.N., Hodgson, A., Adjuik, M., Smith, T., 2002. Mortality in a seven-and-a-half-year follow-up of a trial of insecticide-treated mosquito nets in Ghana. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 96, 597-599.

Bibliography

- Binka, F.N., Kubaje, A., Adjuik, M., Williams, L.A., Lengeler, C., Maude, G.H., Armah, G.E., Kajihara, B., Adiamah, J.H., Smith, P.G., 1996. Impact of permethrin impregnated bednets on child mortality in Kassena-Nankana district, Ghana: a randomized controlled trial. *Tropical Medicine & International Health* 1, 147-154.
- Birley, M.H., Charlewood, J.D., 1987. Sporozoite rate and malaria prevalence. *Parasitology Today* 3, 231-232.
- Black, R.E., Morris, S.S., Bryce, J., 2003. Where and why are 10 million children dying every year? *The Lancet* 361, 2226-2234.
- Booyesen, F., Van Der Berg, S., Burger, R., Von Maltitz, M., Du Rand, G., 2008. Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries. *World Development* 36, 1113-1130.
- Breman, J.G., Alilio, M.S., Mills, A., 2004. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. *Am. J. Trop. Med. Hyg.* 71, 1-15.
- Briët, O.J.T., Vounatsou, P., Amerasinghe, P.H., 2008. Malaria seasonality and rainfall seasonality in Sri Lanka are correlated in space. *Geospat Health* 2, 183-190.
- Bryce, J., Boschi-Pinto, C., Shibuya, K., Black, R.E., 2005. WHO estimates of the causes of death in children. *The Lancet* 365, 1147-1152.
- Burkot, T.R., Graves, P.M., 1995. The value of vector-based estimates of malaria transmission. *Ann Trop Med Parasitol* 89, 125-134.
- Burkot, T.R., Graves, P.M., Paru, R., Wirtz, R.A., Heywood, P.F., 1988. Human malaria transmission studies in the *Anopheles punctulatus* complex in Papua New Guinea: sporozoite rates, inoculation rates, and sporozoite densities. *Am. J. Trop. Med. Hyg* 39, 135-144.
- Central Bureau of Statistics, (Kenya), Ministry of Health (Kenya), ORC Macro, 2004. Kenya Demographic and Health Survey 2003.
- Charlwood, J., Smith, T., Lyimo, E., Kitua, A., Masanja, H., Booth, M., Alonso, P., Tanner, M., 1998. Incidence of *Plasmodium falciparum* infection in infants in relation to exposure to sporozoite-infected anophelines. *Am J Trop Med Hyg* 59, 243-251.
- Charlwood, J.D., Kihonda, J., Sama, S., Billingsley, P.F., Hadji, H., Verhave, J.P., Lyimo, E., Luttkhuizen, P.C., Smith, T., 1995. The Rise and Fall of *Anopheles Arabiensis* (Diptera: Culicidae) in a Tanzanian Village. *Bulletin of Entomological Research* 85, 37-44.
- Coetzee, M., Craig, M., le Sueur, D., 2000. Distribution of African Malaria Mosquitoes Belonging to the *Anopheles gambiae* Complex. *Parasitology Today* 16, 74-77.
- Cook, P.E., McMeniman, C.J., O'Neill, S.L., 2008. Modifying insect population age structure to control vector-borne disease. *Adv. Exp. Med. Biol.* 627, 126-140.
- Craig, M.H., Snow, R.W., le Sueur, D., 1999. A Climate-based Distribution Model of Malaria Transmission in Sub-Saharan Africa. *Parasitology Today* 15, 105-111.
- Cressie, N.A., 1993. *Statistics for Spatial Data*, Revised ed. Wiley, New York.
- Dellaportas, P., Forster, J.J., Ntzoufras, I., 2002. On Bayesian model and variable selection using MCMC. *Statistics and Computing* 12, 27-36.
- Dercon, S., Shapiro, J.S., 2007. "Moving on, staying behind, and getting lost: lessons on poverty mobility from longitudinal data". Global Poverty Research Group Working Paper- 075.
- Dery, D.B., Brown, C., Asante, K.P., Adams, M., Dosoo, D., Amenga-Etego, S., Wilson, M., Chandramohan, D., Greenwood, B., Owusu-Agyei, S., 2010. Patterns and seasonality of malaria transmission in the forest-savannah transitional zones of Ghana. *Malar. J* 9, 314.

Bibliography

- Dhingra, N., Jha, P., Sharma, V.P., Cohen, A.A., Jotkar, R.M., Rodriguez, P.S., Bassani, D.G., Suraweera, W., Laxminarayan, R., Peto, R., 2010. Adult and child malaria mortality in India: a nationally representative mortality survey. *The Lancet* 376, 1768-1774.
- Diallo, D.A., Cousens, S.N., Cuzin-Ouattara, N., Nebié, I., Ilboudo-Sanogo, E., Esposito, F., 2004. Child mortality in a West African population protected with insecticide-treated curtains for a period of up to 6 years. *Bull World Health Organ* 82, 85-91.
- Diggle, P., Moyeed, R., Rowlingson, B., Thomson, M., 2002. Childhood Malaria in the Gambia: A Case-Study in Model-Based Geostatistics. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 51, 493-506.
- Diggle, P., Tawn, J., Moyeed, R., 1998. Model-based geostatistics. *Applied Statistics* 47, 299-350.
- Drakeley, C., Schellenberg, D., Kihonda, J., Sousa, C.A., Arez, A.P., Lopes, D., Lines, J., Mshinda, H., Lengeler, C., Schellenberg, J.A., Tanner, M., Alonso, P., 2003. An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. *Trop Med Int Health* 8, 767-774.
- Drakeley, C.J., Corran, P.H., Coleman, P.G., Tongren, J.E., McDonald, S.L.R., Carneiro, I., Malima, R., Lusingu, J., Manjurano, A., Nkya, W.M.M., Lemnge, M.M., Cox, J., Reyburn, H., Riley, E.M., 2005. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proceedings of the National Academy of Sciences of the United States of America* 102, 5108 -5113.
- Ecker, M., Gelfand, E., 1997. Bayesian Variogram Modeling for an Isotropic Spatial Process. *Journal of Agricultural, Biological and Environmental Statistics* 4, 347-368.
- van Eijk, A.M., Adazu, K., Ofware, P., Vulule, J., Hamel, M., Slutsker, L., 2008. Causes of deaths using verbal autopsy among adolescents and adults in rural western Kenya. *Tropical Medicine & International Health* 13, 1314-1324.
- Fernandes, M.V., Schmidt, A.M., Migon, H.S., 2009. Modelling zero-inflated spatio-temporal processes. *Statistical Modelling* 9, 3 -25.
- Filion, G.J.P., Paul, R.E.L., Robert, V., 2006. Transmission and immunity: the importance of heterogeneity in the fight against malaria. *Trends in Parasitology* 22, 345-348.
- Filipe, J.A.N., Riley, E.M., Drakeley, C.J., Sutherland, C.J., Ghani, A.C., 2007. Determination of the Processes Driving the Acquisition of Immunity to Malaria Using a Mathematical Transmission Model. *PLoS Comput Biol* 3, e255.
- Filmer, D., Pritchett, L.H., 2001. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. *Demography* 38, 115-132.
- Freeman, J.V., Christian, P., Khatry, S.K., Adhikari, R.K., LeClerq, S.C., Katz, J., Darmstadt, G.L., 2005. Evaluation of neonatal verbal autopsy using physician review versus algorithm-based cause-of-death assignment in rural Nepal. *Paediatric and Perinatal Epidemiology* 19, 323-331.
- Gallup, J., Sachs, J., 2001. The economic burden of malaria. *The American Journal of Tropical Medicine and Hygiene* 64, 85 -96.
- Garenne, M., Fontaine, O., 2006. Assessing probable causes of death using a standardized questionnaire: a study in rural Senegal. *Bull World Health Organ* 84, 248-253.
- Garnham, P.C., 1988. History of discoveries of malaria parasites and of their life cycles. *Hist Philos Life Sci* 10, 93-108.
- Gelfand, A., Smith, A., 1990. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85, 398-409.

Bibliography

- Gelfand, A.E., Ravishanker, N., Ecker, M.D., 2000. Modeling and Inference for Point-Referenced Binary Spatial Data.
- Gelman, A., 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statist. Sci.* 7, 457-472.
- Gemperli, A., Sogoba, N., Fondjo, E., Mabaso, M., Bagayoko, M., Briët, O.J.T., Anderegg, D., Liebe, J., Smith, T., Vounatsou, P., 2006. Mapping malaria transmission in West and Central Africa. *Trop. Med. Int. Health* 11, 1032-1046.
- Gemperli, A., Vounatsou, P., Kleinschmidt, I., Bagayoko, M., Lengeler, C., Smith, T., 2004. Spatial Patterns of Infant Mortality in Mali: The Effect of Malaria Endemicity. *American Journal of Epidemiology* 159, 64 -72.
- Gemperli, A., Vounatsou, P., Sogoba, N., Smith, T., 2006. Malaria Mapping Using Transmission Models: Application to Survey Data from Mali. *American Journal of Epidemiology* 163, 289 -297.
- Genton, B., D'Acremont, V., Rare, L., Baea, K., Reeder, J.C., Alpers, M.P., Müller, I., 2008. *Plasmodium vivax* and Mixed Infections Are Associated with Severe Malaria in Children: A Prospective Cohort Study from Papua New Guinea. *PLoS Med* 5, e127.
- George, E.I., McCulloch, R.E., 1993. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association* 88, 881-889.
- Gething, P.W., Patil, A.P., Smith, D.L., Guerra, C.A., Elyazar, I.R., Johnston, G.L., Tatem, A.J., Hay, S.I., 2011. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malaria Journal* 10, 378.
- Gimnig, J.E., Kolczak, M.S., Hightower, A.W., Vulule, J.M., Schoute, E., Kamau, L., Phillips-Howard, P.A., ter Kuile, F.O., Nahlen, B.L., Hawley, W.A., 2003. Effect of permethrin-treated bed nets on the spatial distribution of malaria vectors in western kenya. *Am J Trop Med Hyg* 68, 115-120.
- Gimnig, J.E., Vulule, J.M., Lo, T.Q., Kamau, L., Kolczak, M.S., Phillips-Howard, P.A., Mathenge, E.M., ter Kuile, F.O., Nahlen, B.L., Hightower, A.W., Hawley, W.A., 2003. Impact of permethrin-treated bed nets on entomologic indices in an area of intense year-round malaria transmission. *Am. J. Trop. Med. Hyg* 68, 16-22.
- Githeko, A.K., Service, M.W., Mbogo, C.M., Atieli, F.K., Juma, F.O., 1993. *Plasmodium falciparum* sporozoite and entomological inoculation rates at the Ahero rice irrigation scheme and the Miwani sugar-belt in western Kenya. *Ann Trop Med Parasitol* 87, 379-391.
- Le Goff, G., Carnevale, P., Fondjo, E., Robert, V., 1997. Comparison of three sampling methods of man-biting anophelines in order to estimate the malaria transmission in a village of south Cameroon. *Parasite* 4, 75-80.
- Gosoni, D.G., Vounatsou, P., Kahn, K., Tillé, Y., 2012. Geostatistical modeling of large non-Gaussian irregularly distributed data. *Computational Statistics & Data Analysis*.
- Gosoni, L., Vounatsou, P., Sogoba, N., Maire, N., Smith, T., 2009. Mapping malaria risk in West Africa using a Bayesian nonparametric non-stationary model. *Computational Statistics & Data Analysis* 53, 3358-3371.
- Gosoni, L., Vounatsou, P., Sogoba, N., Smith, T., 2006. Bayesian modelling of geostatistical malaria risk data. *Geospat Health* 1, 127-139.
- Gosoni, L., Vounatsou, P., Tami, A., Nathan, R., Grundmann, H., Lengeler, C., 2008. Spatial effects of mosquito bednets on child mortality. *BMC Public Health* 8, 356.

Bibliography

- Greenwood, B., 2009. Can malaria be eliminated? *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103, S2-S5.
- Greenwood, B., Mutabingwa, T., 2002. Malaria in 2002. *Nature* 415, 670-672.
- Greenwood, B.M., Bojang, K., Whitty, C.J., Targett, G.A., 2005. Malaria. *The Lancet* 365, 1487-1498.
- Greenwood, B.M., Greenwood, A.M., Bradley, A.K., Tulloch, S., Hayes, R., Oldfield, F.S., 1987. Deaths in infancy and early childhood in a well-vaccinated, rural, West African population. *Ann Trop Paediatr* 7, 91-99.
- Griffin, J.T., Hollingsworth, T.D., Okell, L.C., Churcher, T.S., White, M., Hinsley, W., Bousema, T., Drakeley, C.J., Ferguson, N.M., Basáñez, M.-G., Ghani, A.C., 2010. Reducing *Plasmodium falciparum* Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies. *PLoS Med* 7, e1000324.
- Hall, D.B., 2000. Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics* 56, 1030-1039.
- Hamel, M.J., Adazu, K., Obor, D., Sewe, M., Vulule, J., Williamson, J.M., Slutsker, L., Feikin, D.R., Laserson, K.F., 2011. A Reversal in Reductions of Child Mortality in Western Kenya, 2003–2009. *The American Journal of Tropical Medicine and Hygiene* 85, 597 - 605.
- Harbach, R.E., 1994. Review of the Internal Classification of the Genus *Anopheles* (Diptera: Culicidae): The Foundation for Comparative Systematics and Phylogenetic Research. *Bulletin of Entomological Research* 84, 331-342.
- Harbach, R.E., 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull. Entomol. Res.* 94, 537-553.
- Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R.F., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W., 2009. A World Malaria Map: *Plasmodium falciparum* Endemicity in 2007. *PLoS Med* 6, e1000048.
- Hightower, A.W., Ombok, M., Otieno, R., Odhiambo, R., Oloo, A.J., Lal, A.A., Nahlen, B.L., Hawley, W.A., 1998. A geographic information system applied to a malaria field study in western Kenya. *Am. J. Trop. Med. Hyg* 58, 266-272.
- Houweling, T.A., Kunst, A.E., Mackenbach, J.P., 2003. Measuring health inequality among children in developing countries: does the choice of the indicator of economic status matter? *Int J Equity Health* 2, 8.
- INDEPTH network, 2002. Population and Health in Developing Countries: Population, health and survival at INDEPTH sites. IDRC.
- Jambulingam, P., Mohapatra, S.S., Das, L.K., Das, P.K., Rajagopalan, P.K., 1989. Detection of *Plasmodium ovale* in Koraput district, Orissa state. *Indian J. Med. Res.* 89, 115-116.
- Kasasa, S., Rumisha, S.F., Amek, N.O., Diboulo, E., Huho, B.J., Smith, T., Vounatsou, P., Malaria transmission intensity and mortality burden across Africa project; statistical issues and approaches to data analysis. in preparation.
- Kawamoto, F., Liu, Q., Ferreira, M.U., Tantular, I.S., 1999. How prevalent are *Plasmodium ovale* and *P. malariae* in East Asia? *Parasitol. Today (Regul. Ed.)* 15, 422-426.
- Kayla, L.F., Odhiambo, F., Hamel, M.J., Feikin, D.R., Bayoh, N., Amek, N.O., Sewe, M., Ogwang, S., Obor, D., 2009. Health and Demographic Surveillance System, Kisumu, Kenya, Fifth Annual Report, Sixth Year of Registration of Demographic Events, 2007.

Bibliography

- Kazembe, L.N., Appleton, C.C., Kleinschmidt, I., 2007. Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. *Geospat Health* 2, 41-50.
- Kazembe, L.N., Kleinschmidt, I., Holtz, T.H., Sharp, B.L., 2006. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. *Int J Health Geogr* 5, 41.
- Kemp, C.D., Kemp, A.W., 1988. Rapid Estimation for Discrete Distributions. *Journal of the Royal Statistical Society. Series D (The Statistician)* 37, 243-255.
- Kitron, U., Spielman, A., 1989. Suppression of Transmission of Malaria Through Source Reduction: Antianopheline Measures Applied in Israel, the United States, and Italy. *Review of Infectious Diseases* 11, 391 -406.
- Kleinschmidt, I., Bagayoko, M., Clarke, G.P., Craig, M., Le Sueur, D., 2000. A spatial statistical approach to malaria mapping. *Int J Epidemiol* 29, 355-361.
- Kolenikov, S., Angeles, G., 2005. The Use of Discrete Data in Principal Component Analysis for Socio-Economic Status Evaluation. *Population English Edition* 28, 1-52.
- Kolenikov, S., Angeles, G., C, J.C., 2004. The Use of Discrete Data in PCA: Theory, Simulations, and Applications to Socioeconomic Indices.
- Korenromp, E.L., Williams, B.G., Gouws, E., Dye, C., Snow, R.W., 2003. Measurement of trends in childhood malaria mortality in Africa: an assessment of progress toward targets based on verbal autopsy. *The Lancet Infectious Diseases* 3, 349-358.
- Krishna, A., Kristjanson, P., Radeny, M., Nindo, W., 2004. Escaping Poverty and Becoming Poor in 20 Kenyan Villages. *Journal of Human Development* 5, 211.
- Krotoski, W.A., 1989. The hypnozoite and malaria relapse (review). *Prog Clin Parasitol* 1, 1-19
- Kuo, L., Mallick, B., 1998. Variable Selection for Regression Models.
- Lam, K.F., Xue, H., Cheung, Y.B., 2006. Semiparametric analysis of zero-inflated count data. *Biometrics* 62, 996-1003.
- Lambert, D., 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* 34, 1-14.
- Lengeler, 2004. Insecticide-treated bednets and curtains for preventing malaria: a cochrane review. *The Cochrane Library Update Software* CD000363.
- Lim, S.S., Fullman, N., Stokes, A., Ravishankar, N., Masiye, F., Murray, C.J.L., Gakidou, E., 2011. Net Benefits: A Multicountry Analysis of Observational Data Examining Associations between Insecticide-Treated Mosquito Nets and Health Outcomes. *PLoS Med* 8, e1001091.
- Lindblade, K.A., Eisele, T.P., Gimnig, J.E., Alaii, J.A., Odhiambo, F., ter Kuile, F.O., Hawley, W.A., Wannemuehler, K.A., Phillips-Howard, P.A., Rosen, D.H., Nahlen, B.L., Terlouw, D.J., Adazu, K., Vulule, J.M., Slutsker, L., 2004. Sustainability of Reductions in Malaria Transmission and Infant Mortality in Western Kenya With Use of Insecticide-Treated Bednets. *JAMA: The Journal of the American Medical Association* 291, 2571 -2580.
- Lindsay, S.W., Parson, L., Thomas, C.J., 1998. Mapping the ranges and relative abundance of the two principal African malaria vectors, *Anopheles gambiae sensu stricto* and *An. arabiensis*, using climate data. *Proc Biol Sci* 265, 847-854.
- Lines, J.D., Curtis, C.F., Wilkes, T.J., Njunwa, K.J., 1991. Monitoring Human-Biting Mosquitoes (Diptera: Culicidae) in Tanzania with Light-Traps Hung Beside Mosquito Nets. *Bulletin of Entomological Research* 81, 77-84.
- MacDonald G, 1957. *The epidemiology and control of malaria*. Oxford University Press, London.

Bibliography

- Magalhães, R.J.S., Clements, A.C.A., Patil, A.P., Gething, P.W., Brooker, S., 2011. The Applications of Model-Based Geostatistics in Helminth Epidemiology and Control. *Adv Parasitol* 74, 267-296.
- Manda, S., Meyer, R., 2005. Age at First Marriage in Malawi: A Bayesian Multilevel Analysis Using a Discrete Time-to-Event Model. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 168, 439-455.
- Manh, B.H., Clements, A.C.A., Thieu, N.Q., Hung, N.M., Hung, L.X., Hay, S.I., Hien, T.T., Wertheim, H.F.L., Snow, R.W., Horby, P., 2011. Social and environmental determinants of malaria in space and time in Viet Nam. *Int J Parasitol* 41, 109-116.
- Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8, 1235-1246.
- Mathers, C.D., Fat, D.M., Inoue, M., Rao, C., Lopez, A.D., 2005. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull. World Health Organ* 83, 171-177.
- McGuinness, D., Koram, K., Bennett, S., Wagner, G., Nkrumah, F., Riley, E., 1998. Clinical case definitions for malaria: clinical malaria associated with very low parasite densities in African infants. *Trans. R. Soc. Trop. Med. Hyg.* 92, 527-531.
- McKenzie, D.J., 2005. Measuring inequality with asset indicators. *Journal of Population Economics* 18, 229-260.
- Mendis, K., Sina, B., Marchesini, P., Carter, R., 2001. The neglected burden of *Plasmodium vivax* malaria. *The American Journal of Tropical Medicine and Hygiene* 64, 97 -106.
- Midega, J.T., Mbogo, C.M., Mwnambi, H., Wilson, M.D., Ojwang, G., Mwangangi, J.M., Nzovu, J.G., Githure, J.I., Yan, G., Beier, J.C., 2007. Estimating dispersal and survival of *Anopheles gambiae* and *Anopheles funestus* along the Kenyan coast by using mark-release-recapture methods. *J. Med. Entomol* 44, 923-929.
- Molineaux, L., 1997. Malaria and mortality: some epidemiological considerations. *Ann Trop Med Parasitol* 91, 811-825.
- Molineaux, L., Muir, D.A., Spencer, H.C., Wernsdorfer, W.H., 1988. The epidemiology of malaria and its measurement, *Malaria: Principles and Practice of Malariology*. ed, Churchill Livingstone. Wernsdorfer WH, McGregor I.
- Montgomery, M.R., Gragnolati, M., Burke, K.A., Paredes, E., 2000. Measuring Living Standards with Proxy Variables. *Demography* 37, 155-174.
- Mueller, I., Zimmerman, P.A., Reeder, J.C., 2007. *Plasmodium malariae* and *Plasmodium ovale* – the “bashful” malaria parasites. *Trends in Parasitology* 23, 278-283.
- Müller, O., Traoré, C., Becher, H., Kouyaté, B., 2003. Malaria morbidity, treatment-seeking behaviour, and mortality in a cohort of young children in rural Burkina Faso. *Tropical Medicine & International Health* 8, 290-296.
- Murray, C.J., Lopez, A.D., 1997. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 349, 1269-1276.
- Murray, C.J., Rosenfeld, L.C., Lim, S.S., Andrews, K.G., Foreman, K.J., Haring, D., Fullman, N., Naghavi, M., Lozano, R., Lopez, A.D., 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet* 379, 413-431.
- Mutuku, F.M., Alaii, J.A., Bayoh, M.N., Gimnig, J.E., Vulule, J.M., Walker, E.D., Kabiru, E., Hawley, W.A., 2006. Distribution, description, and local knowledge of larval habitats of *anopheles gambiae* s.l. in a village in western Kenya. *Am J Trop Med Hyg* 74, 44-53.

Bibliography

- Muturi, E.J., Burgess, P., Novak, R.J., 2008. Malaria Vector Management: Where Have We Come From and Where Are We Headed? *The American Journal of Tropical Medicine and Hygiene* 78, 536-537.
- Nevill, C.G., Some, E.S., Mung'ala, V.O., Muterni, W., New, L., Marsh, K., Lengeler, C., Snow, R.W., 1996. Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast. *Tropical Medicine & International Health* 1, 139-146.
- Nobre, A.A., Schmidt, A.M., Lopes, H.F., 2005. Spatio-temporal models for mapping the incidence of malaria in Pará. *Environmetrics* 16, 291-304.
- O'Meara, W.P., Mwangi, T.W., Williams, T.N., McKenzie, F.E., Snow, R.W., Marsh, K., 2008. Relationship Between Exposure, Clinical Malaria, and Age in an Area of Changing Transmission Intensity. *The American Journal of Tropical Medicine and Hygiene* 79, 185-191.
- Obrist, B., Iteba, N., Lengeler, C., Makemba, A., Mshana, C., Nathan, R., Alba, S., Dillip, A., Hetzel, M.W., Mayumana, I., Schulze, A., Mshinda, H., 2007. Access to Health Care in Contexts of Livelihood Insecurity: A Framework for Analysis and Action. *PLoS Med* 4, e308.
- Odhiambo F.O, Laserson K.F, Sewe M, Hamel M.J, Feikin D.R, Adazu K, Ogwang S, Obor D, Amek N, Bayoh N, Ombok M, Lindblade K, Desai M, Kuile F, Phillips-Howard P, Eijk A. M, Rosen D, Hightower A, Ofware P, Muttai H, Nahlen B, DeCock K, Slutsker L, Breiman R.F and Vulule J.M . (2012) Profile: The KEMRI/CDC Health and Demographic Surveillance System—Western Kenya. *International Journal of Epidemiology* 2012; 41:977–987
- Okiro, E., Hay, S., Gikandi, P., Sharif, S., Noor, A., Peshu, N., Marsh, K., Snow, R., 2007. The decline in paediatric malaria admissions on the coast of Kenya. *Malaria Journal* 6, 151.
- Ombok, M., Adazu, K., Odhiambo, F., Bayoh, N., Kiriinya, R., Slutsker, L., Hamel, M.J., Williamson, J., Hightower, A., Laserson, K.F., Feikin, D.R., 2010. Geospatial distribution and determinants of child mortality in rural western Kenya 2002-2005. *Tropical Medicine & International Health* 15, 423-433.
- Omumbo, J., Ouma, J., Rapuoda, B., Craig, M.H., le Sueur, D., Snow, R.W., 1998. Mapping malaria transmission intensity using geographical information systems (GIS): an example from Kenya. *Ann Trop Med Parasitol* 92, 7-21.
- Onori, E., Grab, B., 1980. Indicators for the forecasting of malaria epidemics. *Bull World Health Organ* 58, 91-98.
- Oti, S.O., Kyobutungi, C., 2010. Verbal autopsy interpretation: a comparative analysis of the InterVA model versus physician review in determining causes of death in the Nairobi DSS. *Population Health Metrics* 8, 21.
- Phillips-Howard, P.A., Nahlen, B.L., Alaii, J.A., Gimnig, J.E., Kachur, S.P., Hightower, A.W., Lal, A.A., Schoute, E., Oloo, A.J., Hawley, W.A., 2003a. The efficacy of permethrin-treated bed nets on child mortality and morbidity in Western Kenya i. development of infrastructure and description of study site. *Am J Trop Med Hyg* 68, 3-9.
- Phillips-Howard, P.A., Nahlen, B.L., Kolczak, M.S., Hightower, A.W., Ter Kuile, F.O., Alaii, J.A., Gimnig, J.E., Arudo, J., Vulule, J.M., Odhacha, A., Kachur, S.P., Schoute, E., Rosen, D.H., Sexton, J.D., Oloo, A.J., Hawley, W.A., 2003b. Efficacy of Permethrin-Treated Bed Nets in the Prevention of Mortality in Young Children in an Area of High

Bibliography

- Perennial Malaria Transmission in Western Kenya. *The American Journal of Tropical Medicine and Hygiene* 68, 23 -29.
- Po, J.Y.T., Subramanian, S.V., 2011. Mortality Burden and Socioeconomic Status in India. *PLoS ONE* 6, e16844.
- Quigley, M.A., Armstrong Schellenberg, J.R., Snow, R.W., 1996. Algorithms for verbal autopsies: a validation study in Kenyan children. *Bull World Health Organ* 74, 147-154.
- Rajaratnam, J.K., Marcus, J.R., Flaxman, A.D., Wang, H., Levin-Rector, A., Dwyer, L., Costa, M., Lopez, A.D., Murray, C.J.L., 2010. Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. *Lancet* 375, 1988-2008.
- Ramis-Prieto, R., García-Pérez, J., Pollán, M., Aragonés, N., Pérez-Gómez, B., López-Abente, G., 2007. Modelling of municipal mortality due to haematological neoplasias in Spain. *Journal of Epidemiology and Community Health* 61, 165 -171.
- Reeves, B.C., Quigley, M., 1997. A review of data-derived methods for assigning causes of death from verbal autopsy data. *International Journal of Epidemiology* 26, 1080 -1089.
- Ridout, M., Hinde, J., Demétrio, C.G.B., 2001. A Score Test for Testing a Zero-Inflated Poisson Regression Model against Zero-Inflated Negative Binomial Alternatives. *Biometrics* 57, 219-223.
- Riedel, N., Vounatsou, P., Miller, J.M., Gosoni, L., Chizema-Kawesha, E., Mukonka, V., Steketee, R.W., 2010. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malar J* 9, 37.
- Roberts, L., Enserink, M., 2007. Malaria. Did they really say ... eradication? *Science* 318, 1544-1545.
- Ross, A., Maire, N., Molineaux, L., Smith, T., 2006. An epidemiologic model of severe morbidity and mortality caused by plasmodium falciparum. *The American Journal of Tropical Medicine and Hygiene* 75, 63 -73.
- Rowe, A.K., Rowe, S.Y., Snow, R.W., Korenromp, E.L., Schellenberg, J.R.A., Stein, C., Nahlen, B.L., Bryce, J., Black, R.E., Steketee, R.W., 2006. The burden of malaria mortality among African children in the year 2000. *International Journal of Epidemiology* 35, 691 -704.
- Royston, P., Ambler, G., Sauerbrei, W., 1999. The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology* 28, 964 -974.
- Rumisha, S.F., Gosoni, D.G., Kasasa, S., Smith, T., Abdulla, S., Masanja, H., Vounatsou, P., 2012. Bayesian modeling of large geostatistical data to estimate seasonal and spatial variation of sporozoite rate. *Journal of Applied Statistics*.
- Rumisha, S.F., Smith, T., Masanja, H., Abdulla, S., Vounatsou, P., 2011a. Malaria transmission intensity and mortality in older children and adults in Rufiji DSS, Tanzania. in preparation.
- Rumisha, S.F., Smith, T., Masanja, H., Abdulla, S., Vounatsou, P., 2011b. Assessing the relation between child survival and malaria transmission: an analysis of the MTIMBA data in Rufiji DSS, Tanzania. in preparation.
- Salum, F.M., Wilkes, T.J., Kivumbi, K., Curtis, C.F., 1994. Mortality of under-fives in a rural area of holoendemic malaria transmission. *Acta Tropica* 58, 29-34.

Bibliography

- Sartorius, B., Kahn, K., Collinson, M.A., Vounatsou, P., Tollman, S.M., 2011. Survived infancy but still vulnerable: spatial-temporal trends and risk factors for child mortality in the Agincourt rural sub-district, South Africa, 1992-2007. *Geospat Health* 5, 285-295.
- Sasiwongsaroj, K., 2010. Socioeconomic inequalities in child mortality: A comparison between Thai Buddhists and Thai Muslims. *Journal of health research* 24, 81-86.
- Savigny, D. de, Mwageni, E., Masanja, H., Juma, Z., Momburi D., Mkilindi Y, Mbuya C, Kasale H, Reid G., 2002. Household wealth ranking and risks of malaria mortality in rural Tanzania.
- Schellenberg, J.A., Victora, C.G., Mushi, A., de Savigny, D., Schellenberg, D., Mshinda, H., Bryce, J., 2003. Inequities among the very poor: health care for children in rural southern Tanzania. *The Lancet* 361, 561-566.
- Schellenberg, J.R.A., Abdulla, S., Nathan, R., Mukasa, O., Marchant, T.J., Kikumbih, N., Mushi, A.K., Mponda, H., Minja, H., Mshinda, H., Tanner, M., Lengeler, C., 2001. Effect of large-scale social marketing of insecticide-treated nets on child survival in rural Tanzania. *The Lancet* 357, 1241-1247.
- Schur, N., Hürlimann, E., Garba, A., Traoré, M.S., Ndir, O., Ratard, R.C., Tchuem Tchuenté, L.-A., Kristensen, T.K., Utzinger, J., Vounatsou, P., 2011. Geostatistical Model-Based Estimates of Schistosomiasis Prevalence among Individuals Aged ≤ 20 Years in West Africa. *PLoS Negl Trop Dis* 5, e1194.
- Scott, J.A., Brogdon, W.G., Collins, F.H., 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg* 49, 520-529.
- Service, B., Townson, H., 2002. *The Anopheles vector*, Forth ed. Warrell DA and Gilles HM.
- Shaukat, A., Breman, J., McKenzie, F.E., 2010. Using the entomological inoculation rate to assess the impact of vector control on malaria parasite transmission and elimination. *Malaria Journal* 9, 122.
- Singer, J.D., Willett, J.B., 1993. It's about Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics* 18, 155-195.
- Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S.S.G., Cox-Singh, J., Thomas, A., Conway, D.J., 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* 363, 1017-1024.
- Smith, T., Charlwood, J.D., Takken, W., Tanner, M., Spiegelhalter, D.J., 1995. Mapping the densities of malaria vectors within a single village. *Acta Tropica* 59, 1-18.
- Smith, T., Genton, B., Betuela, I., Rare, L., Alpers, M.P., 2002. Mosquito nets for the elderly? *Transactions of the Royal Society of Tropical Medicine and Hygiene* 96, 37-38.
- Smith., T.A., Leuenberger, R., Lengeler, C., 2001. Child mortality and malaria transmission intensity in Africa. *Trends in Parasitology* 17, 145-149.
- Snow, R.W., Bastos de Azevedo, I., Lowe, B.S., Kabiru, E.W., Nevill, C.G., Mwangusye, S., Kassiga, G., Marsh, K., Teuscher, T., 1994. Severe childhood malaria in two areas of markedly different *falciparum* transmission in east Africa. *Acta Trop* 57, 289-300.
- Snow, R.W., Craig, M., Deichmann, U., Marsh, K., 1999. Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. *Bull World Health Organ* 77, 624-640.
- Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., Hay, S.I., 2005. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 434, 214-217.

Bibliography

- Snow, R.W., Marsh, K., 1995. Will reducing *Plasmodium falciparum* transmission alter malaria mortality among African children? *Parasitology Today* 11, 188-190.
- Snow, R.W., Marsh, K., 2002. The consequences of reducing transmission of *Plasmodium falciparum* in Africa, in: *Advances in Parasitology*. Academic Press, pp. 235-264.
- Snow, R.W., Omumbo, J.A., Lowe, B., Molyneux, C.S., Obiero, J.-O., Palmer, A., Weber, M.W., Pinder, M., Nahlen, B., Obonyo, C., Newbold, C., Gupta, S., Marsh, K., 1997. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *The Lancet* 349, 1650-1654.
- Snow, R.W., Winstanley, M.T., Marsh, V.M., CRJC. Newton, Waruiru, C., Mwangi, I., Winstanley, P.A., Marsh, K., Snow, R.W., Forster, D., Winstanley, P.A., Marsh, K., Armstrong, J.R.M., 1992. Childhood deaths in Africa: uses and limitations of verbal autopsies. *The Lancet* 340, 351-355.
- Sogoba, N., Vounatsou, P., Bagayoko, M.M., Doumbia, S., Dolo, G., Gosoni, L., Traore, S.F., Toure, Y.T., Smith, T., 2007. The spatial distribution of *Anopheles gambiae sensu stricto* and *An. arabiensis* (Diptera: Culicidae) in Mali. *Geospat Health* 1, 213-222.
- Spiegelhalter, D.J., Nicola G. Best, Bradley P. Carlin, Angelika van der Linde, 2002. Bayesian measures of model complexity and fit. *Journal Of The Royal Statistical Society Series B*, *Journal Of The Royal Statistical Society Series B* 64, 583-639.
- Steketee, R.W., Sipilanyambe, N., Chimumbwa, J., Banda, J.J., Mohamed, A., Miller, J., Basu, S., Miti, S.K., Campbell, C.C., 2008. National Malaria Control and Scaling Up for Impact: The Zambia Experience through 2006. *The American Journal of Tropical Medicine and Hygiene* 79, 45 -52.
- Stolwijk, A.M., Straatman, H., Zielhuis, G.A., 1999. Studying seasonality by using sine and cosine functions in regression analysis. *J Epidemiol Community Health* 53, 235-238.
- Tanner, M., Savigny, D. de, 2008. Malaria eradication back on the table. *Bulletin of the World Health Organization* 86, 82-82.
- Thomson, M., Connor, S., D'Alessandro, U., Rowlingson, B., Diggle, P., Cresswell, M., Greenwood, B., 1999. Predicting malaria infection in Gambian children from satellite data and bed net use surveys: the importance of spatial correlation in the interpretation of results. *Am J Trop Med Hyg* 61, 2-8.
- Thomson, M.C., Connor, S.J., Milligan, P., Flasse, S.P., 1997. Mapping malaria risk in Africa: What can satellite data contribute? *Parasitology Today* 13, 313-318.
- Todd, J.E., De Francisco, A., O'Dempsey, T.J., Greenwood, B.M., 1994. The limitations of verbal autopsy in a malaria-endemic region. *Ann Trop Paediatr* 14, 31-36.
- Trape, J., Rogier, C., 1996. Combating malaria morbidity and mortality by reducing transmission. *Parasitology Today* 12, 236-240.
- Trouiller, P., Olliaro, P.L., 1998. Drug development output from 1975 to 1996: what proportion for tropical diseases? *Int. J. Infect. Dis.* 3, 61-63.
- UNreport, 2009. World mortality: <http://www.un.org/esa/population/publications/worldmortality/WMR2009.htm>.
- UNICEF, 2008. UNICEF 2009-The State of the World's Children reports [WWW Document]. URL <http://www.unicef.org/sowc/>
- Velema, J.P., Alihonou, E.M., Chippaux, J.-P., van Boxel, Y., Gbedji, E., Adegbin, R., Malaria morbidity and mortality in children under three years of age on the coast of Benin, West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 85, 430-435.

Bibliography

- Vieira, A.M.C., Hinde, J.P., Demetrio, C.G.B., 2000. Zero-inflated proportion data models applied to a biological control assay. *Journal of Applied Statistics* 27, 373.
- Vounatsou, P., Raso, G., Tanner, M., N'goran, E.K., Utzinger, J., 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. *Parasitology* 136, 1695-1705.
- Vyas, S., Kumaranayake, L., 2006. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy Plan* 21, 459-468.
- Wagstaff, A., 2000. Socioeconomic inequalities in child mortality: Comparison across nine developing countries. *Bulletin of the World Health Organization* 78.
- White, G.B., 1974. *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 68, 278-298.
- WHO malaria report, 2009. WHO | World Malaria Report 2009 [WWW Document]. URL http://www.who.int/malaria/world_malaria_report_2009/en/index.html
- WHO 1998) Roll Back Malaria: a Global Partnership. Geneva, World Health Organization: <http://www.rollbackmalaria.org/>
- Win, T.T., Lin, K., Mizuno, S., Zhou, M., Liu, Q., Ferreira, M.U., Tantular, I.S., Kojima, S., Ishii, A., Kawamoto, F., 2002. Wide distribution of *Plasmodium ovale* in Myanmar. *Trop. Med. Int. Health* 7, 231-239.
- Wirtz, R., Zavala, F., Charoenvit, Y., Campbell, G.H., Burkot, T.R., Schneider, I., Esser, K.M., Beaudoin, R.L., Andre, R.G., 1987. Comparative testing of monoclonal antibodies against *Plasmodium falciparum* sporozoites for ELISA development. *Bull World Health Organ* 65, 39-45.
- Woolhouse, M.E., Dye, C., Etard, J.F., Smith, T., Charlwood, J.D., Garnett, G.P., Hagan, P., Hii, J.L., Ndhlovu, P.D., Quinnell, R.J., Watts, C.H., Chandiwana, S.K., Anderson, R.M., 1997. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci. U.S.A* 94, 338-342.
- World Bank, 2001. *Measuring Inequality*.
- Yamey, G., 2000. African heads of state promise action against malaria. *BMJ* 320, 1228-1228.
- Yau, K.K.W., Wang, K., Lee, A.H., 2003. Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. *Biom. J.* 45, 437-452.
- Yé, Y., Sankoh, O., Kouyaté, B., Sauerborn, R., 2008. *Environmental Factors and Malaria Transmission Risk*. Ashgate.
- Zhang, Y., Bi, P., Hiller, J., 2007. Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models. *International Journal of Biometeorology* 52, 179-187.

Curriculum vitae

Personal details

Name: Nyagauara Ombek Amek
Official address: C/o HDSS Section
P.O Box 1578 Kisumu-Kenya
Email address: Official: namek@kemricdc.org
Personal: amekombekl@yahoo.com

Education background

2008-2012 PhD in Epidemiology
University of Basel, Switzerland
(Bayesian spatio-temporal modelling of the relationship between mortality and malaria transmission in rural western Kenya)
2003-2004 MSc. Biostatistics
University of Ibadan
1998-2002 Bachelor of Science (Mathematics)
University of Nairobi

Work experince

June 2005-Todate Senior data manager
KEMRI/CDC HDSS
Nov 2004-2006 Consultant/Facilitator
Data Perspective-Consultany and training firm in data Management and statistics in research
Nov 2004-May 2005 Senior data manager
Kesho Bora Study-WHO/UON/CDC Atlanta funded mulitcentre study

Conferences

2006 “HDSS week” (a conference for all HDSS sites in Kenya), oral presentation on “How big is adult mortality in the rural western Kenya”
2009 INDEPTH AGM- Accra, oral presentation on “ Spatio-temporal modelling of sparse geostatistical data: an experience from KEMRI/CDC HDSS”

Pulications

- 1) **Amek N**, Bayoh N, Hamel M, Lindblade KA, Gimnig J, Laserson, KF, Slutsker L, Smith T, Vounatsou P (2011). Spatio-temporal modeling of sparse geostatistical malaria sporozoite rate data using a zero inflated binomial model. *Spatial and Spatio-temporal Epidemiology* 2, 283-290.
- 2) **Amek N**, Bayoh N, Hamel M, Lindblade KA, Gimnig J, Odhiambo F, Laserson, KF, Slutsker L, Smith T, Vounatsou P (2012). Spatial and temporal dynamics of malaria transmission in rural western Kenya. *Parasites & Vectors*: under review.
- 3) O'Reilly E. C, Jaron P, Ochieng B, **Amek N**, Tate J. E, Parsons M. B, Bopp C.A, Williams K. A, Vinje' J, Blanton E, Wannemuehler K. A, Vulule J, Laserson K.F, Breiman R.F, Feikin D. R, Widdowson M. A, Eric Mintz E (2012). Risk Factors for

Death among Children Less than 5 Years Old Hospitalized with Diarrhea in Rural Western Kenya, 2005–2007: A Cohort Study PLoS Med 7:9

- 4) Odhiambo F.O, Laserson K.F, Sewe M, Hamel M.J, Feikin D.R, Adazu K, Ogwang S, Obor D, **Amek N**, Bayoh N, Ombok M, Lindblade K, Desai M, Kuile F, Phillips-Howard P, Eijk A. M, Rosen D, Hightower A, Ofware P, Muttai H, Nahlen B, DeCock K, Slutsker L, Breiman R.F and Vulule J.M . (2012) Profile: The KEMRI/CDC Health and Demographic Surveillance System—Western Kenya. International Journal of Epidemiology 2012; 41:977–987