Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission

Gillian H. Stresman

Thesis submitted in accordance with the requirements for the degree of
Doctor of Philosophy of the
University of London

February 2015

Department of Immunology and Infection
Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

This research was funded by the Bill and Melinda Gates Foundation through the Malaria Transmission Consortium

Declaration by candidate

I, Gillian H. Stresman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: ___  _____ Date: ____17/02/2015_____

(Gillian H. Stresman)

Abstract

Heterogeneous malaria exposure may result in distinct clusters of higher malaria burden, or hotspots, across space and time. Targeting control programs to these areas may be highly efficient, however, operationally attractive approaches for identifying hotspots are needed for any such program to be sustained by local malaria control programs. The principal aim of this project was to investigate the ability of convenient sampling to identify hotspots of malaria transmission in a low endemic transmission setting in the western Kenyan highlands: 1) The boundaries of hotspots, and associated uncertainties, was determined using a large community survey; 2) The value of convenience sampling to estimate transmission in the community was assessed using cross-sectional surveys of 4964 children in 46 government primary schools and 3042 individuals in five rural-health facilities; 3) The value of compound-level screening of sentinel age groups that are likely to have patent level infections was determined and; 4) The potential use for convenience sampling in hotspot targeted approaches was assessed using spatial information on residences collected during the school and health-facility surveys. The community-based approach was able to detect 77% of the parasite infections in selected hotspots of malaria exposure using field-based tools in sentinel age groups. Both convenience-sampling approaches tested produced similar estimates of malaria transmission to the community when restricted to those residing in the same catchment areas and those testing positive for malaria were more likely to reside in a hotspot. The findings suggest that operationally attractive approaches provide reliable information on malaria transmission and may play an important role in targeted malaria control strategies. Future research on ascertaining what coverage of the hotspot is needed to see sustainable reductions in transmission would provide a threshold with which to gauge the utility of this strategy.

Acknowledgements

First, I would like to offer my appreciation to my supervisors Dr. Teun Bousema and Professor Chris Drakeley for guiding me and supporting me throughout this entire process and serving as a constant source of motivation. I would also like to thank Dr. Jonathan Cox and Professor Peter Diggle for the invaluable advice and interest in my work. I would also like to thank Dr. Jennifer Stevenson for making data checking by candlelight as pleasant an experience as it can be and to Amrish Baidjoe, whose ability to cook gourmet meals on a hotplate after a 14-hour day in the field will remain legendary.

I would also like to thank the Ministry of Health officials at the Rachuonyo District Hospital, in particular Dr. Tom Onyango, whose support of the malaria research study was invaluable. Also, I would like to acknowledge all those who worked long hours to collect and process the thousands of samples collected as part of the work in the Kenyan highlands.

I would like to acknowledge the Bill and Melinda Gates Foundation for funding this research through the Malaria Transmission Consortium and for the support of my colleagues at LSHTM who have provided advice, and friendship during my studies. In particular, I would like to thank Dr. Kevin Tetteh whose enthusiasm for maps and overall intellectual curiosity made working that much more enjoyable; Dr. Nuno Sepulvda who always managed to smile despite my barrage of questions; and Dr. Lynn Grignard for supporting me with the molecular work.

Finally, I would like to thank my friends and family for their unwavering support, love and encouragement and for keeping me motivated throughout this process.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACT: artemisinin combination therapy
AL: artemether-lumefantrine
AMA-1: apical membrane antigen
AOC: area under the curve
BMGF: Bill and Melinda Gates Foundation
DEM: digital elevation model
DI: discrimination index
DOMC: (Kenyan) Division of Malaria Control
EIR: entomologic inoculation rate
ELISA: enzyme-linked immune-sorbent assay
FMDA: focal mass drug administration
GIS: geographic information system
GPS: global positioning system
Hb: haemoglobin
HRP-2: histidine rich protein-2
IgG: immunoglobulin G
IRS: indoor residual spraying
ITN: insecticide treated net
LLIN: long-lasting insecticide treated net
LSHTM: London School of Hygiene & Tropical Medicine
MBG: model based geostatistics
MaIS: malaria indicator survey
MIS: multispectral image segmentation
MOPHS: Ministry of Public Health and Sanitation
MSP-1: merozoite surface protein 1
MSP-2: merozoite surface protein 2
MTC: malaria transmission consortium
NDVI: normalized differences vegetation index
OD: optical density
nPCR: nested polymerase chain reaction
PDA: personal digital assistant
pLDH: Plasmodium lactate dehydrogenase
RDT: rapid diagnostic test
R_O: base reproductive rate
R_C: reproductive rate under control
RMSE: root mean-square prediction error
ROC: receiver operator curve
SCR: sero-conversion rate
TWI: topographic wetness index
WHO: World Health Organization

Chapter 1: Introduction

Malaria is a complex infection. Control and elimination campaigns should be developed based on an understanding of the parasite biology, epidemiology, and transmission dynamics. This chapter provides background to the research presented in subsequent chapters and puts the work in context of the existing literature. Section 1.1 provides an overview of malaria epidemiology and discusses some of the immunological factors that are likely determinants of the observed population level patterns of parasite carriage and morbidity. Next, in section 1.2 and 1.3, heterogeneity in malaria transmission is explored including how to measure it both in terms of spatial statistical methods and the different malaria metrics available. Finally, the need for operational approaches to malaria control and elimination is discussed in section 1.4.

1.1 Burden of malaria and malaria epidemiology in transition

Since the etiology of malaria was first discovered at the end of the 19th century, significant progress has been made in understanding the basic biology of transmission, in reducing the clinical burden, and in controlling and eliminating infections. (1-4) However, despite the considerable progress made, there is still much about malaria biology and transmission dynamics that is not fully understood and malaria is still a significant public health burden. (5) Globally, it has been estimated between 198-451 million new cases and 584,000-1.2 million malaria deaths occur each year, a 47% global reduction since 2000. (5-8) Of the six species of human malaria, *Plasmodium falciparum* is responsible for the vast majority of all malaria deaths. There are an estimated 2.57 billion people at risk of *P. falciparum* malaria worldwide however; the burden of *P. falciparum* is disproportionately distributed with the majority of people at risk in Africa, where transmission of this species is the most intense. (9, 10)

Malaria is a vector-borne infection that has a complex lifecycle with distinct phases of the *Plasmodium* parasite in the human host and vector, the female *Anopheles* mosquito (figure 1-1). Although there are developmental processes unique to each *Plasmodium* species, such as the dormancy phase of *P. vivax* and the two species of *P. ovale* called hypnozoites, the underlying transmission cycle is similar. (3) Briefly,

parasites, in the form of sporozoites are injected into the human hosts through the bite of an infectious female *Anopheles* mosquito. The sporozoites quickly progress to the liver where they mature and are eventually released into the blood stream. Once in the blood, merozoites undergo a cycle of infecting red blood cells, multiplication, bursting out, and re-invading uninfected red blood cells; this cycle of invasion and rupturing of red blood cells is responsible for the clinical symptoms associated with malaria. (4, 11, 12) During the blood stage infection, some of the parasites differentiate into gametocytes, the stage that is infectious to mosquitoes. If a female *Anopheles* mosquito ingests gametocytes while taking a blood meal, the parasite will undergo a maturation cycle in the vector, known as the extrinsic incubation period. Once the sporozoites reach the salivary glands, the mosquito becomes infectious and is capable of transmitting sporozoites through their next human blood meal thereby maintaining the transmission cycle. (3) The focus of this research is *P. falciparum*, and therefore all subsequent discussion and mention of malaria is in reference to this species.

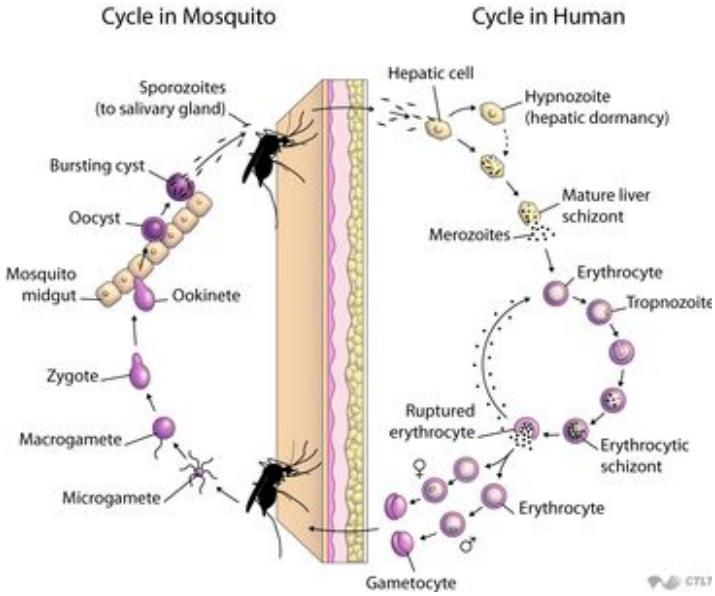

Figure 1-1: Schematic of the malaria life cycle depicting both the human and mosquito stages including the hypnozoite phase unique to *Plasmodium vivax* and the *P. ovales*. Image Credit: "Life cycle of the malaria parasite" from Epidemiology of Infectious Diseases. Available at: http://ocw.jhsph.edu. Copyright © Johns Hopkins Bloomberg School of Public Health.

1.1.1 Malaria Epidemiology – Transmission patterns

Classification of malaria risk has evolved over time to reflect the use of new diagnostic tools or improved understanding of transmission dynamics. (13) Commonly used malaria endemicity classifications include areas with <5% *P. falciparum* parasite rate (PfPR) as having low or unstable malaria transmission, areas with >40% PfPR as having high transmission and those in between being moderate. (14) These guidelines are useful for developing country-level recommendations for malaria control strategies (figure 1-2), however the broad classifications and the reliance on aggregated data may mask variability of local level transmission intensity. (9, 13, 15)

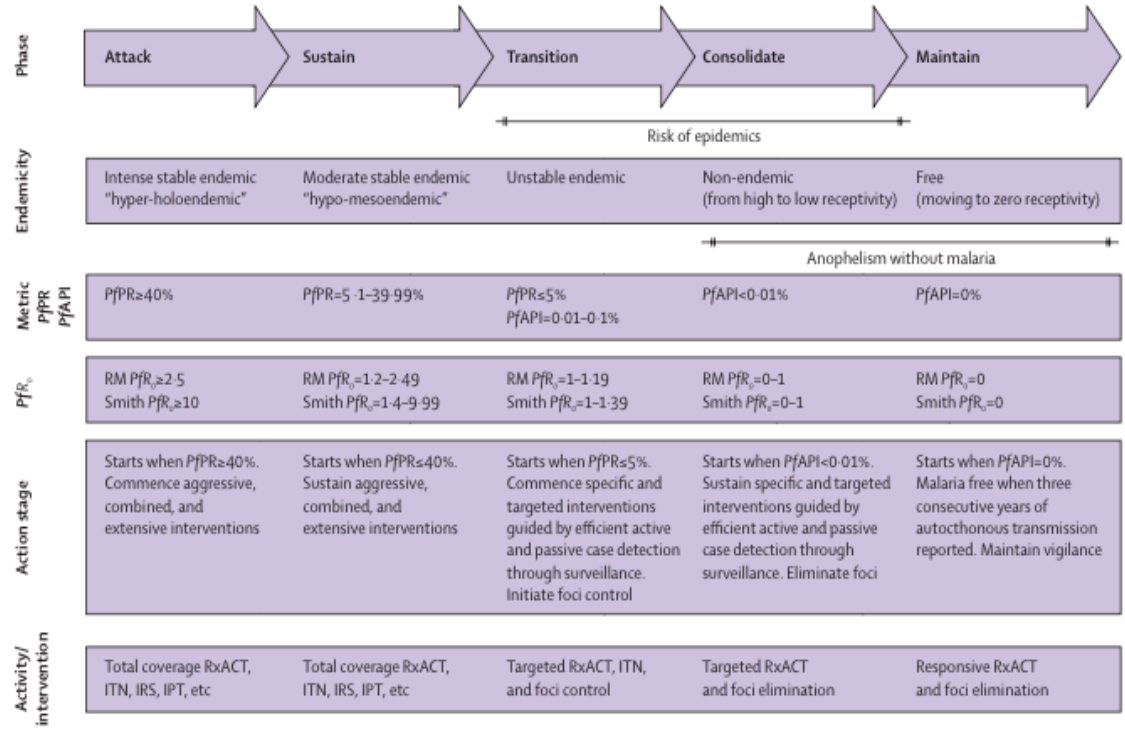

Figure 1-2: Classification of malaria endemicity and recommended phases and action for control and elimination. IRS=indoor residual spraying; IPT=intermittent preventive therapy; ITN=insecticide-treated net; PfAPI=P. falciparum annual parasite incidence; PfPR=P. falciparum parasite rate; RxACT=radical treatment with artemisinin-combination therapy. Image Credit: Hay et al., 2008 (13)

Transmission patterns exhibit high levels of variability and are driven by many factors that affect mosquito-human-parasite contact. (12) At the macro level, malaria transmission is largely driven by ecological conditions that are conducive to mosquito and parasite development. (2) The availability of standing water is essential for mosquitoes to breed and propagate. (2, 16) Therefore, malaria transmission tends to be seasonal with the highest prevalence occurring after the peak of the rainy season. (17) Historically, malaria, especially P. vivax was very widespread and covered the majority of the world whereas P. falciparum was largely confined to areas with higher temperatures. (15, 18, 19) The current range of malaria transmission is largely confined to areas with tropical climates. (20, 21, 22) Tropical temperatures are especially conducive for malaria transmission as the development of the parasite in the mosquito midgut is highly sensitive to ambient temperature; parasites mature more slowly in temperatures that are too cold or too hot as well as incurring a fitness cost for mosquito survival. (23, 24) Inversely, in warmer climates, while temperature is within the optimal range, parasite development is much faster and thereby facilitating transmission. (16)

Next, the ecological niches preferred by each of the 60 species of *Anopheles* identified as being competent for malaria parasite development largely determine the vector composition and the malaria risk profile in an area. (25-27) Each species has a different vectorial capacity, meaning that some are more efficient vectors than others. (28) Each species also exhibits differences in behavior, such as choice of blood meal and whether it feeds indoors or outdoors. (16) The susceptibility of each species to malaria control interventions will also influence the observed transmission patterns. For example, in areas with high ITN use, transmission may be (partly) driven by the vector species that preferentially feed outdoors or are resistant to the insecticide used. (29-31)

There are also several factors that facilitate transmission at the local level, including microclimates and human host, which have been identified as being associated with either increased risk or protection for malaria. (16, 32, 33) Risk factors associated with malaria are complex and will vary based on the microepidemiological characteristics of each region (12, 16) therefore; only the

predominant ones are highlighted here. Firstly, proximity to vector breeding sites such as swamps, irrigation, or floodplains has been found to increase risk of malaria and is likely associated with the mosquito flight range and preference in selecting blood meals. (17, 34) For example, a study in Uganda identified a dose-response relationship between risk of malaria and distance to swamp with those residing closest to the standing water having almost four times the risk of malaria compared to those living the furthest away. (28) House construction materials are also an important determinant of malaria risk. For example, houses with thatch roofs tend to provide easy entry increasing the risk of exposure as well as could protect mosquitoes from predators while resting to digest their blood meal. (35, 36) Similarly, houses with open eaves, a space between the roof and walls, have been associated with increased risk as they provide easy entry and exit points for mosquitoes. (16, 33, 35) Finally, household wealth has also been consistently associated with malaria where those in the poorest households have an estimated twice the odds of infection than the wealthiest groups. (33) The reason for the protective association of wealth on malaria burden is likely multifactorial. Wealthier households have more disposal income to spend on protective measures, prompt treatment, or improved house construction inhibiting mosquito entry. Furthermore, wealthier households also tend to have higher levels of education, which may contribute to a more informed and timely decision-making on malaria prevention and case-management. (37)

Several behavioral and innate factors have also been associated with risk of malaria. Studies have found that those who report sleeping under a bednet have a reduced odds of infection compared to those not sleeping under nets. (38) Those engaging in certain occupational endeavors such as gold mining or forestry, for example, or travel to malaria endemic areas have also been found to be at increased risk for malaria. (39-41) In addition, innate factors also mediate risk of infection. For example, evidence suggests that mosquitoes may be more attracted to malaria-infected individuals harboring gametocytes. (42, 43) Also, the chemical composition of individual odors has been found to repel or attract mosquitoes with the African *Anopheles* mosquitoes being particularly attracted to certain foot odors. (44, 45) Finally, genetic traits such as alpha thalassemia, sickle cell anaemia, and

glucose 6 phosphate dehydrogenase (G6PD) deficiency have also been found to be associated with malaria risk. (46) For example, a study in Uganda found that children with the sickle cell trait had a 32% reduced risk for malaria compared to those without a haemoglobinopathy, adjusting for ITN use. (28) Similarly, it has also been found that certain ethnic groups, such as the indigenous Fulani population, and their ability to mount a strong and efficient immune response reduces their susceptibility to malaria. (47)

1.1.2 Malaria Epidemiology – Impact of Antimalarial Immunity

Immunity to malaria is complex and not yet fully characterized and plays an important role in the observed patterns of disease. (4) Development of antimalarial immunity is associated with exposure to the malaria parasite and age. There are two phases of the immune response, after protection acquired from maternal antibodies wanes, which are important for malaria epidemiology. The first phase is anti-disease immunity, which provides protection against the severe forms of malaria pathology. In areas of intense transmission, anti-disease immunity typically occurs within the first five years of life (48-50) and is acquired with repeated exposure. (1, 51) For example, a study in Uganda found that those exposed to a larger variety of parasite clones, a proxy for transmission intensity, acquired a higher level of protection against symptomatic malaria. (52) The rate of development of anti-disease immunity influences the age distribution of symptomatic cases in a community. The development of anti-disease immunity also means that a large proportion of individuals, including young children, are asymptomatically infected and therefore not likely to present for treatment and be recorded by health facility based surveillance programs. (53) In fact, it has been found that in many settings, including those with low levels of transmission more than 75% of infections are carried in the absence of concurrent symptoms. (54)

As an individual develops protection from the severe forms of malaria, the immune system is also developing a response that is able to regulate parasite densities. (4) The comparably slowly forming anti-parasite immunity is more persistent, once acquired, with an estimated half-life four times greater than that of anti-disease immunity. However, in most infected individuals, even those with high antibody

levels, parasite densities tend to fluctuate with microscopically detectable malaria likely to be present at some points during each infection. (52, 55, 56) Parasite densities are important for malaria epidemiology in two ways. The first is that high parasite densities are a component cause of severe forms of malaria (57) reducing the likelihood that older individuals will present with symptomatic malaria. Second, as discussed below in section 1.3, the sensitivity of commonly used diagnostic tools in malaria endemic settings is limited when parasite densities are low meaning that a significant proportion of infections go undetected unless more sensitive detection tools are used. (58, 59). Furthermore, these difficult to detect reservoirs of parasite populations produce gametocytes and therefore plausibly contribute to maintaining malaria transmission. (53, 58, 60, 61)

1.1.3 Strategies for malaria control – reducing disease burden

Malaria control strategies have largely focused on vector control and prompt and effective treatment of symptomatic cases. Indoor residual spraying (IRS) involves spraying the interior surfaces of all structures using compounds with long-lasting insecticidal properties such as organophosphates (ie. dichlorodiphenyltrichloroethane (DDT)) or pyrethroids (ie. deltamethrin). IRS acts by both repelling mosquitoes away from the sleeping spaces and therefore reducing exposure and by killing mosquitoes that rest and digest their blood meal on sprayed surfaces reducing the potential for onward transmission. (62) The large-scale use of DDT was the primary intervention used in the 1950s and has been cited as enabling elimination in several countries. (2, 15) However, in areas of high endemicity and efficient vectors, the use of IRS is not likely to contribute significantly to transmission reduction. (63) More recently, studies have shown that the use of IRS is associated with a protective efficacy of between 33-92%, although it appears to be most effective in areas with high coverage levels and when the primary vectors are endophilic and endophagic. (64)

Insecticide treated bednets (ITNs) are increasingly been relied on as the primary tool for malaria control and studies suggest that they can provide better protection than IRS. (64, 65) ITNs are hung over sleeping spaces and provide those under the net a physical as well as chemical barrier from exposure to the mosquito. ITNs,

provide protection from malaria in a similar way to IRS with the main difference being that the ITN, if in good condition and hung properly, provides a physical barrier between the mosquito and its' potential blood meal. This barrier acts to reduce exposure to potentially infectious mosquitoes while simultaneously blocking the mosquito from ingesting gametocytes and becoming infectious. (3) Although they have been found to be effective at reducing malaria burden in individuals, ITNs work best when high community coverage is achieved. (66) However, in practice, ITN distribution has not been uniform and the use of bednets, particularly in adolescents and young adults, is poor. (67, 68) However, even with high coverage, the protective efficacy is not ubiquitous. For example, in a high transmission setting that had good public health capacity in conjunction with high ITN use, the burden of malaria, according to slide positivity rate, only decreased from 80.5% to 36.3% over 15 years. (69) The persistent transmission in areas with high ITN use may be due to residual transmission due to outdoor biting mosquitoes or due to people not being under the net at the time when mosquitoes are active. (29, 63) The combined effects of IRS and ITN have also been shown to have a significantly greater impact compared to a single intervention. However, the consistency across a range of transmission intensities and the utility of IRS in areas with universal ITN coverage still needs to be shown. (68, 70) Additional mosquito control measures available include larviciding, or larval source management: due to the difficulty in achieving high coverage of all potential breeding sites, the use of these methods is limited. (71, 72)

In addition to vector control, a strong public health system is an essential component of any malaria control strategy. Ensuring prompt and effective treatment with antimalarial drugs is important for reducing the risk of adverse outcomes and has also been associated with decreases in malaria burden and maintaining low levels of transmission despite reintroduction of parasites. (73, 74) While the recommended first line antimalarial drugs were inexpensive and effective, the policy of presumptive treatment for malaria in children and adults when no diagnostic tool was available was maintained. Resistance to the first line antimalarials became widespread prompting a switch to artemisinin and artemisnin based combination therapies (ACT). (75, 76) The use of ACTs and the

increasing availability in RDTs to provide point of care confirmatory testing lead to a change to a test and treat policy, whereby ACTs would be restricted to those with confirmed malaria. (5)

1.1.4 Strategies for Malaria Elimination - Targeting Transmission

Since it became apparent that the malaria burden was declining in some countries and in others, it was not as high as previously estimated, there has been a renewed focus on malaria elimination. (77) The substantial efforts to improve access and coverage to basic malaria control interventions such as ITNs and IRS has likely contributed to the estimated 30% decline in malaria burden observed between 2004 and 2010 but reductions have not been uniformly distributed. (7, 78-80) Of the countries with endemic malaria one third have reduced transmission substantially and are pursuing an elimination strategy. (30, 81) Factors leading to successful elimination of malaria are complex, not fully understood, and vary per setting. (82) There are also some systemic factors that are difficult to measure that likely have an impact on malaria burden, for example, the increased rate of urbanization and improved socio-economic status in many malaria endemic areas has been associated with a reduction in malaria risk. (33, 83) To achieve and maintain elimination, the infectious lifecycle stages within both the mosquito vector and human are increasingly becoming the focus of interventions. (73, 84)

As discussed, there are several factors that contribute to the capacity of a mosquito to transmit malaria and provide targets for blocking transmission. (2) In addition to the conventional mosquito control tools identified in section 1.1.3 there are also some new approaches being developed. Studies have shown that, during a blood-meal, mosquitoes can absorb drugs administered to people that can have an impact on transmission potential. (85) The use of ivermectin has been found to shorten the mosquito lifespan as well as delay the development of the parasite in the mosquito reducing the probability of onward transmission. (86) For example, a study in Burkina Faso found that ivermectin was associated with a 4- to 7- fold increase in mosquito mortality and up to a 35% reduction in malaria transmission from malaria-infected individuals. (85) In addition to the use of synthetic drugs to impact transmission capacity in the mosquito, there is also evidence of

transmission blocking immunity that develops in infected humans and is related to gametocyte exposure. This anti-gametocyte immunity transferred during a blood-meal may act by impeding the development of parasites within the mosquitos rendering them less infectious, although the overall impact this has on transmission is not yet known. (87) Transmission blocking vaccines that block parasite development in the mosquito are a novel tool being developed to contribute to reducing and eliminating malaria transmission, however potential candidates are still being evaluated. (88)

Understanding human infectiousness could also provide insight for reducing malaria transmission. The likelihood of transmission from human to mosquito has been associated with gametocyte density. Studies have found that the number of mosquitoes and the burden of infection within mosquitoes increase as parasite densities in the human host increases. (9, 61, 89) Also, research has shown that gametocyte densities are highest during the transmission season, (90) and that carriage is greatest in younger children with some studies reporting that 90% of children were gametocyte carriers at the end of the transmission season. (60) Antimalarial drug choice also impacts transmission potential: Fast clearance of asexual parasites inhibits the production of new gametocytes; drugs that clear immature and mature gametocytes reduce the duration and potential for infectiousness; and the prophylactic potential of the drug which is associated with the drugs' half-life (range ~ 3.2 to ~23-28 days) delays reinfection. (91) For example, studies have found that gametocytes can persist for more than one month after clearance of asexual parasiteamia but this can be shortened if treating with a drug with gametocytocidal properties such as artemisinin or primaquine. (92) Also, even with incomplete parasite clearance, the lower density of circulating gametocytes due to treatment can also render an individual less infectious to mosquitoes (93) and simulations suggest that this can contribute to reductions in malaria transmission, particularly in low endemic settings. (91, 94)

1.1.5 Malaria Elimination – Changing Epidemiology

The reductions in malaria burden achieved in African countries have not been uniformly distributed with the largest reductions occurring in areas where malaria

transmission was initially low to moderate with the bulk of the malaria burden now concentrated in only 10 countries. (21, 80) However, the decline in malaria transmission over the past decade has resulted in a shift in epidemiological patterns of malaria burden observed.

A recent review by Cotter et al. (95) has provided a comprehensive assessment of the changes in malaria risk factors in areas where malaria transmission intensity has declined. Briefly, more cases of symptomatic malaria are found in the adult and male populations partly as a result of decreased immunity but also due to occupational and behavioral factors such as travel to endemic areas, which increase risk of exposure. (39, 96) Malaria transmission is also likely to persist in hard-to-reach populations whom typically have less access to malaria control programs through choosing not to participate or because they reside in difficult to access localities. (97) Lastly, as transmission declines, imported malaria and human movement may provide a significant source of parasites that could sustain transmission in the absence of strong public health systems (see appendix 1.2). (98) It is likely that the malaria burden in these populations is not new, however the lingering parasite carriers become more apparent due to the reductions observed in the rest of the populations.

A reduction in malaria transmission also impacts the clinical disease burden. (99) For example, a study in Kenya found that the proportion of severe malaria cases declined first with rates of hospitalization for malaria staying constant for the first decade after transmission was reduced. (100) The change in the age profile of clinical malaria has been observed in several settings. (96, 99, 101, 102) For example, in the Gambia, as malaria transmission declined, the mean age of pediatric malaria admissions changed from 3 to greater than 5 years (103) and in a large prospective study in Tanzania, the mean age of severe malaria increased from 1 to 3 years. (104) However, once extremely low levels of transmission are achieved, evidence suggests that, a greater proportion of infections are likely to be symptomatic due to the lower levels of acquired immunity associated with the lack of exposure. (73) For example, a study in an unstable transmission setting in the highlands of Kenya found that only 0.1% of individuals were infected with malaria

by PCR based on a community survey whereas of patients attending health facilities with suspected malaria based on symptoms, 6.5% had confirmed malaria. (105)

In addition to the shifts in the age and immune profile of malaria in settings with reductions in transmission, changes are also expected to occur in the spatial distribution of disease. Specifically, as transmission is reduced, the distribution of parasite populations becomes more spatially clustered. (95, 106) This shift suggests that understanding this spatial heterogeneity is necessary to fully characterize the epidemiology of declining malaria transmission.

1.2 Heterogeneity of disease transmission

Heterogeneity of disease transmission occurs when a small proportion of the population, either defined based on spatial proximity or population characteristics, is disproportionately affected and experiences the majority of disease. (107) The importance of the spatial distribution in malaria risk is not a recent phenomenon however; due to factors such as availability of data, assessing heterogeneity has traditionally focused on national level estimates. (108, 109) However, national estimates provide a simplified view of the variability present at the scale where transmission is actually taking place and therefore a focus on the local-level heterogeneity is needed to fully characterize malaria epidemiology. The detection of local level clusters of infection has an important role for improving understanding of the microepidemiological patterns of disease transmission, and to ensure that control strategies are tailored to the specific epidemiological characteristics in an area as much as is feasible. (110, 111)

The current hypothesis suggests that these hotspots may be responsible for sustaining transmission from one season to the next. (106) Mosquito exposure is heterogeneous with a small proportion of people receiving the majority of bites. (107) The differential biting rates have been associated with ease of access (ie. not using a bednet), body surface area, body temperature, and body odors. (45, 112) The disproportionate rate of exposure likely results in more parasite carriage in highly exposed individuals, which then facilitates transmission. This

heterogeneous exposure has been estimated to lead to a 1.5 to 4-fold increase in the basic reproductive number of malaria. (84, 106, 113) Although this idea is biologically plausible, there is limited evidence to support this. Therefore, improved understanding of local-level heterogeneity in transmission is also important for identifying areas that may fuel parasite transmission as well as their role in sustaining transmission.

1.2.1 Sources of heterogeneity in malaria transmission

There are many different individual and spatial drivers of heterogeneity in the malaria transmission cycle, as discussed in section 1.1.1. (17, 97, 114) However, the precise combination of innate and ecological factors each person is exposed to will fuel the observed patterns. (97) For example, in a cohort of children observed for over 2 years in urban Uganda, 47% of children experienced no episodes of malaria and only 15% experienced more than 2 per person-year. (28) Similarly, a study in Kenya found that children that were infected at baseline during a cohort study had over 5 times the odds of acquiring another asymptomatic infection over the 3 months of the study follow-up. (48) Understanding these individual level risk factors is a large focus on the epidemiological research conducted and has been useful to inform malaria control programs and policies. However, the spatial heterogeneity is apparent in malaria transmission patterns and is comparably less well studied at the local level. (26) For example, as transmission declines, it is likely to become concentrated spatially around vector breeding sites. (97) Other possible drivers of spatial heterogeneity include non-uniform impacts of malaria control. For example, a study on Bioko Island found that after ITN distribution, malaria risk was associated with not using a net and was consistently high in one part of the island resulting in a new spatial pattern of malaria risk. (115)

1.2.2 Impact of disease clustering

1.2.2.1 Measuring transmission

The reproductive rate (R_0) is an important concept in disease transmission as it provides a measure of the transmissibility of a disease. Formally, R_0 is defined as the number of secondary cases that arise from the introduction of a single case into a naïve population. (73) In practice, the populations in the majority of malaria

endemic countries are not naïve and therefore, estimates of R_0 are typically adjusted downward and reflect the transmission potential in a population with non-sterilizing immunity to malaria under certain control scenarios (R_C). (84) Estimates of R_C less than one suggest that any new infection will result in less than one new case and malaria transmission will eventually disappear. (116) For malaria, R_C is difficult to measure directly, however, available estimates have suggested that it varies between just above one to several hundred. (84)

The heterogeneous distribution of mosquito breeding sites, amongst other factors, is known to impact the estimates of R_C and malaria risk. Estimates derived assuming a homogenous distribution are biased and tend to underestimate the true risk (17) because the humans that are bitten most can amplify transmission. (106, 109, 117) However, in areas where the infected individuals are bitten more often than the rest of the population, they can also serve to absorb some of the infectiousness, which would be missed by standard survey techniques used to estimate R_C. (109) In these areas with increased exposure, the clinical manifestations of disease will likely also be impacted as described in section 1.1.5.

1.2.2.2 Malaria control

The heterogeneous nature of infections suggests that if interventions were applied uniformly, achieving the near 100% coverage of perfectly effective malaria control strategies would be needed. (118, 119) However, it has been proposed that reducing transmission could be made much more efficient if high coverage could be achieved in those that are bitten the most instead of the entire population. (107, 109) In fact, simulations have suggested that such focal targeting of interventions to areas that then fuel transmission to surrounding communities, could double the reductions in malaria burden compared to a uniform strategy. (65) Targeting areas of increased transmission is being done in many countries including Zanzibar, Burkina Faso, and Swaziland. (59, 120, 121) However, despite the popularity and the biological plausibility supporting this approach, there has been little evidence to date supporting the impact of employing such a targeted strategy. There have also been no empirical studies conducted with the stated aim to ascertain the impact of the spatial distribution of disease on the impact of malaria control

interventions despite evidence suggesting that spatial biases exist. (16, 65, 122-126)

Another important consideration with employing a spatially targeted control strategy is at which baseline endemicity do they become important. Heterogeneity in transmission occurs at all spatial scales and transmission intensities. (127) In high endemic settings, heterogeneity in mosquito exposure is likely to be less well defined as due to the overall high levels of exposure experienced by the population. Evidence suggests that even in high transmission settings, malaria incidence is highly over dispersed with one study on the Kenyan coast finding that 23% of the person-time constituted 55% of the clinical malaria episodes. (128) This differential distribution of risk of malaria incidence has also been found to be reflected in the manifestation of hotspots: in areas of high transmission, hotspots are characterized by lower average age of clinical malaria as in such populations immunity is acquired relatively quickly. (127) Although hotspots are apparent in high transmission settings, they tend to be more pronounced and therefore more easily identifiable/targetable in low endemic settings. (129) For example, transmission levels in Zanzibar were historically high leading to the deployment of uniformly distributed control interventions. (130) As transmission declined, and Zanzibar became a country with low endemic transmission, clear hotspots were visible with a recent study finding that 80% of clinical cases are reported in only 20% of the health facilities on the island. (120) Targeting hotspots within high transmission areas is likely not practical given the high risk across the entire population. As transmission declines, such a targeted approach likely becomes more attractive. However, there is currently no known endemicity threshold where spatial heterogeneity becomes important for control or elimination strategies and current approaches are primarily driven by what is operationally feasible. (121)

1.2.3 Tools for detecting spatial heterogeneity

The ability to detect hotspots of malaria has developed with advances in geographical information systems (GIS) and statistical cluster detection methods. (131, 132) Initially, spatial analysis of malaria was restricted to visual comparisons

of the differences in malaria burden across space. (15, 133) For example, malaria prevalence estimates are plotted by district or by health facility catchment area to provide a picture of spatial heterogeneity in transmission. (26, 134) A recent study in Sudan showed that despite an overall slide prevalence of <1.0%, the population within a single area had a prevalence of 70%. (135) However, in most areas where heterogeneity is less dichotomous more robust methods are needed to interpolate malaria risk as well as to identify disease clustering. (136)

1.2.3.1 Spatial prediction surfaces

The ecological and innate drivers of malaria transmission suggest that risk is expected to be more similar to those areas located in close proximity and less like those areas further away. (26) The continuous spatial variation in risk can be measured using the mean, variance and a spatially defined correlation structure quantified using a semivariogram which plots the variance between points by distance; these values can then be applied to interpolate risk across space. (26, 132, 137) Advances in spatial statistical methods have facilitated the creation of risk maps, which are being increasingly applied to malaria and their application in parasite epidemiology, have recently been reviewed. (14, 26, 137-139) This section will therefore focus on those approaches most commonly applied to malaria.

Malaria risk maps are typically developed using available covariates such as distance to breeding sites, (140) a measure of the propensity for water to accumulate and pool to form breeding sites, (34) ambient temperature, (6, 18) or land use, typically measured using normalized difference vegetation index (NDVI). (141) Different geostatistical approaches have been developed such as kriging or model-based geostatistics in both frequentist (142) and Bayesian (143) frameworks and tend to be informed using village-level prevalence estimates. (14, 144) Such spatial predictions of malaria risk are useful for capturing patterns in malaria at different scales and account for uncertainty in the estimates. However, despite rapid improvements in the scale of data being generated, due to the availability and resolution of the majority of geocoded data currently accessible, use of this modeling approach is largely restricted to national and regional-scale

mapping of malaria risk and is not able to capture heterogeneity on the village or local level. (6) Risk surfaces for malaria have been used to identify priority areas for interventions or to quantify the population at risk of malaria transmission. (6, 9, 13, 145, 146) Due to the granularity of such maps, which are typically constructed at resolutions of several kilometers, the utility for identifying hotspots at the local level tends to be limited.

1.2.3.2 Cluster detection

To identify areas that may disproportionally contribute to malaria risk, statistical approaches that detect spatial clustering are useful. When testing for the presence of disease clusters, the assumption is that the points, or locations of the cases, are distributed completely at random meaning that risk is consistent across space. (26) Methods therefore assess whether the points are distributed not at random and can be considered as clustered. These techniques account for the non-uniform distribution of the population at risk and therefore require data on both infected and non-infected individuals. (26, 147, 148) Disease clusters tend to result from local level spatial variation and therefore identifying such foci are useful to study the processes driving disease transmission and could facilitate targeting interventions to areas that contribute disproportionately to the spread of malaria. (110) Although several spatial clustering methods have been developed, including kernel density, Kulldorff's spatial scan statistic, or the cumulative X^2 test, the premise for cluster detection are similar and are based on the likelihood ratio test. (148, 149)

The most popular cluster detection approach used in malaria research is the Kulldorff's spatial scan statistic, (26, 148) which has been made accessible through the software package SatScan (Harvard Medical School, Boston, USA). The statistical process behind Kulldorff's statistic uses a series of circles or elliptical shaped windows of incremental sizes centered on each data point. A likelihood ratio test is then conducted comparing the rate inside the window to outside. Monte Carlo simulations, which generate permutations of the data across the area, are used to test the null hypothesis that points are distributed randomly. (26, 150, 151) Hot- or coldspots are delineated using either a point representing the center

and the radius of the cluster size or the coordinates of the points with significantly greater or lesser risk, respectively. (127, 152) Kulldorff's statistic has been extended to allow different types of outcome variables such as count or binomial data and have also added the flexibility to include covariates to adjust estimates of malaria risk. This method has been used to identify hotspots in many malaria endemic settings and has been found to successfully predict malaria risk in subsequent years. A study on Bioko Island in Equatorial Guinea used SatScan to identify areas that had significantly higher malaria burden and identified areas suggestive of residual transmission. (115) Similarly, in a study in Tanzania, areas identified as significant clusters of malaria have been associated with increased malaria prevalence in subsequent years suggesting that some of the underlying spatial process is being captured. (148)

1.2.3.3 Spatial-temporal heterogeneity

Options for extending both continuous risk surfaces and point cluster detection tools to include a temporal element have also been developed. (26, 127, 153) Temporal options for continuous risk surfaces developed include stationary and the more complex anisotropic versions, whereby spatial autocorrelation is dependent on both location and direction. (9) Although spatio-temporal methods are more computationally intensive, they can provide more accurate predictions of risk when spatial data is available at different time points and also enables a better prediction by accounting for historical trends as well as changing risk due to factors such as malaria control interventions or changes in climate patterns.

Methods incorporating temporal dimensions for cluster detection using SatScan have also been applied to malaria data. One approach is to analyze each year (or unit of time) of data independently and visually examine any observed trends between the images. For example, Bejon et al used SatScan to detect clusters for each year of a 12-year study and were able to identity stable hotspots and hotspots that varied from year to year. (129) The second approach is to employ the space-time model extension to Kulldorff's spatial scan statistic where the moving window extends into a cylindrical shape with the height reflecting the temporal element. (26, 154, 155) For example, using this approach, a study in South Africa

detected 5 spatial and 2 time clusters with the time clusters corresponding to the localized outbreaks recorded. (156)

1.2.4 Spatial scale and targeting transmission heterogeneity

Heterogeneity in malaria transmission is apparent at all spatial scales as has been demonstrated by recent work in Swaziland (157) and Kenya. (127) Although heterogeneity in malaria transmission is increasingly being recognized as an important component of malaria epidemiology, current guidelines to deploying interventions focus on the country or district level (12, 13) or are necessarily ambiguous to ensure tailoring to local circumstances. (158, 159)

The generation of malaria burden maps using risk surfaces described above has been a useful tool for informing decision making at the national level, or increasingly the regional scale, which are operationally attractive units for deploying interventions. (118, 160) However, as transmission declines, the national/district level prevalence estimates become less accurate and the uneven distribution of malaria transmission becomes more prominent: local level estimates tend to be highly heterogeneous and some areas within a district can be hyperendemic whereas other may be extremely low. (161, 162) For example, current data suggest that 80% of all malaria infections in Zanzibar are reported in 20% of health facilities on the island. (120)

Spatial scales that are relevant for malaria transmission largely consist of where mosquitoes come into contact with humans. (163, 164) A recent study has suggested that the optimum scale relevant for surveillance strategies that can target >60% excess of new cases and can identify smaller sub-hotspots within primary foci involves targeting an 8 km area. (129) However, it is unknown how this translates into identification of hotspots or the relevance and impact of targeting such areas with control strategies.

Operationally, elimination strategies have used an iterative approach whereby high burden districts are initially targeted. (59) As transmission declines, the focus shifts to sub-districts, and finally to clusters of households as the spatial

distribution becomes progressively patchier. (120, 127) For example, P. vivax malaria elimination in China relied on a mass drug administration approach targeting administrative units with high reported clinical incidence and then as transmission dropped, they refined the approach to focus on sub-units with persisting levels of higher incidence. (165) Similarly, in low transmission settings, identifying and targeting foci of parasite populations at the individual or household level, through active or re-active case detection, discussed further in section 1.4.2.2, is being employed to target local transmission dynamics.(13, 97, 166) However, such an approach is only likely to become operationally feasible once caseloads get low and the number of people required to follow-up is minimized although the impact on reducing the parasite reservoir has yet to be confirmed. (167, 168) Regardless of the approach used, all strategies for reducing or stopping transmission are reliant on being able to accurately quantify malaria transmission and ensuring robust datasets are available with which to gauge the effectiveness of programs.

1.3 Approaches to quantify malaria transmission

In order to understand the heterogeneous distribution of malaria and identifying the areas most at risk, transmission intensity must be measured. As the direct measure of transmission, R_C, is not easily quantified, different malaria metrics have been developed that measure different stages in the malaria lifecycle and provide indirect measurements of transmission many of which have been used to detect hotspots. (59, 127, 169, 170)

1.3.1 Metrics - Overview

1.3.1.1 Entomological

The entomological inoculation rate (EIR) is correlated with R_C and is considered the gold standard for measuring transmission. EIR assesses the number of infectious bites that a person in a given area is expected to receive, typically measured over one year and provides a measure of the degree of exposure to malaria in a population. (117, 169) The EIR is notoriously difficult to calculate directly as it involves catching host-seeking mosquitoes and identifying the proportion of these mosquitoes that are harboring sporozoites, the transmissible

stage of the malaria parasite. (119) Using human landing catches is considered to be the ideal method for identifying host-seeking mosquitoes. This technique involves people sitting awake all night, when *Anopheles* mosquitoes are typically active, aspirating all mosquitoes that land on the person for counting, and assaying for the presence of sporozoites. (171) As human landing catches are laborious and involve risk of the workers being exposed to malaria, alternative methods have become more widely utilized including the use of traps where mosquitoes are caught inside or while leaving the house. (31, 63) However, in addition to being laborious, there is little standardization in methods employed across studies and sites. EIR is highly seasonally variable, and is difficult to measure in areas of low transmission intensity where the density of mosquitoes is low. (172) Therefore, despite it being considered the gold standard for estimating R_C, due to questionable precision and accuracy, EIR is not extensively used and is particularly difficult to assess over small spatial scales. (169)

1.3.1.2 Parasitological

Microscopy

Microscopy has been considered the gold-standard malaria diagnostic tool in the field since the 1960s when it replaced the use of spleen rate. The *P. falciparum* parasite rate (PfPR) estimates are currently the most widely reported metric of malaria burden. (2, 13) The widespread use of this metric is in part because of the ease in collection in field conditions and it being the diagnostic tool recommended in clinical settings. Blood slides are typically prepared and are visualized under an oil immersion microscope where the presence of parasites can be viewed and quantified. (173) Microscopy is able to consistently detect as few as 5 parasites/μl of blood however, the sensitivity has been reported to vary considerably with some estimates suggesting a limit of detection closer to 100-200 parasites/μl of blood is more consistent for routine microscopy in clinical settings. (58, 170, 174)

Microscopy quality is notoriously low at rural health centers, where the largest burden of clinical malaria is seen. (174) Factors that impact performance include training and monitoring of microscopy quality, fluctuations of staff, patient load and the time available to read each slide, the quality of equipment, and the

technicians' skill in preparing slides. (175, 176) Estimates have suggested that parasite prevalence using microscopy could be negatively biased by at least 20%, due to factors, such as fluctuating parasite densities, described in section 1.1.2 above. (11) However, other studies suggest that although individuals do shift from detectable and submicroscopic, the rate is relatively consistent and is predictably related to the total parasite population therefore does not have substantial impacts on prevalence estimates. (58, 177) A recent study in Kenya found that through consistent monitoring for quality assurance, high sensitivity and specificity can be achieved, however in most malaria endemic settings this is not the case. (178)

To mitigate some of the biases with using parasite prevalence to measure malaria transmission intensity, age standardized rates, usually in those 2-10 years ($PfPR_{2-10}$), have been employed to improve comparability between sites. (109) As discussed in section 1.1.3, estimates in this population tend to be more robust as children are most likely to present with detectable parasite densities. (58) The $PfPR_{2-10}$ has been shown to be a sensitive metric of transmission intensity: The relationship between EIR is strongly correlated with estimates of $PfPR_{2-10}$ (117, 179) and therefore $PfPR_{2-10}$ is considered to be a reasonable proxy for transmission intensity until very low levels (ie. ~<3%) when it become unreliable due to the large number of people that are needed to achieve a reasonable sensitivity. (13)

Rapid Diagnostic Tests

Despite the utility of microscopy, many facilities where malaria is endemic do not have the necessary equipment, reagents, or skilled personnel. Therefore, the malaria diagnosis largely relies on clinical signs and symptoms, which are highly non-specific. (170, 174) Malaria RDTs are increasingly popular as they provide an easy diagnostic tool that has similar sensitivity to conventional field based microscopy while being less technically demanding therefore achieving better consistency between operators. (180) Most malaria RDTs detect the presence of histidine rich protein 2 (HRP2), an antigen secreted exclusively by the *P. falciparum* parasite. Other tests have been developed that are based on the enzyme lactate dydrogenase (pLDH), which is produced by all *Plasmodium* species. (181)

RDTs are typically distributed in a cassette format and involve adding a fixed volume of blood, usually 5 µl, and a reagent, which then reacts with the parasite antigen if present. (182) Over the past decade the use of RDTs have become more widespread in rural health centers in endemic countries and have also been trialed to be used as part of a community based strategy with community health workers. (183, 184) The sensitivity and specificity of RDTs varies by brand and is related to the lowest limit of parasite densities that can reliably be detected with most being inconsistent when parasite densities are below 200 parasites/µl of blood. (182)

There are also some potential measurement biases associated with the use of RDTs. First, HRP2 has been found to persist for up to 6 weeks after treatment of an infection suggesting that some false positive cases are likely. (185, 186) Also, as transmission declines and parasite densities in persisting infections tend to be lower, the utility of RDTs becomes questionable. (120) For example, when transmission was high in Zanzibar, sensitivity was reported to be 92% (against microscopy) but when transmission became low, sensitivity dropped to 79%. (187) Similarly, in Swaziland, another setting undergoing an elimination agenda, RDTs were able to achieve a high specificity, but missed the 2 cases detected by more sensitive methods. (157) Furthermore, in settings where malaria transmission has declined the importance of non-P. *falciparum* infections may increase. In such circumstances, pLDH based tests that detect any *Plasmoidum* species could be useful whereas RDTs that detect HRP2 may become less relevant. (188)

The increased inter-facility consistency and the availability of RDTs in a greater number of clinical and research settings suggests that RDTs can provide more robust estimates with which to inform decision-making and to identify heterogeneity in malaria burden. (180) The use of RDTs, particularly in a clinical setting, has also been found to reduce the overuse of antimalarial drugs and therefore, despite the higher cost associated with the test, lead to cost savings by reducing the amount of drugs administered. (189) For example, a study in Uganda found that the introduction of RDT's into health facilities resulted in greater than 2-fold reduction in antimalarial drug prescription. (190) The scaling up of RDTs

will also improve the quality of the routine data collected at health facilities and malaria surveillance. (191)

Molecular methods

Molecular methods for detecting the presence of *Plasmodium* DNA in a sample was developed in the 1990's (192) and are considered to be highly valuable for determining positivity in malaria research. (53) The standard method, polymerase chain reaction (PCR), involves extracting DNA from a sample and using a series of reactions where any parasite specific DNA present is amplified creating exponentially more copies. The higher number of copies present in the reaction makes it easier to detect visually using an ultraviolet transilluminator. (193) PCR-based methods are extremely sensitive and specific and have been found to detect as few as 1 or 2 parasites/μl of blood. (54, 194) The increased sensitivity of PCR has been instrumental to quantify the full extent of the parasite reservoir and overall detects 50% more infections than microscopy or RDTs. (53, 58) Molecular methods are also better able to identify mixed infections and non-P.*falciparum* species compared to microscopy. (195) However, the technical complexity and high cost of the assay as well as the length of time required to process samples, limits the application of PCR in the field or as a point of care diagnostic tool in endemic settings. (174)

Molecular based diagnostics are being developed to provide tools that are more operationally attractive for malaria control programs. (196) Loop mediated isothermal amplification (LAMP) is the most advanced and may provide the required balance between a sensitive molecular based diagnostic tool that is operationally attractive and that can be used in malaria endemic countries. (197, 198) LAMP methods are similar to those of PCR whereby malaria specific DNA in a sample is extracted and selectively amplified. After the reaction has developed, positive samples can be visualized by a change in the color intensity. (199) The LAMP system is less technically demanding, requires less time to obtain a result, and can achieve similar sensitivity to PCR making in an attractive alternative in areas where submicroscopic infections are of interest. (200)

Despite the high potential for use of molecular based methods in malaria control and elimination settings, as a point of care test, it has been found to offer little benefit over conventional RDT or microscopy. In clinical settings, symptomatic malaria cases tend to have parasite densities at levels detectable by RDT or microscopy and molecular methods may not be as easily justified. (201) However, when the priority shifts to reducing or eliminating transmission, detecting subpatent infections becomes a priority as they do contribute to maintenance of transmission. (53) In such settings, LAMP and other molecular assays have the potential to become a useful tool to quickly detect areas of focal transmission.

1.3.1.3 Serological

An alternative approach to measure malaria transmission is to detect anti-malarial antibodies, which provides a marker for exposure to malaria. (51, 202) The use of serology to describe malaria epidemiology has become more common since the refinement of the enzyme linked immunosorbent assay (ELISA) for the detection of malaria specific antibodies. (203) Briefly, this assay works by binding antigens to specific plates and all non-malarial antibodies are blocked. The bound-antibodies are then detected with an enzyme-linked secondary antibody. The presence of the target antigens (bound-antigen) is visualized through a color change in the reaction, and quantified using a spectrometer. (203) It has been shown that anti-malarial antibodies can be effectively extracted from filter paper blood spots and can be processed as a high-throughput technique making it an operationally attractive tool for quickly assessing malaria exposure in large populations. (204)

Different malaria-specific antigens have been identified and although some are associated with clinical protection, here they are discussed as markers of exposure. (49, 52) The presence of antibodies to any anti-malarial antigens has become a popular metric to assess the cumulative burden of malaria exposure in a population. (204) The rate of acquisition of antibodies with age can be calculated using a reverse catalytic conversion model and provides an estimate of the seroconversion rate (SCR), or the number of people in a population expected to seroconvert each year. A study in Tanzania found that SCR was strongly correlated with the EIR and provides a proxy measure for estimating the force of infection in

a community. (205) A different study in Brazil found that SCR correlated with the annual parasite index, an alternative measure of transmission intensity, collected by the malaria surveillance program. (39) Also, a study on Bioko Island, Equatorial Guinea found that changes in SCR were correlated with changes in parasite rate and with reductions in all cause child mortality. (115) Seropositivity is also able to rank areas by endemicity. (164)

Serological tools are more sensitive than conventional parasitological diagnosis in areas of low or highly seasonal transmission. (157) Antibody responses are longer lived and are therefore able to detect past infections as individuals can maintain stable levels in the absence of recent exposure to parasites. (202) Furthermore, conventional diagnostics are particularly affected by fluctuating parasite densities, which affect the sensitivity of the diagnostic tool. Serological tools are particularly useful in low endemic settings where the sensitivity of field friendly parasitological tools is inadequate.(202, 206, 207)

Serology has shown to be able to confirm or suggest interruption in transmission based on the age-adjusted serological profiles of the population, and particularly the exposure in younger children. (203) A recent serological assessment in Swaziland found that presence of anti-malarial antigens were virtually absent in those under 20 years (1.9%) of age while they were 10 times greater in adults (11.7%) suggesting that there is little ongoing transmission at present while adults were historically exposed to more intense malaria transmission. Seropositivity observed in the adult population may also be influenced by travel to neighboring endemic countries. (157) Serological tools have also been shown to identify heterogeneity in the impact of interventions. For example, on Bioko Island seroconversion rates for AMA1 in certain parts of the island were significantly lower compared to others combined with the lower seroprevalence observed in children suggest that the impact of control interventions were heterogeneously distributed. (115) Therefore, the use of serology has the potential to become a powerful tool for assessing malaria burden and heterogeneous transmission intensity in a population.

1.3.2 Metrics for measuring spatial heterogeneity

Work to identify spatial heterogeneity in malaria burden has relied on different metrics to spatially delineate areas of high malaria burden. The generation of risk surfaces typically rely on $PfPR_{2-10}$ data generated using microscopically confirmed cases (6, 9, 145, 208) For example, in Burkina Faso, using data on parasitaemia in children, risk maps were generated to characterize the heterogeneity in malaria risk across the country with a range in predicted prevalence between 11 and 92%. (209) Other metrics such as clinical incidence (139) or RDT positivity (210) have also been used to inform risk models but their application to delineate spatial heterogeneity is limited due to quality or the availability of sufficiently large and geocoded datasets as well as resolution of model covariates to inform models.

Despite the reliance of $PfPR_{2-10}$ to model malaria risk, a variety of other metrics has been used to inform spatial clustering for the detection of hotspots. For example, a study in Ethiopia used clinical case data recorded by hospitals to detect clusters of infections that identified several hotspots. (211) A second example of clustering of clinical malaria is a study conducted by Bejon et al in Kenya that identified hotspots of symptomatic malaria in children reporting to health facilities. (127) Other indicators that have been used include malaria deaths (133) and although application is more limited, PCR has also been used to inform cluster detection. (148, 157) Finally, testing for clusters of confirmed malaria infections according to microscopy or RDT parasite rate is commonly used. (135, 155, 156) For example, a country-wide survey conducted by Ashton et al. in Ethiopia identified significant clusters of microscopically confirmed infections using SatScan. (212)

Serological data is increasingly being used to identify hotspots of malaria however; the interpretation of these results is less straightforward than with metrics of current parasite infection. (127, 152, 154) Studies using serological tools for detection of hotspots have used different antigens, or combinations of antigens. Similarly, different approaches to modeling the data have also been used with some relying on the binary seropositivity metric (213) and others using antibody density (115) or a combined estimate. (129) Furthermore, SCR has also been used in some settings to inform clustering of malaria transmission. (110) Complicating

the interpretation of serological indicators for detecting hotspots is that evidence suggests that antibody domains within the same antigen do not all exhibit the same spatial clustering patterns, likely due to the differences in individual responses to antigens. (213)

The different malaria metrics provide different approximations of malaria transmission intensity making explicit comparisons between results difficult. However, some consistency has been observed between hotspots defined using different metrics (213) but this is not always the case. (110) How these differences impact resulting hotspot boundaries or how hotspots can be delineated in a way that is operationally feasible has never been formally assessed.

1.4 Operational research and malaria

Operational research focuses on translating current knowledge into routine health programs and facilitating more informed decision-making based on local circumstances. (214) Although definitions of what constitutes an operational strategy in the public health context differ, the primary focus is to identify strategies that can improve health in a way that is tailored to local capacity to maximize health outcomes in the affected populations. (215, 216) An ideal strategy is flexible enough to ensure targeting to local conditions, that the responsibilities and actions of each party is clearly delineated and considers the current realities of the community involved including the availability of infrastructure, transportation, communication, and human capacity. (216, 217) Operational feasibility has been associated with three factors: government stability, effectiveness and commitment; the capacity of health systems; and the size and ease of access to the populations at risk. (118) Operational research is useful to identify best practices and to tailor programs to meet local needs but is also linked to good monitoring and evaluation programs that facilitate updating practices based on changing circumstances such as the development of drug resistance or targeting interventions to high-risk areas. (214)

There is a deficit in translating research into policy and improving health-care practice, and malaria programming is no exception. The importance of identifying

best practices and how malaria control interventions can be implemented to achieve their maximum level of efficacy has long been highlighted as a priority for research. (218) However, only an estimated 3% of all malaria literature can be considered to be operational research and what exists primarily focuses on strategies for malaria control in the African setting with little focus on low transmission or elimination settings. (219) Despite the lack of an operational focus, studies cite the importance of community involvement as reasons for successfully programs and achieving high coverage rates for interventions. (59, 120)

Malaria transmission is largely concentrated in settings that may be difficult to access, have limited infrastructure, public health capacity, and have limited resources. (220-222) Operational approaches and the necessary research to identify the optimum strategies and the best way to implement them by local malaria control programs is highlighted as one of the critical areas for sustainability of the current malaria control and elimination agenda. (95)

1.4.1 Current strategies for monitoring malaria transmission

As discussed, R_C is the most direct measure of malaria transmission. (84) Operationally tractable tools have been developed to estimate R_C based on the proportion of locally acquired and imported cases using surveillance data, (40) however, this tool only becomes relevant when the number of new cases is low. Therefore, the majority of malaria surveillance for transmission will likely be informed by a combination of both passive and active case detection designs using either microscopy, RDTs, or clinical malaria, depending on the setting.

Passive case detection is the primary source of historical data on malaria burden and typically involves aggregated data routinely collected at health facilities. (20, 223, 224) The metric typically used is the number or proportion of malaria cases or deaths, diagnosed based on clinical symptoms or increasingly through test-positivity rate for confirmed cases of malaria. (22) Confirmed malaria incidence may provide a useful metric in very low transmission environments where the majority of infections would likely be symptomatic and report to health facilities;

however, as discussed in section 1.1.5, in the majority of malaria endemic countries passive case detection may not necessarily reflect malaria transmission due to the pervasiveness of silent infections and poor reporting rates, as discussed further in chapter 3. (5, 167, 219)

To supplement health facility based reporting on malaria burden, many countries conduct countrywide, community-based surveys. Malaria indicator surveys (MaIS) are typically designed using a two-stage cluster design to ensure that nationally representative data are collected while including a focus on high-risk areas. (225) These surveys have traditionally assessed malaria burden in children under five years of age and confirming malaria infections using microscopy or RDT. (226, 227) With the increased focus on elimination and recognition of the importance of the silent parasite reservoir, surveys are increasingly including all ages in their sampling framework and employing more sensitive diagnostic tools, including serology. (228) The MaIS surveys are useful to obtain a current picture of malaria burden across the country. However, due to the high expense and logistical difficulties, they are typically conducted sporadically with some countries engaging in MaIS every one or two years while others have only conducted a single survey. (225) These large surveys become even less operationally attractive for assessing malaria burden when transmission levels are low. Not only are large numbers needed to achieve sufficient power, the operationally attractive diagnostic tools, as discussed above, are not sufficiently sensitive to provide an accurate picture of malaria transmission. (229, 230)

1.4.2 Convenience sampling: An operationally tractable approach

1.4.2.1 Malaria Surveillance

Alternative sampling strategies that rely on sentinel populations can provide useful alternatives for malaria surveillance that could provide more robust and reliable data compared to routinely reported records yet are more operationally attractive than the large community based approaches such as MaIS. (231, 232) Health facility surveys whereby all attendees instead of just suspected malaria cases are sampled or primary school surveys are two such options that have been used. (232, 233) For example, a study in The Gambia found that health facility

surveys were able to identify heterogeneity in malaria transmission and provided similar estimates to those obtained in the surrounding community. (134) Similarly, school surveys conducted in Ethiopia were able to identify heterogeneity in malaria transmission however; no concurrent community estimates were available. (212) However, the bias of such convenience sampling approaches for malaria surveillance is not well characterized.

1.4.2.2 Identifying hotspots of transmission

One of the goals of malaria surveillance is to identify areas with increased risk and to provide the evidence base needed to effectively tailor malaria control efforts. (159) However, to inform malaria elimination a new framework is needed to detect areas with residual transmission so that interventions can be targeted accordingly. (63, 222) Shifting the timing of surveys to focus on the low transmission season will likely be useful to identify residual parasite infections. (161, 225)

Despite the paucity of operational research on how best to do it, several countries are currently engaging in malaria elimination activities. The elimination-focused activities are predicated on identifying and treating the residual parasite population in the community. Two main types of approaches have been reported: active and re-active case detection. (97) Active case detection involves seeking out cases in the community using a mass treatment (234) or screening and treatment (120) of areas thought to be at increased risk; typically defined using routinely reported health facility data. Due to the risk of drug resistance, mass drug administration has not been extensively trialed, and although simulations suggest that they will result in a reduction in burden, (94) in low transmission settings field trials have found little impact. (234) Screening and treatment in the community over time has been found to reduce risk of malaria infections (235) but this has not been consistent in all settings and is not operationally attractive due to the repeated community sampling required. (59) The most cited reason for the limited impact of a screen and treat approach is due to the sensitivity of diagnostic tools used and the proportion of subpatent infections that are missed. (53, 120)

In low transmission settings, the most commonly implemented strategy to tackle residual transmission is a re-active approach with ongoing programs currently reported in several of the Asian-pacific countries, Swaziland, and a pilot study in Zambia. (121, 166, 217) Re-active case detection to identify reservoirs of infections in the community typically use a symptomatic index case to identify areas where additional parasite carriers are likely due to the tendency of cases to cluster in space and is analogous to contact tracing for directly transmitted infectious diseases. (121) Such an approach can be complicated in areas where the risk for imported cases to re-ignite transmission and alternative methods to identify possible networks or snowball sampling may become necessary. (40, 167, 236) However, the utility of such an approach, best practices to ensure optimum coverage of the parasite populations, and what proportion of the reservoir is actually targeted are still relevant questions for study. Furthermore, despite the use of re-active case detection in many countries, implementation can be logically difficult and more work is needed to determine if these efforts are worth the time. (121)

1.4.2.3 Operationally attractive hotspot detection

For hotspot targeted approaches to malaria control and elimination strategies to become operationally feasible, convenient ways of identifying hotspots are needed. Health facility and primary school based populations could provide a useful means for malaria surveillance; however, it is not known how representative these populations are of the community and potential bias needs to be better understood so that malaria burden is then interpreted accordingly. Furthermore, the utility of convenience sampling to identify hotspots in the community may provide an operationally attractive method however, evidence is required to better understand the sensitivity and therefore potential of such an approach. Moreover, the sensitivity of a community based re-active case detection approach is not well characterized and is important to gauge the potential impact of a targeted approach to identify the entirety of the parasite population and therefore justifying the large efforts involved.

Chapter 2: Study Rationale and Objectives

This chapter will introduce the rationale for this research project as well as introduce the main and specific objectives. The overview of the study rationale highlights the two unifying themes of this work: i) identifying hotspots of transmission and their potential as part of control and elimination strategies and ii) operational approaches that can provide logistically attractive options for local malaria programs. This section concludes by outlining the aims of this research.

2.1 Study Rationale

Malaria risk is not distributed equally, and heterogeneity in malaria exposure is especially pronounced in areas with low and moderate transmission intensities. Heterogeneity in malaria exposure suggests that a small proportion of people are not only carrying the majority of the malaria burden, but are also contributing disproportionately to onward transmission. (97, 106) It has been suggested that individuals experiencing the majority of the malaria burden tend to cluster in space and form hotspots of infection. (129)

Identification of hotspots could be extremely useful, as it would allow targeting of malaria control, which would reduce costs of deploying interventions, as only a subset of the population would be targeted. (107) Because hotspots may fuel malaria transmission in larger regions, it is conceivable that such a targeted approach could lead to a greater reduction in malaria transmission in areas surrounding malaria hotspots. (106) Modeling exercises have shown that hotspot targeted interventions that achieve 90% coverage with insecticide treated nets and indoor residual spraying in areas with a baseline parasite prevalence of ~15% could result in a reduction of parasite prevalence and vector densities inside the hotspot to less than 1%. (65) Such an approach could then lead to a reduced reservoir of infection or even local malaria elimination by inhibiting the spread of malaria to surrounding communities. (97)

In an era where malaria elimination is possible in many settings, it is these persistent hotspots of infection that may prove to be a formidable challenge if these cannot be adequately targeted. (81, 118) The first issue is how to measure

and define what constitutes a hotspot of transmission so that boundaries drawn reflect the true nature of transmission in the community: if hotspots are incorrectly specified and a proportion of the hotspot is missed it is likely that transmission will be sustained and the targeted approach will be incorrectly deemed to be ineffective. (127, 160) The second major challenge with employing a hotspot-targeted strategy is that the current approach for identifying reservoirs of infection involves costly and time-consuming community based surveys. The ability to easily and accurately identify hotspots in a timely manner is essential to ensure that this method can be integrated into local malaria control programs. Therefore, there is a need for strategies that can identify hotspots of malaria that are effective and easily sustained by local public health infrastructure. (121, 216, 222)

2.2 Main Objective

The overall objective of this study was to determine if operationally attractive approaches for the identification of hotspots of malaria transmission in the western Kenyan highlands are possible and can provide viable alternatives to a community based survey approach.

2.3 Specific Objectives

1) Define hotspot of malaria transmission in the community
2) Determine if surveys conducted at primary schools and health facilities result in comparable transmission indices compared to community surveys
3) Identify the sensitivity of primary schools and health facilities to provide a reliable metric to identify hotspots of malaria transmission intensity in the broader community.
4) Determine the proportion of parasite carriers that can be identified using an intensive but operationally tractable community-based sampling approach within hotspots.

Chapter 3: Study Design Overview

To address the outlined objectives, this research combines data collected across several different studies. This chapter outlines the framework depicting how the different studies fit together in relation to the specific aims outlined in chapter 2. To avoid ambiguity in terminology, definitions of important terms are provided and gold standards that are relevant to this work are outlined. A brief overview of the study site, including the study population and malaria epidemiology is discussed with more detailed descriptions provided in subsequent chapters. Finally some background on the primary school and health systems that are present in the area is reviewed.

3.1 Research Framework

Each component of this work offers essential information by providing insight on the 'true' state in the community, the alternative approaches to identify the 'true' state of malaria epidemiology, as well as looking into the bias associated with the alternative approaches. The relationship between the different studies described in subsequent chapters and how they relate to the specific objectives presented in chapter 2 are provided in figure 3-1. Briefly, three community-based surveys were conducted to act as the gold standard and represent the true malaria transmission in the community. The main studies were then conducted and compared with the relevant dataset to address each of the objectives of this research. For example, detecting malaria heterogeneity (specific objective 1) involved using the baseline school zone community survey (baseline) to provide data on the 'true' transmission levels and was compared with results from the school surveys (main studies).

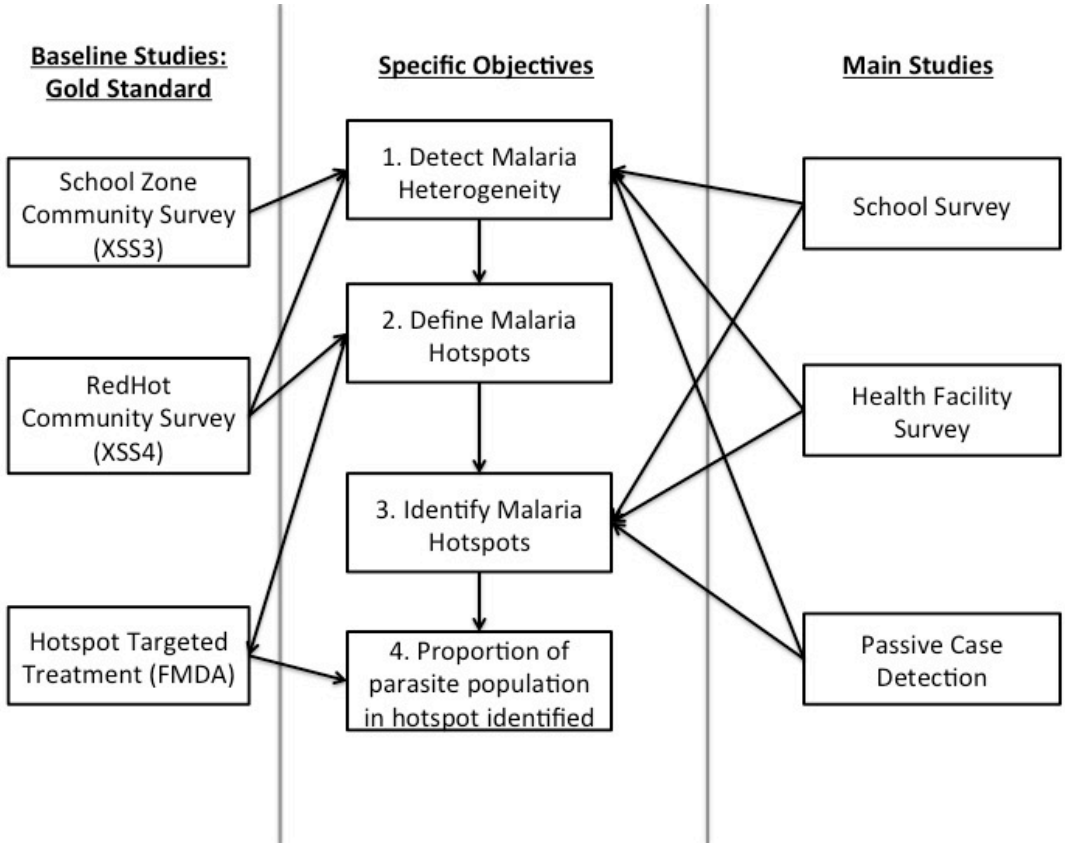

Figure 3-1: Schematic overview of studies conducted and how they relate to the specific objectives of this research.

Malaria transmission can vary dramatically over a small geographical area. (17, 129, 237) To ensure comparability over space, the surveys informing this research were all conducted within the same study site ensuring as much spatial overlap as was possible (figure 3-2). The initial study area consisted of one larger area extending into Kisii, the neighboring district, for the school surveys. The extremely low malaria prevalence observed in this region prompted the study area to be restricted to a smaller area of approximately 200 km^2 for the health facility surveys. Based on data collected during the previous studies, the community work was revised further to concentrate on a 100 km^2 area to ensure that sufficient heterogeneity could be detected. Finally, within this 100 km^2 community study area, five smaller areas of focal transmission were subsequently identified.

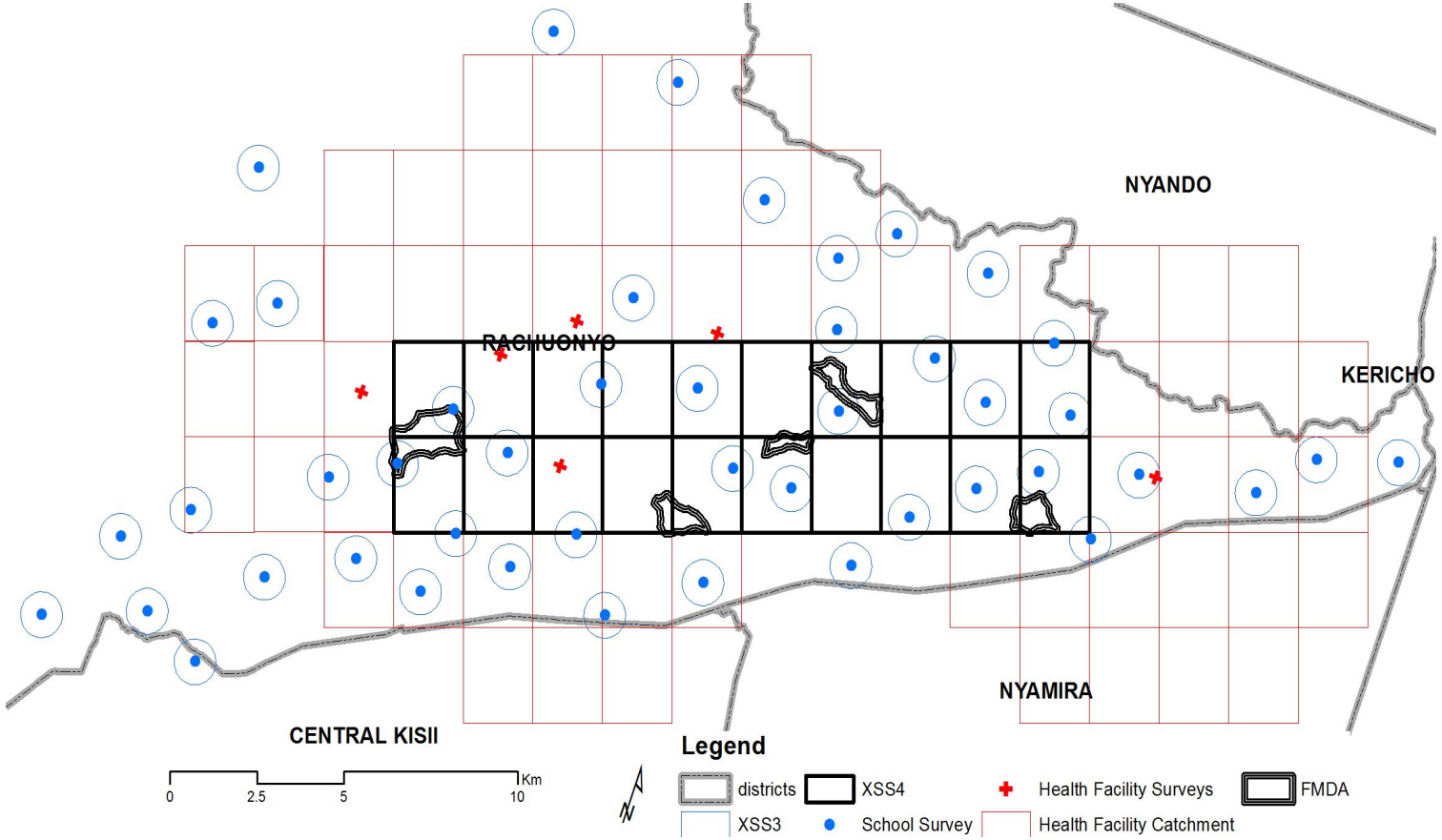

Figure 3-2: Map of study area showing the degree of spatial overlap achieved by the different studies of which this research is composed including the schools and health facilities included in this survey and the community area targeted for the community work.

Similarly, malaria can also vary over time with differences between and within transmission seasons. (129, 238) Therefore, temporal overlap is also important to consider when comparing different approaches to measure malaria transmission and was accounted for in designing the research framework when possible. However, due to the intensive research operations in the area and the seasonal nature of transmission, it was not always feasible. Specific protocols followed in each survey are described in their respective chapters, but key differences are highlighted (table 3-1). The variations of note include the temporal differences between the surveys. When possible work was conducted in the main transmission season, but due to the specific objectives of the individual studies, the season and year of data collection were not always synchronized. The second main difference between surveys was due to changing to more sensitive diagnostic tools. In 2012, the Paracheck RDT was replaced with the more sensitive First Response version and auxiliary digital were replace with infrared tympanic thermometers. The variance does not affect the internal validity of the individual studies but does affect what information is available for comparison between studies as well as the strength of the interpretation of the findings and therefore must be acknowledged.

3.2 Definitions and Terms of Reference

Several key definitions and gold standards are discussed below and terms are used as consistently as possible throughout this work. Firstly, although the terms heterogeneity, foci and hotspot tend to be used interchangeably in the literature (97, 106, 107, 148, 156, 239) as part of this research they are considered as important and distinct terms and have important conceptual differences.

3.2.1 Heterogeneity

In the strictest sense, heterogeneity simply refers to something that shows diversity as opposed to homogeneity, which suggests an even distribution of exposure, risk, or other metric under investigation. (107) In the malaria field, the term typically refers to differences in the magnitude of the malaria risk given the metric used and typically has a spatial dimension but can also refer to individual exposure. (162) For example, if the malaria prevalence estimates observed at two neighboring schools were different by a certain magnitude, this would be considered to exhibit a heterogeneous malaria burden. Similarly, heterogeneity

Table 3-1: Overview of methodological differences between surveys.							
	FMDA	XSS 4	XSS 3	School	Health Facility 1	Health Facility 2	Passive Case Detection
Year	Mar 2012	July 2011	July 2010	July 2010	October 2011	August 2012	Feb-Oct 2012
Season*	Low	High	High	High	Medium	High	All
Target Population	Community	Community	Community	School Children	Patients and accompanying people attending outpatient department		Suspected Malaria Cases
No. Clusters	10	N/A	46	46	5	5	3
% Geolocated	100	100	100	98	25	30	0
Temperature	All Tympanic	All Auxiliary	None	None	All Auxiliary	All Tympanic	All Tympanic
RDT	First Response	Paracheck	Paracheck	Paracheck	Paracheck	First Response	First Response
PCR	Yes	Yes	No	No	Yes	Yes	No
Serology	No	AMA1 MSP2	AMA1 MSP2	AMA1 MSP2	AMA1 MSP2	AMA1 MSP2	AMA1 MSP2
* high = peak during long rains, medium = peak during short rains, low=between peaks							

can exist at local, regional, national and international scales. (9, 110, 212) In this study, the term heterogeneity is defined as any unit of study exhibiting biologically relevant variability in malaria transmission intensity, as quantified by any parasitological or serological metric. The unit of study can, for example, comprise a school catchment area or be an individual compound with biologically relevant differences suggesting variable intensities of malaria transmission and is not restricted to statistically significant differences.

3.2.2 Foci

Malaria transmission is sustained in areas where environmental conditions are conducive to mosquito breeding and has human settlements that can sustain and transmit the malaria parasite within the dispersal range of the vector, typically around 1 km in African settings. (16) The size of the foci is dependent on the dispersal range of the mosquito and availability of blood meals: Precise borders are difficult to measure and are likely fluid. (17) Entomological data may provide the best way to measure foci but as mentioned, are difficult to obtain, particularly in areas of low transmission intensity with obtaining data to measure other metrics for transmission potential such as the effective reproductive rate (R_C) also being problematic. (117, 119) Therefore, as part of this research, foci are considered to be areas that can sustain or support transmission of malaria and provides more of a conceptual framework with which to inform discussions on hotspots of malaria. (106)

3.2.3 Hotspot

Where foci are considered the entire area where malaria transmission is possible, hotspots are considered to be those areas within foci that experience higher than average malaria transmission intensity. (148) More precisely, hotspots are areas where transmission intensity or risk is greater than the average for the area with the size of the hotspot smaller than the typical dispersal range of mosquitoes. Hotspot borders are defined based on predictive models of malaria exposure as detailed in section 4.2 based on data collected as part of community-based surveys. (106) Theoretically, it is malaria hotspots that are likely to be fueling transmission to the larger foci.

3.2.4 Operational

A major component of this work focuses on operational research to identify hotspots of transmission. Operational research means studies to inform implementation of programs in a way that optimizes the process including ensuring a cost-effective and sustainable approach. (216) When applying the term operational to the identification of hotspots of malaria transmission in the community, it is defined as providing evidence to support the use of a simpler yet effective approach to accurately identify hotspots of malaria in the community so they can be targeted for control that can be sustained by the capacity in malaria endemic regions.

In addition to establishing clear definitions for key terminology, there are several gold standards that have been established to facilitate this research.

3.2.5 Current malaria infection

A current malaria infection is considered those individuals that have live *Plasmodium* parasites, at any stage of development, in their bloodstream or liver. Nested polymerase chain reaction (nPCR) detecting the presence of the 18S ribosomal subunit of *P. falciparum* DNA is one of the most sensitive methods for detecting the presence of blood stage malaria infection and can detect parasite densities as low as 1 to 5 parasites/µl of blood. (192) In contrast, as discussed in section 1.3, rapid diagnostic tests (RDTs) or microscopy are not able to consistently detect parasites at such low densities. (174, 240) Therefore, for the purposes of this research, nPCR is considered to be the gold standard for establishing a current malaria infection.

3.2.6 Malaria Transmission

Malaria transmission is defined as the active transmission of the malaria parasite by the anopheline mosquito vector to the human population. (61) Similarly, the frequency of the transfer of *Plasmodium* from mosquito to human is associated with transmission intensity such that areas of high transmission intensity experience a higher frequency of infectious bites compared to areas of low transmission. (117) Measuring transmission intensity is ideally conducted using the entomological inoculation rate, a measure of the number of infectious mosquito bites a person receives over a given unit of time. (169) However, as

mentioned, collecting entomological data is challenging and parasite measures are subject to several biases or logistical difficulties. (51, 205, 241) Therefore, the gold standard for malaria transmission in our studies is the seroconversion rate (SCR) to $AMA1_{42}$ and/or $MPS1_{19}$ antigens as estimated from community-based surveys.

3.3 Rationale for selected operational approaches

3.3.1 Primary schools

The use of school surveys to monitor malaria burden is not a new phenomenon with the earliest reports of using this strategy dating back to the 1920's. (232) School surveys have recently been emphasized as a potential tool for quickly obtaining population based estimates, particularly once transmission has reduced to the point where community based approaches are no longer viable. (230, 233) Children provide a convenient sentinel population: the proportion of cases expected in the school age population has been estimated to be less stochastic than other age groups, tends to comprise between 20-40% of cases, and is highest in areas of moderate transmission intensity. (231) Also, school-aged children tend to have sufficient levels of immunity to prevent clinical disease but are not able to fully control parasite densities suggesting that they will be more likely to have infections that are detectable by microscopy or RDTs, further highlighting the utility of this group as a sentinel population. (54, 242, 243)

Primary schools have available infrastructure with which to conduct screening and could provide a secure location to store supplies for testing. The use of existing school supply distribution networks could also be used to facilitate distribution of malaria testing supplies. Schools also tend to be accessible to the local community and are used as community-meeting points in several settings, which may contribute towards improve community acceptability. (244) Malaria testing with RDTs is a simple process that can be easily taught (245). Therefore, teachers or regular community volunteers can provide the human capacity required to conduct regular malaria screenings.

Levels and equity of school attendance will limit the utility of school surveys in malaria surveillance. (232) Although many countries are making progress towards universal free primary education, there are still many barriers to uptake. It is

known that malaria risk is lower in wealthier households, which are those also most likely to attend school in areas where school fees may be prohibitive of attendance. (33) The second potential source of bias is the overlap between school catchment areas. In many areas, children do not necessarily attend the closest school to their home, some of which travel great distances. (246, 247) Therefore, the school-derived estimates of malaria burden may not always accurately reflect the immediate community and this bias must be better characterized.

Despite these potential biases, school surveys have been shown to identify heterogeneity in malaria transmission in different settings using both parasite and serological based diagnostic tools. (67, 161, 247) Therefore, given the cost-effectiveness and ease of integrating such an approach into a program at the national scale, school surveys have the potential to be an operationally attractive approach for malaria surveillance. (248)

3.3.2 Health facilities

Health facility based cross-sectional surveys where all attendees are included in the screening process in contrast to passive case detection where only those suspected of malaria are included, provide an operationally attractive means for collecting data on malaria transmission. (233) By including all attendees in the sampling framework, data is easily collected on a cross-section of the population instead of just those suspected of having malaria. A study in Tanzania found that SCR estimates obtained from health facility surveys were similar to those obtained in the community and were able to accurately capture the heterogeneity in malaria transmission in the area. (249) Similarly, a study in The Gambia found that seroprevalence estimates were able to detect the heterogeneity in malaria risk and were similar to estimates obtained from the surrounding community. (134) Therefore, the use of periodic all-attendee malaria surveys in health facilities may provide an alternative approach to malaria surveillance that could easily be integrated into the existing public health infrastructure.

Including all health-care seeking individuals in the malaria transmission surveillance program could mitigate against some of the known biases associated with passive case detection. Also, as mentioned, those referred for a confirmatory

test, as part of routine clinical practices is dependent on whether malaria is suspected and in some settings whether the patient is willing to pay for the testing. (250-252) Sampling all individuals with sensitive diagnostic tools mitigates these issues and may also ensure better reporting of data, another known issue with routine reporting. (134, 223) However, the use of the health care seeking population for malaria surveillance will likely have some inherent biases including differences in health care seeking behavior which may arise due to physical access or cost barriers, for example. (253-255) In malaria endemic areas, it is not uncommon for people to travel for more than one hour to access health services and influences the type and severity of cases presenting for care. (126, 256) Therefore, how these biases affect the utility of health facility surveys for malaria surveillance require further investigation.

Health facilities also provide an operationally attractive approach for malaria surveillance. Similar to school surveys, health facilities infrastructure is already established and sampling could easily be incorporated into existing activities: with trained personnel, established drug administration capacity, and supply delivery routes. (134, 257) Also, health facilities are commonly used distribution points for bednets, maternal care, vaccination campaigns, and other health initiatives, and in most settings are acknowledged as the health care provider by the community. Health facilities therefore provide a locally appropriate institution for a malaria surveillance initiative. (256, 258) Using health facility attendees for malaria surveillance also ensures a captive population and rapid sampling ensuring that large amounts of data can be quickly generated. (134)

3.4 Study Site

3.4.1 Population

This work was conducted in the western Kenyan highlands in the Rachuonyo South district, Nyanza Province, Kenya, centered on the town of Ringa (latitude: -0.47076, longitude: 34.853449). At the time of the study, Nyanza province, like all provinces in Kenya, was under the jurisdiction of a provincial commissioner, with authority then subdivided into districts overseen by a district commissioner and district officer. Each district was further sub-divided into locations, sub-locations and villages, which were administered, by chiefs and assistant chiefs. At the time of

this research, the administrative boundaries were not used to define the study site but are useful spatial reference to compare with other work conducted in the region.

The study area is described in more detail in subsequent chapters but briefly, the majority of the site falls between 1400 and 1600 m above sea level and is intersected with rivers and the terrain is marked with several rolling hills and valleys. (259) There is a bimodal rainfall pattern with the heaviest rains typically occurring between April and June and a smaller peak between October and December each year. The highest rainfall recorded in 2012 was in April with 431 mm and the driest month was January with only 1 mm of precipitation. Temperatures can range between lows of 14°C to highs of 32°C on average, with the hottest days typically occurring around January and coldest in July.(29, 260)

The study area is primarily rural with a few small towns dotted along the main highway that runs through the site. The study population is largely made up of the Luo ethnic group who are predominantly subsistence farmers with maize being the principal crop. The majority of people live in compounds of extended family units consisting of multiple houses in proximity to their fields. The terms compound and household are used interchangeably. The houses are typically constructed with mud walls, open eaves, and iron sheet roofing, however there are a few houses with brick walls and tiled roofs. (259) The town centers are much more densely population and people typically live in cement, multi-unit housing.

3.4.2 Malaria Epidemiology

In the western Kenyan highlands, malaria is biennial with 2 peaks in transmission following the 2 rainy seasons. This site is characterized as being in the highland fringe, or the area between highland epidemic prone and lowland endemic transmission settings. (261) The area is classified as having stable, mesoendemic transmission and has a mean parasite prevalence by RDT between 12 and 16% however prevalence is highly heterogeneous and can range from 0 to over 70%. (259) The predominant species of malaria is *Plasmodium falciparum* and the main vector species include *Anopheles arabiensis* and *An. funestus*. Both species are known to feed indoors and outdoor during the night, however, *An. funestus* is more

antropophilic while *An. arabiensis* is known to feed on cattle and other animals. (16, 27) Recently, an additional, currently incompletely characterized vector species with a high sporozoite rate was identified in the area. (29) Malaria vector control programs have been active with both indoor residual spraying (IRS) and long lasting insecticide treated bednet (ITN) distribution. An annual IRS program has been in place in the study area since 2005. Spraying is targeted to take place at the beginning of the long rainy season in April, however in practice spraying is sometimes implemented as late as August. Insecticides used are rotated and include the use of Lambdacyhaolthrin (Icon) in 2011 and both Icon and deltamethrin in 2012. IRS coverage is reported to be over 90%. A mass distribution of ITNs took place in 2011 where one Permanet 2.0 embedded with deltamethrin was distributed for every two people in a house. Ongoing ITN distribution is also provided through health facilities as part of their maternal health care clinic where one net is distributed to each pregnant mother. (260)

3.4.3 School Structure

Kenya is one of the many countries committed to free primary education and is under the jurisdiction of the provincial administrative director of education. A head teacher is responsible for the administration of each school. (248) In 2003, fees for government run primary schools were eliminated, which lead to a spike in enrollment of over 1 million children. (262) Despite the abolishment of school fees and improved access, there are estimates that approximately half of the costs to educate children still fall to the parents, including money for uniforms and sitting exams therefore there are still subsets of the population to which basic education is still unattainable. (262) Even with the continued presence of barriers to education primary school enrollment in the study area was good and estimated at 97.8% in 2007. In the larger study area, and including both public and private institutions, there were approximately 289 primary, 99 secondary, 9 post-secondary and 13 non-formal schools, of which 76 primary, 35 secondary, 3 post-secondary and 4 non-formal schools were located in the restricted study area.(263) There is currently a precedent for national school-based de-worming and malaria surveillance programs across the country. (244, 263)

3.4.4 Health Facility System

In Kenya, the health care system comprises government, private, and faith-based facilities and all report to the district medical officer of health and district public health officer. (178, 264) Health facilities are stratified according to size and services offered with national and provincial level facilities, serving as referral points for the smaller units. Within each province, district and sub-district hospitals are present which offer comprehensive medical care including surgery facilities. (265) Health centers are smaller than sub-district hospitals and are typically staffed with at least one clinical officer and provide for most primary care services. (266) The dispensaries offer the lowest level of health services and are run by a nurse in charge but are supervised by the clinical officer at the nearest health center. A network of community health and support volunteers is also present within most health facilities in Kenya and tasked with outreach, data collection, and other activities, depending on the needs of the facility. (265)

The 12 functional health facilities in the area were identified based on community consultation and included five government, five faith-based, and two private institutions. The district hospital is located approximately 15 km west of the study area in the town of Ouygis and a sub-district facility is approximately five km to the northeast in Kabondo. The policy in Kenya is consistent with current WHO guidelines that all malaria cases should be diagnosed before treatment. (75) However, in this area, no facilities had supplies of rapid diagnostic tests (RDT). At the time of this survey approximately seven of the 12 facilities in the area had a working laboratory capable of testing malaria by microscopy, although the quality of facilities varied dramatically. The first line treatment for uncomplicated malaria infection is artemether-lumefantrine combination therapy (AL; Coartem©) with quinine injections used for severe malaria and as a second line therapy. (260)

Ultimately, the study site in the highlands of western Kenya provided an ideal setting to explore the many important research questions addressed in this thesis. The landscape and endemic and highly heterogeneous transmission (259, 267) present a unique opportunity to define hotspots of malaria as well as to explore operational alternatives to identify these areas that may be critical to identify so that they can be targeted with malaria control interventions.

Chapter 4: Results - Defining hotspot of malaria transmission

In order to apply a hotspot targeted strategy to malaria control and elimination practices, the ability to accurately define areas in the community that experience higher transmission intensity is critical. This chapter explores the impact of different practical issues present when identifying hotspots of malaria transmission in the community including the choice of spatial statistical approach, malaria metric, and the importance of sample sizes on where hotspot boundaries are drawn. To address the objective of defining hotspots of malaria (specific objective 1), data from a large community cross-sectional survey was used as described in section 4.2.

4.1 Background and rationale

As geographic information systems (GIS) have become more accessible over the past two decades for both research and programmatic disease surveillance, spatial analysis of malaria has become common practice, and has lead to the concept of 'shrinking the malaria map'. (15, 136, 145) Identifying the country-level spatial heterogeneity in malaria burden has been useful for prioritizing areas for malaria control (9) and to identify thresholds of endemicity for gauging change in transmission intensity. (13) However, the high heterogeneity in malaria transmission at the local level is increasingly being recognized (17, 120) with theory suggesting that there are specific individuals or defined areas that experience a disproportionate burden of malaria. (97, 106) Identifying these highly exposed populations at the local level could be extremely useful for malaria control programs to employ a targeted strategy directing resources to those experiencing the highest burden. The first step to any hotspot targeting strategy however is to define the populations of interest. In the context of this research, defining hotspots is considered to be the more theoretical exercise to determine where the hotspot boundaries are drawn. In contrast, identifying hotspots, as discussed further in chapter 6, is considered to be the more practical component related to operationally detecting the areas determined to be hotspots of transmission for subsequent intervention.

4.2 Hotspots of malaria: Impact of geostatistical methods on determining boundaries of hotspots of malaria (P1)

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk
Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic pier review?	

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	Nature Communications
Please list the paper's authors in the intended authorship order.	Gillian H Stresman*, Emanuele Giorgi*, Amrish Baidjoe, Phil Knight, Wycliffe Odongo, Chrispin Owaga, Shehu Shagari, Euniah Makori, Jennifer Stevenson, Chris Drakeley, Jonathan Cox, Peter J Diggle, Teun Bousema *Authors Contributed Equally
Stage of publication	Pending in country publication committee approval

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceived the research questions behind this paper, collected the field data, conducted all data analysis except for MBG, and wrote the manuscript.

Student Signature: _____ **_____ Date: ___17/02/2015_____**

Supervisor Signature: ____ **_____ Date: ___17/02/2015_____**

Abstract

The spatial heterogeneity of many diseases suggests that a targeted approach to control and elimination programs could be an attractive option to maximize impact while minimizing intervention costs. However, in the malaria literature in particular, there has been little consistency in how such areas of focal transmission are defined. Here we assess the impact of different malaria metrics (parasitological and serological), sample size, and hotspot detection technique (model-based geostatistics and spatial scan statistics) on the delineation of hotspot boundaries. Using data from a large community-based malaria survey in the western Kenyan highlands, we show that the choice of malaria metric, sample size, and statistical method have a significant impact of the size and location of hotspots, with only poor to moderate agreement in the households identified as being part of a hotspot. Our results provide the first comprehensive assessment of the challenges associated with applying hotspot theory to practice.

Introduction

Malaria is an important cause of global morbidity and mortality with an estimated 3.4 billion people at risk. [1] The past decade has seen a large reduction in the malaria burden in some areas and an estimated 47% reduction in mortality compared to 2000. [2] As country policy shifts from control towards an elimination agenda [3] new approaches are needed to supplement existing tools. [4] Increasingly, research and programmatic activities are focusing on the heterogeneous nature of malaria transmission at the community level. Identifying and targeting 'hotspots' (ie. local foci of intense malaria, which may also fuel transmission in surrounding areas) with control interventions, could lead to a more sustainable reduction in malaria burden .[5, 6] Several studies have been conducted that have identified 'hotspots' of malaria at various spatial scales. However, methodologically, there has been little consistency as studies have used a range of different malaria metrics, cluster detection methods, or assumptions within the same method. [7-10]

As hotspots of malaria increasingly become the focus of malaria research and control practices [11] it is important to ensure that we are able to identify the most robust and meaningful approaches for defining where hotspot boundaries lie. For example, there is evidence that certain spatial statistical approaches are able to

consistently detect 'hotspots' over time [10, 12] and that hotspots are more likely to have malaria cases in subsequent years. [13] However, the impact of different statistical techniques on the delineation of hotspot boundaries, the likelihood of identifying false hotspots, and consequently the likelihood of individual households being inappropriately targeted for (or excluded from) any subsequent intervention, are unknown. Also, although there is evidence to suggest that the sensitivity of the malaria metric used to inform hotspot detection is important, [14] there is little understanding of the impact of this on hotspot delineation. Model-based geostatistics (MBG) [15] can predict areas of increased disease prevalence and has been effectively applied in other disease systems that exhibit both large and small-scale variation in disease transmission. [16] However, in the context of malaria to date, these methods have mainly been applied at the national or provincial scale, which have limited relevance for informing community level control practices. [8, 17, 18]

Here we use data collected in a large cross-sectional malaria survey carried out in the western Kenyan highlands to demonstrate a novel MBG-based approach for the detection of local-level hotspots of malaria while accounting for uncertainty in their estimated severity and spatial extent. We ran models based on two different malaria metrics: parasitaemia (measured by polymerase chain reaction [PCR]) and seropositivity in order to determine the level of consistency in the output. The models for each malaria metric were then applied to random subsets of the data to determine the impact of sample size on the number, size and location of identified hotspots. The MBG maps for both malaria metrics were also compared to results obtained using SatScan, the most commonly applied cluster detection technique currently used in malaria research, to determine the degree of consistency between the two methods. The results generated are not intended to provide a gold standard for hotspot detection, but to illustrate the difficulties in translating theoretical concepts of malaria transmission hotspots to practices that can effectively be applied to reduce malaria transmission as part of malaria control programs.

Methods

Data Sources

EPIDEMIOLOGICAL DATA

Epidemiological data were obtained from a community cross-sectional malaria survey conducted in July 2011 in a 100 km^2 area situated in a rural area in the western Kenyan highlands (0°28'S, 34°51'E). [11] This population is primarily comprised of people from the Luo ethic group with subsistence farming being the main occupation. The area is characterized by low but spatially heterogeneous malaria transmission, with *Plasmodium falciparum* being the predominant malaria species. [19] Firstly, all houses in the study area were digitized using high-resolution satellite imagery (Quickbird, DigitalGlobe Services Inc, USA) [11] and was used as a proxy for the total population size and distribution. Briefly, 17,503 individuals residing in 3,213 randomly selected households selected from the list of digitized houses were sampled with each participant providing three filter paper blood spot samples. Key data relating to characteristics of individuals and households were obtained using questionnaires. Filter paper blood spots were assayed by PCR to detect the presence of a current malaria infection. [20, 21] Age-seroprevalence curves were fitted for antibody response to AMA1 and/or $MSP1_{19}$ measured by enzyme linked immunosorbent assay to provide a measure of malaria exposure. [22, 23] Ethical approval for the collection of the epidemiological data was granted by the ethical committees of the London School of Hygiene & Tropical Medicine (LSHTM 5721) and the Kenya Medical Research Institute (SSC 1802).

ENVIRONMENTAL DATA

Mean elevation for each compound was derived from Version 2 of the ASTER global digital elevation model (DEM) (NASA, USA). The normalized-difference vegetation index (NDVI) was calculated for the study area using the Quickbird satellite imagery. The mean, minimum, and maximum NDVI values were calculated for a 500 m circular buffer around each compound.

To obtain land classification data, multispectral image segmentation (MIS) of the Quickbird imagery was conducted with eCognition (v 4.0, Trimble Geospatial Imaging, Germany) software and the proportion of tree cover within the 500 m circular window surrounding each compound was determined. Other land cover

classifications such as roads, buildings, and bare earth, were also extracted but were not associated with either outcome tested. Fishponds were identified using a refined MIS procedure capable of detecting smaller features and manually verified against the satellite imagery. The distance from each compound to the nearest fishpond was calculated in ArcGIS (ESRI, USA).

Two variables associated with water were included in the model: a topographic wetness index (TWI) and distance to streams. To calculate TWI, the DEM was used to generate a surface representing the flow direction and accumulation of water and corresponding TWI for the study area. [24] The maximum and mean TWI values for the 500 m surrounding each compound were then calculated. Finally, the locations of all streams in the area were determined by first locating the likely location of streams using the topographic data and then manually digitizing the more precise stream path using the satellite imagery and classified according to the Strahler method for determining stream order. [25] The distances of each house to all stream orders were calculated.

Determining hotspots of P. falciparum infection and exposure

A model-based geostatistical approach was used to model the spatial variation in malaria prevalence. [15, 16] Two models were generated: malaria infection was assessed using PCR prevalence and exposure to malaria was assessed using seroprevalence estimates from the community cross-sectional survey. Model covariates were restricted to the environmental variables described above. Surfaces of the predicted prevalence for both outcomes were generated and used to determine informative thresholds of risk for what would be considered as a hotspot of malaria. Thresholds were determined by assessing the predicted prevalence that encompassed 20% of the population. [26] Next, the probability that any given area exceeded this threshold was determined and those with greater than 80% probability that malaria prevalence exceeded the predetermined threshold were considered to be hotspots. This process was repeated for both outcomes to generate separate surfaces for hotspots of current infection and exposure to malaria (see appendix 1.4 for detailed methodology). All analyses were conducted in R v.3.0.2 (R-Project, USA).

Model validation

Model validation is directed at ascertaining whether the fitted model adequately reproduces the spatial correlation structure of the data. To achieve this, we first fit a simple logistic regression model to the data, i.e. adjusting for the regression effects of the environmental variables but ignoring any spatial correlation. We then calculated the empirical semi-variogram of the residuals from this model, which provides an estimate of the underlying spatial correlation structure of the data. Next, we repeat the logistic regression fitting and variogram calculation from each of 10,000 datasets simulated under the fitted model. From the simulated variograms, we calculate pointwise 95% tolerance bounds for the semi-variogram under the assumption that the model generated the data. An adequate fit to the data is indicated if the empirical semi-variogram falls within the tolerance bounds. Secondly, cross-validation was conducted by fitting the model to a random sample of 70% of the dataset, calculating the root-mean-square prediction error (RMSE) of prevalence over the 70% sample, and comparing this with the RMSE achieved when the fitted model was used to predict prevalence at the locations of the remaining 30% of the data.

Impact of Sample Size

The impact of sample size on model estimates was assessed by the change in two summaries of predictive performance: the integrated mean square error (IMSE) for the predicted surface; and the discrimination index (DI) for the exceedence probabilities at the determined prevalence threshold. (see appendix 1.4). [15] First, to estimate what level of performance would have been achieved if 100% of the population had been sampled, we imputed a complete population data set using the predicted malaria risk surfaces generated using the available data and assigning the corresponding household prevalence for the complete set of digitized households. [11] Next, we selected a random sub-set of the imputed data for each of the sampling fractions 10-90% and fitted the geostatistical models to each sub-set. The corresponding IMSE and DI values were calculated and plotted as functions of the sampling fraction.

The next step was to determine the impact of a reduced sample size on hotspot boundaries. For this we delineated hotspot areas using the definition given above

in conjunction with the geostatistical model fitted to the actual data and classified each compound accordingly as being within a hotspot or not. We then re-fitted the geostatistical model to random samples of the data, with sampling fractions between 10-90%. The resulting exceedance probability surfaces were imported into ArcGIS and hotspot boundaries were determined. The sensitivity and specificity of the houses correctly identified, using the complete sample as the reference, were calculated and compared using the area under the curve (AOC) from a receiver operator curve (ROC) analysis. [27]

Comparison with spatial scanning tool

Hotspots identified using MBG were compared to those generated using a spatial scan statistic (SatScan, USA), currently the most commonly used approach for cluster detection within the malaria hotspot literature. [10, 28-30] For this, we first determined spatial scan statistics for the entire field. Using the Bernoulli model, scans were conducted for both PCR and seropositivity outcomes, using both circular and elliptical scanning windows and allowing the scanning window size to be a maximum of 25% and 50% of the total population. The expected prevalence consisted of the global mean prevalence (ie. the mean of the entire study area). [31] All households that were found to have a significantly greater (α=0.05) prevalence than expected were identified as being part of a hotspot. Hotspots derived using the various scanning approaches were visualized using ArcGIS. A second set of SatScan analyses was conducted using a locally weighted expected prevalence with the different scanning assumptions as has been previously described. [11] Results were compared between metrics, methods, and assumptions at the structure level using the Pearson correlation coefficient and the kappa statistic. The hotspots consistently identified with at least partially overlapping boundaries as well as the number of hotspots identified using MBG that were missed by SatScan were also assessed. To determine the sensitivity of the uncertainty in the exceedence prevalence threshold set to define a hotspot in the MBG-based method, a comparison was also conducted assuming any structures within the areas that had greater then 50% probability of exceeding the defined threshold.

Results

MBG Models

The results of the geostatistical model suggested a positive association between PCR prevalence and maximum and mean NDVI and a negative associated with mean elevation, distance from fishponds and the proportion tree cover. The optimum model fit for seroprevalence also indicated a negative association with mean elevation, distance from fishponds and tree cover. In addition, maximum TWI, minimum NDVI and distance to 2nd and 3rd order streams also had a negative association while mean TWI had a positive association with seroprevalence (table P1-1).

The predicted prevalence for PCR infection obtained from the spatial model suggests that there is spatial heterogeneity of malaria infection in this 100 km^2 area (figure P1-1A). The areas with a predicted PCR prevalence greater than 28% encompassed 20% of the total population providing a threshold for what is considered a hotspot of current malaria infection in this area. Next, the seroprevalence estimates generated by the model also suggest that heterogeneity in exposure is present within this small study area and that overall exposure levels were much higher than those of current infection (figure P1-1B). The threshold for what is considered a hotspot of malaria exposure was determined to be those areas where predicted seroprevalence exceeded 70%.

Table P1-1: Final adjusted mixed effects logistic regression models for both outcomes. NDVI=normalized difference vegetation index; TWI=topographic wetness index

PCR Prevalence				Seropositive			
	Estimate	Std. error	p.value		Estimate	Std. error	p.value
Intercept	5.430	3.272	0.097	Intercept	7.972	2.165	0.0002
Mean Elevation (m)	-0.007	0.002	<0.0001	Mean Elevation (m)	-0.005	0.001	<0.0001
Maximum NDVI	1.532	1.030	0.137	Max TWI	-0.011	0.011	0.297
Mean NDVI	5.132	2.934	0.080	Mean TWI	0.230	0.104	0.028
Distance from Fish Pond (m)	-0.001	0.000	0.000	Minimum NDVI	-0.227	0.229	0.320
Tree Cover (%)	-3.094	1.473	0.036	Distance from Fish Ponds (m)	-0.0005	0.0001	<0.0001
				Distance 3rd Order Stream (m)	-0.0001	0.000	0.039
				Distance 2nd Order Stream (m)	-0.0002	0.0001	<0.0001
				Tree Cover (%)	-2.921	0.8194	0.0004

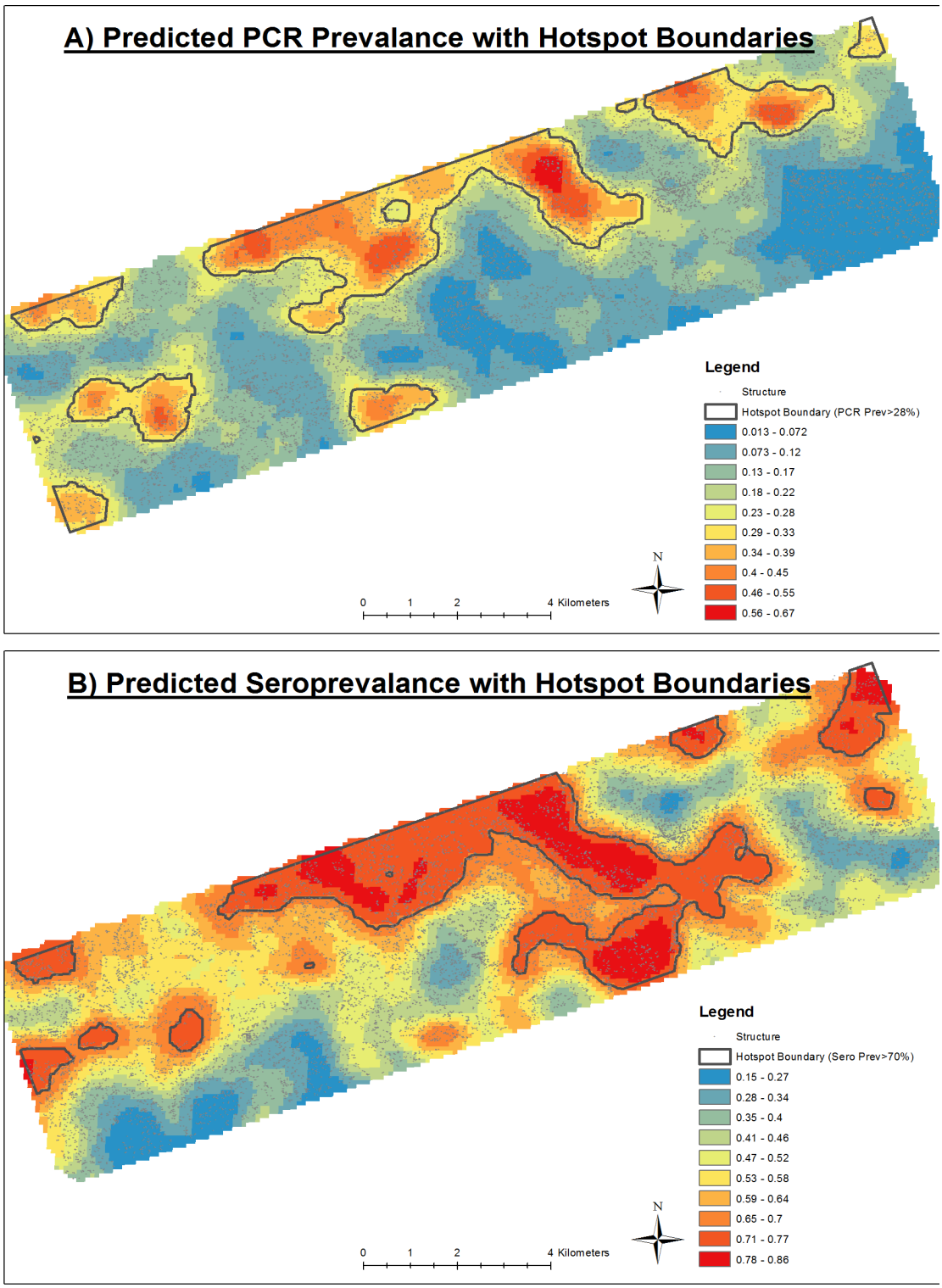

Figure P1-1: Results of the modeled predicted prevalence of A) current malaria infection with overlaid hotspot boundaries showing the area that has a predicted PCR prevalence greater than 28% and B) malaria exposure as measured by seroprevalence with overlaid hotspot boundaries showing the area that has a predicted seroprevalence greater than 70%.

Next, the probability of predicted prevalence exceeding the 28% and 70% thresholds was mapped for both PCR (figure P1-2A) and seropositivity (figure P1-2B), respectively. The hotspot boundaries for PCR infection and seropositivity (ie. areas that had a probability >80% of exceeding the threshold) encompassed 6.0% and 8.3% of the population, respectively. The percent agreement between PCR and seroprevalence at this probability threshold was 92.3% (Kappa=0.424). Boundaries corresponding to areas with greater than 50% probability of exceeding the threshold included 17.9% and 21.6% of the population for PCR and seroprevalence, respectively. With the more relaxed definition of hotspots consisting of areas where the probability of exceeding the threshold was >50%, the percent agreement was 83.4% (Kappa=0.478).

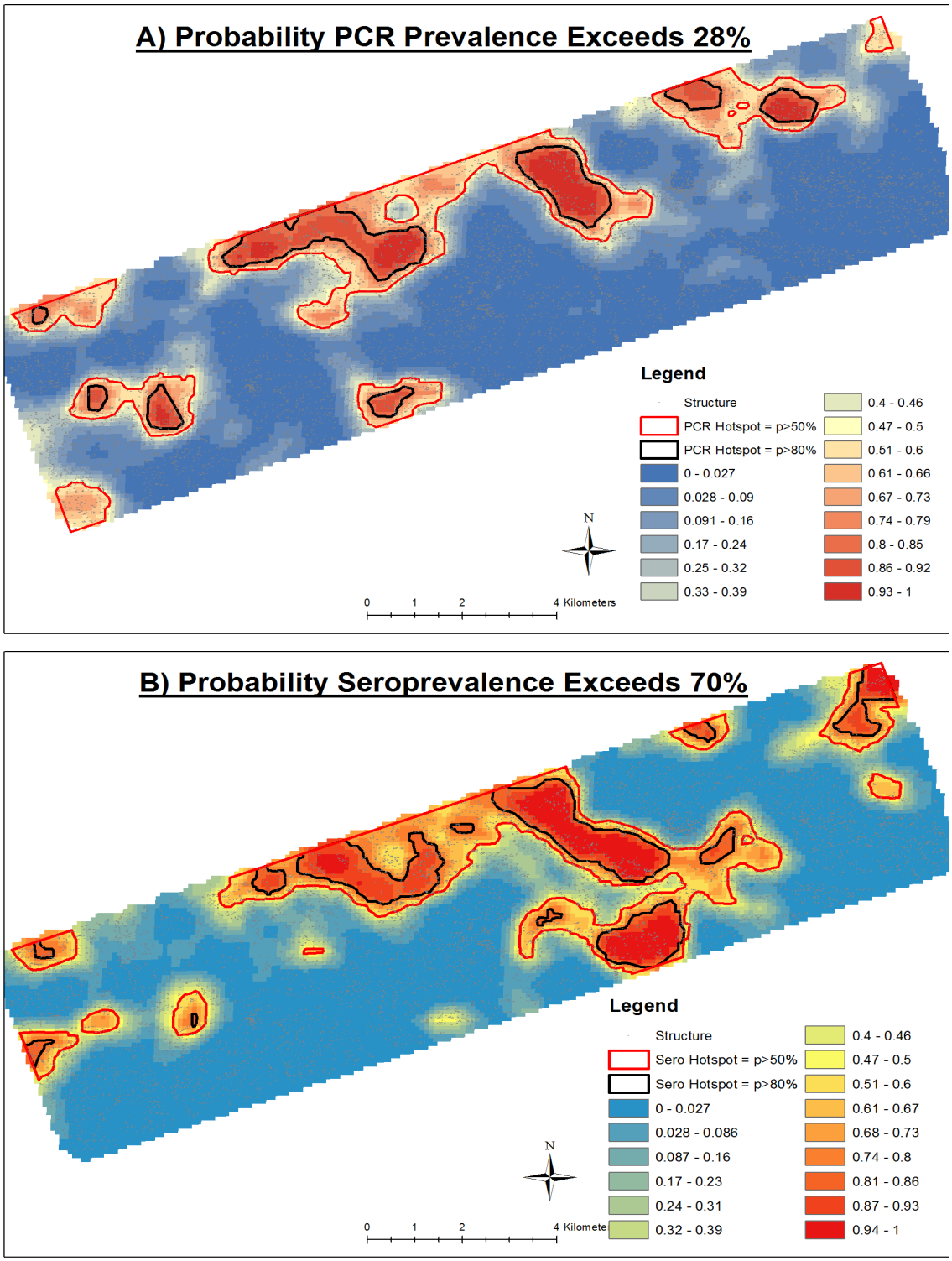

Figure P1-2: Probability contour maps of the study area indicating the probability that the prevalence of malaria A) infection by PCR and B) exposure by seroprevalence exceeds 28% and 70%, respectively with the corresponding hotspot boundaries using both 50% and 80% thresholds.

MBG Model Validation

Two methods to validate the model were used. The results of the probabilistic model validation for both the PCR (figure P1-3A) and seroprevalence (figure P1-3B) outcomes suggest that the model fits well, as in each case the empirical semi-variogram lies well within the 95% tolerance limits throughout its range. The semi-variograms also suggest that there is residual spatial dependence in both PCR and seroprevalence up to 1.5km. Secondly, the results of the cross-validation also suggest well-fitting models. For the PCR model, the MSE of the fitting and validation sub-sets dataset were both 0.259 whereas the MSE for the seroprevalence model were also similar at 0.278 and 0.242 for the fitting and validation sub-set, respectively.

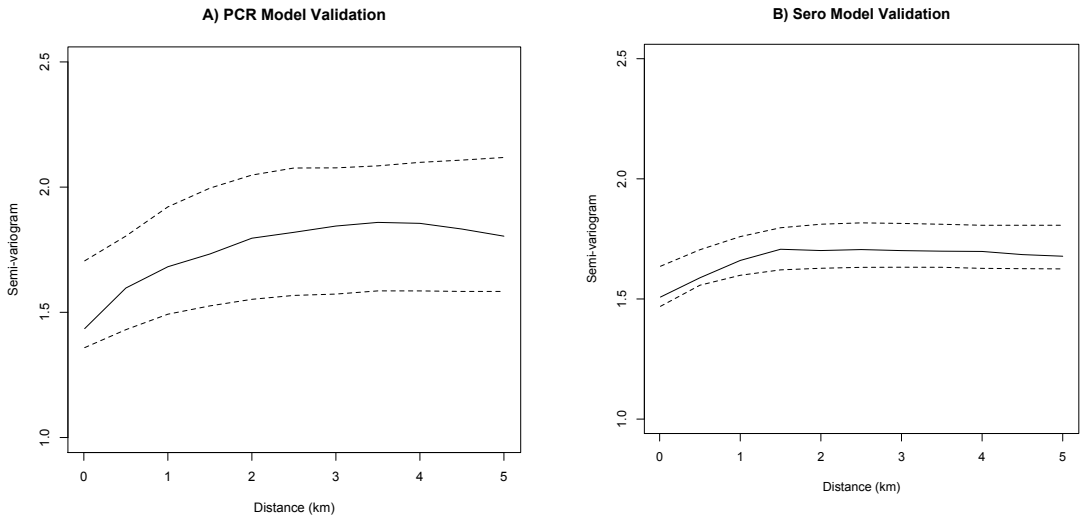

Figure P1-3: Semi-variogram (solid line) and predicted 95% tolerance bounds (dashed lines) for probabilistic model validation for A) PCR prevalence and B) Seropositivity

Impact of Sample Size on MBG Estimates

We found a negative relationship between the sample size and the relative increase of IMSE in the MBG. The rate of relative increase was similar between PCR (Figure P1-4A) and seroprevalence (Figure P1-4B) models and had a greater impact on the efficiency in modeling the predicted surface compared to the probability contour map. Based on this imputed dataset, the proportion of the population that was sampled as part of the survey resulted in a relative increase of 0.4 in IMSE for the predictive surface.

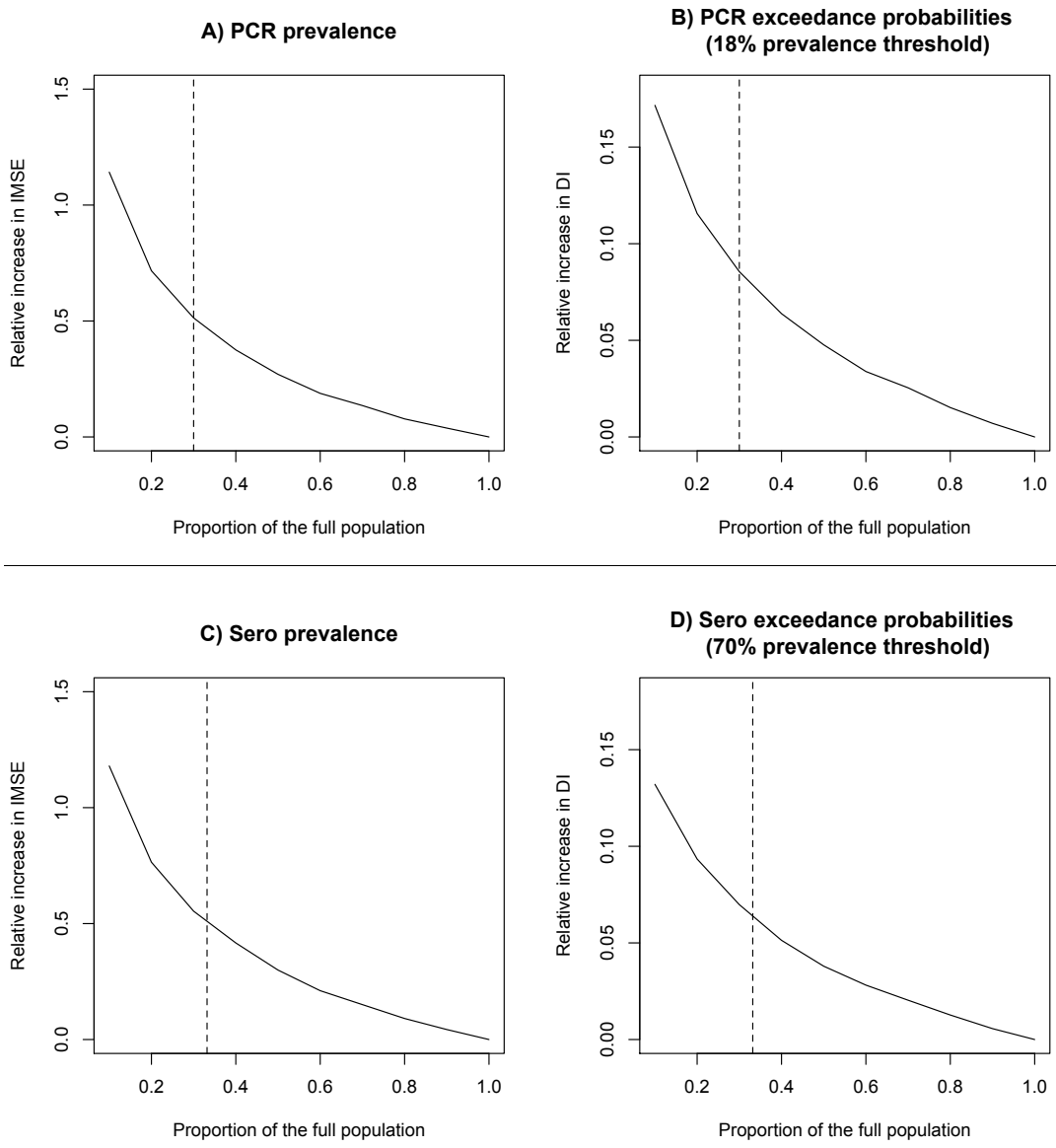

Figure P1-4: The impact of reduced sample size on model efficiency for both the predicted and probability surfaces for both PCR (A,B) and seroprevalence (C,D), respectively (solid line) with the dashed vertical line representing the sample size achieved during the community survey.

Next, the difference in hotspot boundaries, as determined by the structures that were consistently identified to be part of hotspots was assessed (table P1-2). The PCR prevalence model showed a change in the number of structures identified as being within or outside hotspots when sample size was reduced to 70% of the dataset, or 20.9% of the total population. A second significant change in AOC was observed with 30% of the dataset, or 9.0% of the total population. The geostatistical model for PCR prevalence showed no ability to reliably detect

hotspots with fewer than 10% of the sampled, or 3.0% of the total population. The impact of sample size on the geostatistical models for seroprevalence showed similar results. A significant reduction in the ability of the model to consistently draw hotspot boundaries occurred at 70% of the sampled population, or 23.4% of the total population. A second drop in hotspot consistency occurred at 40% or 13.3% of the sampled and total populations, respectively. When considering these results in combination with the increase in IMSE (figure P1-4), these are likely to be underestimates.

Table P1-2: Results of the impact of sample size on the ability to consistently detect the same structures as being located inside hotspots of malaria infection (PCR prevalence) and exposure (seroprevalence).

	PCR Prevalence				Seroprevalence			
% of Sample	% of Total Pop	AOC	Std. Error	95% CI	% of Total Pop	AOC	Std. Error	95% CI
100	29.9	1.0	-	-	33.2	1.0	-	-
90	26.9	0.923	0.0048	0.914-0.933	29.9	0.926	0.0039	0.918-0.934
80	23.9	0.896	0.0054	0.885-0.906	26.6	0.913	0.0042	0.905-0.921
70	20.9	0.847	0.0061	0.835-0.859	23.4	0.859	0.0050	0.849-0.869
60	17.9	0.812	0.0065	0.799-0.824	19.9	0.855	0.0050	0.845-0.865
50	14.9	0.819	0.0064	0.807-0.832	16.6	0.866	0.0049	0.856-0.875
40	12.0	0.834	0.0062	0.821-0.846	13.3	0.773	0.0056	0.761-0.784
30	9.0	0.739	0.0067	0.726-0.752	10.0	0.804	0.0054	0.793-0.815
20	6.0	0.693	0.0066	0.680-0.706	6.6	0.706	0.0056	0.695-0.717
10	3.0	-	-	-	3.3	0.744	0.0057	0.733-0.755

Comparing Methods: MBG vs. SatScan

Overall, the results of the two methods tested for detecting hotspots of PCR prevalence suggested poor overlap in the structures consistently detected as being part of a hotspot of malaria infection. First, the global SatScan approach tended to overestimate the size of hotspots compared to MBG and this was consistent throughout all scanning assumptions tested (Figure P1-5A). Although the percent agreement between the MBG and global SatScan for PCR prevalence for both

exceedance probability thresholds tested was high, the Kappa statistic and correlation coefficient suggest that agreement is moderate and was better when the less conservative definition of MBG hotspot (p>50%) was used (table P1-3). Also, between 1 and 4 hotspots that were detected using MBG were missed by SatScan, depending on the assumptions used. Next, the different assumptions applied to the locally weighted SatScan approach resulted in fewer hotspots missed when compared to MBG (figure P1-5B). However, the SatScan approach identified additional hotspots that were not detected by MBG. The locally weighted scans also resulted in poor agreement when compared to MBG (table P1-3).

Table P1-3: Multiple Satscan comparisons with MBG as the gold standard for PCR outcome showing results for defining the exceedence threshold as areas with probability greater than 80% (per protocol) as well those areas with any increased probability (p>0.5) of exceeding the defined threshold (Any). Satscan comparisons included adjusting the window shape and size as well as a combined metrics for being identified as having significantly greater risk by any of the assumptions. Satscan results are shown using a global scanning assumption encompassing the entire study population as the comparison and a locally weighted scanning assumption determining increased risk compared to the surrounding area. HS=Hotspot; Kap=Kappa statistic; Corr=correlation coefficient

Per Protocol - p>0.80						Any – p>0.5				
PCR-POSITIVE	# HS overlap	# HS miss	Agrmt (%)	Kap	Corr	# HS overlap	# HS miss	Agrmt (%)	Kap	Corr
Global Scan										
Circular Window										
Max 50%	5	3	73.0	0.197	0.300	3	4	76.3	0.376	0.402
Max 25%	7	1	71.4	0.188	0.297	5	2	76.3	0.398	0.432
Elliptical Window										
Max Size 50%	7	1	75.6	0.209	0.301	4	3	79.9	0.441	0.461
Max Size 25%	7	1	80.0	0.246	0.325	4	3	83.2	0.490	0.496
Combined	7	1	69.0	0.162	0.263	5	2	75.7	0.399	0.441
Locally Weighted Scan										
Circular Window										
Max 1k	6	2	81.5	0.227	0.284	7	0	79.3	0.332	0.333
Max 250	6	2	79.6	0.199	0.258	7	0	77.4	0.299	0.301
Elliptical Window										
Max Size 1k	7	1	80.0	0.222	0.288	6	1	78.4	0.332	0.335
Max Size 250	6	1	80.0	0.221	0.287	7	1	78.4	0.331	0.335
Combined	7	1	76.8	0.207	0.289	6	1	75.6	0.300	0.309

Finally, the hotspots defined using the seroprevalence geostatistical model showed a reasonable overlap with the results obtained from SatScan for both the global (figure P1-5C) and locally weighted (figure P1-5D) scanning approaches. Visually, the global scanning approach for all SatScan assumptions tested appear to define larger hotspots compared to those detected by MBG while between 1 and 4 hotspots detected by MBG were missed by SatScan (table P1-4). Results of the global scans suggest that there was poor agreement between the methods with Kappa ranging between 0.133 and 0.252 when the conservative definition of MBG hotspot was used ($p>0.80$) and improving to moderate agreement (kappa range: 0.376-0.546) with the relaxed MBG hotspot definition ($p>0.50$) (table P1-4). Finally, the comparison of the locally weighted scans for seroprevalence also suggest moderate to low agreement between methods. However, when considering the relaxed definition of MBG hotspot ($p>0.50$) hotspots detected by SatScan overlapped, at least partially, with all hotspots defined by MBG (figure P1-5D).

Table P1-4: Satscan comparison with MBG as the gold standard for Sero outcome showing results for defining the exceedence threshold as areas with probability greater than 80% (per protocol) as well those areas with any increased probability (p>0.5) of exceeding the defined threshold (Any). Satscan comparisons included adjusting the window shape and size as well as a combined metrics for being identified as having significantly greater risk by any of the assumptions. Satscan results are shown using a global scanning assumption encompassing the entire study population as the comparison and a locally weighted scanning assumption determining increased risk compared to the surrounding area. HS=hotspot; Kap=Kappa statistic; Corr=correlation coefficient

	Protocol - p>0.80					Any – p>0.5				
SERO-POSITIVE	# HS overlap	# HS miss	Agrmt (%)	Kap	Corr	# HS overlap	# HS miss	Agrmt (%)	Kap	Corr
Global Scan										
Circular Window										
Max 50%	11	1	73.6	0.245	0.333	5	4	81.8	0.541	0.560
Max 25%	11	1	74.3	0.252	0.339	5	4	82.1	0.546	0.563
Elliptical Window										
Max 50%	8	4	59.7	0.155	0.267	7	2	69.8	0.376	0.442
Max 25%	8	3	64.0	0.183	0.295	7	2	73.2	0.418	0.471
Combined	11	1	55.6	0.133	0.248	7	2	67.4	0.358	0.442
Locally Weighted Scan										
Circular Window										
Max 1k	10	2	72.8	0.204	0.273	9	0	73.0	0.312	0.321
Max 250	9	2	73.5	0.210	0.276	9	0	73.2	0.306	0.313
Elliptical Window										
Max 1k	8	3	77.8	0.246	0.303	9	0	76.7	0.355	0.357
Max 250	8	3	75.3	0.236	0.306	9	0	75.1	0.345	0.351
Combined	9	3	71.8	0.225	0.314	9	0	73.2	0.342	0.358

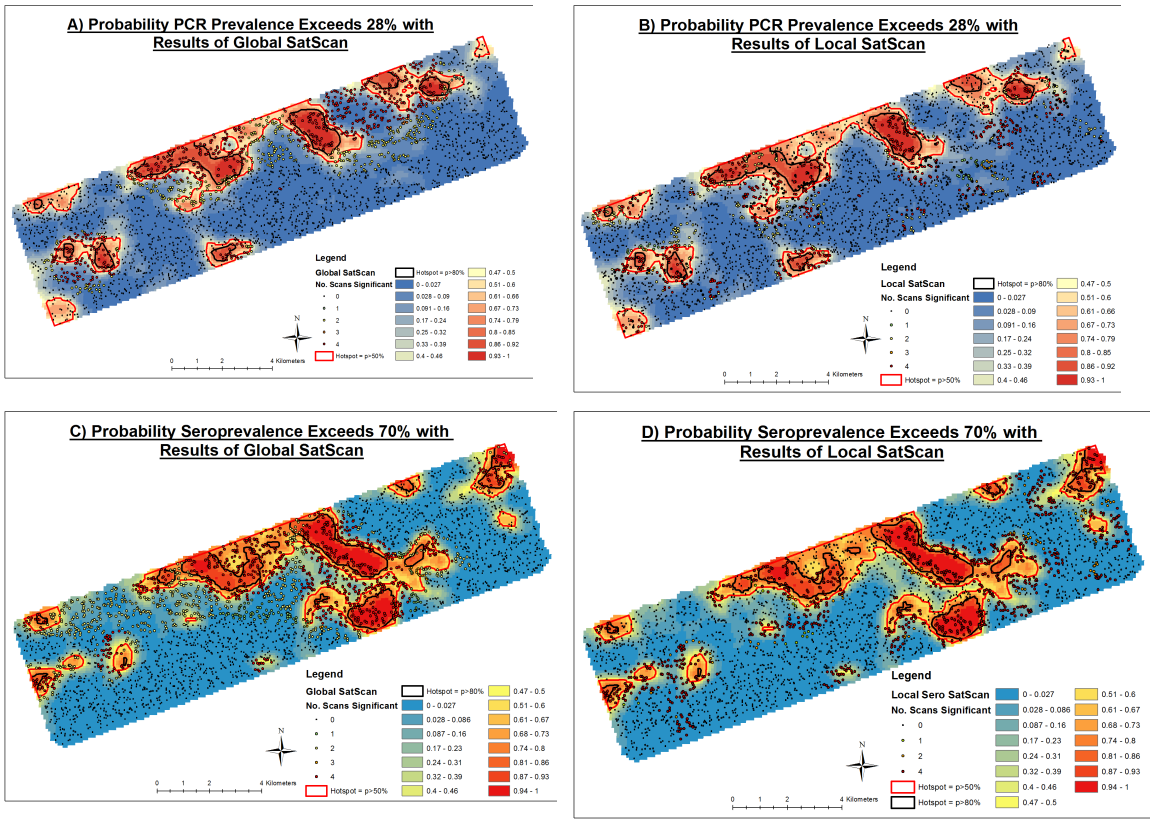

Figure P1-5: Probability contour map showing the probability that prevalence of malaria infection exceeds the defined threshold with Satscan results superimposed showing the households that were located within a hotspot based on the different assumptions tested for PCR and seroprevalence for both A, C) global and D, B) locally weighted scanning approaches, respectively.

Discussion

We detected highly heterogeneous transmission in the western Kenyan highlands using different spatial analytical approaches. Comparing the outcomes of these analytical approaches we have identified several challenges in our ability to consistently detect hotspots of malaria infection as different methods and metrics produced different results. These uncertainties in our ability to define hotspots of transmission have implications for implementing control or elimination strategies.

We used two different malaria metrics, parasite prevalence by PCR and seropositivity to antimalarial antigens providing measures of current infection and malaria exposure. The different metrics do not provide the same picture of where

hotspots are located in the community. The two malaria metrics included in this study are inherently measuring different things in terms of malaria transmission and it is possible that they are also measuring different facets of hotspot dynamics. [32, 33] For example, hotspots of current infection that were missed by serological measures may reflect a more epidemic or temporary hotspot whereby seroprevalence estimates have yet become high enough to be considered 'relatively' higher compared to the other areas of more consistent exposure. [9] For analytical methods to successfully target the true extend of hotspots it may be necessary to combine both metrics, or to use alternative measures like the reproductive rate (R_0), for example. [34] Regardless of which metric is used, the resulting hotspots should be interpreted accordingly (i.e. hotspots of infection or exposure).

In addition to malaria metric, sample size also resulted in significant changes in where the hotspot boundaries were drawn using the MBG approach. The purpose of this study was not meant to carry out robust assessment of the sample sizes required to conduct spatial analysis but to illustrate the point that sample size matters in determining hotspot boundaries. We determined that the first significant change in the structures identified as being part of hotspots occurred when one-fifth of the total population was sampled. When this is considered in conjunction with the baseline relative increase in IMSE given that we only had a random selection of the population to start with, this is likely an underestimate. Although a more rigorous sample size analysis would be useful for both MBG and SatScan to determine both the optimal number and distribution of points, [35] our result suggests that the number and/or distribution of points available will impact where the boundaries are drawn.

Although we cannot determine which statistical approach is better able to accurately identify and define true hotspots in the community, these results indicate that the approach used will affect the resulting map. The two approaches tested are very different in their use and are based on different assumptions. The MBG approach is generally used to fit a spatial residual risk surface and allows for a greater understanding of the nature of malaria hotspots by letting the overall risk surface depend on both measured and unmeasured risk factors for malaria.[15] The

alternative, SatScan, is a simple testing procedure aimed at assessing whether a disease shows spatial clustering and provides much less information on the dynamics of the disease. Additionally, almost every disease shows spatial clustering; hence testing this very general hypothesis is not usually of much interest.[36, 37] Some of the differences observed between the two cluster detection approaches are likely due to factors such as edge effect, the subjectivity in defining boundaries around points that were identified to have a statistically significantly greater risk of being part of a hotspot vs. defining thresholds and probabilities of increased prevalence, or how the models were able to address the complex dynamics of malaria transmission in this area.[6, 16, 38] Regardless of the reasons for the differences between methods, it is essential that those using these approaches acknowledge the limitations of the methods and our understanding of what constitutes a hotspot of malaria and not to over interpret the results.

There are several gaps in our ability to reliably detect hotspots of malaria that have been recognized as part of this research. Firstly, as part of the MBG approach to define hotspots of malaria, thresholds where malaria is predicted to be above a set prevalence must be determined. In other applications of this geostatistical methodology, a predefined policy threshold existed facilitating its application. [16] However, for defining hotspots of malaria such a threshold does not readily exist at a local level and due to the microepidemiology of malaria and the variability that exists between areas, thresholds will likely differ depending on the setting.[10, 26] The next challenge will be to determine the ideal probability threshold to classify areas as hotspots. In an ideal scenario, i.e. with unlimited amounts of data, the model will produce a probability surface this is polarized into areas with 100 or 0% probability of exceeding a specified threshold. A more realistic scenario, albeit one that may still be unachievable because of resource constraints, is a multi-phase approach. Typically, a MBG analysis in conjunction with specified prevalence threshold and critical probability p (e.g. p=0.95) will divide the study-region into three sub-regions: those almost certainly in a hotspot (predictive probability > p), almost certainly not in a hotspot (predicted probability < p) and uncertain (all others). Given such a map can be created in a time-scale relevant for malaria transmission dynamics, additional data-collection could then be restricted to the uncertain sub-region. In our study we used two (p=0.5 and p=0.8) critical

probabilities to define hotspots, which inevitably gave different results. How the increased uncertainty corresponds to true hotspots in different settings merits further investigation. The probabilistic approach used by MBG could also inform malaria control practices by defining areas where limited resources could be targeted to the areas with the highest burden irrespective of whether true hotspots are accurately identified in their entirety: probabilistic thresholds could then be used to prioritize the structures to be targeted given available resources.

Here, we compared MBG to SatScan using a range of scanning assumptions for the latter. Other approaches to define clustering, such as Moran's I or kernel density analysis, have also been used to define hotspots of malaria infection.[13] However, all methods carry inherent assumptions with how they process data. For example, while SatScan adjusts for sampling density, it assumes that clusters will either be circular or elliptical in shape and, based on current practices in the malaria literature, assumes that unadjusted prevalence data is sufficient to represent the complex nature of malaria hotspot dynamics.[5] In contrast, while MBG provides a more statistically robust picture of malaria risk and accounts for uncertainty in the estimates, it assumes that data conform to the generalized linear geostatistical modelling framework including selection of an appropriate form for the regression and residual spatial correlation components of the model, and that model covariates are available at all locations for which estimates of prevalence are required.[15] Nevertheless, as we have shown, because the MBG approach is embedded within a general statistical modeling framework, well-established principles can be used to build the model and to assess its' goodness-of-fit to the available data.

The most important difference between MBG and SatScan lies not in the precise details of their respective implementations in any particular example, but in their underlying inferential philosophies. SatScan is rooted in a significance-testing paradigm, whereby a priori each location is or is not part of a hotspot. In contrast, MBG uses an estimation/prediction paradigm, whereby prevalence is modeled as a spatially continuous function of location, $r(x)$, the analysis assigns a predictive probability distribution to the unknown value of $r(x)$ at each location, and the notion of a hotspot is regarded as no more than an operational convenience. Put

another way, MBG is concerned not with how likely it is that a location has an above-average prevalence, but with how likely it is that a given location has a prevalence sufficiently high to be of practical concern in a specific setting. In our opinion, the estimation/prediction paradigm is inherently better suited to the problem at hand than is the testing paradigm. Nevertheless, we acknowledge that determining which specific spatial statistical models best represent malaria hotspot dynamics need to be further explored.

Ultimately, the use of cluster detection methods has highlighted the highly heterogeneous nature of malaria transmission that occurs over a small area.[9, 10, 30] The next natural step is to target these hotspots of malaria with control programs as, theoretically, a greater impact on malaria transmission can be achieved if interventions can be targeted to the right place at the right time.[5, 11] If control or elimination programs are targeted to hotspots that do not accurately reflect the true nature of transmission in the community, malaria can quickly resurge and this approach could then incorrectly be perceived as ineffective. Therefore, it is important to obtain a deeper understanding of hotspots of malaria transmission at the local level and the best way of detecting them in terms of both statistical methodology and malaria metric. A better understanding of malaria hotspots will ensure that efforts are not wasted and that control policies can be informed by evidence, as well as to determine how close we need to be to detecting the true hotspot boundaries so as to achieve a sustainable reduction in malaria burden.

4.3 General Chapter Discussion

4.3.1 Overview of Findings

The results of this chapter suggest that defining hotspots that can accurately reflect malaria transmission dynamics in the community is complex. Different malaria metrics and cluster detection approaches offered some consistency and were able to identify similar areas of increased burden. However, the specific correlation in where the boundaries were drawn was only poor to moderate according to both the kappa statistic and the correlation coefficient. Therefore, for an accurate assessment of malaria transmission, as well as to establish benchmarks with which to make unbiased comparisons for the efficacy of interventions, the role of spatial clustering in malaria transmission is needed. A

better understanding of local-level transmission dynamics would provide a baseline of true of malaria hotspots and to better inform the optimal methods for capturing these events in practice.

The inconsistency in delineating boundaries precipitates two main challenges for applying hotspot theory to practice. Firstly, SatScan is the most commonly used and operationally attractive malaria hotspot detection tool and is useful to identify areas of relatively higher burden. (127) However, these results indicate that the different assumptions used impacts the hotspots identified within the same study area. Therefore, comparing results between study areas becomes difficult unless the same assumptions are employed. These assumptions may be setting dependent. For example, the size and shape of breeding sites will determine the optimal scanning window size and shape. (16) It becomes difficult to translate a hotspot targeted approach to malaria control programs if it is unclear which hotspot detection method best reflects the true malaria transmission dynamics in different settings. Therefore, a better understanding of the transmission dynamics in a variety of settings would be useful to inform guidelines on which approach to use where.

4.3.2 Implications of Spatial Methodology for Hotspot Detection

Some of the assumptions inherent in the analytical approaches presented are highlighted above, however elaborating on these differences and the potential risks and benefits inherent are merited. Firstly, the MBG approach would ideally produce a scenario where the exceedence probabilities show a dichotomous result around the defined threshold. However, due to the large sample sizes required, uncertainty is inevitable, and an iterative sampling approach is then recommended to increase sampling to further refine the resulting map. (146) Although accounting for uncertainty in model estimates is a benefit, such an iterative sampling approach is not very practical and such an exercise would likely be confounded by the seasonal and stochastic nature of malaria transmission. Similarly, until guidelines for predictive thresholds are established for defining what constitutes high transmission areas, there is a risk that predictive thresholds will be over-interpreted or arbitrarily set and not based on potential impact or identifying populations that are a priority for control activities. Ideally, thresholds

could be defined to define those areas in low transmission settings where Rc is greater than 1 suggesting that transmission can be sustained. In high transmission settings, it is possible that the majority of households would have a Rc value greater than 1 suggesting that those areas with higher than average Rc could be of practical concern and constitute the gold standard for transmission hotspots; although in high transmission settings it is unlikely that such an approach would be practical given the uniform high risk. However, measuring Rc is currently difficult and the only proxies currently available (EIR, SCR, see chapter 1) are restricted to generating population level estimates. The ability to obtain a household-level measure or proxy of Rc would provide a gold standard for the operationalization of thresholds for malaria hotspots. If adequate data to inform these models, such as an individual metric of exposure, becomes available, a more complete picture of malaria is obtained by accounting for risk not explained by the modeled covariates, which could result in more robust delineation of hotspots if correct thresholds for each metric can be identified.

In contrast, SatScan has the benefit of being a simple and operationally tractable tool to define clusters of infection and in this study was shown to identify the areas of increased risk for malaria, although provided more generous estimates of hotspot boundaries. As discussed, SatScan is a tool that identifies clusters of increased malaria. Although the simple hypothesis testing approach clearly captures some of the transmission dynamics present in the community, there is a risk that results are over interpreted without consideration to fundamental concepts. For example, studies have identified clusters of malaria with small sample sizes and imprecise measures such as clinical incidence. (268) The selection of control populations used to compare the distribution of cases may not be appropriate leading to biased results. (156) Lastly, as mentioned, SatScan inherently identifies whether the cases of malaria cluster, which is known as malaria clustering has been repeatedly observed at all spatial scales. (127) Until studies have shown that the clusters identified using SatScan (or MBG) are adequately capturing malaria transmission dynamics and are important for maintaining or fueling transmission, the results should be interpreted accordingly.

All methods and metrics to detect areas with high malaria burdens assessed were able to consistently identify the same regions as 'hot'. The precise delineation of hotspot boundaries is likely more essential in a research context where evidence of the effectiveness of such a targeted approach will require the entire hotspot to participate in intervention campaigns. As all methods were able to identify what are likely the main hotspots, regardless of model assumptions, achieving a 'fuzzy' definition with less precise smoothed estimates, in an operational context may be good enough. By expanding the target region to a defined political unit, such as village or enumeration area, which is not only a more operationally attractive unit to deploy interventions, but would also minimize the risks of missing sections of the hotspot which could then refuel transmission despite a successful intervention. (96)

4.3.3 Defining Hotspots of Malaria Transmission

Due to the lack of a gold standard approach for hotspot detection, the Satscan method was identified as the working definition of malaria hotspots for the purpose of this research. Despite the more robust MBG approach available, the use of Satscan to define hotspots in this setting was taken due to uncertainties in the ideal threshold for areas of increased risk, to provide consistency with other studies in the area, as well as consistency with other research projects as this method is currently the most widely used cluster detection technique in the field at the moment. (148) Therefore for the purpose of this research, a hotspot of malaria transmission is operationally defined as areas with statistically significantly higher seroprevalence than the surrounding community using the SatScan approaches defined in section 4.2 above (objective 1).

Targeting interventions to hotspots of malaria transmission is operationally attractive as it suggests that a similar reduction in malaria transmission can be achieved by focusing on smaller areas compared to a universal approach. (65) Therefore, exploring the utility of this strategy and the potential application in a locally feasible and accepted manner is needed. Historically, malaria surveillance and treatment programs have employed the use of convenience sampling through primary schools or health facility sampling approaches. Therefore, the use of

convenience samples to detect heterogeneity and inform such targeted strategies is a logical next step.

4.4 Conclusions

The main conclusions that can be drawn from the results in this chapter include:

1) The choice of cluster detection techniques for defining hotspots of malaria has an impact on the resulting map both in terms of hotspots identified and where the boundaries are drawn;
2) Hotspots of malaria based on a metric of current malaria infection and malaria exposure provide different pictures of where hotspots and hotspot boundaries are drawn, but all identify similar regions of high burden;
3) The sample size used to inform the cluster detection technique has an impact on where hotspot boundaries are drawn with larger sample sizes resulting in narrower hotspot boundaries.

Chapter 5: Results – Convenience Sampling for Measuring Malaria Transmission Intensity

The ability of convenience samples to reliably gauge transmission intensity in the surrounding community is critical to supporting the use of these more operationally attractive approaches as a viable alternative to community based surveys. This chapter addresses the ability of both school and health facility surveys to provide a reliable metric of malaria transmission in the broader community. To address this objective, concurrent school and community surveys were conducted as described below in section 5.2 and health facility surveys outlined in section 5.3.

5.1 Background and rationale

Historically, measuring malaria transmission intensity has primarily relied on convenience samples as a way to monitor the disease burden using data collected passively at health facilities and to a lesser extent school surveys. (13, 232) As the understanding of malaria epidemiology evolved and the pervasiveness of asymptomatic and subpatent infections and their contribution to transmission recognized (53, 58), there has been a shift to incorporating community based data, such as the malaria indicator surveys, to obtain a more accurate assessment of the true nature of malaria burden. (225, 228) However, as community surveys are operationally unattractive due to their logistical difficulties, and being time consuming and expensive to conduct, (269) it is important to assess how reliable the more operationally attractive alternatives are at measuring transmission in the community (specific objective 2).

5.2 Convenience Sample: Primary Schools

5.2.1 Primary school surveys as a metric for malaria transmission

School-aged children provide a useful sentinel population for measuring malaria transmission intensity (section 3.3.1). Programs are already in place that use schools to screen for prevalence of other diseases such as helminthes and therefore extending them to include malaria would not require many additional resources. Similarly, training of existing school staff to conduct malaria testing can easily be done making this approach feasible and operationally attractive. (212,

263, 270) However, evidence is required to determine if malaria estimates in school-aged children are representative to those of the surrounding communities.

5.2.2 Reliability of school surveys in estimating geographic variation in malaria transmission in the western Kenyan Highlands (P2)

Authors: Gillian H Stresman*, Jennifer C Stevenson *, Caroline W Gitonga, Jonathan Gillig, Chrispin Owaga, Elizabeth Marube, Wycliffe Odongo, Albert Okoth, Pauline China, Robin Oriango, Simon J Brooker, Teun Bousema, Chris Drakeley, Jonathan Cox *Authors Contributed Equally

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?	**PLoS One**		
When was the work published	**15/10/2013**		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes, CCA License	Was the work subject to academic pier review?	**Yes**

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	
Please list the paper's authors in the intended authorship order.	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conducted all data analysis and wrote the manuscript.

Student Signature: _____  **_____ Date: ___17/02/2015_____**

Supervisor Signature: ____  **_____ Date: ___17/02/2015_____**

Abstract:

Background: School surveys provide an operational approach to assess malaria transmission through parasite prevalence. There is limited evidence on the comparability of prevalence estimates obtained from school and community surveys carried out at the same locality.

Methods: Concurrent school and community cross-sectional surveys were conducted in 46 school/community clusters in the western Kenyan highlands and households of school children were geolocated. Malaria was assessed by rapid diagnostic test (RDT) and combined seroprevalence of antibodies to bloodstage *Plasmodium falciparum* antigens.

Results: RDT prevalence in school and community populations was 25.7% (95% CI: 24.4-26.8) and 15.5% (95% CI: 14.4-16.7), respectively. Seroprevalence in the school and community populations was 51.9% (95% CI: 50.5-53.3) and 51.5% (95% CI: 49.5-52.9), respectively. RDT prevalence in schools could differentiate between low (<7%, 95% CI: 0-19%) and high (>39%, 95% CI: 25-49%) transmission areas in the community and, after a simple adjustment, were concordant with the community estimates.

Conclusions: Estimates of malaria prevalence from school surveys were consistently higher than those from community surveys and were strongly correlated. School-based estimates can be used as a reliable indicator of malaria transmission intensity in the wider community and may provide a basis for identifying priority areas for malaria control.

Introduction

Obtaining accurate estimates of malaria transmission can be an intensive process, especially when transmission is low [1]. As malaria transmission levels continue to decline in many malaria endemic areas [2], developing robust, cost, and time effective approaches to measure and monitor changes in transmission intensities becomes more urgent. The issue is particularly relevant to national malaria control programs as they largely carry the responsibility for malaria surveillance and for whom the more extensive approaches (ie. large population surveys, longitudinal

entomological surveillance) are likely to be logistically and financially burdensome [3], [4].

In most malaria-endemic settings, children experience the highest incidence of clinical malaria and highest parasite prevalence [5]. Although much focus has centred around children under 5 years [6], [7], older, school-aged populations also provide a valuable source of information on malaria burden. The school-aged population has been shown to carry higher parasite prevalence and densities compared to adults [8], [9] and also tend to have a lower reported rate of bednet use [10]. The lower net use combined with the higher parasite densities suggest that school-aged children experience a high malaria burden and may also be important sources for onward transmission of parasites [11].

In areas with malaria transmission, malaria-specific antibody prevalence increases with age as a consequence of cumulative exposure to malaria antigens and consequently, the rate at which individuals become antibody positive, the seroconversion rate, is strongly associated with transmission intensity [12]. Antibody responses in school-aged children are important in defining the slope of the age-dependent seroconversion curve [13] and therefore constitute a highly informative sentinel population both for monitoring variations in parasite prevalence [14] and the rate of acquisition of age-dependent antibodies [13] over time.

There are a number of logistical advantages associated with sampling children in schools [11], [15], [16]. School surveys provide a convenient location to sample large numbers of children in a shorter timeframe than equivalent sampling in the community and can be easily integrated into routine public health programming. However, sampling school populations also has inherent biases that can make their generalizability problematic. For example, healthy and more affluent children may be more likely to attend school, children may attend a school outside of their immediate community, and they may be more likely to be positive for malaria by RDT than adults due to their higher parasite densities and therefore school estimates may not reflect community prevalence [5], [11], [14]. Therefore, the suitability of sampling children at school for estimating community-wide

transmission intensity requires direct comparisons of school and community surveys to assess whether school-based estimates of malaria can provide accurate estimates of community-based transmission.

Here, we investigate the concordance in paired school and community based estimates across a range of malaria transmission intensities measured by infection prevalence using RDTs and seroconversion rates. Households of school- and community survey participants were mapped to determine the impact of the spatial overlap between the two populations on the reliability of school surveys in an area of highly heterogeneous transmission intensity in western Kenya.

Methods

Ethics Statement

This study was approved by the ethical committees of the London School of Hygiene & Tropical Medicine and the Kenya Medical Research Institute and was part of a larger government-lead, national school survey programme [15]. Approval was also provided by the Permanent Secretary's office of the Ministry of Education (MoE) and the Division of Malaria Control, Ministry of Public Health and Sanitation. Prior to the school surveys, meetings were held with the teachers, parent-teachers' association, as well as the broader community including parents, caretakers, and guardians. Information sheets describing the survey were distributed at all community meetings and additional copies were left at the schools, education office, and the chief's and assistant chief's offices. Parents/guardians who did not want their children to participate were given the option to opt-out of the study. Participating children provided assent: if a child refused, the next randomly selected child would be approached [15]. Individual written parental consent was not sought because the survey was conducted under the authority of the Division of Malaria Control, Ministry of Public Health and Sanitation, which have the legal mandate to conduct routine malaria surveillance. Two independent ethical review committees approved this approach.

For the community survey, individual informed consent was sought from all

residents of the compound above the age of 6 months by signature or thumbprint accompanied by the signature of an independent witness. Consent for children under the age of 18 was provided by a parent/guardian and children between 14 and 17 years also provided written assent by signature or thumbprint accompanied by the signature of an independent witness. As defined in the Kenya national guidelines, participants below 18 years of age who were pregnant, married, or a parent were considered "mature minors" and consented for themselves [19].

Study site and recruitment of study participants

This study was carried out in July 2010 in a rural, highland fringe area (1400-1600 m above sea level) of Rachuonyo South and Kisii Central districts, Nyanza Province, Kenya [17]. The predominant ethnic groups in Rachuonyo South and Kisii Central districts are the Luo and Kisii, respectively. Compounds are distributed broadly across a rolling landscape intersected with small streams and rivers. The main malaria vectors are *Anopheles funestus*, and *An. arabiensis*, and *Plasmodium falciparum* is the predominant malaria parasite. There are two seasonal peaks in malaria transmission reflecting the bimodal rainfall pattern, with the heaviest rainfall typically occurring between March and June, with a smaller peak in October/November each year.

A census of government primary schools in the study area was conducted (n=122) and the numbers of pupils per school determined. A sample of 46 schools with at least 100 pupils was randomly selected using an iterative process to limit the odds of selecting schools with overlapping catchment areas. At each school, 11 boys and 11 girls per class from classes 2 to 6 were selected using random number tables [15]. Corresponding “communities” were defined as all residences (called compounds) falling within 600 m of each school. Compounds were enumerated and their geographical location recorded using a Personal Digital Assistant (PDA) equipped with a Global Positioning System (GPS) receiver. An unstratified random sample of all enumerated compounds within the 600 m buffer were selected for inclusion in the study. The 600 m radius was chosen to minimize the possibility of overlap between the catchment areas of schools. All residents of the randomly

selected compounds above the age of 6 months were eligible for the community survey.

The power of the study was calculated to detect significant equivalence in malaria prevalence estimates between the school and community populations. The average number of people sampled in each survey was 4300 with a mean of 90 people per cluster. The mean baseline malaria prevalence by RDT was estimated to be 20% in the school and community populations. With a 5% absolute tolerance limit, there is greater than 99% power to detect equivalence between the school and community surveys at an alpha level of 0.05 [18]. The design effect was calculated to be 16.9, for each calculated value of **ρ.** When the correlation within clusters is taken into account, the adjusted power is 82%.

Survey Procedures

For both surveys, participants were asked to provide a finger-prick blood sample for detection of malaria by rapid diagnostic test (RDT) (Paracheck, Orchid Biomedical Systems, India). The same finger prick sample was used to measure haemoglobin concentrations using a HemoCue photometer (HemoCue, Angelhom, Sweden) and to provide three blood spots on Whatman 3 mm filter paper (Maidstone, UK). Questionnaires were administered to assess wealth indices, use of preventative measures for malaria, travel history, and household characteristics [15]. Individuals found to be positive for malaria were treated with artemether-lumefantrine (AL; Coartem®, Novartis) and haematinics were provided to individuals found to be anaemic, according to the national guidelines at the time of the survey. In the school survey, treatment was not given directly to children. If a child was positive the child had to bring their parent/guardian to the school to receive the drugs. If the parent was not available, the drugs were left with the teacher and the child was asked to come to school the next day with the parent/guardian to receive them. The compound of each child sampled at school was located and mapped using a PDA with GPS receiver.

Laboratory Analysis

Filter paper blood spots were stored with desiccant at room temperature until transport to -20 °C for long-term storage. Antibodies to *P. falciparum* Apical

Membrane Antigen-1 (AMA-1) for Merozoite Surface Protein-1 (MSP-1) were detected by Enzyme Linked Immunosorbent Assay (ELISA) as previously described [12]. Antibody prevalence was determined after defining a cut-off optical density (OD) using the mixture model [20], [21].

Statistical Analysis

All analysis was conducted in STATA 12.0 (StataCorp, Texas, USA) and Quantum GIS 1.8 (Open Source Geospatial Foundation Project). Age-specific seropositivity rates were used to estimate seroconversion rates (SCR). A person was considered seropositive if they were positive for at least one of the antigens tested [12], [13]. Hypothesis testing for means (t-test) and proportions (z-test) were used to compare the difference between proportions from the school and community populations with the null hypothesis being that there is no difference. Crude agreement between the school/community pairs was assessed using Spearman's rank sum agreement, and Youden's index was used to determine the optimum cut-off point for delineating high and low transmission intensities [22]. As correlation is a measure of association, and not of agreement [23], concordance was determined using Lin's concordance correlation coefficient (r_c) [24] and the Bradley-Blackwell F test was used to test if the concordance was statistically significant [25]. Total least squares regression was used to determine if the school and community estimates are concordant. The reduced major axis (RMA) is the line of best fit calculated from the data using the total least squares regression. Concordance is achieved when the slope of the RMA is not significantly different from the line of perfect concordance, which has a slope of one signifying that a change in one unit in one measure has a corresponding one-unit increase in the second metric. Comparisons were calculated for both the community versus all school survey participants as well as for the community versus a restricted sample of school survey participants living within 600 m of the school (<600m population) to ensure that both populations being compared resided in the same area.

Results

Study Population

A total of 4964 individuals were sampled at school, of which 4888 (98.5%) could be traced to their compounds and were included in subsequent analysis.

In the community survey, 3742 participants were sampled in 46 communities (table P2-1). Due to the random sampling, 4.4% of children were sampled at school and had their compound visited by a field team during the community survey. These children were included in analysis for both populations.

Table P2-1: Demographic characteristics of the community and school study populations. Prevalence of demographic, reported malaria control, and outcome measures of malaria infection, seroprevalence and anaemia in the community and school populations, as well as the school populations stratified by distance to school.

		Community	School	School Children by Distance to School		
				0 to 600m	601 to 1000 m	>1000m
Sample Size	N	3742	4888	1780	1717	1391
N per Cluster	Median	80	108	37	38	30.5
	Range	72-96	81-111	17-94	4-60	8-47
Sex	Male %	44.1	49.9	49.2	48.9	52.0
Age	Mean (SD)	21.1 (20.6)	11.8 (2.2)	11.7 (2.2)	11.8 (2.2)	11.8 (2.2)
	Range	0.5-100.7	6-25	6.4-20.5	6-25.5	6-22.6
Bednet Use	% (95% CI)	57.1 (55.4-58.7)	32.5 (31.2-33.8)	33.4 (31.2-35.6)	31.3 (29.1-33.5)	32.9 (29.1-33.5)
	Range*	22.1-95.3	12.2-77.8	5.9-75.7	0-80.6	5.9-80
IRS in Past Year	% (95% CI)	73.8 (72.3-75.2)	70.4 (68.9-71.5)	68.3 (66.1-70.4)	70.7 (68.5-72.8)	72.9 (68.5-72.8)
	Range*	10.4-100	11.3-93.6	9.2-95.8	11.4-95.5	12.5-100
Recent Travel	% (95% CI)	12.3 (11.2-13.3)	16.1 (15.0-17.1)	14.9 (13.2-16.5)	17.5 (15.7-19.3)	16.0 (15.7-19.3)
	Range*	0-31.5	0-37.9	0-41.7	0-44.4	0-43.5
SES**-% (Range*)	1	19.1 (0-57.7)	21.4 (5.7-38.5)	22.5 (3.4-51.8)	22.0 (4.9-50)	20.0 (0-50)
	2	15.3 (0-42.5)	23.6 (9.3-41.4)	24.5 (4.4-50.0)	22.9 (0-56.7)	23.5 (0-53.6)
	3	19.7 (0-52.3)	15.0 (4.5-28.1)	13.6 (0-33.3)	15.1 (0-40.5)	17.4 (4-37.5)
	4	20.3 (0-	19.6 (11.2-	17.7 (2.8-	21.4 (4.4-50)	19.4 (0-

		72.8)	39.2)	41.2)		38.9)
	5	18.9 (0-48.8)	19.8 (5.4-40.9)	21.3 (0-50.0)	18.5 (0-50)	18.5 (0-61.5)
RDT	% +ve (95% CI)	15.5 (14.4-16.7)	25.7 (24.4-26.8)	25.5 (23.5-27.5)	26.9 (24.8-29.0)	24.3 (24.8-29.0)
	Range*	0-51.2	0-71.4	0-88.2	0-75	0-78.4
SeroPrevalence	% +ve (95% CI)	51.5 (49.5-52.9)	51.9 (50.5-53.3)	51.5 (49.2-53.8)	55.3 (52.0-57.7)	48.2 (52.9-57.7)
	Range*	22.6-85.9	5.6-87.4	12.5-90.6	0-91.9	2.8-96.9
Haemoglobin (g/DL)	Mean (95% CI)	12.7 (12.5-22.1)	13.4 (13.4-13.5)	13.4 (13.3-13.5)	13.4 (13.3-13.4)	13.4 (13.3-13.4)
	Range	2.9-25.0	4.4-19.7	4.4-17.7	4.9-18.3	6.3-19.7
***Range of cluster level summaries** ****Socioeconomic Status (SES) is divided into quintiles with 1=Low and 5=High**						

The range of the number of people sampled in each cluster was 72-96 and 81-111 in the community and school surveys, respectively (table P2-1). Compound net ownership in the school population was reported to be 66.1% (95% CI: 64.7-67.4) and 78.6% (95%CI: 77.3-80.0) in the community ($p<0.0001$). The school population reported a significantly lower bednet use (32.5%) compared to the community (57.1%) ($p<0.0001$). The age distribution of the participants in the community survey was, as expected, markedly different than that in the school survey (figure P2-1a). Analysis of the spatial distribution of residences of the school children sampled showed that 36.4% of children lived within 600 m of their school (figure P2-2), with a mean distance of 793 m (IQR: 465-1040 m) (figure P2-1b). The proportion of school children residing within the community catchment area varied per school and ranged from 16 to 89% (figure P2-1c). Due to differences in sample sizes it was not possible to directly compare malaria outcomes between school children and school-aged children sampled in the community.

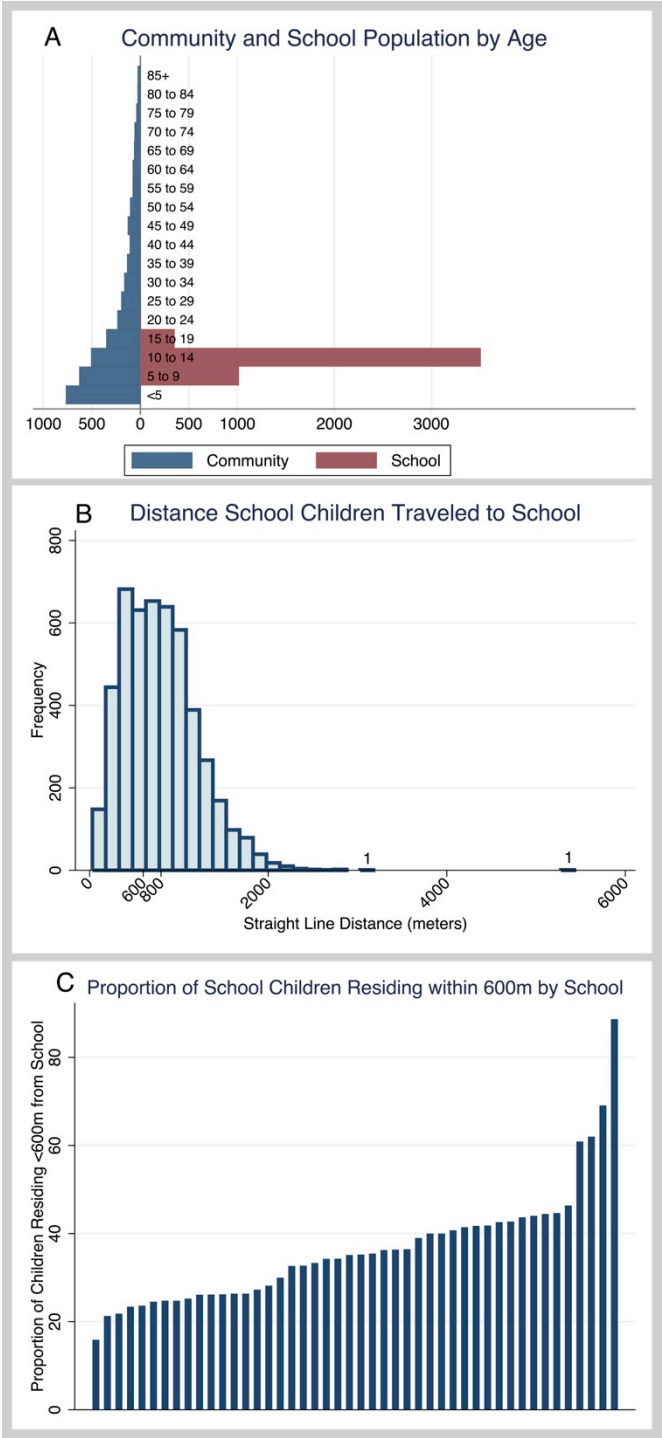

Figure P2-1: Characteristics of the study population – age and distance travelled to school. (A) A population pyramid showing the age distribution of those sampled in the community survey compared to those sampled during the school survey. (B) Histogram depicting the distance between the school and compound where each child resides. (C) The proportion of children sampled at each school that reside within 600m of the school.

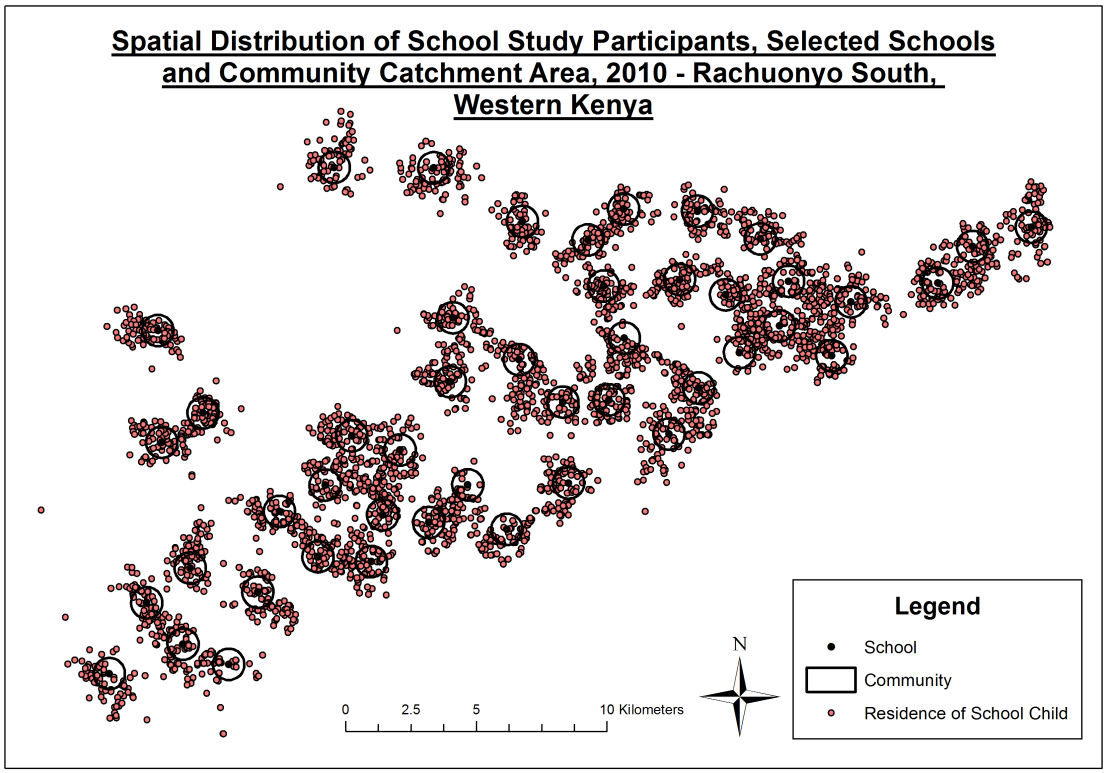

Figure P2-2: Spatial distribution of school study participants, location of the schools, and community catchment area. Each point represents the compound of a child included in the study. The black crosses indicate the location of each school that was included in the survey. The black circular outline corresponds to the area with a 600m radius around each school and thus represents the community catchment area sampled during the community survey.

Malaria Prevalence

P. falciparum infection prevalence by RDT was significantly higher in the school population at 25.7% (95%CI: 24.4 – 26.8) compared to 15.5% (95%CI: 14.4 – 16.7) in the community (p<0.0001). RDT prevalence ranged from 0 to 71.4% in the schools and from 0 to 51.2% in the communities with the higher prevalence in schools and communities typically located in areas of lower elevation (test for trend p=0.026 and p=0.035, respectively) (table P2-1). School and community parasite prevalence rates were strongly correlated (r=0.77; p<0.0001) (figure P2-3A). Restricting the school sample to < 600 m population strengthened this correlation (r=0.83; p<0.0001) (figure P2-3B).

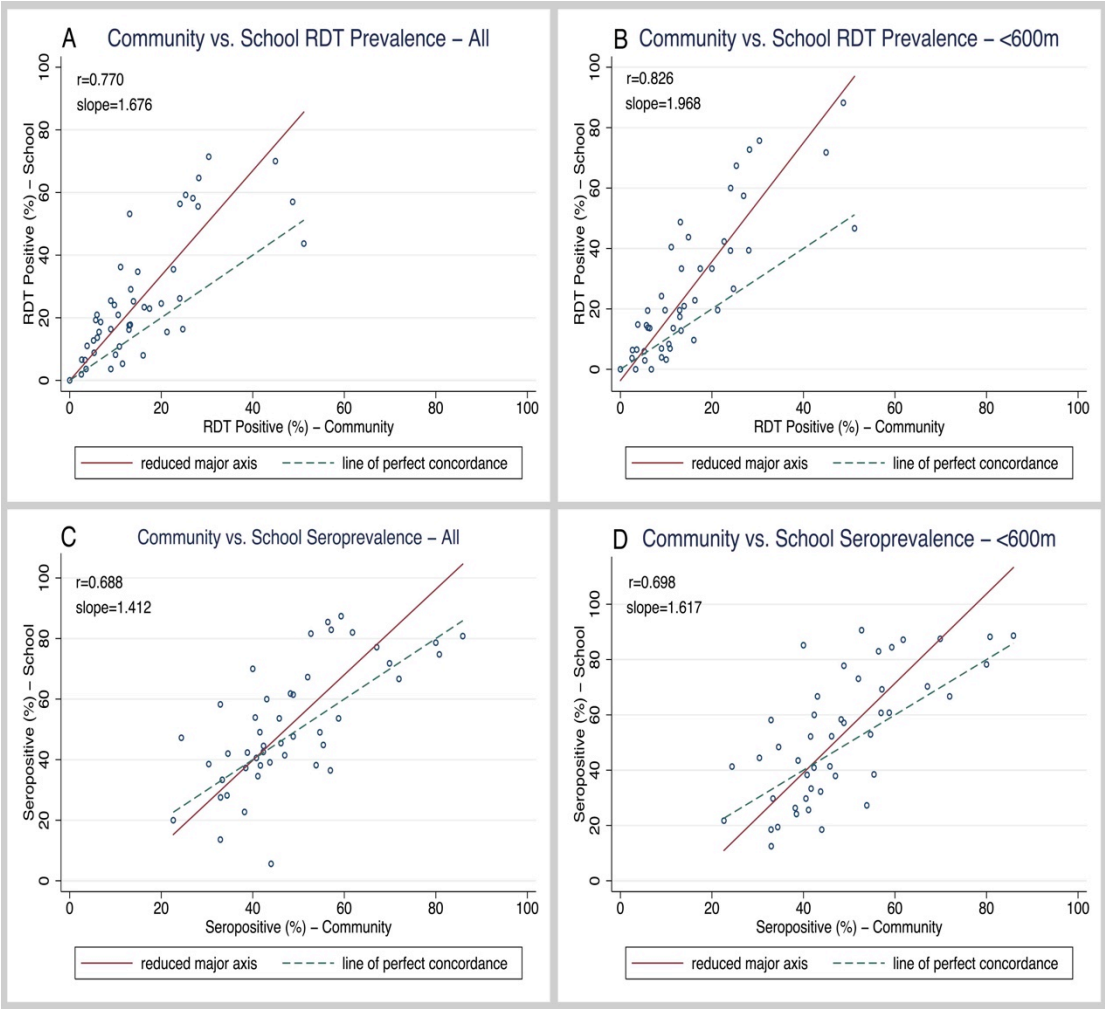

Figure P2-3: Prevalence of malaria infection in school vs. community surveys in 46 clusters by RDT and serology. Scatter plots are shown with the line of perfect concordance (x=y) and the data's reduced major axis using total least squares regression. (A) RDT prevalence per cluster in community vs. all school children. (B) RDT prevalence per cluster in community vs. school children residing within 600m from school. (C) Seroprevalence per cluster in community vs. all school children. (D) Seroprevalence per cluster in community vs. school children residing within 600m from school.

Despite this strong correlation, prevalence rates from school surveys were, as expected, higher than corresponding rates in the community and as such the two measures were statistically discordant (Bradley-Blackwood $p<0.0001$, RMA slope=1.676) (figure P2-3A). When the analysis was restricted to include the < 600

m population, the two populations were still statistically discordant (r_c =0.56, Bradley-Blackwood p<0.0001, RMA slope=1.97) (figure P2-3B).

Seroprevalence estimates ranged from 5.6 to 87.4% and 22.6 to 85.9% in the school and community surveys, respectively. Seroprevalence in the two populations did not differ significantly with 51.5% (95% CI: 49.5–52.9) in the community and 51.9% (95% CI: 50.5–53.3) in the school population (p=0.39) (table P2-1). The cluster-level paired estimates of seroprevalence exhibited good correlation (r=0.69, p<0.0001) (figure P2-3C) in the community and all school population. Restricting the analysis to the <600 m population had little impact on correlation with the community (r=0.70, p<0.0001) (figure P2-3D).

Seroprevalence estimates from school and community surveys were positively correlated (r=0.69, p<0.0001), but statistically discordant (r_c =0.64, Bradley-Blackwood p=0.0035, RMA slope=1.41) (figure P2-3C). When restricting the school survey population to the < 600 m population, there was little improvement in the concordance (r_c =0.61), or RMA slope (1.62) and the measures were still significantly discordant (Bradley-Blackwood p<0.0001) (figure P2-3D).

Table P2-2: Prevalence of malaria by rapid diagnostic test in community and school populations by transmission zone. RDT prevalence rates and corresponding 95% confidence intervals in the community, all school children, and school children restricted to within the community catchment area (<600m from school). Transmission intensity defined based on RDT prevalence in the community - low=0-10%; moderate=10.1-20%; high=>20% RDT.

	Community	**School**	**School (<600m)**
Low - % (95% CI)	5.8 (4.4-7.2)	12.0 (8.2-15.9)	8.9 (5.1-12.6)
Moderate - % (95% CI)	13.9 (12.4-15.3)	23.0 (16.3-29.8)	24.3 (16.8-31.8)
High - % (95% CI)	30.8 (24.6-37.0)	48.4 (36.8-60.1)	54.4 (42.0-66.8)

Agreement in Transmission Intensity

Transmission intensity strata in this study area were defined based on approximate terciles of community RDT prevalence: 0-9.9% (low), 10-19.9% (moderate) and ≥20% (high). When the school RDT prevalence estimates were

stratified according to community transmission intensity, malaria prevalence rates by RDT in the school showed a clear increasing trend in malaria prevalence (table P2-2). Overall, the seroconversion rates based on school surveys (λ=0.07; 95% CI: 0.071-0.078) were similar to those in the community (λ=0,07; 95% CI: 0.066-0.075). When stratified by transmission intensity, school surveys produced similar seroconversion rates to those of the community in both the high and low transmission settings (figure P2-4).

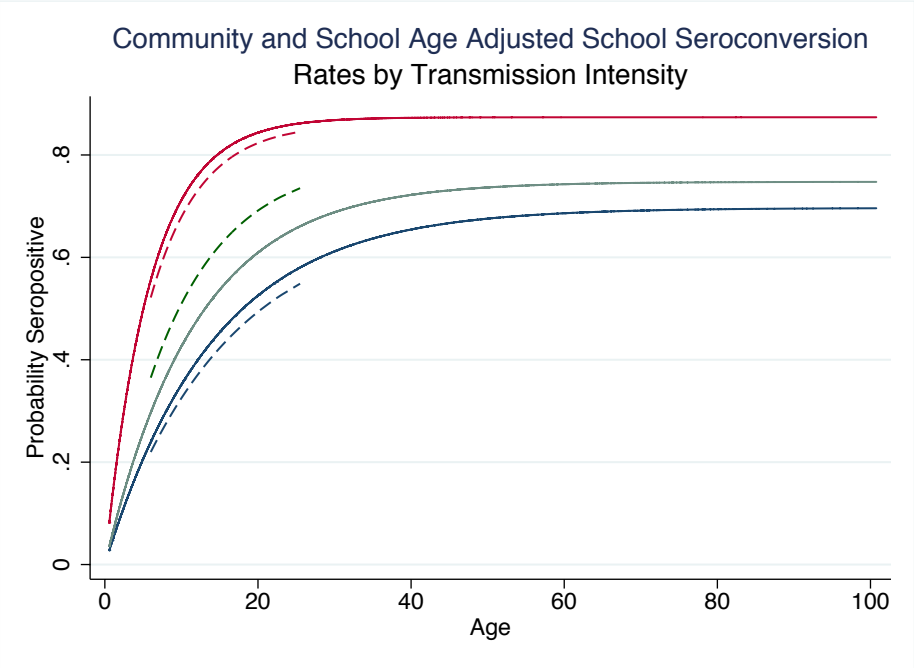

Figure P2-4: Age-adjusted seroprevalence in community and school surveys (all children) by transmission intensity. The age-adjusted community seroconversion curves (solid) and school aged population (dashed lines). The different transmission intensities are represented as: high (red) moderate (green) and low (blue).

A Spearman's rank test was used to determine whether estimates of parasite rates obtained through school surveys can provide a guide to transmission intensity in the community. Agreement in school and community RDT prevalence resulted in a Spearman's correlation of 0.78 and 0.84 in the all school children and the <600 m population, respectively, and both correlations were statistically significant ($p<0.0001$). When stratifying these results by transmission intensity, there was

good correlation of the rank of RDT prevalence between school and community clusters in the low (ρ=0.59, p-value=0.01) and high (ρ=0.61, p-value=0.02) transmission intensities. In the <600 m population, there was a strong correlation between ranks in the high transmission setting (ρ =0.67, p-value=0.01). Seroprevalence only showed agreement between the community and the <600 m population in high transmission settings (ρ =0.64, p-value=0.01).

As the <600 m school population showed better correlation with the community, the optimum cut-off point for what was considered a low and high transmission area based on school RDT prevalence in the <600 m population compared to the community was ascertained. Based on this data, RDT prevalence estimates of less than 7% (95% CI: 0-19%) and greater than 39% (95% CI: 25-49%) in school survey represented areas in the community with low and high transmission levels, respectively. This cut-off point resulted in a sensitivity of 58.8%, 66.7% and 78.6% to correctly identify schools in low, medium, and high transmission areas in the community, respectively (overall sensitivity of 68.0%). The specificity using the cut-points for low, medium, and high transmission in the school and community was 93.5%, 69.2% and 91.2%, respectively.

Cluster Specific Agreement

To obtain better concordance between each school/community pair, school estimates were adjusted based on the linear regression coefficient of the cluster level prevalence estimates. RDT prevalence per school was adjusted by 0.55 (95% CI: 0.48-0.62) in the all school children population and by 0.51 (95% CI: 0.45-0.57) in the <600 m school populations. The adjusted all school data showed better concordance (r_c =0.76) with the community data, the RMA slope was 0.92, and the two measures were significantly concordant (Bradley-Blackwood p=0.36) (figure P2-5A). Concordance in the <600 m school population was stronger (r_c=0.82), had a RMA slope of 1 and the measures were statistically concordant (Bradley-Blackwood p=0.23) (figure P2-5B). Adjustment of the seroprevalence did not change concordance between the community and school measures.

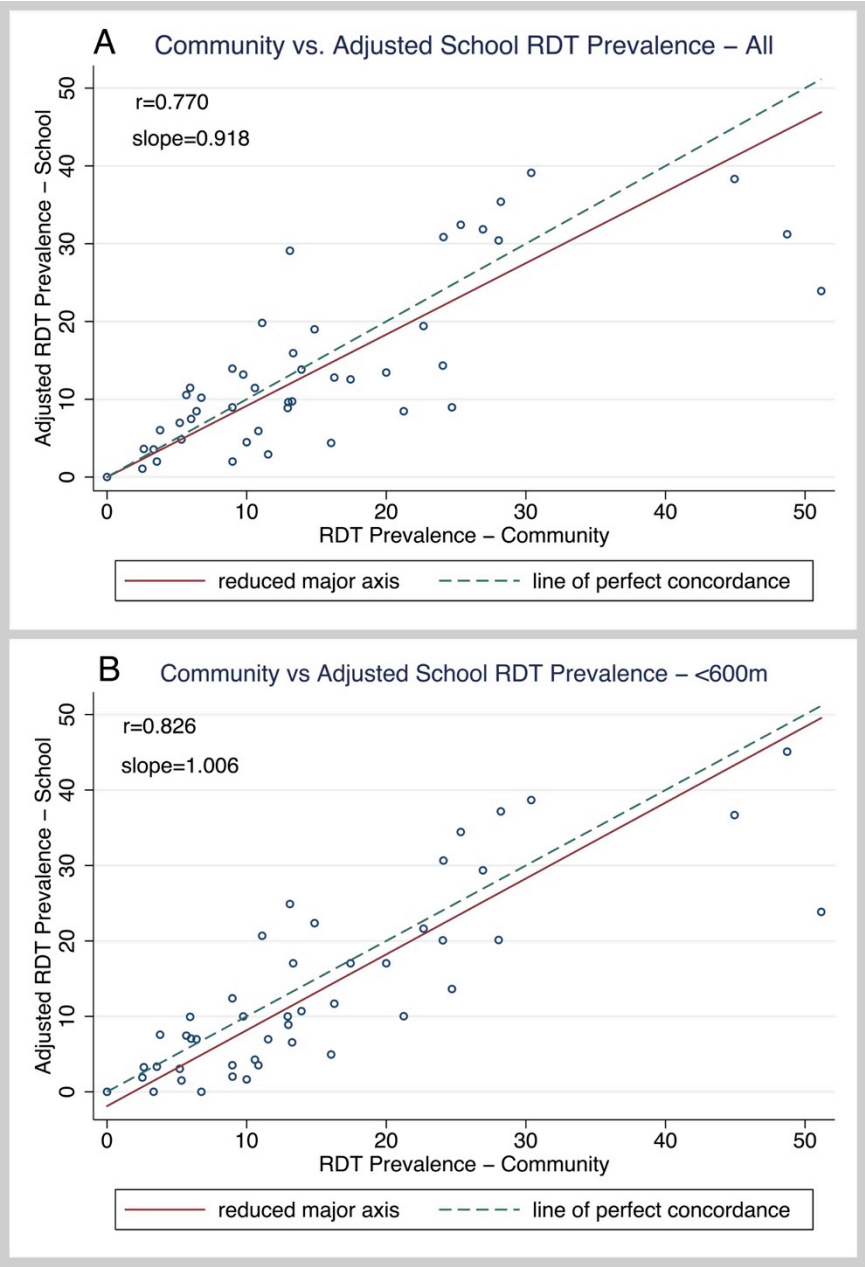

Figure P2-5: Prevalence of malaria infection: adjusted school vs. community surveys in 46 clusters by RDT. Scatter plots are shown with the line of perfect concordance (x=y) and the data's reduced major axis using total least squares regression. (A) RDT prevalence per cluster in community vs. adjusted prevalence in all school children. (B) RDT prevalence per cluster in community vs. adjusted school prevalence restricting to children residing within 600m from school.

Discussion

This is the first time that malaria prevalence rates measured during school surveys have been directly compared to those in the surrounding community as a means of assessing the accuracy of school surveys for providing an alternative approach to monitor and/or target malaria control [26]-[28]. The data show that school surveys exhibit good correlation with the community measures of infection and exposure. As expected RDT prevalence was higher in the school surveys [3], [11]: the two survey designs effectively sample populations with markedly different age distributions and, at least in areas of moderate to high malaria transmission, school-aged children are more likely to be parasitaemic than the broader community [14]. Irrespective of this higher prevalence, school surveys were able to rank malaria prevalence according to their endemicity in a similar way to the community surveys.

School surveys can identify areas of high or low transmission intensity in a way that is simple, cost-effective and can quickly assess a large geographical area [15], [16], [26]. Results indicate that schools in the highest RDT prevalence strata (in this area, > 39% school RDT prevalence) correspond to areas where there is high transmission in the community. These areas would therefore be expected to be a high priority for malaria control. Conversely, schools with the lowest RDT prevalence (in this area, <7% school RDT prevalence) could be assumed to indicate either areas with low priority for control (in high endemic settings), areas that have potential for implementing elimination strategies, or as a threshold to identify where malaria control has been successful. A crude measure, identifying priority areas, is operationally attractive for local malaria control programs and could result in more effective targeting of scarce resources. A more accurate reflection of malaria transmission in the community is also possible if the higher RDT malaria prevalence expected in school-aged children is acknowledged [20]. This is the first attempt to quantify the overestimation of malaria prevalence expected in the use of school surveys as a means to gauge malaria transmission in the community [11].

The relationships described here may differ in different malaria transmission or epidemiological settings [26], [29]. For example, in high transmission settings, the

parasite profile will be different as more children under 5 years of age are likely to be infected with malaria [30]. The different transmission settings are not likely to have an impact on the ability of school surveys to reflect areas of high or low prevalence in the community; however, prevalence strata would obviously be different. Also, the numeric factor used to adjust school RDT prevalence for malaria in the different transmission strata will vary between settings. Correction factors have been proposed to account for bias when using operationally attractive, yet imperfect methods for surveillance of a wide range of public health problems including helminths, HIV, and fractures [31]-[33]. If validated, a similar approach for malaria could be useful as adjusted school measures for malaria could facilitate monitoring changes in malaria transmission intensity.

One important consideration in using school surveys is knowledge of the catchment from which the students derive. In this survey, the community was defined as the area within a 600 m radius of the school: an arbitrary but pragmatic decision influenced by the distance between schools, and the spatial heterogeneity of transmission. After determining the location where the children sampled at school resided, only 36.4% actually lived within this catchment area, with the mean straight line distance from the child's compound to school being just under 800 m. The variability in the distance that some children travelled to school differed per school and was likely related to factors such as the size and reputation of the school, proximity to other schools and environmental factors that affect access. In our study the catchment area of schools influenced the concordance with community estimates: despite the reduced sample size, both correlation and concordance improved when restricting the comparison to the school children residing within 600 m of the school.

School surveys may be biased due to absenteeism and characteristics of the children that actually attend school, like health and SES. The healthy child effect, a selection bias where healthier children are present and sick children are absent from school, may impact prevalence rates as it suggests that the school malaria prevalence rates would be lower than the true value. However, this may only be an issue in low transmission areas where school-aged children would not have had

the opportunity to build up sufficient immunity to reduce the likelihood of clinical malaria and therefore be more likely to stay home due to malaria infection.

Similarly, the equal opportunity for children to attend school is also not likely a factor due to the government of Kenya instituting free primary school education in 2003 [34]. In our study we found that children came from all SES classes and previous work has shown that 97.6% of children in Rachuonyo district have attended school [35]. Although the above mentioned factors may have an impact on the estimates of malaria infection obtained during the school survey in this study site, they are likely to be non-differential and of little consequence in the application of this approach as an operational strategy to use school surveys to target or monitor malaria transmission. In areas that do not have universal primary education, or have low attendance rates, school-based surveys may not be as representative as has been shown in this setting.

Other factors may have an impact on the observed concordance between the school and community surveys including altitude and age and these are likely to be site specific. However, restricting the school children to those that resided in the same altitude range as the community had little impact on the results (data not shown). Similarly, the age range of people sampled in the community survey is much broader than in the school population. When the results of the community survey were restricted to the school-aged population, no impact was observed. The lack of impact using this population may have been the result of the very low sample sizes in the age-restricted community population.

Despite the inherent uncertainty in the cluster estimates, the sample size per cluster in the community and all school surveys were similar and therefore the error would not be expected to have a large impact on the results. In the <600 m population there was more variability in sample sizes, however there were only 14 schools with fewer than 30 people sampled. When the analysis was repeated with these clusters removed, there was little impact on the results (data not shown). Similarly, the prevalence data were not normally distributed, which violates the assumptions inherent in the Bradley-Blackwood F test [25]. To determine the impact of this, prevalence data were log additive transformed [36] to obtain a

normal distribution and analyses were rerun. However, there was little impact on the interpretation of the results with similar statistics of concordance

This study provides evidence that school surveys are able to inform malaria control strategies and be used to measure or monitor changes in transmission intensity. As local malaria control programs continue to take increased ownership of the operational and financial elements of the malaria control and elimination agenda, the ability to obtain accurate metrics on malaria transmission in an efficient way will be essential for informed decision-making and long-term sustainability. If these findings are shown to be consistent in other settings, school surveys for malaria could provide such an operationally attractive tool for assessing malaria transmission in the surrounding community.

5.2.3 Primary school surveys as a metric for malaria transmission: Unpublished results

In addition to the analysis presented in section 5.2.2, a more detailed exploration of the data was conducted. The more in depth analysis of the data not only corroborates the findings presented above but also provides important insight to factors associated with the malaria epidemiology in this study population.

5.2.3.1 Concordance of Seroconversion Rates

Concordance between SCR estimates was examined in the all school population (figure 5-1A) and the school children residing within the community catchment area (figure 5-1B). Despite the imprecise estimates of SCR within each cluster due to the low sample size, in the school population, there was moderate correlation (r=0.549) in SCR with the community estimates and borderline significant concordance was observed (Bradley-blackwood F=2.559; p=0.089). Similar to the other malaria metrics tested (section 5.2.2), when restricting the school population to those residing within the community catchment, correlation improved (r=0.613) and there was strong concordance between the estimates (Bradley-blackwood F=0.162; p=0.851). Therefore, in this population, transmission intensity ascertained using the convenience sampling approach was a reliable gauge of malaria and could be a useful tool for malaria surveillance.

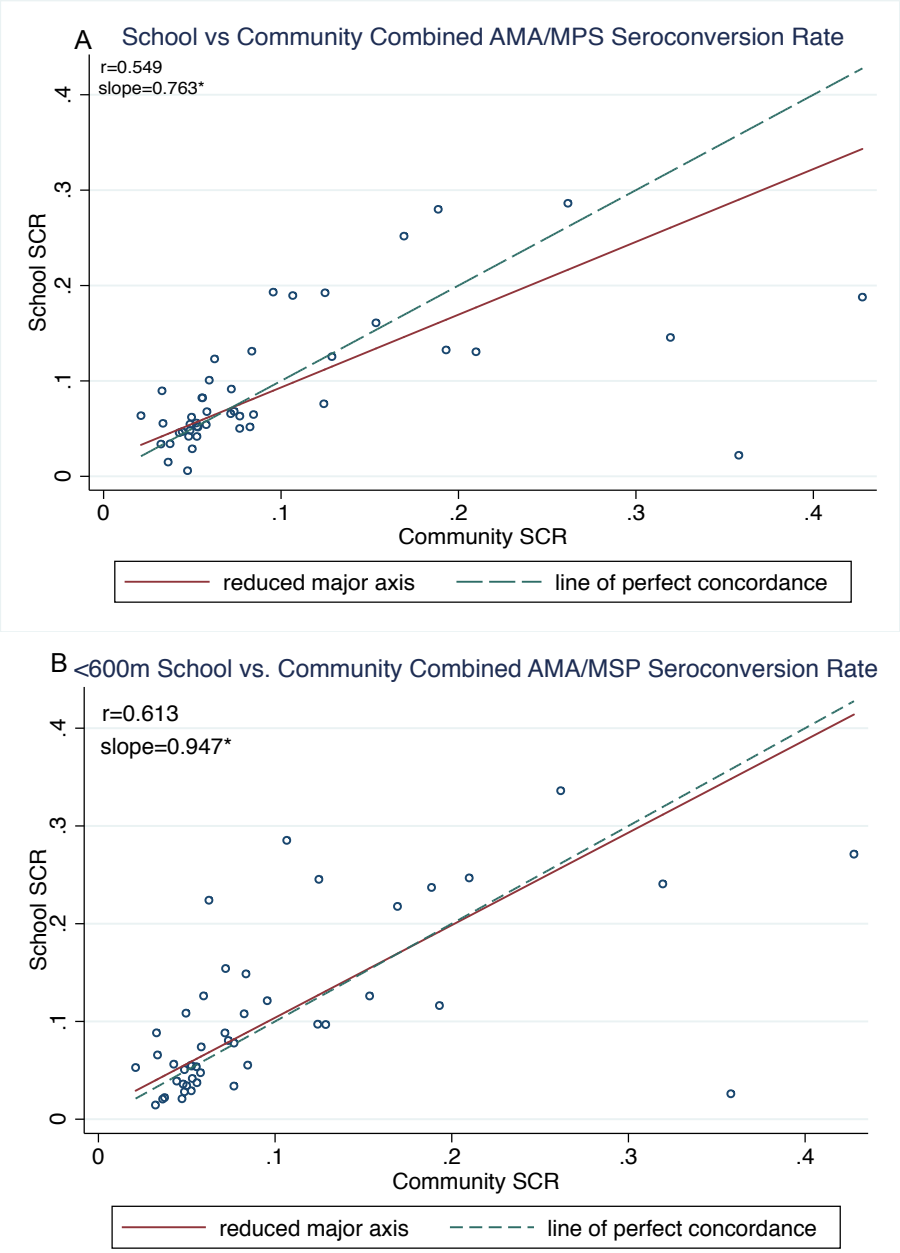

Figure 5-1: Prevalence of malaria infection in school vs. community surveys in 46 clusters by seroconversion rates (SCR). Scatter plots are shown with the line of perfect concordance (x=y) and the data's reduced major axis using total least squares regression A) SCR in the community vs. all school children; B) SCR in the community vs. school children residing within the community cluster (<600 m from the school).

5.3 Convenience Sample: Health Facilities

In addition to school surveys, health facilities provide another convenient source of data on the disease burden in the community. (100) Historically, estimates from

passively collected clinical data and more recently health facility based cross-sectional surveys have been used to assess trends in malaria burden. (134) However, as discussed in section 3.3.2, the quality of health facility data is affected by the obvious limitations of restricting sampling to those with suspected malaria, to those who are willing and able to attend the clinic, and the limited quality of recorded data and diagnosis. (223, 224, 250, 271) Therefore, the ability of use of routinely collected data as well as purposefully conducted health facility surveys to provide an accurate representation of malaria transmission in the community is needed to assess the sensitivity of these convenient sampling approaches for monitoring malaria transmission (specific objective 2).

5.3.1 Passive Case Detection

5.3.1.1 Introduction/Background

Passive case detection (PCD) systems have been used to monitor disease burden and is the main source of information on malaria morbidity and mortality over time. (100) When the focus was on controlling malaria and alleviating clinical disease and mortality, the use of PCD systems provided the most sensible data for monitoring and surveillance. With the decline in transmission and a shift to malaria elimination, there is evidence that data on suspected and/or confirmed cases may not provide the best marker to assess the broader community level transmission however, direct comparisons with community surveys have rarely been done. (100) The value of PCD to reflect heterogeneity in community malaria transmission needs confirmation in areas with a large proportion of asymptomatic infections (53), changing malaria endemicity (99), or areas with significant spatial variation in risk. (127) Therefore, the aim of this study was to assess if data collected during PCD in this highly heterogeneous setting was consistent with community level transmission.

5.3.1.2 Methods

A PCD study was set up in five health facilities whose catchment populations fall within the study site to maximize the spatial overlap with ongoing community work and therefore provide insight on how this data represents malaria burden in the community population: Ober Health Centre, Omrio Health Centre, Oriang Catholic Dispensary, Nyandiwa Baptiste Dispensary, and Wire SDA Dispensary

were identified for inclusion. Facilities were identified based on the presence of a working laboratory, and a full-time laboratory technician. The PCD study took place between February and September 2012 for a total of 27 weeks. Meetings were held with the District Ministry Of Health (DMOH) and facility in charges to obtain their consent for participating in the study, the best ways to integrate the PCD into daily operations, and for training on all procedures. No remuneration was offered. Each facility was provided a tympanic thermometer (Braun ThermoScan Compact Ear Thermometer, Kronberg, Germany) and supplies of RDTs (First Response, Premier Medical Corporation Ltd., India) to improve diagnostic capacity and case management. Spot checks took place at least twice a month on randomly selected days. The PCD began with a four-week pilot phase where adherence to research protocols was monitored. After which, it was decided to suspend work in two facilities: Nyandiwa had only worked a total of 4 days during that period and staff in Oriang was found to be falsifying data.

For other facilities, all patients referred to the laboratory for a blood slide (BS) for malaria were included in the PCD study. All patients receiving a BS also received a test for fever using the provided ear thermometer. Those with a tympanic temperature greater than 37.5 °C were tested for malaria with both BS and RDT. If patients were not febrile, they received a BS only. All individuals referred for a malaria test were recorded in a provided record book to record name, age, sex, head of compound, nearest primary school, temperature, BS and RDT result. All houses involved in the ongoing community work (appendix 1.3) were provided with cards that had the study house identification number recorded to identify those residing within the community study area. The participants were asked to carry their card any time that they attended the facility throughout the year. The card did not entitle them to free treatment but the technician asked the patient if they had a card and recorded the house identification number or the color of the card if they did not carry it with them, which represented a known area in the community.

5.3.1.3 Results

Over the 27-week period of the PCD study, a total of 8783 patients attended the facilities and were recorded in the register with Ober receiving the most patients

overall and Wire the fewest (table 5-2). Approximately one third of patients were children under five years. In total, 3357 (38.2%) patients were suspected of having malaria and were referred to the laboratory for testing, but this ranged across facilities with 27.5%, 47.4%, and 67.7% at Ober, Omiro and Wire, respectively.

Malaria positivity ranged per facility and per month. Overall the slide positivity of suspected malaria cases was 45.2%, but this ranged by month from 7.9% to 90.0% (table 5-2). Omiro had the highest slide positivity rate (74.6%). Across all facilities there was higher slide positivity in children under five years of age compared to those five years of age and older. Similar trends were observed for RDT positivity: overall 47.7% of patients suspected of having malaria and were febrile tested positive by RDT and a similar trend was observed to that of slide positivity with Omiro experiencing the highest and Ober having the lowest prevalence by RDT positivity (table 5-2).

Attendance trends over the 27-week period of the PCD survey showed minimal seasonal variability with the peak rains occurring between March (week 10) and June (week 25) (figure 5-2 A, B). Treatment despite a negative test and treating patients without a laboratory test for malaria infection was observed in all facilities. The number of patients treated for malaria without a confirmatory test result was much higher in Ober Health Center compared to the other clinics. Given the rate of overtreatment, using the number of people treated for malaria as an indicator of burden does not appear to be a reliable metric that can be consistently compared across facilities.

Table 5-2: Demographic characteristics and malaria results of the patient population (i.e. suspected malaria cases referred for blood slide) ordered according to transmission intensity as quantified by seroconversion rate in section 5.3.3.

	Omiro	Wire	Ober	Combined
Number of Patients Attending Facility				
N - All*	1864	1416	5503	8783
% <5	35.2	27.8	29.8	33.2
Number Suspected Malaria Cases Sent to Lab				
N - All*	883	959	1515	3357
% <5	46.0	36.5	34.6	38.2
Age – Mean (IQR)				
All	12.9 (3-17)	16.7 (3-22)	17.5 (3-25)	16 (3-23)
<5	2.5 (1-4)	2.3 (1-4)	2.4 (1-4)	2.4 (1-4)
≥5	21.7 (9-27)	25.0 (12-32)	25.4 (14-35)	24.4 (12-32)
Sex - % Male				
All	43.8	43.3	44.2	43.8
<5	52.5	48.7	48.0	49.7
≥5	37.0	40.3	42.2	40.4
Slide Positivity - % (Monthly Range)				
All	74.6 (50-90)	50.2 (25.4-69)	24.9 (7.9-43.8)	45.2 (7.9-90)
<5	77.3 (0-100)	51.3 (15.4-81.8)	31.7 (0-60.0)	51.6 (32.8-67.7)
≥5	72.5 (53.8-100)	47.7 (21.1-75.0)	22.3 (3.6-39.0)	41.3 (27.9-56.3)
RDT Positivity - % (Monthly Range)				
All	63.8 (0-100)	48.4 (0-78.6)	34.7 (0-62.5)	47.7 (0-100)
<5	69.7 (0-100)	44.0 (0-100)	40.8 (0-100)	51.0 (0-100)
≥5	58.5 (0-100)	50.3 (0-100)	30.7 (0-83.3)	51.7 (0-100)

*175 entries missing age data

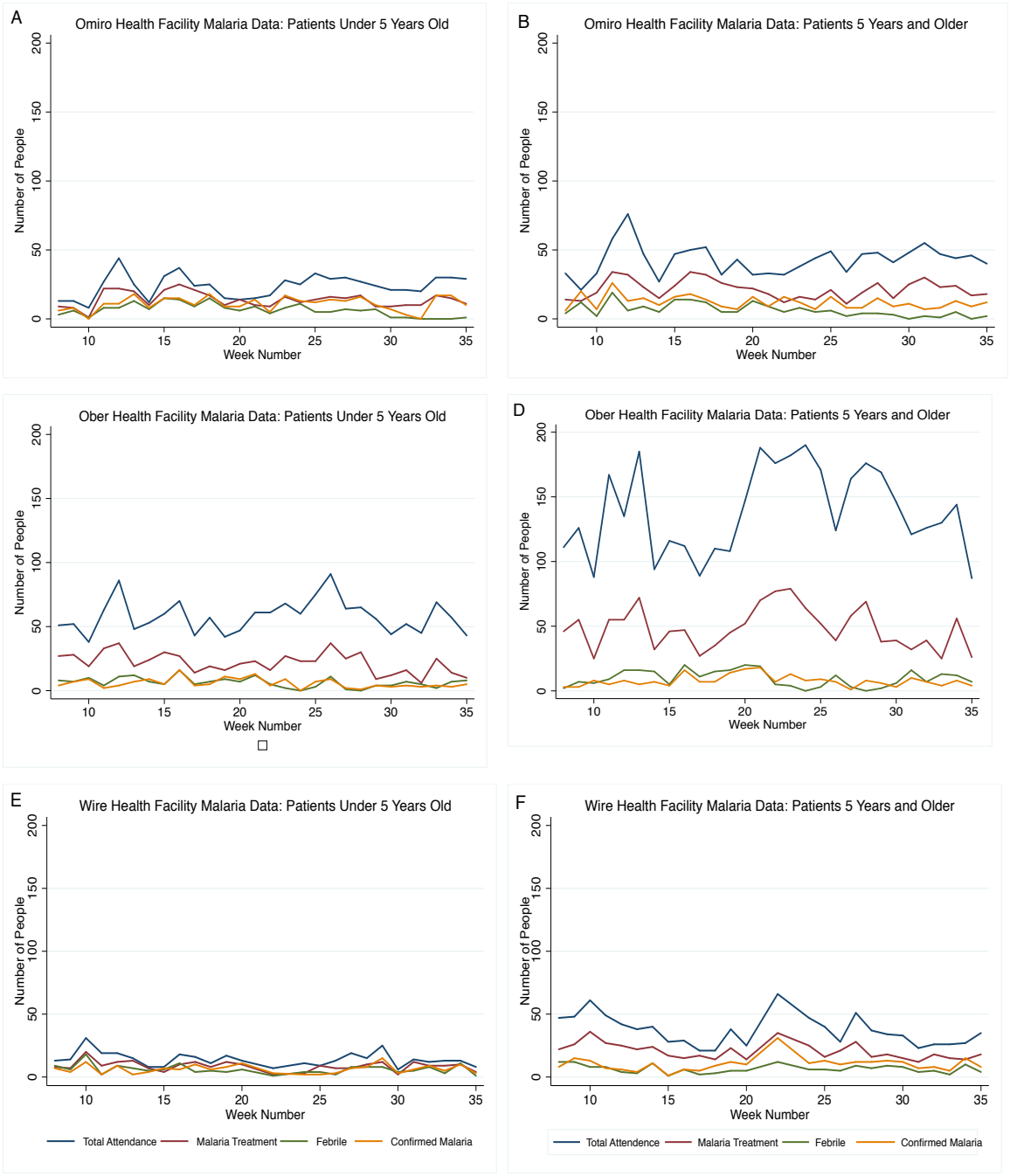

Figure 5-2: Overview of weekly malaria PCD data collected in each health facility. Panels A, C, E show results for those under 5 years of age and B, D, F for those aged 5 years and older for Omiro, Ober, and Wire health facilities, respectively. Lines show total attendance, number of febrile individuals, number treated for malaria, and the number with a confirmed malaria diagnosis according to either blood slide or RDT per week.

To provide an accurate assessment of whether a PCD approach is able to provide reliable estimates of malaria transmission in the community, ID cards were distributed to the community to identify those attending the facilities that resided in the area where community based estimates were available. ID cards were distributed to approximately 25% of the community (appendix 1.3). However, of the 8.2% of attendees that reported having received a card, only 14.1% had remembered to bring it with them to the facility. Due to the small sample size of individuals with a known location in the community, a meaningful comparison could not be made.

5.3.1.4 Section Discussion

The PCD survey was a simple and cost effective approach to obtain information on the malaria burden. Both RDT and BS positivity showed similar trends. It was not possible to determine how estimates of malaria directly compared with the community due to the small proportion of people that could be geolocated. However, the rank of the facilities as determined by the PCD was similar to that from the cross-sectional surveys, as discussed below in section 5.3.3, suggesting that this convenience sample provides a reasonable relative measure of intensity of malaria transmission in the broader population. The utility of routinely collected data to monitor malaria transmission is therefore likely to be more useful at the regional scale: Over a broader geographic scale than used in this study, routine data could be useful to identify areas of relatively high or low endemicity and/or identify areas that are of interest for further investigation.

Due to the lack of spatial information on patient residence, a comparison at a more granular level could not be made. One of the strengths of routinely collected data is that malaria trends can be analyzed in real-time to quickly identify any changes in transmission intensity. (272) Such systems are integral to any malaria surveillance program to enable a quick response to contain any potential outbreak or to ensure that control operations are targeted where and when they would be more effective. However, as mentioned, their ability to detect the presence of emergence of hotspots will depend on the availability of spatial information of the residence of cases.

This PCD system in our setting involved continuous monitoring of both adherence to study protocols and for data quality. We found that this was needed and one facility had particularly unreliable data. This continuous support will not be available in routine practice and is a known problem with relying on routinely collected data to monitor malaria transmission. (178) For research purposes, facilities could only be included if they had a functioning laboratory: in this setting, of the 12 health facilities in the region, only seven had functioning laboratories at the time of the survey and of those only three were consistently staffed and therefore retained in the study. For PCD systems to become a reliable tool to monitor malaria transmission, the availability of confirmatory testing to replace clinical diagnosis, the variation in the clinical subjectivity, and more specifically, the variation inherent in the rate of referrals for suspected malaria must be addressed before meaningful comparisons between facilities can be made.

Ultimately, the use of routinely collected data is the most convenient source of data to monitor malaria transmission in the community. This study showed that data generated during PCD are able to rank regions according to transmission intensity, but the commonly used malaria metrics (ie. slide positivity) overestimate true prevalence and the relationship is likely to be highly seasonally variable (ie. 100% slide positivity rate during the low transmission season). Therefore, PCD systems are useful for identifying areas that have higher or lower transmission intensity and would therefore merit for a more detailed examination using alternative approaches such as school or all attendee health facility surveys with corresponding geolocation exercises, to obtain more precise estimates of transmission dynamics in the community.

5.3.2 Health facility surveys as a metric for malaria transmission

One approach to minimize the biases associated with using routinely collected data in health facilities is to conduct a cross-sectional survey of the health care seeking population. Testing all patients and accompanying individuals for malaria regardless of the presence of malaria symptoms could minimize bias inherent with focusing on suspected cases alone and therefore has the potential to provide a more robust estimate for transmission intensity in the broader community than

sampling approaches that include symptomatic individuals only (specific objective 2).

5.3.3 High Levels of Asymptomatic and Subpatent Plasmodium falciparum Parasite Carriage at Health Facilities in an Area of Heterogeneous Malaria Transmission Intensity in the Kenyan Highlands (P3)

Authors: Gillian H Stresman, Jennifer C Stevenson, Nnenna Ngwu, Elizabeth Marube, Chrispin Owaga, Chris Drakeley, Teun Bousema, Jonathan Cox

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?	**American Journal of Tropical Medicine & Hygiene**		
When was the work published	**14/10/2015**		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes, CC-BY**	Was the work subject to academic pier review?	**Yes**

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	
Please list the paper's authors in the intended authorship order.	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed and conducted the data collection, contributed to laboratory analysis, conducted all statistical analysis and wrote the manuscript.

Student Signature: _____  **_____ Date: ___17/02/2015_____**

Supervisor Signature: ____  **_____ Date: ___17/02/2015_____**

Abstract

In endemic settings health facility surveys provide a convenient approach to estimating malaria transmission intensity. Typically, testing for malaria at facilities is performed on symptomatic attendees, yet asymptomatic infections comprise a considerable proportion of the parasite reservoir. We sampled individuals attending five health facilities in the western Kenyan highlands. Malaria prevalence by rapid diagnostic test was 8.6--32.9% in the health facilities. Of all PCR positive participants, 46.4% (95% CI: 42.6--50.2%) had infections that were RDT negative and asymptomatic of which 55.9% consisted of multiple parasite clones as assessed by merozoite surface protein-2 genotyping. Subpatent infections were more common in individuals reporting the use of non-artemisinin based antimalarials in the two weeks preceding the survey (OR 2.49, 95% CI: 1.04--5.92) compared to individuals not reporting previous use of antimalarials. We observed a large and genetically complex pool of subpatent parasitemia in the Kenya highlands that must be considered in malaria interventions.

INTRODUCTION

In order to allow national programs to effectively tailor malaria control strategies to local transmission dynamics it is essential that existing surveillance systems are capable of providing accurate, spatially-specific measures of malaria transmission intensity. [1, 2] Most malaria surveillance systems, including the system in Kenya, are predicated on passive detection of cases at health facilities that use either clinical diagnosis alone, or clinical diagnosis with parasitological confirmation by microscopy or rapid diagnostic tests (RDT). [3, 4, 5, 6] However, estimates of malaria burden from passive case detection data are subject to a number of potential biases that can vary considerably between health facilities, including the occurrence of non-malarial fevers, variations in accessibility of health services, willingness to pay any ancillary costs, and the diagnostic test used. In addition, the experience of the laboratory and clinical personnel, quality of microscopy or particular brand or availability of RDTs, and time dedicated to malaria testing are also important potential sources of bias making results difficult to compare. [6, 7]

Health facility based cross-sectional surveys that sample from all individuals presenting at the facility, as well as any accompanying individuals (as distinct from

sampling only among individuals with suspected malaria) have been shown to be a useful tool for measuring malaria transmission intensity. [8, 9] Health facility surveys provide an operationally attractive method to estimate malaria prevalence in the wider catchment population, as the inclusion of all health facility attendees mitigates against some of the biases associated with passive case detection. [7, 10] However, most health facility malaria surveys have relied on diagnosis by microscopy or RDT, both of which have a limited ability to detect parasitemia at low parasite densities. [8, 11, 12] The number of malaria infections detected through these surveys is therefore likely to have been substantially lower than would have been achieved using a more sensitive diagnostic approach such as polymerase chain reaction (PCR). [11, 13, 14] The potentially large proportion of infections that are undetected poses a significant challenge for malaria surveillance, control, and elimination strategies: transmission is likely underestimated and reservoirs of infection missed. As a result, control programs may only target a subset of the actual parasite population or campaigns may be implemented before the parasite reservoir is at or below the threshold where elimination is feasible. [13, 15, 16, 17]

In the current study, two cross-sectional surveys were carried out in five rural health facilities in the highlands of western Kenya to 1) assess the utility of this type of survey approach for measuring malaria transmission; 2) identify the prevalence and complexity of asymptomatic and subpatent infections and; 3) to evaluate factors associated with having asymptomatic and subpatent infections.

METHODS

Study Site and Population

This study was conducted in health facilities in a highland fringe area covering a region of approximately 200 km^2 in Rachuonyo South, Nyanza Province in the western Kenyan highlands. The area is situated between 1400 and 1600 m above sea level and the landscape is characterized by rolling terrain intersected with rivers and streams. The population is predominantly people from the Luo ethnic group with subsistence farming being the main occupation. [18] Malaria in the area is spatially heterogeneous with prevalence estimates in primary schools ranging between 0 and 71% and transmission follows a bimodal seasonal pattern associated with the long and short rainy season typically occuring between April

and June and October and December, respectively. [19, 20] The predominant malaria vectors in the area are *Anopheles funestus* and *An. arabiensis* and *Plasmodium falciparum* is the principal malaria parasite species present. [21] Two surveys were conducted in five rural health facilities representing all government facilities in the area in collaboration with the District Ministry of Health. Sampling took place in Agawo, Ober, Omiro, and Tala health facilities in both surveys. In the second survey, Othoro Health Center was replaced with Wire Dispensary, a faith-based facility, to achieve maximum overlap with the ongoing community work (figure P3-1). The surveys were conducted in October 2011 and July 2012 to correspond with a period of low and high transmission, respectively and examine the sensitivity of these surveys to changes in transmission intensity. [18]

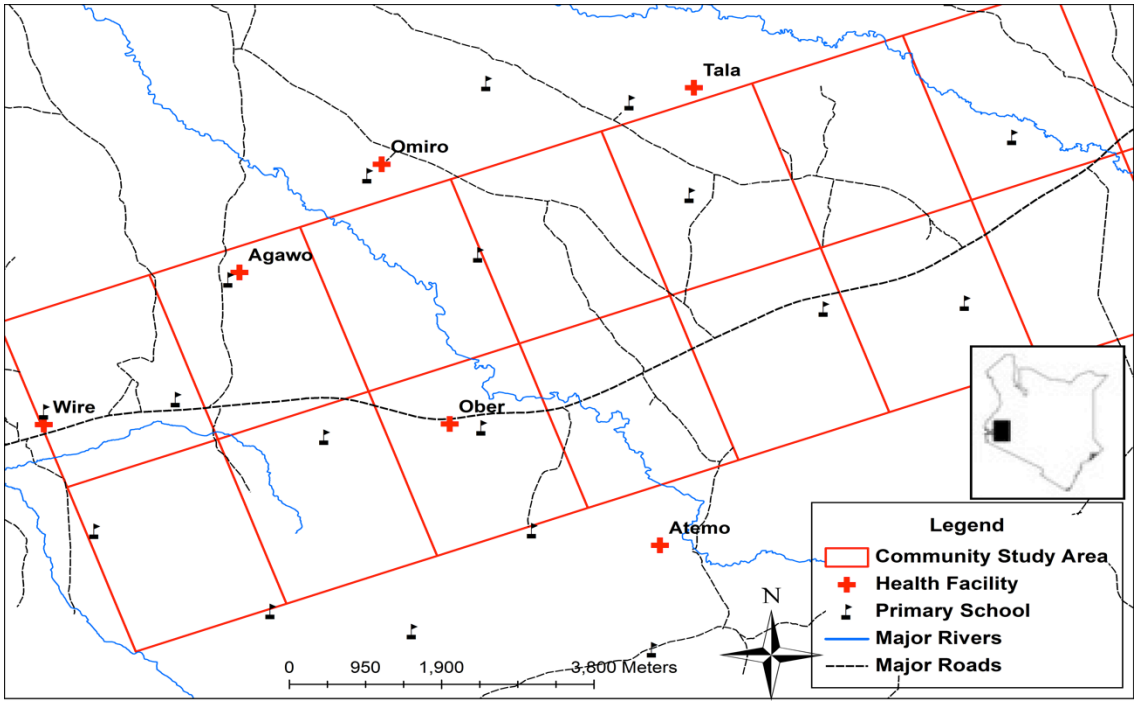

Figure P3-1: **Locations of rural health facilities included in the study as well as government primary schools and boundaries of the community survey. Note: Othoro Health Center is located along the main road approximately 20 km to the west of this area.**

Consenting and Sample Collection

All consenting patients and those accompanying them that attended the outpatient department during the four-week survey period were eligible for inclusion. At each facility, a maximum of 150 people from each of three age categories (0.5-5, 6-15,

>15 years old) were included. Recruitment within an age category was stopped once the target had been reached. Individuals were excluded if they were extremely ill and required immediate medical attention, if they were less than 6 months of age, if they were attending a scheduled clinic or other ward of the health facility, if they were unwilling or unable to provide consent (e.g. under 18 without being accompanied by a suitable guardian), or if they had been previously sampled at that same facility during the current study.

Two field workers were stationed at each facility and data collection activities were integrated into the normal day-to-day operations as far as possible. A field worker would approach each potential eligible participant and explain the study while they were waiting to visit the clinician. After the consenting process, a short questionnaire was administered on participant demographics, malaria history and control behaviors, whether they were a patient or accompanying person, current and recent symptoms, recent drug use, and travel history. Each participant was screened by RDT to determine the presence of current patent infections; three blood spots were collected on filter paper (3MM Whatman, Maidstone, UK) for subsequent molecular and serological analysis. Filter papers were dried and stored with desiccant at -80 °C. In the first year of the survey, axillary temperature was measured using a digital thermometer and those with temperature > 37.2 °C were considered febrile. [18] In the second year, tympanic thermometers were used due to the increased accuracy and shorter time to result. For those tested with the tympanic thermometers, only those with temperatures >37.5 °C were considered febrile. In the second survey, the RDT was changed from Paracheck (Orchid Biomedical Systems, India) to the more sensitive First Response kit (Premier Medical Corporation Ltd., India). [22] All diagnostic information was made available to the clinician for clinical decision-making. The final diagnosis and any drugs prescribed by the clinician to study participants were also recorded.

Research Ethics

The ethical committees of the London School of Hygiene & Tropical Medicine (Ref: LSHTM 5956) and the Kenya Medical Research Institute (Ref: SSC 1589) approved this study. Individual informed consent was obtained from all participants by signature or thumbprint accompanied by the signature of an independent witness.

Consent for children under the age of 18 was provided by a parent/guardian and children between 14 and 17 years also provided written assent by signature or thumbprint accompanied by the signature of an independent witness. As defined in the Kenya national guidelines, participants below 18 years of age who were pregnant, married, or a parent were considered 'mature minors' and consented for themselves. [23]

Laboratory Analysis

Filter paper blood spots were used to test for antibodies to malaria to ascertain malaria exposure and transmission intensity. Antibodies to *P. falciparum* Apical Membrane Antigen-1 (AMA1) and Merozoite Surface Protein-1 (MSP1-19) were detected by Enzyme Linked Immunosorbent Assay (ELISA). Briefly, two blood spot sections per sample were punched and antibodies eluted according to Baidjoe et al. [24] Antibody prevalence for each antigen was determined after defining a cut-off optical density (OD) based on a standard curve of known antibody concentration using the mixture model and normalized across plates. [20, 25] A person was considered to be seropositive if they had normalized OD values above the cutoff for at least one of the antigens tested. Age-adjusted seroconversion rates (SCR) were calculated. [25]

Nested polymerase chain reaction (nPCR) was used to test for the presence of parasite DNA to provide a gold standard measure for current infection. A Chelex-saponin approach was used to extract DNA as described by Baidjoe et al. [24] and the nPCR assay targeting the 18S ribosomal subunit of *P. falciparum* was used as previously described. [26] Samples that were positive by nPCR were then selected for subsequent analysis to identify allelic diversity using the polymorphic MSP2 region to provide an alternate measure of transmission intensity. [24, 25, 27] An additional nPCR reaction was conducted to amplify the block-3 region of the MSP2 domain targeting the FC27 and IC3D7 allelic variants. [28] The product of the MSP2 PCR was viewed on 1.5% agarose gel to determine the dilution factor necessary to prepare samples for capillary electrophoresis: intense bands were diluted at 1/100, moderate bands at 1/40 and faint bands at 1/10. Electropherograms were viewed using Peak Scanner (v1.0) and all discrete peaks greater than 500 florescent units were considered to be distinct allelic types. [29]

Case Definitions

Subpatent malaria infections were those who tested positive for malaria by nPCR but negative for malaria by RDT; patent infections were defined as those who were positive by both nPCR and RDT. Individuals who were positive by RDT but negative by PCR (n=267) were considered to be false positives likely attributable to residual HRP2 antigen and were not included in the analysis exploring subpatent infections (they were included in estimates of RDT prevalence, however). [30] Asymptomatic infections were individuals who tested positive for malaria by nPCR but were afebrile at the time of sampling and did not report history of fever in the 24 hours prior to sampling. [14]

Statistical Analysis

Statistical analysis was conducted using Stata 12.1 (STATACorp LP, USA) and R V3.02. Comparisons of parasite prevalence estimates between facilities, between years, and between age categories were performed using a two-sided test for proportions and the corresponding exact binomial 95% confidence intervals (CI). To assess the ability of health facility surveys to provide reasonable estimates of the community, data from a large community cross-sectional survey conducted in July 2011 in the same study area was used. [18] Data was restricted to those sampled as part of the community survey that resided within the health facility catchment areas as defined by cost-distance analysis and SCR was calculated as described above. [31] The health facility samples were restricted to those collected in July 2012 to minimize any potential seasonal bias. Multiplicity of infection (MOI) was calculated for all positive samples and 95% CI were calculated assuming a zero truncated Poisson distribution to account for all samples containing a minimum of one clone. Allelic richness (Rs), a metric for allelic diversity, was calculated using FSTAT v 2.9.3.2 software as previously described. [32]

Random effects logistic regression was used to assess factors associated with having a subpatent as well as asymptomatic malaria infection. Explanatory variables tested included year, sex, age, whether the individual was a patient or accompanying person, reported taking an antimalarial drug in the past 2 weeks, reported taking an anti-pyretic drug, reported using a bednet the previous night,

reported living in a household where indoor residual spraying had taken place in the previous 6 months, and number of infecting parasite clones. Due to the non-specificity of malaria symptoms it was not possible to further stratify patients by reason for attending facility. The final adjusted models were generated retaining all variables that were significant at the 0.05 level in a backwards fashion and the AIC values were used to confirm the optimum model fit.

RESULTS

Population Demographics

In total, 1598 and 1444 people were sampled in the first and second surveys, respectively, the majority of which were patients (table P3-1). There were similar proportions of males and females sampled in the <5 and 6-15 age categories, but significantly more females than males were sampled in the >15 year age group ($p<0.0001$). Most of the accompanying people were >15 years of age. Also, the majority of individuals reported that they had slept under a bednet the previous night, although in both surveys participants aged 6-15 were less likely to have reported using a net than younger children ($p<0.0001$) or adults ($p<0.0001$) (table P3-1). The majority of patients (63.4%; 95% CI: 61.4--65.3%; Facility Range (Range): 25.5--79.0%) reported having a fever in the previous 24 hours compared to 19.0% of accompanying people (95% CI: 15.9--22.4%; Range: 0--37.7%) but only 23.2% (95% CI: 21.5--24.9%; Range: 18.4--37.0%) and 7.5% (95% CI: 5.4--9.7%; Range: 0--19.7%) of patients and accompanying people, respectively had a current fever at the time of their health visit. Overall, 30.6% (95%CI: 28.9--23.2%; Range: 15.6--39.6%) of participants reported having taken antipyretic drugs and 13.7% (95% CI: 12.5--15.0%; Range: 8.8--21.9%) reported taking an antimalarial drug in the past 2 weeks.

Table P3-1: Demographics of the study population in health facility surveys in five rural health facilities carried out during the short and long malaria transmission seasons.

	Low Transmission Season (October 2011)			High Transmission Season (July 2012)		
	Mean	95% CI	Range	Mean	95% CI	Range
N						
All	1598	-	284-388	1444	-	203-379
6m-5 years	537	-	76-147	514	-	52-150
6-15 years	304	-	32-90	249	-	28-79
> 15 years	767	-	149-150	681	-	104-150
Sex - % Male						
All	37.5	35.2-40.0	33.8-38.9	38.7	36.2-41.3	34.6-40.1
6m-5 years	49.0	44.7-53.3	43.7-53.9	52.3	47.9-56.7	44.4-58.0
6-15 years	47.0	41.3-52.8	42.9-54.2	46.6	40.3-53.0	39.7-54.4
> 15 years	25.6	22.5-28.9	20.6-31.8	25.5	22.3-29.0	22.4-31.5
Patient/ Accompanying status - % Patient						
All	81.4	79.4-83.3	66.9-93.0	79.5	77.3-81.5	53.7-90.5
6m-5 years	96.5	94.5-97.8	91.6-91.7	93.8	91.3-95.7	88.5-98.0
6-15 years	96.0	93.2-97.9	90.6-100	97.2	94.3-98.9	92.9-100
> 15 years	64.9	61.4-68.3	43.9-85.3	62.3	58.5-65.9	30.4-80.8
Bednet - % reported sleeping under net previous night						
All	87.2	85.5-88.9	82.2-94.0	90.4	88.8-91.9	89.0-91.8
6m-5 years	86.8	83.6-89.6	82.6-92.1	94.0	91.5-95.9	88.7-97.5
6-15 years	82.1	77.3-86.3	69.6-92.3	81.1	75.7-85.8	76.0-84.8
>15 years	89.6	87.2-91.7	84.1-96.7	91.2	88.8-93.2	89.6-93.3
Recent IRS - % reported having IRS in past 12 months						
All	77.8	75.4-80.4	70.1-87.4	76.9	74.6-79.0	70.6-81.0
Recent Travel - % reporting having travelled in past 3 months						
All	32.5	30.0-35.1	26.7-39.9	20.1	18.1-22.3	10.7-29.8
6m-5 years	27.9	23.8-32.4	17.3-50.0	16.1	13.1-19.6	6.0-25.9
6-15 years	21.9	17.1-24.4	0-32.6	6.8	4.0-10.7	2.0-10.3
>15 years	40.7	36.7-44.8	22.2-49.0	28.0	24.7-31.6	14.4-39.3

Malaria Transmission Intensity

All metrics tested were able to detect a change in malaria burden between the two surveys. Seroprevalence estimates increased from 37.6% (95% CI: 35.2--40.0; Range: 24.5--53.0%) during the first survey to 46.8% (95% CI: 44.2--49.4%; Range: 34.4--62.0%) in the second survey (p<0.0001). Similarly, malaria parasite prevalence by RDT increased from 16.9% (95%CI: 15.1--18.8%; Range: 8.6--30.1%) to 22.4% (95% CI: 20.3--24.6%; Range: 9.5--32.9%) and by PCR from 20.4% (95%CI: 18.4--22.4%; Range: 9.5--40.3%) to 25.5% (95%CI: 23.2--27.7%; Range: 8.7--51.5%) during the first and second survey, respectively (table P3-2). Prevalence within age categories also increased between surveys with the highest estimates in the 6--15 years age category and the lowest in adults (p<0.001) (table P3-S1).

Table P3-S1: Malaria prevalence estimates by rapid diagnostic test (RDT) and nested polymerase chain reaction (PCR) per year with the range of estimates across health facilities stratified by age category.

	Short Transmission Season			**Long Transmission Season**		
	% Pos	**95% CI**	Range	**% Pos**	**95% CI**	Range
RDT						
All	16.9	15.1-18.8	8.6-30.1	22.4	20.3-24.6	9.5-32.9
6m-5 years	18.9	15.6-22.3	7.6-44.0	27.6	23.7-31.5	5.3-54.3
6-15 years	30.2	25.0-35.4	13.0-56.5	44.6	38.4-50.8	21.5-59.6
>15 years	10.1	25.0-35.5	2.2-13.1	10.4	38.4-50.8	7.2-14.8
PCR						
All	20.4	18.4-22.4	9.5-40.3	25.5	23.2-27.7	8.7-51.6
6m-5 years	25.9	22.2-29.7	14.3-53.2	27.8	23.9-31.7	8.0-67.9
6-15 years	26.1	21.2-31.1	7.8-57.6	34.1	28.2-40.1	11.4-60.3
>15 years	14.2	11.7-16.7	5.4-30.4	20.5	17.4-23.5	8.0-39.3

Table P3-2: Prevalence of malaria per facility for all malaria metrics including seroconversion rate (SCR), polymerase chain reaction (PCR) and rapid diagnostic test (RDT) prevalence, multiplicity of infection (MOI) and allelic richness (Rs), ordered from highest to lowest transmission intensity. TA=Tala; OM=Omiro; AG=Agawo;

Low Transmission Season (October, 2011)											
	SCR	95% CI	Sero (%)	95% CI	PCR (%)	95% CI	RDT (%)	95% CI	MOI	95% CI	Rs
TA	0.076	0.06-0.10	53.0	47.1-58.8	35.0	29.4-40.6	29.4	24.7-35.5	2.33	2.07-2.65	30.9
OM	0.069	0.05-0.09	49.1	43.1-55.0	40.3	34.4-46.1	16.9	13.7-23.2	1.99	1.79-2.24	24.1
AG	0.054	0.04-0.07	42.8	37.0-48.5	14.8	10.7-18.9	19.8	15.3-24.5	1.97	1.68-2.40	29.2
OB	0.028	0.02-0.04	25.8	21.4-30.2	9.5	6.6-12.5	11.6	8.4-14.8	1.72	1.44-2.19	27.0
OT	0.025	0.02-0.03	24.5	19.9-29.0	9.5	6.4-12.6	8.6	5.7-11.6	1.84	1.56-2.28	29.0
High Transmission Season (July 2012)											
TA	0.114	0.09-0.14	62.1	56.3-67.7	51.6	45.8-57.3	32.9	27.4-38.3	2.29	2.09-2.52	37.1
OM	0.113	0.09-0.15	55.6	48.9-62.2	31.3	25.1-37.5	27.6	21.6-33.6	1.85	1.63-2.15	28.0
WI	0.069	0.05-0.09	52.2	45.3-59.1	28.8	22.5-35.0	18.0	12.7-23.6	1.5	1.34-1.75	20.5
AG	0.061	0.05-0.07	39.5	34.2-44.5	16.2	12.4-20.1	27.1	22.4-31.7	2.12	1.84-2.50	39.5
OB	0.048	0.04-0.06	34.2	29.4-39.0	8.7	5.8-11.5	9.5	6.6-12.5	1.95	1.60-2.53	18.0

Similarly, SCR indicated a range of transmission intensity between facilities and an increase in transmission intensity between the two surveys (figure P3-2A). Also, based on this small sample of 5 facilities, SCR estimates from the health facility survey during the high transmission season were strongly correlated (r=0.96) with estimates obtained from a community cross-sectional survey in the same area conducted the previous year (figure P3-3). With the exception of allelic diversity (p=0.62), the malaria metrics tested were able to consistently rank health facilities according to transmission intensity, as quantified by SCR. The intensity of malaria transmission (indicated by the SCR) experienced by individuals attending the selected health facilities during the first survey was associated with health facility level parasite prevalence by both RDT (p=0.04) and PCR (p=0.05) as well as multiplicity of infection (p=0.04). Despite the association of RDT and transmission intensity, it is worth noting that one facility (Agawo) would have been misclassified as being in a high transmission setting based on RDT results in symptomatic patients alone (figure P3-2B). SCR during the first survey was also strongly associated with SCR in the second survey (p<0.001) and ranks between transmission intensity and all malaria metrics showed similar trends (data not shown).

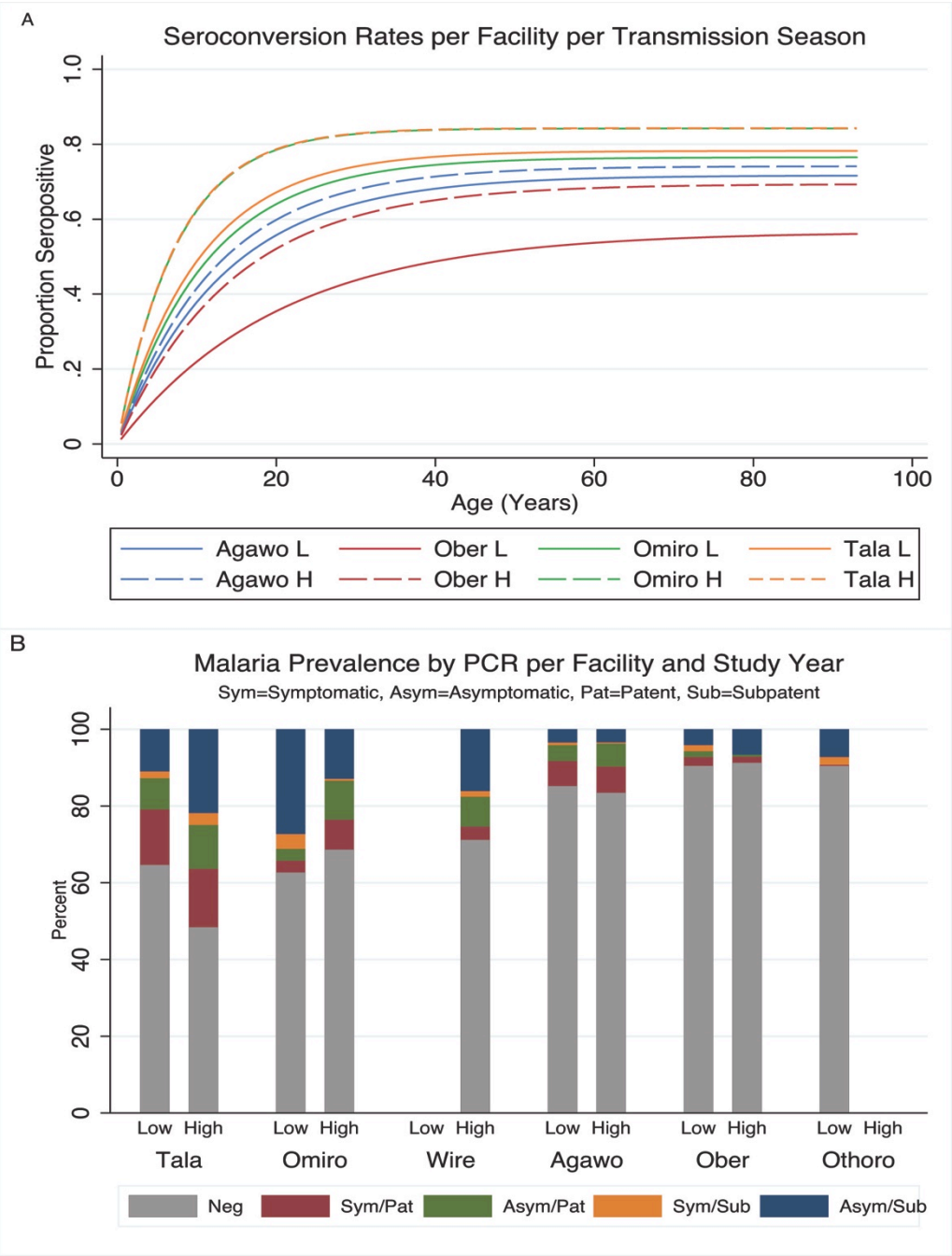

Figure P3-2: A) Seroconversion rates per health facility and transmission season for facilities sampled in both surveys. Note: Omiro2 and Tala2 curves overlap B) PCR prevalence ordered according to transmission intensity including subpatent and asymptomatically infected individuals per health facility and transmission season.

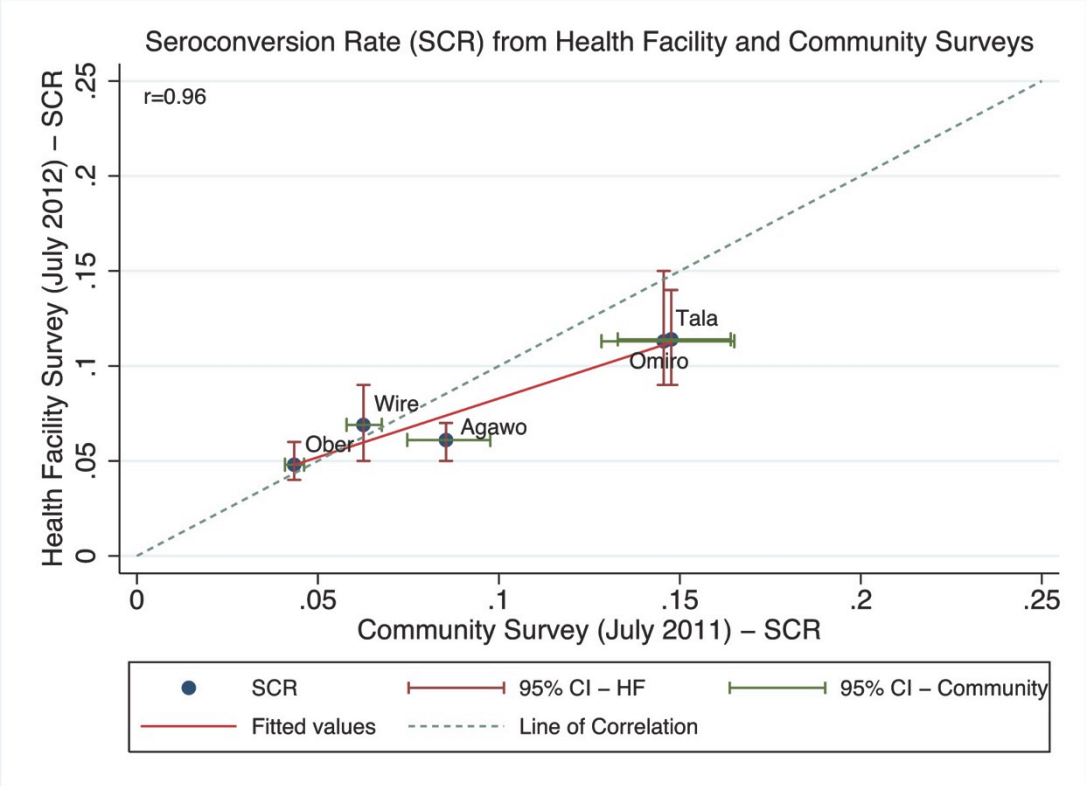

Figure P3-3 Comparison of health facility and community: Comparison of transmission intensity estimates based on seroconversion rate (SCR) from health facility and community surveys and corresponding correlation coefficient (r). Health facility estimates were restricted to sampling that occurred in the high transmission season and community estimates were restricted to those residing in the health facility catchment area to minimize spatial or seasonal biases as much as was possible.

Subpatent and Asymptomatic Infections

Overall, 586 infections were detected by RDT, 54.4% of these were confirmed by PCR. PCR identified an additional 358 infections (12.0% of the total study population). In total, 52.9% (Range: 24.7--97.0%) and 67.5% (Range: 27.3--81.4%) of the PCR positive individuals had sub-patent and asymptomatic infections, respectively; the majority of which were found in adults ($p<0.0001$) (table P3-S2). Based on the clinical records, the majority of subpatent infections (83.8%; 95% CI: 79.6--87.5%) were not provided treatment while 95.1% (95% CI: 93.0--96.7%) of RDT positive individuals were prescribed an antimalarial drug.

Table P3-S2: Prevalence of subpatent and asymptomatic infections of all PCR positive individuals stratified by age category with the range of estimates across health facilities.			
	% Pos	**95% CI**	Range
% Subpatent & Asymptomatic Infections			
6m-5 years	33.3	27.8-39.2	14.6-73.3
6-15 years	31.5	24.4-39.2	6.1-83.3
>15 years	71.5	65.4-77.2	56.5-85.7
% Subpatent & Symptomatic			
6m-5 years	9.4	6.2-13.5	4.4-20.0
6-15 years	1.2	0.15-4.4	0-16.7
>15 years	6.7	3.9-10.6	3.3-25.0
% Patent and Asymptomatic			
6m-5 years	20.3	15.7-25.5	0-38.5
6-15 years	30.9	23.8-38.6	0-46.1
>15 years	15.5	11.1-20.7	0-30.4
% Patent and Symptomatic			
6m-5 years	37.0	31.2-42.9	6.7-53.9
6-15 years	36.4	29.0-44.3	0-51.5
>15 years	6.3	3.5-10.1	0-11.9

Of all PCR positive participants, 26.0% (Range: 3.0--42.6%) were patent and symptomatic, 21.1% (Range: 0--32.7%) had patent and asymptomatic infections, while 46.4% (Range: 21.8--75.7%) were subpatent and asymptomatic for malaria. In total, 6.5% (Range 3.0--21.2%) of PCR positive individuals were subpatent and symptomatic: 38.6% (17/44) of these individuals were diagnosed with malaria while 10 of the 17 participants as well as the 27 participants not treated for malaria were diagnosed with another fever-inducing illness such as flu or typhoid (figure 2B).

The majority of infected individuals had one (43.2%) or two (29.4%) allelic types with the most diverse samples showing evidence of 7 different parasite clones. The FC27 subtype was most prevalent with 57 distinct allelic types identified compared to 31 unique types from the 3D7 family. The MOI in the study population was low with a mean of 2.05 (95% CI: 1.92--2.19; Range: 1.7--2.3) and 2.02 (95% CI: 1.91--2.15; Range: 1.5--2.3) clones per person in the first and second survey, respectively. Estimates of MOI were slightly higher in the 6--15 year old population but no difference was observed between patent and subpatent and symptomatic and asymptomatic infections (table P3-3).

Table P3-3: Unadjusted multiplicity of Infection (MOI) and range per facility, number of distinct alleles (A) and allelic diversity (R_s) for PCR positive samples; combined results for both health facility surveys.

	MOI	**95% CI**	Range	**A**	**R_s**
Age					
6m-5y	1.98	1.85-2.13	1.46-2.36	70	67.59
6-15y	2.23	2.03-2.46	1.75-2.45	67	67.0
>15y	1.97	1.84-2.13	1.39-2.5	58	56.77
Malaria Drugs					
No Drug	2.02	1.93-2.13	1.56-2.31	80	47.45
ACT	2.02	1.74-2.42	1.33-2.5	37	36.39
non-ACT	2.26	1.89-2.78	1.96-2.75	32	32.0
Detectable Parasites					
Patent	2.06	1.93-2.21	1.67-2.31	78	78.0
Subpatent	2.01	1.89-2.14	1.32-2.79	62	62.85
Symptoms					
Symptomatic	2.03	1.92-2.16	1.40-2.34	78	76.14
Asymptomatic	2.03	1.89-2.18	1.52-2.51	62	62.0

Factors Associated with Subpatent/Asymptomatic Infections

In adjusted models, individuals older than 15 years had 2.55 (95% CI: 1.50--4.30) times the odds of having an asymptomatic infection compared to those less than 5

years old. The odds of asymptomatic infections also being subpatent compared to patent were 7.53 (95% CI: 4.88--11.62). If a person was attending the health facility seeking care or were sampled during the first survey, they were more likely to be symptomatic (table P3-4).

Table P3-4: Unadjusted and adjusted results for fixed effects of mixed effects logistic regression using health facility as random effects for variables associated with having an asymptomatic malaria infection compared to those with symptomatic infections.

Outcome: asymptomatic infection	Unadjusted		Adjusted	
	OR	95% CI	OR	95% CI
Study Year	**1.3**	**0.92-1.83**	**1.67**	**1.13-2.47**
Age Category				
6m – 5 years	1.00	1.00	1.00	1.00
6-15 years	1.64	1.09-2.47	1.98	1.26-3.11
>15 years	6.14	3.89-9.71	2.55	1.50-4.30
Patient (vs. accompanying person)	0.11	0.05-0.25	0.26	0.10-0.67
Subpatent (vs. patent)	8.64	5.81-12.83	7.53	4.88-11.62

Similarly, those over 15 years had over 3 times the odds of having a subpatent infection (OR: 3.53; 95% CI: 2.23--5.59) compared to the youngest age group and older children were half as likely to be asymptomatic (OR: 0.54; 95% CI: 0.33--0.90). Those who had reported taking antimalarial drugs in the past two weeks had greater odds of having a subpatent infection: participants reporting having taken non-artemesinin based antimalarial drugs (ie. quinine, sulphadoxine-pyramethanime) had a 2.49 greater odds of being subpatent (95% CI: 1.04—5.92) and those reported having used artemesinin combination therapy (ACT) had

almost twice the odds of being subpatent, although this was not significant (table P3-5).

Table P3-5: Unadjusted and adjusted results for fixed effects of mixed effects logistic regression using health facility as random effects for variables associated with having a sub-patent malaria infection compared to patent infections.

Outcome: subpatent infection	Unadjusted		Adjusted	
	OR	95% CI	OR	95% CI
Age Category				
6m – 5 years	1.0	1.0	1.00	1.00
6-15 years	0.79	0.51-1.23	0.55	0.33-0.90
>15 years	6.00	3.91-9.20	3.53	2.23-5.59
Asymptomatic	9.08	5.97-13.80	7.65	4.86-12.04
Antimalarial drug (2 weeks)				
No Drug	1.0	1.0	1.0	1.0
ACT	1.58	0.83-3.01	1.81	0.84-3.89
non-ACT	1.64	0.81-3.29	2.49	1.04-5.92

DISCUSSION

This is one of the few studies, and the first in Kenya, to assess the utility of surveys in health facilities as a means of measuring malaria transmission intensity in an area where transmission varies over a small geographical area. [9, 10, 33] The results of this study indicate that health facility derived serological, parasitological and molecular measures can detect differences in transmission intensity at a small geographical scale and are sensitive to seasonal changes. These findings suggest that health facility surveys are able to provide a reasonable measure of community level transmission, are capable of delineating areas of high or low malaria transmission, and that the use of serology and PCR added useful information to assessing transmission levels in the sampled populations that would have been missed if sampling focused solely on those cases suspected of having malaria. [8, 9, 20]

Similar to other studies, subpatent and asymptomatically infections were detected in this setting. It is likely that over half of malaria infections would have been missed had testing been restricted to use of RDTs for symptomatic cases. [11, 12, 13] The proportion of asymptomatic and subpatent infections differed by health facility, the main implication of which being that variations in transmission intensity will affect the proportion of infections missed using RDTs. The underestimation of malaria burden can have significant implications for malaria surveillance or developing control or elimination strategies based on clinical data. [16, 30, 34] For surveillance programs to capture the complete burden of malaria in a region the proportion of infections missed should be taken into account. Firstly, more robust data could be collected through use of more sensitive diagnostic tools such as PCR or through use of a high quality surveillance system targeting sentinel populations to get a more comprehensive picture of malaria transmission. [34, 35, 36] Alternatively, the limited sensitivity of RDT/microscopy can be acknowledged and adjusted for to estimate true prevalence or to modify policy guidelines on an expectation of missed infections. [11, 37]

Obtaining a better understanding of subpatent and asymptomatic infections is key to identifying which individuals are most likely to be missed by the current malaria surveillance practices. Similar to other studies, [14] our results suggest increased odds of having sub-patent and asymptomatic infections in older age groups. These findings align with the current theory that in areas with stable transmission, older individuals will have sufficient immunity to tolerate infections and maintain parasite densities below the limit of detection of RDTs. [30, 38] Also, reporting taking malaria drugs in the two weeks prior to the survey was associated with having a subpatent malaria infection. The increased odds of being subpatent in those reported taking antimalarial drugs may be associated with residual parasitemia shortly after treatment, or the detection of DNA from persisting gametocytes. [39, 40] An alternative explanation for our finding is drug resistance: resistance to sulphadoxime-pyrametamine (SP) is highly prevalent in western Kenya and while the use of this drug is officially limited to intermittent treatment for pregnant women it is widely available in most private retailers. [41, 42] Another possible

explanation includes suboptimal or self-dosing with malaria drugs. Compliance to drug regimens in this area has not been studied to our knowledge, but it is possible that if people are not completing their regimen properly the drugs may only reduce parasite densities to subpatent levels without completely clearing the infection. Bias in recalling when or if they took that specific drug is also a possibility.

We also explored the complexity of malaria infections to gain further insight on the molecular epidemiology of this study population. MOI has been shown to be a marker of transmission intensity that may have advantages in relatively high transmission settings where parasite prevalence may saturate. [3] Although MOI has proved to be a useful metric of malaria transmission intensity in certain settings [27, 32], no significant difference was found between facilities. This may be due to the spatial overlap of the health facility catchment areas, confounding factors not accounted for in the unadjusted analysis such as age, or due to the small sample sizes. However, lower allelic diversities were observed in subpatent and asympatomic infections, as well as in older individuals and those who reported taking antimalarial drugs. The lower allelic richness observed in facilities experiencing lower transmission intensity could be related to lower parasite densities expected in these populations or that certain low-density allelic forms were missed due to the PCR process.

The study design had some important limitations. The introduction of more sensitive diagnostic tools during the second survey may have reduced the proportion of subpatent and asymptomatic infections in that season. This was, however, incorporated in the statistical analysis and had little impact on the model results. Also, due to the cross-sectional nature of this survey, misclassification of participants on asymptomatic/subpatent status could have occurred. [14] It is possible that some individuals may have developed fever in subsequent days and this may have impacted our estimates of asymptomatic malaria. Similarly, the few studies that have looked at misclassification of patent/subpatent over time suggest that a small proportion of infections will shift between states but the overall proportion detected does not shift dramatically, suggesting that it is unlikely that

following these individuals over time would have a significant impact on these findings. [28, 43] Finally, to obtain a specific understanding of how well health facilities are able to gauge transmission intensity in the surrounding community, health facility estimates need to be explicitly compared to those of the community population which they are supposed to represent. In this study, we have made use of an existing community sample from the same area but collected the year before. Despite the temporal difference, the results indicate that a strong correlation in SCR between the convenience and community sampling strategies suggesting that the health facility provides a reasonable proxy for transmission intensity in the surrounding community.

Ultimately, health facility surveys provide an attractive tool to measure and detect heterogeneity in malaria transmission. In terms of sampling they include a broader sample of the health care seeking population instead of being restricted to those suspected of having malaria while at the same time are more operationally attractive compared to community-based surveys, in terms of the time and cost required to collect samples. [9, 20] However, more work is required to determine how these estimates compare to the surrounding community. Estimates based on routinely used diagnostic tools, such as RDTs, are likely to underestimate malaria prevalence due to the presence of sub-patent and asymptomatic infections but, in our study, correctly identified those health facilities with the highest transmission intensity in their catchment area. More research is needed to further explore the molecular epidemiology of malaria infections and to develop strategies that can easily identify these populations to ensure that malaria control decisions are based on a complete picture of malaria transmission.

5.3.4 Health facility surveys as a metric for malaria transmission: Unpublished Results

In addition to the results presented in section 5.3.3, additional work was done to explore the reliability of the diagnostic tools used as well as comparing results to explore agreement with community based surveys.

First, the diagnostic performance of the RDTs used in the cross-sectional surveys was calculated using nPCR as the gold standard. As expected, the First Response kit (Premier Medical Corporation Ltd., India) used in the second year showed a higher sensitivity and better predictive values compared to paracheck (Orchid Biomedical Systems, India) (table 5-3). Interestingly, when diagnostic performance was stratified by age category, results were more variable. Overall, the RDTs had the lowest sensitivity in the adult population at 20.0 and 23.3 in year 1 and 2, respectively. The RDTs performed much better in children, although the sensitivity was still only around 70%. The variability of RDT performance by age category is not surprising, as parasite densities in infected individuals would be expected to be lowest in the adult population who have developed sufficient immunity to control parasite populations and therefore are more likely to have infections below the limit of detection by RDT. (49, 54)

Table 5-3: Diagnostic performance of RDTs in year 1 (Paracheck) and year 2 (First Response) using nPCR as the gold standard for comparison stratified by age category

	Year 1			Year 2		
	≥5	6-15	>15	≥5	6-15	>15
Sensitivity	44.7	62.3	20.0	69.2	71.8	23.2
Specificity	89.8	81.1	91.4	88.4	69.5	92.0
PPV	59.6	53.3	27.0	69.7	54.9	45.7
NPV	89.9	86.1	87.8	88.2	82.6	82.5

Next, results from the health facility cross-sectional surveys were compared with estimates from the community surveys that were conducted in the same area. Results were restricted to those individuals that resided within the health facility catchment areas, as defined using the cost-distance algorithm described in appendix 1.1. Separate comparisons were made for each malaria metric using the most recent survey available with comparable data collection protocols. The surveys used for comparisons are described above: 1) community survey (XSS3) conducted in 2010 (section 5.2.2); and 2) community survey (XSS4) conducted in

2011 (appendix 1.3). Due to the small number of health facilities included in the survey (n=6), calculations for concordance could not be reliably estimated. Therefore analysis was restricted to correlation, which provides a crude measure as it does not account for error and bias (273), but provides useful insight as to the agreement between the surveys being compared.

Firstly, RDT prevalence estimates from the health facility surveys were compared with the community survey conducted in 2010 (XSS3). Although confidence intervals for the community estimates are wide, there was a strong correlation with the health facility survey (r=0.80) with the fitted values falling along the line of perfect correlation (Figure 5-3).

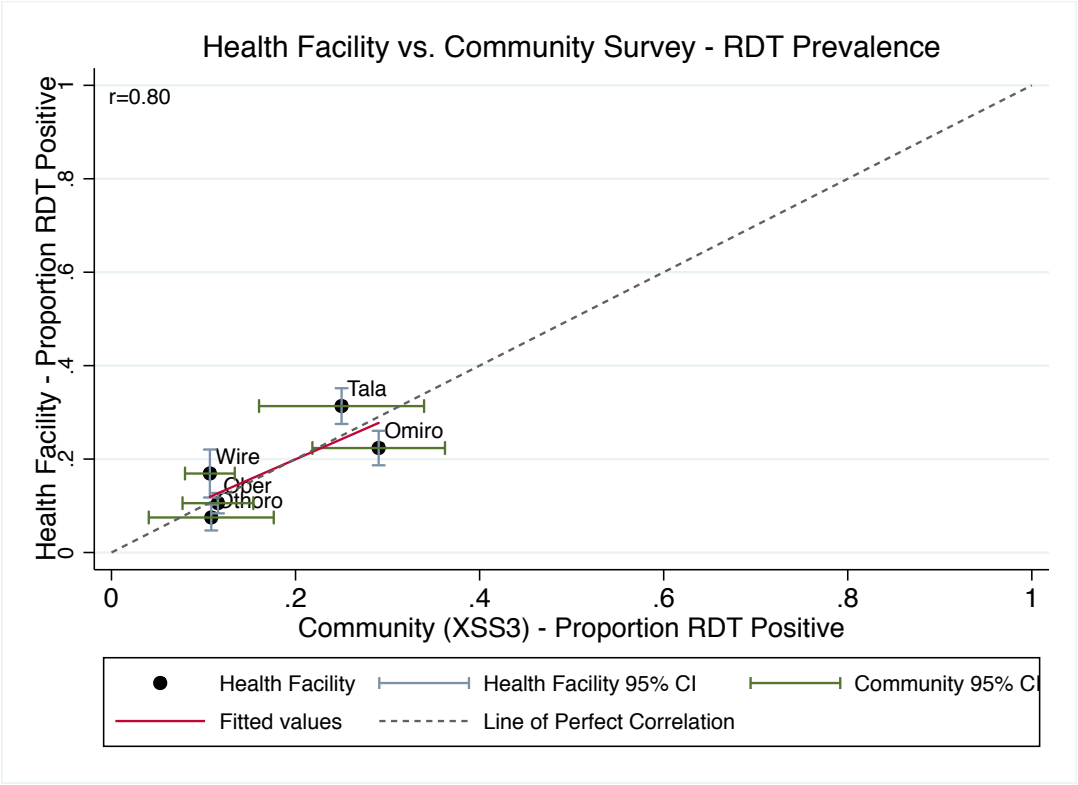

Figure 5-3: Scatter plots showing correlation of RDT prevalence from health facility and community surveys

The second outcome measure used for comparison was seroprevalence and was compared with the most recent community survey conducted in 2011 (XSS4). There was a good correlation observed between the health facility and community results (r=0.78). The community estimates were slightly higher than those

observed in the health care seeking population (figure 5-4). Although these findings may be different if restricted to the patient or suspected malaria populations, health facility surveys appear to slightly underestimate seroprevalence in the community. This finding may be due to bias due to health care seeking populations differing to those in the community. (134) An alternative explanation for this observation could be an association with malaria immunity. In areas with higher malaria transmission, individuals develop antimalarial immunity at a younger age and would be less likely to attend the health facility because of malaria and therefore seropositive individuals would be under-represented in the health facility survey. (127)

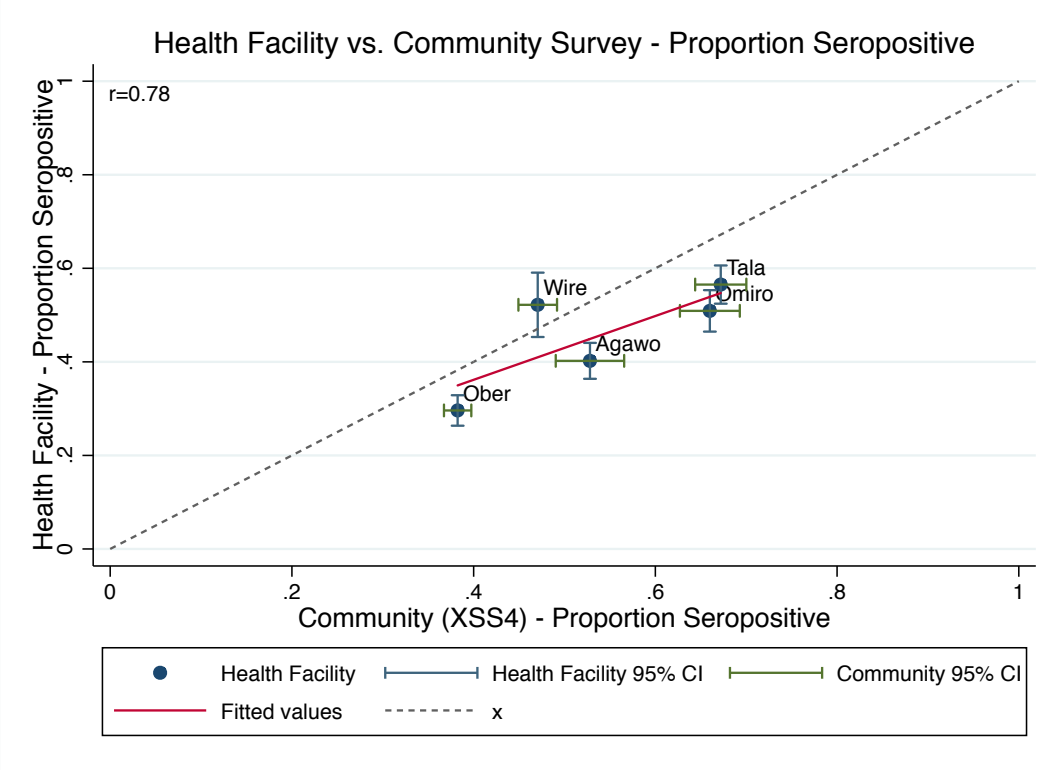

Figure 5-4: Scatter plots showing correlation of seroprevalence from the health facility and community survey (XSS4) restricted to those residing in the health facility catchment areas with corresponding 95% confidence limits.

Finally, PCR results were available for comparison between the health facility survey and the community survey conducted in 2011 (XSS4). Correlation between PCR prevalence estimates obtained during the health facility survey and those

sampled in the community that resided in the health facility catchment areas was 69% (figure 5-5). Unlike RDT and seroprevalence estimates, PCR estimates at the health facility were greater than those obtained during the community survey, particularly in areas with higher transmission intensity. This overestimation of community prevalence may be driven by the health facility population consisting of individuals that are more likely to have a current malaria infection and compounded by the increased likelihood that those accompanying a malaria positive individual are also carrying parasites. (121, 166)

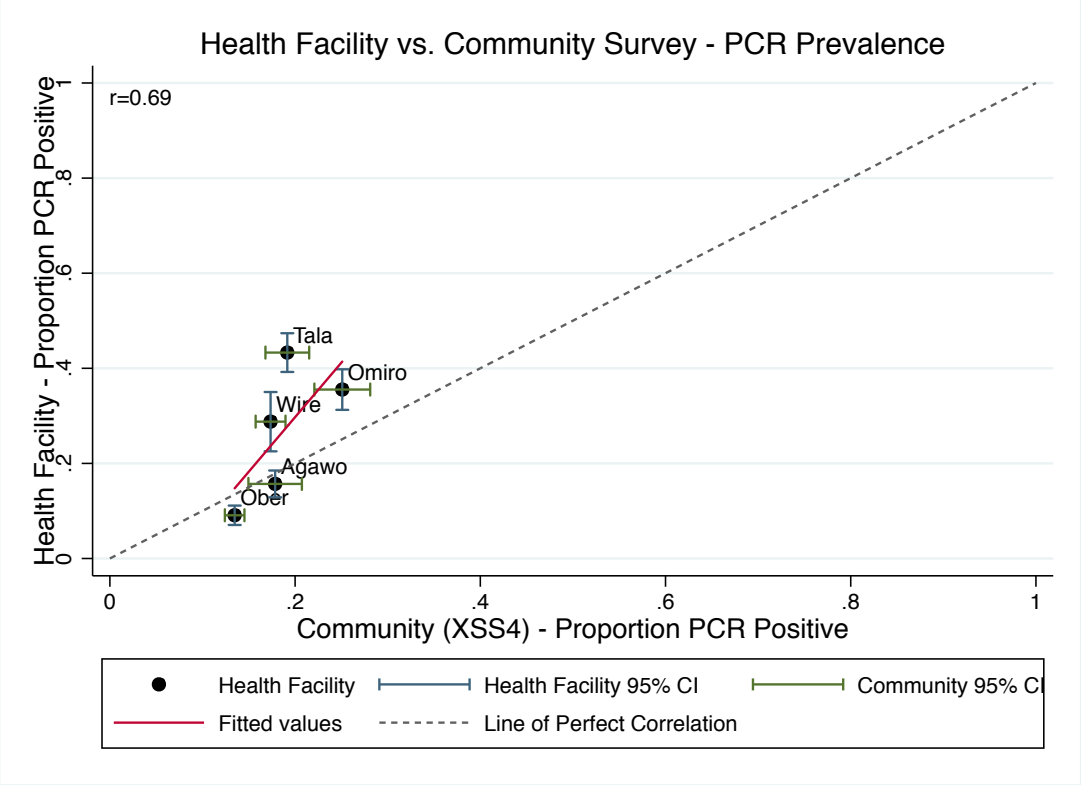

Figure 5-5: Scatter plots showing correlation of PCR prevalence from health facility and a community survey conducted in 2011 (XSS4) restricted to those residing in the health facility catchment areas with corresponding 95% confidence limits.

Comparing the health facility estimates to those of community surveys suggests that bias does exist, and the direction and magnitude of bias is dependent on the diagnostic tool used. There are limitations present in this comparison, particularly due to the small number of facilities and that the surveys being compared were conducted in different years. However, the use of RDT or seroprevalence estimates

in particular, showed strong correlation with the community suggesting health facility surveys can act as a useful convenience sample for measuring malaria transmission intensity in the community.

5.5 General Chapter Discussion

5.5.1 Overview of findings - utility of convenience samples

In this chapter, the utility of convenience samples to provide a reliable proxy metric for malaria burden in the surrounding community was explored. Despite the biases present in the convenience sampling approaches, the results from both the school and health facility surveys were strongly correlated with estimates of malaria burden obtained during community surveys. The strength of the correlation was dependent on the malaria metric used; the spatial and temporal overlap of the studies being compared; and baseline transmission intensity. For example, when comparing the health facility and community surveys, estimates using PCR showed a weaker correlation than those obtained by RDT. (49) The distribution of symptomatic individuals within the populations being compared is likely to influence this association. For example, a greater proportion of symptomatic malaria cases would be expected in the health facility sample compared to the community or school populations.

Also, SCR estimates in both convenience-sampling approaches were consistently and strongly correlated with those obtained in the community; this suggests that the school and health facility survey designs provide good estimates of transmission intensity in the surrounding community. In the school surveys, the concordance with the surrounding community improved when the school sample was restricted to those residing in the community catchment area. As expected in areas with spatially heterogeneous transmission, the exact locality of residence is of great importance in reliably estimating local transmission.

In this setting, the PCD study was able to identify heterogeneity in transmission intensity using confirmed malaria cases by either microscopy or RDT. However, the use of routinely collected data for malaria surveillance required continuous monitoring, having confirmatory diagnostics, and is consistent with findings in

other settings. For example, work from health facilities in Kenya showed that underreporting of malaria burden is common with an estimated overall reporting rate of 35% across 84 months for the 2165 facilities include in the analysis. (274) Similarly, a study in Tanzania found that despite very low transmission, the majority of suspected malaria patients were treated despite less than 2% of slides confirmed to be positive (275) Notably, the rate of testing for suspected malaria cases can be low and therefore those reported tend to be clinical cases with malaria suspected due to fever or other symptoms. The subjectivity and inconsistencies present in malaria testing suggests that relying on health facility based data to accurately estimate transmission levels, as part of a routine malaria surveillance program, may be problematic. (276) Therefore, due to the inconsistencies in reporting and quality of testing, even the most robust of analysis using passive case data tend to be inconclusive. (277)

The utility of convenience sampling to become a viable approach for malaria surveillance will rely on their acceptability by the communities where they will be implemented. Health facilities as points of disease screening and treatment are accepted in areas where there is sufficient trust in the health systems and staff capacity. Although in this setting, the patients were largely willing to partake in the testing, this may not be the case in all settings. Similarly, the use of school surveys to monitor malaria transmission or to screen for other parasitic diseases has been used in different settings (212, 232, 270) but it may not be applicable everywhere, particularly areas with low school enrollment or resistance to testing for disease by non-health care workers. Including community members in consultative processes when deciding on surveillance strategies using convenience sampling strategies is needed.

5.5.2 Biases in convenience sampling

A more detailed discussion of bias in school and health facility surveys are discussed in section 3.3, however it is important to acknowledge the key biases associated with the alternative surveys to ensure that caution is observed when interpreting the results. Firstly, the populations sampled at primary schools and health facilities are non-random subsets of the population and the extent of this

bias will be dependent on many factors including access and costs, among others. (126, 232, 256) It is expected that the parasite prevalence would be higher in the primary school and symptomatic patient population, and any accompanying person residing within the same household. (49, 134, 205) Similarly, the different age structure in the convenience samples compared to the general population also likely impact certain metrics, particularly SCR and seroprevalence. (51, 204) However, despite the sources of bias present, the utility of convenience samples as a metric for malaria heterogeneity and transmission intensity to represent a similar trend in the community appear to be able to provide good estimates of malaria burden in the community therefore making them attractive options for malaria surveillance especially considering the operational advantages.

5.5.3 Limitations

Multiple studies informed this assessment of the utility of convenience samples to measure malaria transmission in the community and each had their own specific limitations as discussed above. There were also some important limitations to the overall analysis and comparison between surveys that are worth noting. It has been found that the malaria transmission in this setting is relatively stable (appendix 1.3) However, some temporal variations in transmission are possible, which would impact the associations observed when comparing surveys that were not conducted concurrently. (129) The utility of the health facilities to provide a reliable estimate of community level transmission was part of the initial proposal, but due to the ongoing community work (appendix 1.3), it was not possible to obtain concurrent community estimates. Nevertheless, the comparison of health facility and community samples restricted to those residing in the same catchment area provided useful insights despite the lack of temporal overlap.

5.5.4 Implications for hotspot detection

The results of this work suggest that health facility and primary school surveys are able to provide reasonable estimates of malaria transmission in the community and could inform a targeted strategy at the regional level. (212) The scale of spatial data available will likely determine intervention strategies but also at what transmission intensities such approaches would likely be useful. For these convenience-sampling strategies to be useful to identify local-level hotspots of

malaria or any residual cases, it is essential to obtain more precise pictures of heterogeneity of transmission at the more local level. Without individual level spatial information, the spatial resolution of transmission estimates is restricted to those of the catchment areas of the unit of sampling. In this study, the mean distance traveled to school was 793 m and to health facilities was 2 km, which would define the level of spatial granularity that could be achieved using these sampling approaches. The utility of convenience samples to detect hotspots of infection in the community at a spatial resolution able to reflect the local level heterogeneity of malaria transmission will depend on establishing an operationally feasible way of identifying where in the community the case resides (appendix 1.1). However, the smaller catchment areas of the schools, relative to those of the health facilities, may be more useful for identifying priority areas for control while transmission intensity is still at pre-elimination levels. As transmission declines or in areas with well-established health infrastructure (ie. active community health workers) health facilities may be more useful at identifying the local pockets of transmission: therefore, the incorporation of convenience sampling strategies would provide useful additions to any local malaria control programs.

5.6 Conclusions

The main conclusions that can be drawn from the results in this chapter include:

1) School and health facility surveys provide a reasonable proxy for measuring malaria transmission in the community
2) Seroconversion rates provide the less biased malaria metric to gauge community level transmission using convenience samples
3) Routinely collected data at health facilities are able to stratify areas based on high or low transmission intensity IF data recorded is reliable and based on cases with confirmed malaria by rapid diagnostic test or quality microscopy.

Chapter 6: Results - Identifying Hotspots and Targeting the Parasite Reservoir

For a local-level hotspot targeted strategy to be incorporated into successful control strategies approaches that can be sustained by local malaria control programs are needed. This chapter examines operationally attractive strategies to identify hotspots as well as to target parasite populations within hotspots. Section 6.2 explores the utility of primary school and health facility based sampling approaches for identifying hotspots in the community (specific objective 3) while, section 6.3 discusses a method for optimizing the targeting of interventions to sub-microscopic parasite carriers in the community (specific objective 4). To address the specific objectives discussed in this chapter, data from several sources was used including the large community cross-sectional survey described in section 4.2, the schools (section 5.2) and health facility surveys (section 5.3), as well as data from a population survey in defined hotspots of malaria exposure in the community as described below in section 6.3.

6.1 Background and Rationale

Currently, programs employing malaria elimination strategies typically include a re-active case detection component. (121) Such systems are predicated on evidence that asymptomatic malaria infections cluster at the household level and it has been shown that such pockets of parasites can be identified using symptomatic cases at health facilities. (166) The re-active approach is a system that can easily be integrated into local health capacity. However, given the subjectivity and inconsistencies inherent with relying on symptomatic and patent malaria cases presenting at health facilities, other convenience sampling approaches to determine index cases could provide alternatives. The convenience sampling approaches discussed in chapter 5 suggest that these study designs provide a reasonable estimate of malaria transmission in the community and therefore it is reasonable to postulate that they can also be useful in identifying not only regional scale heterogeneity but also hotspots of transmission within communities (specific objective 3). As discussed in chapter 5, the limited spatial information obtained during the PCD study inhibited including this convenience sampling design as part of the hotspot identification analysis as was originally planned. If hotspots can be

identified, understanding what proportion of the parasite population can easily be targeted provides a useful benchmark with which to gauge the effectiveness of this approach (specific objective 4).

6.2 Hotspots and Convenience Sampling

As the convenience samples tested provide reasonable outlets for malaria surveillance, these populations may also provide an attractive means for identifying hotspots of malaria transmission, as part of a re-active case detection approach or even potentially to define hotspot boundaries. (54, 174, 180) (166)(96, 120, 121)However, the potential of these alternative sampling strategies for identifying hotspots of infection in the community has not yet been assessed. With this objective in mind, the sensitivity of primary school and health facility surveys to identify known hotspots of infection in the community was evaluated.

6.2.1 Methods

6.2.1.1 Convenience Samples

Data from the school and health facility surveys described in section 5.2 and 5.3, respectively, were used to assess any associations present with malaria positivity and residing in a hotspot of malaria transmission. Briefly, 46 government primary schools were selected and ~100 children in each school were randomly selected for inclusion in the survey which, occurred in 2010 during the high transmission season (July). All children were tested for malaria by RDT (Paracheck, Ochrid Biomedial Systems, Goa, India) and provided blood spots on filter paper (3MM Whatman, Maidstone, UK), which was tested for the presence of antimalarial antibodies to AMA1 and $MSP1_{19}$ by ELISA. (51) Subsequently, 98.5% of the 4964 children sampled were traced to the compound where they resided and spatial coordinates recorded using a hand-held device. Next, in October 2011 (low peak) and July 2012 (high peak), 5 rural health facilities were selected and all patients and people accompanying patients attending the outpatient department were recruited for the study. All participants were tested for malaria using an RDT (Paracheck –Low; First Response [Premier Medical Corporation Ltd., India] – High) and provided blood spots on filter paper (Whatman 3mm, Maidstone, UK) for testing for antimalarial antibodies to AMA1 and $MSP1_{19}$ by ELISA and the presence

of current infection by nPCR. (192, 205, 278) Subsequently, 30% of the 3042 participants were randomly selected and traced to their compound where spatial coordinates were recorded by a handheld device.

6.2.1.2 Hotspot Definition

Hotspots of malaria transmission in the community were defined as described in chapter 4 and appendix 1.3. Briefly, a large community-cross sectional survey was conducted in July 2011 where approximately 30% of the population was sampled and spatial coordinates of their compounds recorded. All samples were then assayed by ELISA for the presence of antimalarial antibodies to AMA1 and $MSP1_{19}$. (204) A locally weighted SatScan was used to identify areas with statistically significantly higher seropositivity and/or serodensity using both circular and elliptical scanning windows to provide the standard for hotspots with which to compare the sensitivity of convenience sampling to target these foci of malaria in the community. Compounds that were part of a significant ($p<0.05$) cluster by at least one scan were considered to be part of a malaria hotspot.

6.2.1.3 Data Analysis

Participants with spatial coordinates available were plotted in ArcGIS (v12.1) and those residing within the community survey study area were selected and retained for further analysis. From the subset of school children and health facility attendees that resided within the study area, those located within hotspots were then identified using the built-in join function in ArcGIS (v, 10.2, ESRI, California, USA). Data was imported into STATA (v12.0) for statistical analysis. Next, the ability of convenience sampling to target hotspots of infection in the community was assessed. The sensitivity, specificity and positive and negative predictive value was calculated. A person was considered to be correctly identified if they tested positive for malaria during the convenience sampling survey and resided in an identified hotspot of transmission in the community. True positive individuals are considered to be those who tested as malaria positive and reside within the hotspot. False negatives are those individuals who tested negative for malaria and resided within a hotspot. Conversely, true negative individuals are those whom tested negative for malaria and reside outside of a hotspot. Finally false positive cases are those who tested positive for malaria yet live outside of hotspot

boundaries. Logistic regression was conducted to identify factors associated with each of the malaria outcome measures available for each study: RDT and seropositivity for the primary school survey and RDT, PCR and seropositivity for the health facility survey. Factors assessed included residing within a hotspot, age, sex, mosquito control behaviors and other relevant epidemiological variables collected during the survey. All standard errors were adjusted for clustering at the school and health facility level.

6.2.2 Results - Convenience Samples to Target Hotspots of Malaria

6.2.2.1 Primary School Survey

Of the 4889 children traced to their compound, 1606 resided within the community cross-sectional survey study area and were retained for analysis. Of the school children residing in the study area, 708 (44.1%) resided in a hotspot of malaria transmission. Overall, 29.6% (95% CI: 27.4-31.9%) of children were RDT positive however; RDT prevalence in those residing within a hotspot was significantly higher (40.8%; 95% CI: 37.2-44.4%) compared to those not residing in a hotspot (20.8%, 95% CI: 18.2-23.5%; $p<0.001$) (figure 6-1A). The sensitivity and specificity of using RDT positivity in primary school children as a tool to identify hotspots of transmission was 40.8% and 79.2%, respectively. The positive and negative predictive values were 60.7% and 62.9%, respectively. Next, several variables were tested to identify associations with RDT positivity in the school children. Those who were RDT positive had over twice the odds of residing in a hotspot (AOR: 2.53; 95% CI: 1.64-3.92) compared to living outside of a hotspot. The only other factor retained in the adjusted model was residing in a house with open eaves (AOR: 2.15; 95% CI: 1.32-3.50) compared to closed eaves (table 6-1).

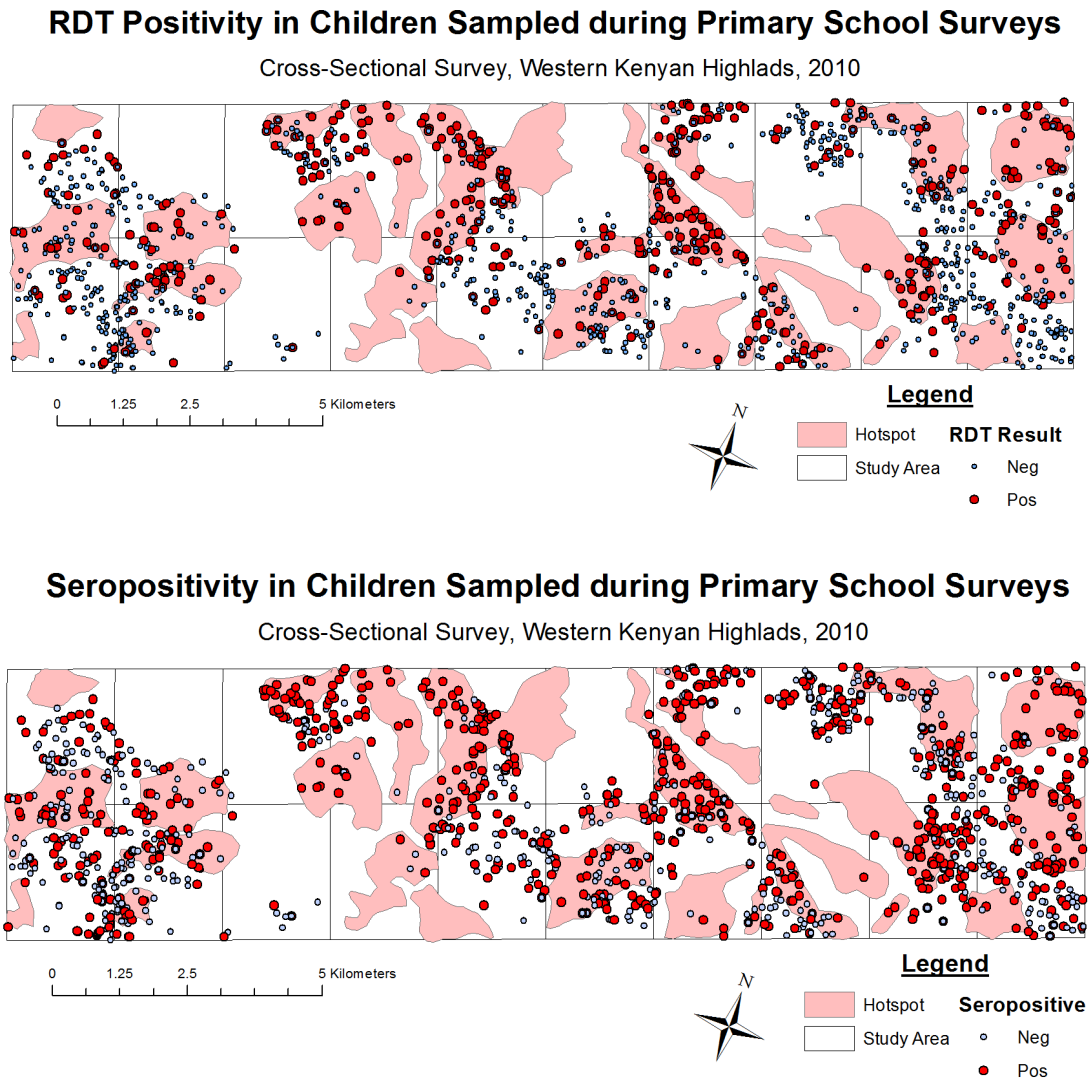

Figure 6-1: Map showing the areas in the community identified as being a hotspot of malaria transmission and the points showing the location of the residence of the children sampled during the school survey with those coloured red indicating those that tested positive by A) rapid diagnostic test and B) seropositivity and those testing negative in blue

Overall, 55.3% (52.9-57.8%) of the school children in the study area were positive for antimalarial antibodies and seroprevalence was significantly higher in children residing inside (64.1%; 95% CI: 60.6-67.6%) compared to children living outside (48.4%; 95% CI: 45.2-51.7%; p<0.001) of hotspots of malaria transmission (figure 6-1 B). The sensitivity and specificity of using seropositivity in primary school children as a tool to identify hotspots of transmission was 64.1% and 51.6%, respectively. The positive and negative predictive values were moderate at 64.1%

Table 6-1: Results of the univariate and multivariate logistic regression for RDT and seropositivity in children sampled at school that resided in the community survey study area including the association with residing in a known hotspot of infection in the community.

	RDT Positivity						Seropositivity					
	Univariate			Multivariate			Univariate			Multivariate		
	OR	95% CI	p-value	AOR	95% CI	p-value	OR	95% CI	p-value	AOR	95% CI	p-value
Reside in Hotspot	2.62	1.69-4.07	<0.001	2.53	1.64-3.92	<0.001	1.90	1.30-2.78	0.001	1.95	1.39-2.75	<0.001
Open Eaves	2.52	1.46-4.33	0.001	2.15	1.32-3.50	0.002	1.85	1.24-2.77	0.003	1.61	1.07-2.40	0.021
Gender - Female	0.87	0.72-1.06	0.173				1.13	0.96-1.34	0.133	1.25	1.06-1.46	0.007
Age	1.0	0.94-1.05	0.913				1.15	1.09-1.21	<0.001	1.16	1.10-1.22	<0.001
Fever in 24 hours	1.01	0.70-1.45	0.965				1.30	0.94-1.80	0.117			
Elevation	0.99	0.98-0.99	<0.001				0.99	0.99-1.00	0.192			
IRS in past year	0.97	0.68-1.39	0.885				0.97	0.71-1.31	0.840			
Net Use	0.88	0.68-1.13	0.319				0.70	0.50-0.98	0.038	0.70	0.50-0.99	0.041
SES												
1 (low)	1.0	-	-				1.0	-	-			
2	0.95	0.70-1.30	0.773				0.81	0.56-1.17	0.265			
3	0.91	0.69-1.21	0.540				0.77	0.47-1.28	0.315			
4	0.88	0.60-1.28	0.501				0.69	0.43-1.13	0.139			
5 (high)	0.55	0.35-0.88	0.013				0.63	0.43-0.92	0.018			
Travel 3 Months	0.97	0.62-1.52	0.883				0.99	0.74-1.33	0.978			

and 51.6%, respectively. After adjusting for age, gender, and bednet use, those children that were seropositive had almost twice the odds of residing in a hotspot of malaria transmission (AOR: 1.95, 95% CI: 1.39-2.75) compared to those residing outside of a hotspot (table 6-1).

6.2.2.2 Health Facility Survey

In total, 3042 patients and accompanying people were sampled as part of the two health facility surveys and 829 (27.2%) were traced to their compound. Of those participants with spatial coordinates available, 508 resided within the study area and were retained for analysis. 34.1% (n=173) of the health facility attendees in the study area resided within a hotspot of malaria transmission. Of those health facility attendees residing within the study area, 17.9% (95% CI: 14.6-21.2%) were positive for malaria by RDT. RDT prevalence in participants that resided in a hotspot of malaria transmission was greater (27.2%; 95% CI: 20.5-33.8%) compared to those not residing in a hotspot (13.1%; 95% CI: 9.5-16.7%, $p<0.001$) (figure 6-2A). The sensitivity and specificity of using RDT positivity in health facility attendees as a tool to identify hotspots of transmission was 27.2% and 86.9%, respectively. The sensitivity and specificity was similar when calculated for each transmission season separately (data not shown). The positive and negative predictive value of RDTs for identifying participants that resided in hotspots was 51.6% and 69.8%, respectively. In adjusted analysis, those who were RDT positive had over twice the odds of residing in a hotspot (AOR: 2.43, 95% CI: 1.85-3.19). Other factors associated with RDT positivity included having a fever at the time of sampling (AOR: 3.52, 95% CI: 1.35-9.13) and having taken antipyretic drugs in the two weeks prior to sampling (AOR: 1.84, 95 CI: 1.22-2.77) (table 6-2).

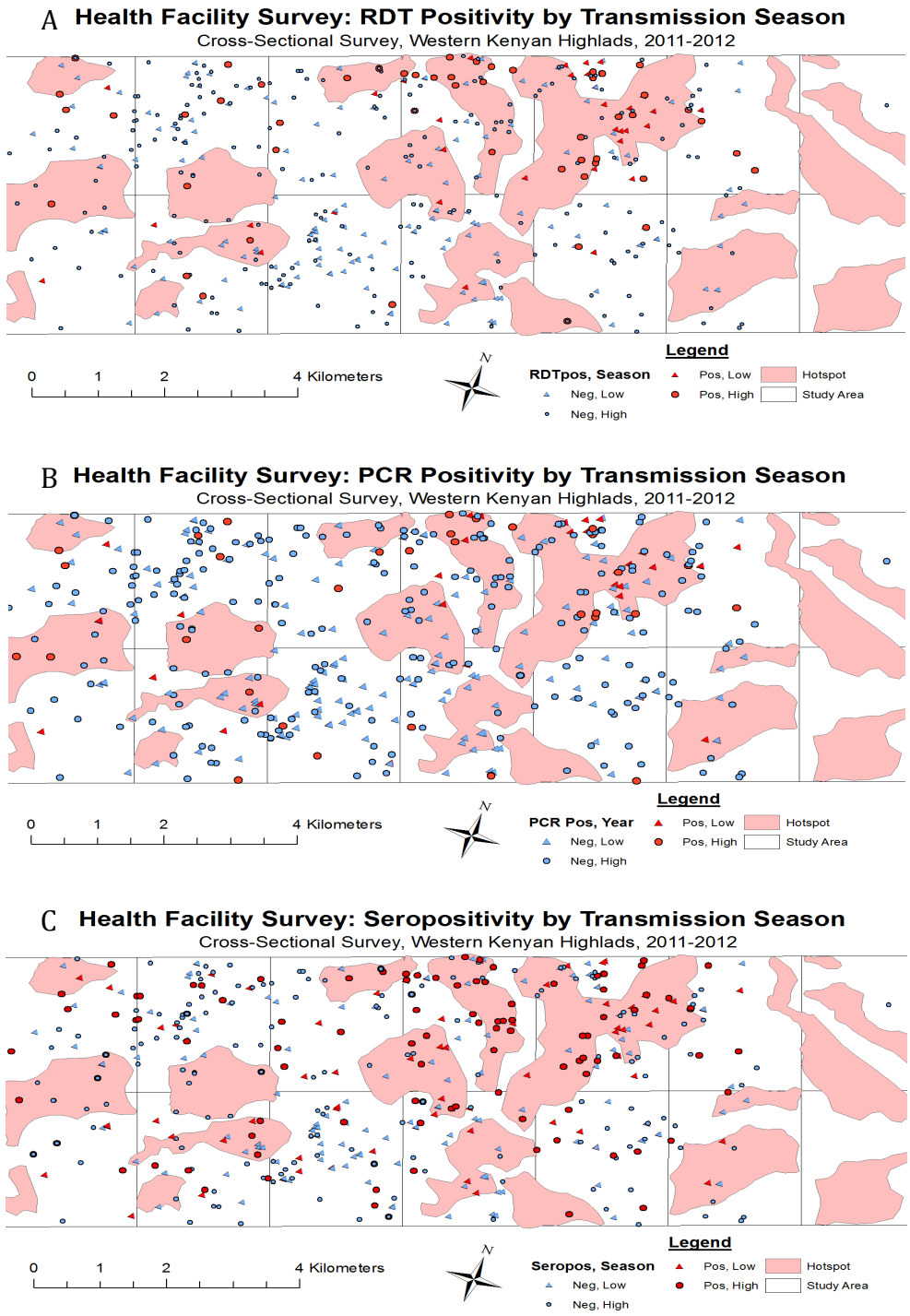

Figure 6-2: Map showing the community hotspots of malaria transmission. The points show the location of the residence of the participants sampled during the health facility survey with those coloured red indicating those that tested positive by A) rapid diagnostic test and B) PCR and C) seropositivity and those testing negative in blue. The shape of the symbol represents the year of sampling with triangles and circles identifying those sampled during the low peak and high peak transmission seasons, respectively.

Malaria prevalence by PCR was then assessed as the main outcome measure. Of the health facility attendees residing within the study area, 14.4% (95% CI: 11.3-17.4%) were positive for malaria by PCR. Similar to the RDT results, PCR prevalence in those residing in hotspots (20.2%; 95% CI: 14.2-26.2%) was significantly greater than those not residing in hotspots (11.3%; 95% CI: 7.9-14.7%; p=0.007) (figure 6-2B). The sensitivity and specificity of using PCR positivity in health facility attendees as a tool to identify hotspots of transmission was 20.1% and 88.7%, respectively. The positive and negative predictive value of PCR diagnosis for correctly identifying residence in hotspots was similar to that of RDT at 47.9% and 68.3%, respectively. Interestingly, stratifying results by transmission season had an impact with PCR being more sensitive in the high (25.3%) compared to the low (16.3%) transmission season, although the difference was not statistically significant (p=0.51), possibly due to the low sample size. Adjusted analysis suggests that residing in a hotspot of malaria transmission was the single covariate tested that showed increased odds (AOR: 1.97; 95% CI: 1.26-3.07) of being PCR positive compared to not residing within a hotspot in this health care seeking population (table 6-2).

Table 6-2: Results of the univariate and multivariate logistic regression for RDT and PCR positivity in participants sampled during the health facility surveys that resided in the community survey study area including the association with residing in a known hotspot of infection in the community.

	RDT Positive						PCR Positive					
	Univariate			Multivariate			Univariate			Multivariate		
	OR	95% CI	p-value	AOR	95% CI	p-value	OR	95% CI	p-value	AOR	95% CI	p-value
Hotspot	2.47	2.17-2.80	<0.001	2.43	1.85-3.19	<0.001	1.98	1.32-2.97	0.001	1.97	1.26-3.07	0.003
High Season	1.02	0.61-1.73	0.927				0.68	0.33-1.39	0.290			
Gender – Female	0.80	0.55-1.15	0.231				0.68	0.52-0.90	0.008			
Age	0.96	0.94-0.99	0.004	0.98	0.96-0.99	0.008	0.97	0.94-0.99	0.40	0.97	0.95-0.99	0.049
Patient	2.50	0.93-6.76	0.070				1.58	0.33-7.68	0.567			
Current Fever	4.48	1.69-11.85	0.003	3.52	1.35-9.13	0.010	2.53	0.94-6.81	0.067			
Net User	0.67	0.30-1.49	0.327				0.56	0.27-1.15	0.115			
Recent IRS	1.78	0.81-3.88	0.149				0.73	0.26-2.04	0.551			
Recent Travel	0.46	0.20-1.07	0.073	0.44	0.24-0.81	0.008	0.45	0.27-0.76	0.003	0.50	0.25-1.01	0.053
Recent Antipyretic	2.02	1.31-3.13	0.002	1.84	1.22-2.77	0.003	1.25	0.49-3.16	0.640			
Recent Antimalarial	1.35	0.75-2.43	0.323				1.35	0.86-2.14	0.193			

The last outcome measure available for analysis in this health care seeking population was the presence of antimalarial antibodies. Of the people residing in the study area, 35.4% (95% CI: 31.2-39.6%) were seropositive for malaria. Seroprevalence was significantly greater in those residing inside (43.9%; 95% CI: 36.5-51.3%) compared to those living outside of hotspots of malaria infection (31.0%; 95% CI: 26.1-36.0%; p=0.004). The sensitivity and specificity of using seropositivity in health facility attendees as a tool to identify hotspots of transmission was 43.9% and 69.0%, respectively. The positive and negative predictive values were also similar to the other diagnostic measures tested at 43.9% and 69.0%, respectively. Finally, similar to the other outcomes, residing in a hotspot was found to be a significant predictor of seropositivity (AOR: 2.11, 95% CI: 1.28-3.50) after adjusting for other risk factors such as age and mosquito control (table 6-3).

Table 6-3: Results of the univariate and multivariate logistic regression for seropositivity in participants sampled during the health facility surveys that resided in the community survey study area including the impact of residing in a known hotspot of infection in the community.

	Seropositive					
	Univariate			Multivariate		
	OR	95% CI	p-value	AOR	95% CI	p-value
Hotspot	1.74	1.20-2.51	0.003	2.11	1.28-3.50	0.004
High Season	1.11	0.71-1.76	0.640			
Gender – Female	1.47	1.10-1.96	0.009			
Age	1.03	1.02-1.04	<0.001	1.03	1.02-1.03	<0.001
Patient	0.35	0.25-0.47	<0.001	0.38	0.30-0.48	<0.001
Current Fever	0.83	0.47-1.46	0.516			
Net User	1.69	0.82-3.48	0.151			
Recent IRS	1.65	1.37-1.98	<0.001	1.51	1.14-2.00	0.004
Recent Travel	0.88	0.59-1.32	0.550	0.66	0.46-0.94	0.022
Recent Antipyretic	1.04	0.86-1.25	0.668	1.50	1.16-1.94	0.002
Recent Antimalarial	1.01	0.41-2.45	0.984			

6.2.3 Section Discussion

To determine if convenience sampling could provide a viable option for identifying hotspots of malaria transmission, the residence of participants sampled during primary school and health facility surveys was geolocated and overlaid on known hotspots of infection in the community. Residing in hotspots was significantly associated with an increased prevalence of malaria infection and measure of malaria exposure in school children and health facility attendees suggesting that convenience sampling could be used to identify and target hotspots of malaria transmission in the community (specific objective 3). Of the two convenience sampling methods assessed, school children showed a greater sensitivity in detecting hotspots of malaria transmission in the community when compared to health facility surveys. These results suggest that there may be opportunities to incorporate the use of such an approach into a hotspot targeted control program.

Malaria control strategies have traditionally made use of both school and health facilities for distribution of interventions. For example, primary schools have been used for malaria surveillance (232), for the distribution of bednets or malaria treatment (244) and for delivering malaria education campaigns. (67) Similarly, health facilities have traditionally been the point of distribution of chemoprophylaxis or bednets targeting pregnant women and children attending antenatal clinics. (258, 279) Employing convenience samples as a sentinel population and incorporating into a hotspot-targeted strategy is a logical next step as these populations, particularly school children, are likely to be most sensitive to changes in malaria transmission dynamics.

Operational research is needed to determine the best way to go about integrating convenience sampling into a hotspot-targeted malaria control strategy. Options to explore include whether a system of re-active case detection in primary schools, similar to that used in health facilities in some settings, could be employed whereby positive cases are traced to their compound where malaria control strategies or treatment campaigns could be employed. (121) For re-active case detection to be a viable option in this study area, parasite prevalence is likely to be too high as too many individuals would need to be traced to their compound to

justify such an approach. However, given that the sensitivity and predictive values of using convenience sampling to target hotspots were good, a re-active approach is promising for areas of lower transmission intensity. Also, the added benefit of screening all attendees instead of just suspected malaria cases to health facility-based re-active case detection should also be further explored and compared to the current approach of relying on confirmed malaria cases. Other avenues to explore include ways to maximize the sensitivity and specificity of hotspot targeting with convenience sampling. For example, one strategy could include spatially plotting positive cases identified during convenience samples over time and visually assess when cases aggregate to define likely hotspots of infection. Alternatively, achieving a predefined number of cases within a set distance or a defined prevalence within a village or other aggregate unit to confirm the presence of a hotspot before justifying a response could be employed. (121) Given the potential for integrating convenience sampling into a hotspot-targeted approach demonstrated in this study, further studies to explore operational research questions for how best to achieve optimum hotspot coverage in different settings and transmission intensities is warranted.

The advantages of using convenience-sampling approaches were discussed in section 5.5 however, there are some important challenges associated with translating this approach into an effective and locally sustainable hotspot targeting campaign. Most notably, the ability to obtain spatial information on residences of individuals sampled is required to identify spatial patterns of disease transmission in the community. Geolocating individuals, or obtaining spatial coordinates for their place of residence, is challenging and in most malaria endemic areas currently relies on physically locating and visiting the compound with a handheld GPS device which, is a time and labor intensive exercise. (280) However, operationally attractive strategies are available to obtain spatial information on cases (Appendix 1.1) and these could easily be adapted to facilitate a hotspot targeting control strategy. For example, locating compounds could be facilitated in primary schools by accompanying each child with confirmed malaria to their compound or a geolocation exercise could be incorporated into the curriculum through production of crude maps providing the necessary spatial reference to

inform malaria control strategies. Similarly, where community health workers or other members of the community familiar with the area are available, they could be engaged to provide or obtain spatial information on cases as is already being done in some settings. (281)

While sampling protocols between studies were kept as consistent as was feasible this was not always possible. Data for the three surveys discussed in this chapter were collected during the high transmission season providing seasonal consistency however, the studies took place over several years which may introduce some temporal bias as malaria transmission is highly stochastic and fluctuates between years. (129) However, the consistency in the results between measures of current infection and exposure and the significant associations observed despite the surveys taking place in different years suggests that the temporal bias may not have resulted in considerably different results. It is acknowledged that different results may have been observed if sampling had been conducted in the low transmission season when hotspots may be more pronounced therefore the sensitivities observed are likely to be underestimates. Secondly, during the primary school surveys children were randomly selected per class for inclusion and all were targeted for tracing to their compounds. In contrast, during the health facility surveys, which have a larger catchment area (appendix 1.1) all patients and accompanying people attending the outpatient department were targeted for inclusion while only a random subset were selected for tracing to their compounds and therefore could be included in this analysis. The differences in sensitivity and strength of association observed between primary school and health facility surveys may be associated with the different spatial densities of compounds available for inclusion in the analysis.

In conclusion, in this low endemic and highly heterogeneous setting, the use of convenience sampling could be used as an operationally attractive method for identifying and targeting hotspots of infection in the community and more robust and operationally driven studies to further tease apart how these findings can be applied by local malaria control programs as well as it's utility in other settings are needed. Similarly, identifying and defining hotspots of transmission in the

community is the first challenge while developing strategies to target the parasite populations within the hotspot must also be explored.

6.3 Operational Approach for Targeting the Submicroscopic Parasite Reservoir in Hotspots

In malaria endemic areas, a substantial proportion of infections may be at submicroscopic parasite densities making them difficult to detect using current field-based tools. (54) In hotspots of malaria transmission the occurrence of submicroscopic parasite carriers may be exacerbated due to the higher exposure experienced by these populations leading to more adept immune responses at a younger age. (127) Therefore, insight in the detectability of pockets of infections once these foci can be identified is of great importance.

6.3.1 Focal Screening to identify the subpatent parasite reservoir in an area of low and heterogeneous transmission in the Kenya highlands (P4)

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic pier review?	

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	Journal of Infectious Diseases
Please list the paper's authors in the intended authorship order.	Gillian H Stresman*, Amrish Y Baidjoe*, Jennifer Stevenson, Lynn Grignard, Wycliffe Odongo, Chrispin Owaga, Victor Osoti, Euniah Makori, Shehu Shagari, Elisabeth Marube, Jonathan Cox, Chris Drakeley, Teun Bousema * Contributed Equally
Stage of publication	Under Review

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Designed protocols and collected data in field. Contributed to laboratory analysis. Conducted all data management and analysis. Prepared manuscript.

Student Signature: ____ _____ **Date: ___17/02/2015_____**

Supervisor Signature: ____ _____ **Date: ___17/02/2015_____**

Abstract

Background: Mass screening and treatment currently fails to identify a considerable fraction of low parasite density infections while mass treatment exposes many uninfected individuals to antimalarial drugs. Here we test a hybrid approach to screen a sentinel population to identify clusters of subpatent infections in the Kenya highlands with low, heterogeneous malaria transmission.

Methods: 2082 inhabitants were screened for parasitaemia by nested polymerase chain reaction (nPCR). Children ≤ 15 years of age and febrile adults were also tested for malaria by rapid diagnostic test (RDT) and served as sentinel members to identify subpatent infections within the household. All parasitaemic individuals were assessed for multiplicity of infections by nPCR and gametocyte carriage by nucleic acid sequence based amplification.

Results: Households with RDT-positive individuals in the sentinel population were more likely to have nPCR-positive individuals (OR: 1.71, 95% CI: 1.60-1.84). The sentinel population identified 64.5% (locality range: 31.6-81.2%) of nPCR-positive households and 77.3% (locality range: 24.2-91.0%) of nPCR-positive individuals. The sensitivity of the sentinel screening approach was positively associated with transmission intensity (p=0.037).

Conclusion: In this low endemic area, a focal screening approach with RDTs prior to the high transmission season was able to identify the majority of the subpatent parasite reservoirs.

Key Words: Malaria transmission, screening, subpatent infection, elimination

INTRODUCTION

Heterogeneity of infectious agents, including malaria is apparent at all spatial scales and levels of transmission intensity although it is most pronounced where transmission is low. [1, 2] Across all levels of transmission intensity, a substantial proportion of malaria infections are asymptomatic and often present at densities below the threshold for detection by microscopy or rapid diagnostic tests (RDT) [3-7]. Whilst not associated with (acute) clinical symptoms, a proportion of these infections may progress to clinical disease [8] and can also produce gametocytes and thereby contribute to onward malaria transmission. [2, 9, 10] It has been argued that for programs to sustainably reduce or eliminate malaria transmission,

the asymptomatic and subpatent reservoir must be detected and targeted. [7, 9, 11, 12] The detectability of malaria infections in malaria endemic countries is related to malaria parasite density that is associated with acquired malaria immunity. [13-15] Consequently, infections are most likely to be detected by microscopy or RDT in children and in symptomatic infections, whilst asymptomatic adults are more likely to carry infections at subpatent densities. [4, 15, 16]

There are two commonly advocated approaches to include asymptomatic malaria-infected individuals in treatment campaigns: mass screening and treatment (MSAT) and mass drug administration (MDA). MSAT campaigns typically test all individuals, using either RDT or microscopy, and treat individuals that test positive. [17] The success of MSAT campaigns is greatly influenced by the sensitivity of the diagnostic. In low transmission settings in particular, a considerable proportion of infections is missed during MSAT campaigns because many infections are present at densities below the detection limit of the diagnostic methods commonly used. [10, 18, 19] Onward malaria transmission from these subpatent infections was considered the most plausible explanation for a recent failure of RDT-based MSAT campaigns to sustainably reduce malaria transmission in the pre-elimination setting of Zanzibar. [7]

Community-based mass drug administration (MDA) campaigns avoid the problem of imperfect diagnostics by treating without prior diagnosis. However, MDA in low endemic settings would administer medication to individuals whom are not infected with malaria nor will have any benefits of the prophylactic effect of drugs due to the low exposure. [20, 21] Based on the limited success of MDA approaches under research conditions and the risk of increasing drug pressure that is associated with the spread of drug resistant strains of parasites, [20, 22] the use of MDA in malaria has received limited support. [1, 21] Alternative strategies are required that are capable of targeting the entirety of the parasite population while being operationally feasible in malaria endemic communities.

Malaria infections are known to cluster at the household level and it has been shown that asymptomatic parasite carriers are more likely to reside in households

when a symptomatic case occurs in the same household. [17, 23-25] For example, in Zambia, it was found that prevalence of malaria in households with a symptomatic case was 8.0% compared to <1.0% in households without a symptomatic case. [25] Similarly in Senegal, it was found the risk of being parasite positive was more than three times higher when residing in a household with a symptomatic case. [24] This suggests that a hybrid approach in which focal mass drug administration is guided by the occurrence of positive (index) cases detected by screening of a sentinel population may represent an efficient method of maximizing the number of infections treated whilst limiting the total number of antimalarials distributed and thereby drug pressure. [12, 17, 22] We aimed to determine the potential of this approach and identify the most appropriate definition of a sentinel population that balances the number of individuals screened against the proportion of the parasite reservoir identified and to ascertain factors associated with its sensitivity.

METHODS

Study area

This study was undertaken in a previously described study site in Rachuonyo South District, western Kenyan highlands [34.75 to 34.95°E, 0.41 to 0.52°S] with elevation between 1400-1600 m. The landscape is intersected with rivers and rolling hills and is characterised by marked variations in elevation within a small area. [26] Malaria transmission intensity is generally low but is highly heterogeneous. [27] *Plasmodium falciparum* is the predominant malaria parasite and transmission follows a bimodal pattern associated with the peaks in rainfall. Five areas within this 100 km^2 area with evidence of on going malaria transmission [26] were selected for the current study (figure P4-1).

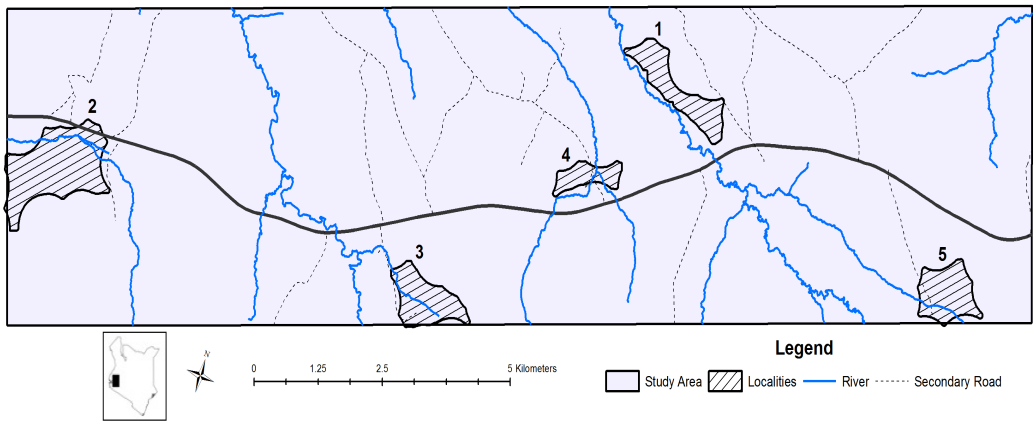

Figure P4-1: Study Area Map of study area showing the main roads (black lines) and rivers (blue lines). The five selected localities where this survey occurred are shown.

Ethical Review and Approval

The study was approved by the ethical committees of the London School of Hygiene & Tropical Medicine (Ref: LSHTM 5956) and the Kenya Medical Research Institute (Ref: SSC 2163 & SSC 2495). Individual informed consent was sought from all eligible participants. Consent for children under the age of 18 was provided by a parent/guardian and children between 14 and 17 years also provided written assent. Participants below 18 years of age who were pregnant, married, or a parent were considered "mature minors" and consented for themselves. [28]

Data Collection

All residents were enumerated and households were assigned spatial coordinates with handheld global positioning system receivers (Garmin 62S; Garmin International, Inc., Olathe, KS, USA). In March 2012, prior to the main malaria transmission season [26], all households were visited and information obtained on standard malaria indicators and socio-economic factors. Tympanic temperature was measured (Braun Thermoscan, Braun, USA); those with a temperature >37.5 °C were considered febrile. All individuals between 6 months and 15 years as well as febrile adults were tested for malaria infections using a RDT (First Response, Premier Medical Corporation Ltd., Kachigam, India). This definition of the sentinel

population was based on previous evidence that these groups have the highest proportion of infections with detectable parasite densities.[4, 15, 29] All RDT-positive cases were provided treatment according to national guidelines. Blood spotted on filter paper (Whatmann 3MM, Maidstone, UK) was collected from all consenting participants ≥ 6 month of age and stored at room temperature. For gametocyte detection, 100 µl of whole blood in nucleic acid stabilizer (Angora buffer, Avantor Performance materials, Deventer, the Netherlands) was collected from all individuals in three of the five localities and stored for up to one week at -20 ˚C and subsequently at -80˚C.

Laboratory Analysis

Filter paper samples were analysed for malaria infection using nested polymerase chain reaction (nPCR) targeting the *P. falciparum* 18S rRNA gene [30, 31] after Chelex-saponin extraction. [30] All nPCR-positive samples were tested for the presence of multiple clonal infections based on the Merozoite Surface Protein-2 (MSP2) using capillary electrophoresis; [32] samples were analysed with Peak Scanner (Applied Biosystems, CA, USA, version 1.0) and unique clones were determined to be any discrete peaks greater the background noise for each plate. For the three localities where whole blood samples were collected, total nucleic acids were extracted for all nPCR-positive samples using the MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche, Switzerland) on an automated extractor (MagnaNA Pure LC 2.0). The presence of gametocytes was determined by detection of gametocyte specific PfS25 mRNA by Nucleic Acid Sequence-Based Amplification (NASBA). [33]

Statistical Analyses

The sensitivity and specificity of different sentinel definitions was determined at the household level (i.e. the proportion of all households with nPCR detected infections that was correctly identified) and the individual level (i.e. the proportion of all individuals with nPCR detected infections that was correctly identified). Five sentinel populations for RDT screening were defined: i) all household occupants ≤5 years; ii) all occupants ≤15 years; iii) all occupants ≤5 years and any febrile individual in other age groups; iv) all occupants ≤15 years and any febrile

individual in other age groups; v) only febrile individuals. If at least one individual in this sentinel population was found to be parasitaemic by RDT, all household members were considered 'infection positive' (and therefore eligible for treatment). In addition, the sensitivity of a sixth approach was determined in which RDTs were not used but all household members were considered infection positive if there was ≥ 1 febrile household member.

Analysis was conducted using STATA (v12.0, STATA Corps, Texas USA) and R (v. 3.0.2, The R Foundation, Boston, USA). Principal component analysis was used to determine socio-economic status for each household and resulting scores were divided into quintiles.[34] Buffer zones of 50, 100, and 250 m around each household were calculated using ArcGIS (version 10.2, ESRI, California USA). The mean number of allelic types present in each infected individual was determined and corresponding 95% CI were calculated assuming a zero-truncated poison distribution. Logistic regression was used to assess associations with RDT positivity in the nPCR positive sentinel population adjusting for clustering within localities. A finite population correction factor was applied to the standard error for all statistics and the corrected 95% confidence intervals (CI) were calculated.

RESULTS

In total, 2082 individuals were sampled in 401 households (locality range [range]: 233-635), representing 94.2% (range: 90.8-100.0%) of all households (table P4-1). There was no significant difference in age, proportion of females, or reported recent travel between the localities. Bednet use the previous night was reported by 71.5% of participants (range: 53.4-77.4%). Overall, 1203 individuals were screened for malaria by RDT based on their age (≤ 15 years, n=1158) or febrile status (n=45) and 24.9% (95% CI: 24.3-25.5%; range: 6.7-48.6%) were RDT-positive.

Of all participants, 23.5% (95% CI: 23.1-24.0, range: 11.7-38.9) were parasitaemic by nPCR, the mean number of allelic forms per infection (MOI) was 2.22 (95% CI: 2.18-2.25, range 1.61-2.61) and 65.0% (n=249; Range: 64.0-84.6%) of the 383 nPCR-positive individuals tested harboured gametocytes. Parasite prevalence

Table P4-1: Population Demographics. Demographics of the study population including the number of people sampled overall, the range of values per locality and for parasite metrics, the 95% confidence interval.

	Mean	Locality Range
Population Characteristics		
Households Sampled (%)	94.2	90.8 - 100.0
N Sampled	2082	233 - 635
Sex (% Male)	45.3	43.1 - 47.3
Reported Net Use	71.5	53.4 - 77.4
Parasite Metric		
nPCR prevalence	23.5	11.7 - 38.9
<5	17.3	9.8 - 32.1
5-15	35.3	14.0 - 59.5
>15	16.4	8.2 - 25.5
MSP2 – MOI*	2.22	1.61 - 2.61
<5	2.24	1.33 - 2.60
5-15	2.34	1.54 - 2.81
>15	1.95	1.72 - 2.23
* N=489 nPCR-positive		

(35.6%, 95% CI: 34.8-36.4 vs. 12.3%; 95% CI: 11.7-12.9%; p<0.001) and MOI (2.33, 95% CI: 2.28-2.38 vs. 1.95, 95% CI: 1.86-2.04; p<0.001) were significantly higher in the 5-15 year old population compared to those 16 years of age and older, respectively. Parasite prevalence in 5-15 year olds was also significantly higher compared to those <5 years (17.3%, 95% CI: 16.3-18.2%; p<0.001) but MOI was not significantly different (2.24, 95% CI: 2.21-2.34; p=0.105). For individuals tested by RDT, MOI was significantly greater in patent infections (2.48, 95% CI: 2.42-2.53) compared to subpatent infections (1.95, 95% CI: 1.87-2.02; p<0.001). Of

all nPCR-positive children ≤15 years old, 29.9% (95% CI: 28.7-31.1%) had subpatent infections.

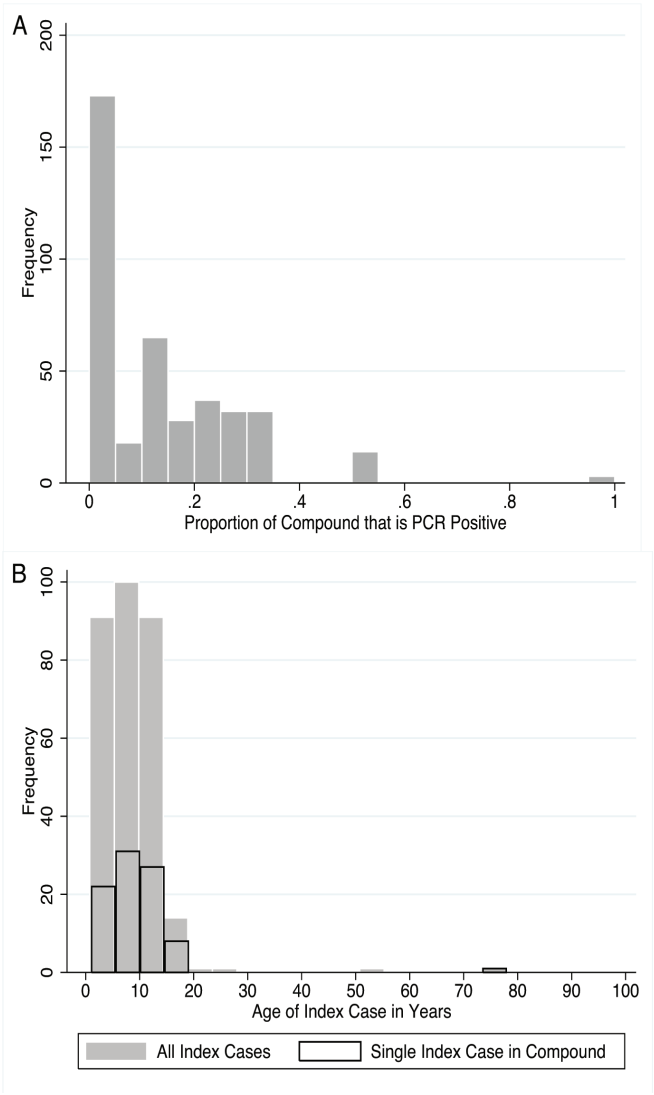

Figure P4-S1: Parasite prevalence per household (A) and age characteristics of index cases (B) A) The frequency distribution of the proportion each sampled household that was positive for malaria parasites by nPCR. B) Frequency distribution of the ages of the RDT-positive index cases for the entire sentinel population as well as for the households where only one RDT-positive individual was found.

Identifying parasite positive households through sentinel populations

Overall, 42.4% of the 401 households surveyed had no infections and there were 8 households (2.0%) where all members were nPCR-positive (supplementary figure P4-S1A). Individuals who were RDT-negative or not screened by RDT were significantly more likely to be nPCR-positive if there was a RDT-positive individual in their household (OR: 1.71, 95% CI: 1.60-1.84); the odds of being nPCR-positive increased with the number of RDT-positive individuals within a household (table P4-S1). Of the six definitions of sentinel population, testing those ≤15 years or febrile adults achieved the highest sensitivity and specificity for detecting the parasite reservoir (table P4-2): at the household level, sensitivity was 64.5% (range: 31.6-81.2) with a specificity of 90.6% (range: 82.3-94.9); at the individual level sensitivity was 77.3% (range: 24.2-91.0) and specificity was 55.7% (range: 31.3-85.1) (table P4-2).

Table P4-2: The sensitivity and specificity of sentinel populations to detect the parasitaemic reservoir. The sensitivity and specificity of different sentinel populations at the household level (i.e. the proportion of all households with nPCR detected infections that was correctly identified) and at the individual level (i.e. the proportion of all individuals with nPCR detected infections that was correctly identified). All household occupants in the sentinel population were screened and all household members were considered infection positive if one individual in the sentinel population recorded a positive result (either RDT or fever). The approach with the highest sensitivity is indicated in bold.

Definition of Sentinel Population	RDT test ≤5 years		RDT test ≤15 years		RDT test ≤5 years and any febrile cases		RDT test ≤15 years and any febrile cases		RDT test febrile cases		No RDT, febrile cases	
	%	Range	%	Range	%	Range	%	Range	%	Range	%	Range
Household Level												
Sensitivity	26.0	10.5-36.5	64.1	31.6-80.0	28.1	10.5-41.2	**64.5**	**31.6-81.2**	6.5	0-10.6	13.8	5.3-15.0
Specificity	95.3	93.9-97.4	90.6	82.3-97.4	95.3	93.9-97.4	**90.6**	**82.3-94.9**	99.4	97.4-100.0	91.8	87.2-95.8
Individual Level												
Sensitivity	39.1	9.1-51.3	76.9	24.2-90.1	41.7	9.1-56.3	**77.3**	**24.2-91.0**	10.4	0-13.9	18.0	12.1-
Specificity	81.3	71.3-93.9	55.8	31.6-85.1	81.8	71.4-93.9	**55.7**	**31.3-85.1**	95.1	91.9-98.3	86.0	81.1-90.7

Table P4-S1: Household level clustering of nPCR-Positive Individuals. The odds of being nPCR-positive in those who were not RDT-positive in households where RDT-positive infections were identified stratified by the number of RDT-positive individuals identified in the household.

No. RDT-Positive In Household	Odds nPCR-Positive	95 % Confidence Interval
0	1.0	-
1	1.47	1.36-1.60
2	1.71	1.54-1.90
3	2.34	2.02-2.70
4	3.64	2.97-4.46
≥5	3.51	2.74-4.49

Correctly and incorrectly classified households appeared evenly distributed throughout the area with little variation between the best (locality 1) (figure P4-2A) and worst performing locality (locality 5) (figure P4-2B). Because 82 households with 244 nPCR-parasite positive individuals were not identified as infection positive, we determined the impact of extending the focal treatment response to include buffer areas around RDT-positive sentinel cases on sensitivity and specificity. Based on the 94.2% of eligible compounds sampled, the median distance for households that were incorrectly classified as parasite-free (parasite positive individuals by nPCR but not by RDT in the sentinel population) from the closest household with an RDT-positive sentinel case was 85.1 m (IQR: 56.9 – 147.3 m). The addition of buffer zones around targeted households improved the sensitivity, but the specificity was greatly reduced beyond 50 m and the addition of a 150 m buffer resulted in the inclusion of nearly every household as infection positive; thereby resulting in an approach similar to MDA if this definition was used to target antimalarial drugs (figure P4-S2).

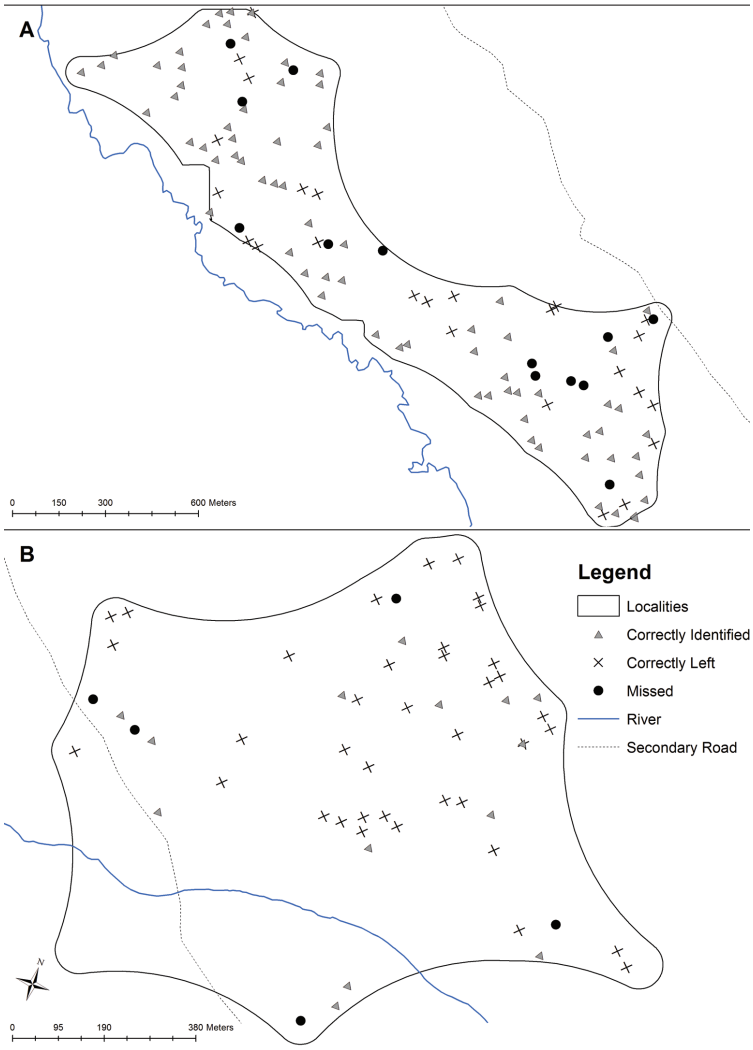

Figure P4-2: Maps of FMDA Sensitivity. Locality with the A) highest and B) lowest proportion of the parasite reservoir identified using the FMDA approach. Dots represent each household screened as part of this study including those households that had nPCR-positive individuals but were missed by the best definition of sentinel population (black circle), those that were correctly identified as infection positive (grey triangle) and those that were correctly classified as infection negative (black cross).

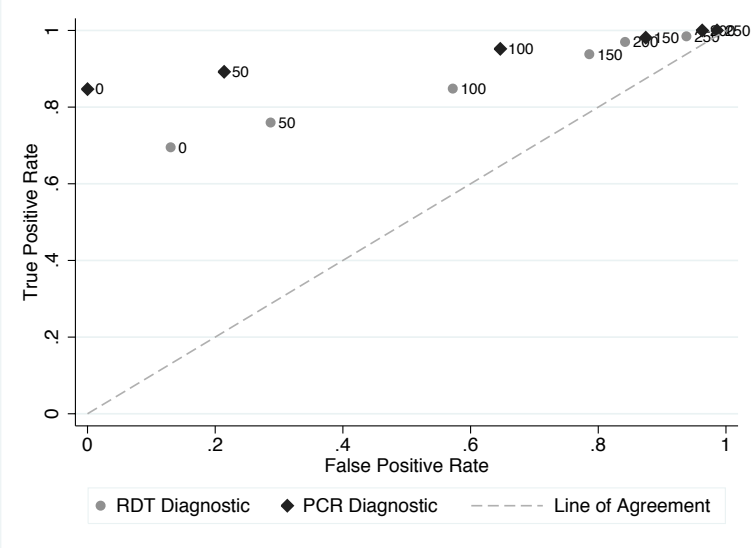

Figure P4-S2: Impact of including all households located within a defined buffer distance (numeric labels) around households identified as infection positive. The true and false positivity rate for coverage of the parasite populations was estimated at each buffer distance (50 – 250 m) if positive households were considered those with an RDT- (light grey circle) or nPCR- (dark grey diamond) positive case in the sentinel population.

Factors associated with parasite carriers being correctly identified vs. missed

The number of RDT-positive individuals in the sentinel population ranged between 1 and 12 per household and larger households were more likely to have RDT-positive cases ($p<0.001$). In the 165 households with RDT-positive individuals, 52.1% had one and 31.5% had two RDT-positive individuals. The majority of RDT-positive individuals were under 15 years old (96.7%), but the age ranged from 6 months to 82 years (figure P4-S1B). Of the households with a single index case, the median age of the RDT-positive individual was 13 (range: 6 months to 82 years). Overall, PCR prevalence (35.0%, 95% CI: 34.3-35.6% vs. 11.1%, 95% CI: 10.7-11.6%; $p<0.001$) and MOI (2.35, 95% CI: 2.31-2.40 vs. 1.76, 95% CI: 1.71-1.81; $p<0.001$) was higher in correctly identified compared to missed households, respectively. Sensitivity was associated with the average nPCR parasite prevalence

in the locality and with the proportion of infections in the sentinel population that were subpatent (table P4-3).

Table P4-3: Locality level factors associated with screening approach sensitivity. Parasite and demographic data per locality ordered by the sensitivity at household/individual level for the optimum definition of sentinel population (testing those ≤ 15 years and febrile adults).

Locality	1	2	3	4	5	p-value
N Households	109	120	73	42	57	-
N Individuals	571	635	365	233	278	-
Sensitivity - nPCR						
Household	81.2	66.1	55.0	44.0	31.6	-
Individual	91.0	76.2	69.9	61.5	24.2	-
Parasite Metrics						
nPCR prevalence (95% CI)	38.9 (38.0-39.9)	19.4 (18.6-20.1)	20.0 (19.1-21.0)	16.8 (15.6-18.0)	11.7 (10.8-12.6)	0.037
Estimated microscopy parasite prevalence [4] (95% CI)*	16.5 (13.6-20.0)	6.1 (4.6-7.9)	6.3 (4.6-8.7)	5.0 (3.3-7.5)	3.2 (2.0-5.0)	-
Subpatent infections in sentinel population (%) **	39.2 (37.6-40.7)	48.4 (46.2-50.5	65.7 (63.1-68.4)	59.0 (55.2-62.7)	81.8 (78.6-85.0)	0.037
MOI (95% CI)	2.61 (2.55-2.67)	1.89 (1.83-1.96)	1.88 (1.79-1.96)	2.32 (2.20-2.45)	1.61 (1.51-1.72)	0.188
Demographic Indicators						
Mean altitude (range)	1449.0 (1417-1478)	1422.8 (1396-1443)	1495.9 (1458-1518)	1465.1 (1447-1477)	1535.2 (1512-1560)	0.104
Economic status, % in lowest SES quintile	19.1	12.6	14.7	18.2	23.5	0.505
Age, % ≤15y	54.5	57.7	54.2	57.9	53.5	0.391
Fever, % febrile individuals	2.6	2.4	2.5	1.7	2.5	0.492

* Microscopy parasite prevalence was estimated based on nPCR data based on a published mathematical model to illustrate transmission intensity in the different localities; statistical testing was not performed since this was a derived variable.

** nPCR positive infections that were RDT negative

In the households where parasite carriers were missed due to no RDT-positive test result, 43.5% (95% CI: 47.5-51.7% range: 25.0-73.3) of nPCR-positive individuals were ≤ 15 years of age (figure P4-3), indicating subpatent parasite carriage in the sentinel population. The odds of individuals in the sentinel population being correctly identified increased if they reported fever in the past 24 hours (Adjusted Odds Ratio [AOR] 1.56, 95% CI: 1.25-1.95); had higher temperature (AOR 1.81 [per °C], 95 CI: 1.58-2.07) and; had a greater number of parasite clones (AOR 1.26 [per clone], 95% CI: 1.16-1.37) (table P4-4). Whereas, females (AOR 0.73, 95% CI: 0.57-0.92) and those reporting having taken antimalarial drugs in the two weeks prior to the survey (AOR: 0.69, 95% CI: 0.53-0.90) were more likely to have a subpatent infection and therefore be missed.

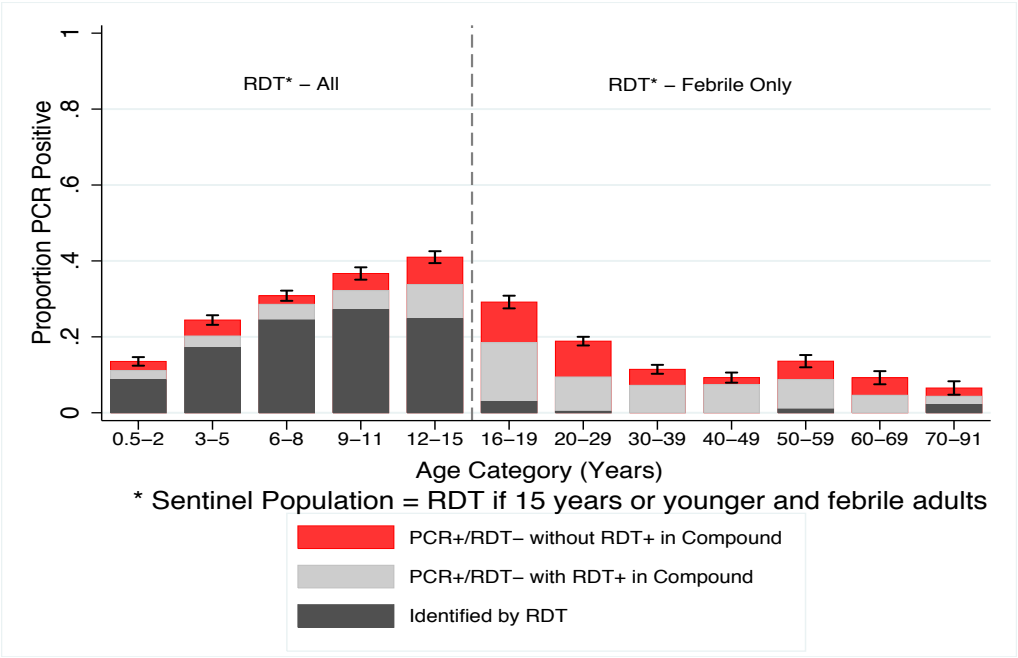

Figure P4-3: Detectability of infections in the sentinel age population. Prevalence of nPCR infection by age group in the sentinel population (aged 0.5-15 years and febrile adults). Bars indicate whether these infections were detected by RDT (black) or whether these were RDT-negative but present in households with RDT-positive individuals (light grey) or RDT-negative without RDT-positive household members (dark grey). Error bars indicate the 95% confidence interval for total nPCR parasite prevalence by age.

Table P4-4: Individual factors associated with RDT-positivity in nPCR-positive individuals. Results of logistic regression to identify factors associated with RDT-positivity in nPCR-positive individuals in the sentinel population, adjusted for clustering within localities.

	Univariate			Adjusted		
	OR	95% CI	p-value	AOR	95% CI	p-value
Gender						
Male	1.0	-	-	1.0	-	-
Female	1.06	0.99-1.13	0.095	0.73	0.57-0.92	0.008
Fever in preceding 24 hours	1.66	1.52-1.82	<0.001	1.56	1.25-1.95	<0.001
Bednet use previous night	0.89	0.82-0.96	0.005	2.31	1.86-2.87	<0.001
Open Eaves	1.88	1.62-2.19	<0.001	1.82	1.33-2.49	<0.001
Temperature in °C	1.44	1.36-1.53	<0.001	1.81	1.58-2.07	<0.001
Reported use of antimalarials in the preceding 2 weeks	1.11	1.03-1.19	0.003	0.69	0.53-0.90	0.007
Complexity of infection, mean number of MSP-2 clones	1.28	1.23-1.34	<0.001	1.26	1.16-1.37	<0.001

CONCLUSIONS

Including the asymptomatic and subpatent parasite reservoir of infection in control measures is predicted to considerably augment efficiency [4, 7] but sensitive and operationally attractive strategies to identify these individuals are needed. Here, we determine the value and limitations of a viable operational

approach in which individuals who are most likely to harbour parasite densities detectable by conventional diagnostics (ie. in this setting, children and those with fever), are screened by RDT to identify foci of subpatent parasite carriage in residents of the same household.

Focal MDA campaigns have been used in areas of heterogeneous malaria transmission [35] but there have not been any attempts to determine the value of this approach in guiding household-level treatment where considerable clustering of malaria infections is likely. [36] The presence of RDT-positive individuals in the sentinel population was highly predictive of nPCR prevalence in individuals who were RDT-negative or not screened by RDT, confirming household-level clustering of malaria. [17, 37] We showed that the majority of individuals that were nPCR-positive for asexual parasites also had concurrent gametocytes; illustrating their potential role in onward malaria transmission. [38]

We showed that in our setting, screening those ≤15 years and febrile adults with a conventional RDT identified over 75% of the patent and subpatent parasite infections while minimizing the administration of antimalarial drugs to non-infected individuals. If we had conducted focal MDA in all identified households, only one-third of all uninfected individuals would have received treatment. This approach would have been considerably more sensitive than strategies used in Zanzibar and Burkina Faso where infection was detected at an individual level by RDT and no attempt was undertaken to target subpatent parasite carriage in household members of RDT positive individuals. [7, 39] By comparison, if a full MDA approach had been used in our setting, three quarters of the total population would have received treatment despite not having a current infection. There are known risks and expenses associated with overtreatment and the use of antimalarial drugs should ideally be targeted to those with infection or at risk of infection. [20, 22]

Our approach to screen individuals most likely to have infections at densities detectable by RDT [40, 41] did not result in the detection of all parasite positive households or individuals. Although RDT screening of all age groups might be

advocated, recent studies have shown that this approach is unlikely to result in complete uptake [19] and will not detect all infections. [39] Furthermore, individuals with subpatent infections in the sentinel population were more likely to be younger and have less complex infections suggesting that the infections that were missed had lower parasite densities. Although the association of subpatent parasite carriage with reported drug use suggests that a fraction of RDT- and nPCR-positive individuals may be older infections with persisting gametocyte populations [42], this is unlikely to have affected our main outcomes and the HRP-2 based RDT is likely to have a similar issue with positive results after clearance of asexual infections. The risk of missing infections due to fluctuating parasite densities and the single time-point of sampling was also minimised by use of an HRP-2 based RDT. [43]

In this study it was striking that there was a high prevalence of subpatent infections in the youngest age groups;[7] particularly in localities with the lowest average parasite prevalence. These findings indicate that even in low endemic settings and young age groups, molecular or alternative diagnostics may be required to detect all parasitaemic individuals. [4, 7, 18] If we had used nPCR to test the sentinel population (≤ 15 years and febrile individuals), we would have achieved a sensitivity of 89.2% and a specificity of 50.9% to detect all nPCR-positive individuals in our study setting. It is currently unknown what coverage of the parasite populations is needed to achieve sustainable reductions in transmission. In addition, it is unknown to what extent that the impact of our screening approach may have been maximized by an iterative approach where repeated screening with RDTs followed by focal MDA may progressively reduce the parasite biomass in the population. However, our findings indicate that any screening and treatment approaches to reduce malaria transmission would benefit from field-deployable molecular diagnostics.

6.4 General Chapter Discussion

Overview of Findings – operational strategies for targeting hotspots and submicroscopic infections

These results suggest that the operationally attractive sampling strategies tested, particularly the use of primary school surveys could provide additional benefits for malaria control programs in terms of using confirmed malaria cases to target hotspots of malaria transmission in the community. Such a re-active case detection approach is being implemented using cases presenting at health facilities, however this work suggests that schools or all attendee health facility surveys should also be assessed as a possible alternative approach. (121) The use of sentinel populations that are likely to harbor infections that are detectable by microscopy or RDT can help identify households with subpatent infections in the community once hotspots are identified. The use of sentinel populations has the potential to supplement hotspot targeted control programs by improving the precision of where interventions are targeted thereby offering a more effective use of resources.

Limitations

There are important limitations to this work. As discussed, the use of convenience sampling for hotspot targeted control strategies is reliant on operational approaches to geolocate the residence of the each participant, which inhibited the use of the PCD data for the hotspot identification analysis. The current need for tracing individuals to their compounds to obtain precise spatial information has the potential to limit the attractiveness of convenience sampling for hotspot detection. However, the importance of this limitation is dependent on the objectives of the program and the desired spatial resolution. Spatial heterogeneity and hotspots of malaria are more likely to be apparent in areas where transmission is low. (106) Therefore, the smaller number of cases that would need to be geolocated and, given that the index case could assist in locating their residence, minimizes the impact of this limitation and increases the utility of convenience sampling to target hotspots. Furthermore, as people are traced, a database of spatial information could be created thereby making future geolocation exercises more efficient.

Next, all of the surveys used to inform the definitions of hotspots and convenience samples were cross-sectional studies. It is known that some hotspots are consistently detected over time whereas others are temporary. (129) Using seropositivity to define hotspots of infection in the community incorporates historical exposure (205) however, non-permanent hotspots may be less likely to be identified using the SatScan approach as individuals within these areas may have lower average antibody levels and therefore be less likely to be seropositive using the population based cut-off inherent in the mixture models used. (282, 283) These temporary hotspots are more likely to be detected by measures of current infection such as RDT, PCR, and clinical indicators or potentially a combination of metrics adjusted for age may be ideal. (127) Similarly, the convenience sampling was not conducted in the same year as the community survey, which was used to define the 'true' hotspots. It is possible that some of the positive cases observed during the school or health facility surveys originated from a hotspot that was not present or detected during the community survey. Therefore, the sensitivity of convenience sampling to target hotspots is likely to be an underestimate.

Implications for malaria control programs

Policies in many malaria elimination and pre-elimination setting have incorporated a system of re-active case detection using confirmed cases at health facilities as index cases to identify potential clusters of infection in the community. (96, 121) The research on convenience sampling and hotspots suggest that the use of school surveys may provide a more sensitive means with which to target clusters of parasite carriers in the community that is also a financially attractive option. (232, 284) However, for the use of school-based surveys to be incorporated into control programs there is a need to validate this strategy in other settings and to identify at which transmission settings such an approach can be effective.

Some important operational questions should be further explored to identify the best ways to incorporate such a strategy into routine malaria control programs. Firstly, the utility of reactive case detection relies on parasites clustering within the compound of index cases. A subsequent important question is whether a pre-

defined area around each index case should also be included in a response and if that response should be a mass treatment campaign, or if the FMDA approach would be useful in this setting.

Conclusions

The main conclusions that can be drawn from the results in this chapter are:

1) Individuals testing positive for malaria by all metrics during convenience sampling approaches are associated with residing in a hotspot of malaria transmission;
2) The community-based focal strategy identified a substantial proportion of the parasite population but still missed over one-fifth of infections;
3) School surveys may be robust alternatives for targeting hotspots of malaria transmission in the community.

Chapter 7: Discussion

This thesis has used data collected in the western Kenyan highlands to assess operationally attractive approaches for identifying hotspots of malaria infection for subsequent targeted interventions. This chapter provides a general discussion, highlights the main findings, and discusses future research directions. Section 7.1 provides a general overview of the potential role of malaria hotspots in the context of malaria elimination and a summary of the principal findings of this research are discussed in section 7.2. Finally, the implications of the research findings and future directions are discussed in section 7.3.

7.1 Research in context

Malaria is still a major public health burden with at least 0.5 million deaths per year. (8) However, substantial progress has been made in the last decade with one third of endemic countries now in elimination or pre-elimination states. (12, 95) As discussed in section 1.1.5, when the malaria burden declines in an area, the transmission dynamics shift; therefore control strategies must adapt so decision-making is relevant to the new state. Importantly, once transmission achieves a low level, malaria becomes increasingly focal with certain high-risk populations and spatially defined areas experiencing a disproportionate amount of the malaria burden. (121, 127) Eliminating and maintaining a malaria free state is expected to be challenging in malaria endemic countries with the current tools available as well as the constant risk of re-importation of the parasite into an area that is conducive to malaria transmission. (222) If these spatial patterns of transmission can be identified by local malaria control programmes for targeting with interventions, it could provide an opportunity to have an economically attractive and more sustainable impact on reducing malaria. (97, 222)

Currently, identifying hotspots of malaria transmission has largely been supported by either clinical (121, 139) or community based data. (152, 282) Current studies have primarily focused on demonstrating the presence of spatial clustering of malaria infections. (9, 110, 127) The next steps are now to focus on methods for applying this information and integrate the spatial dynamics of malaria transmission into effective control and elimination programs that allow the

assessment of the impact of targeted interventions. Although some malaria control programs have adopted re-active case detection as part of their targeted elimination strategies, little research has been conducted on how best to implement such a program or what kind of coverage is achieved or is necessary to achieve. Similarly, it is currently unknown if it will be sufficient to rely on using clinical cases to target local level clusters or if other convenience sampling approaches may provide alternative and viable options.

The aim of this study was to address this need and explore ways to identify hotspots in the community and particularly, the role of convenience sampling to provide an alternative to community based approaches for targeting hotspots of malaria transmission.

7.2 Overview of findings

Heterogeneity in malaria transmission is known to exist at all spatial scales. (127) To date, there has been little consistency in the approaches used to define regional-level heterogeneity, or determine the presence of local-level transmission hotspots. However, the operational feasibility of the methods by which data are generated decreases as the desired granularity increases (figure 7-1). When heterogeneity at different scales can be characterized and utilized, malaria interventions can first be prioritized to regions of higher transmission intensity and, once transmission has been reduced, focus on local areas of residual high transmission intensity to accelerate the path to malaria elimination.

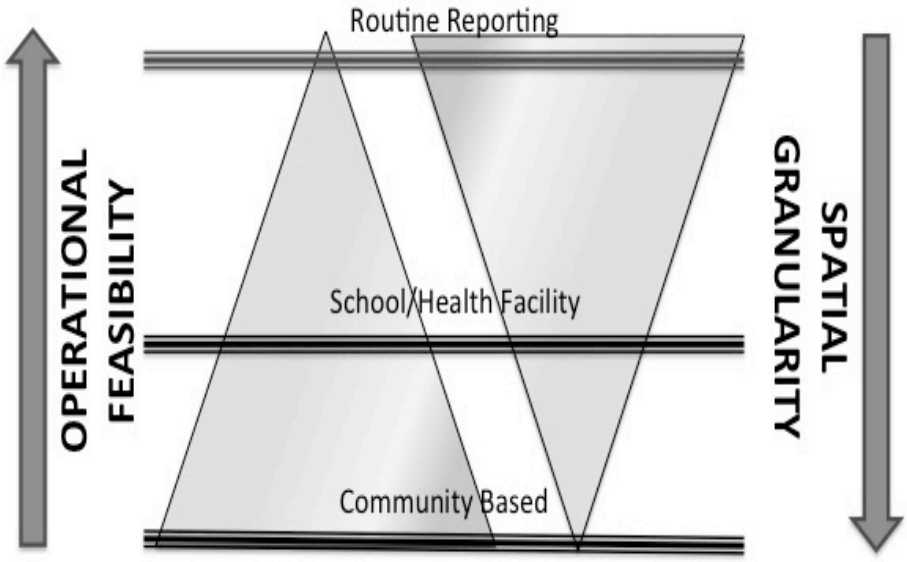

Figure 7-1: Schematic depicting the relationship between an approach for identifying and targeting hotspots that is operationally attractive and the granularity of the spatial information that is typically available: as an approach becomes operationally more difficult, the more spatial precise becomes available. Where the different survey approaches discussed would likely fall along this continuum are marked.

Determining regional-scale variation in malaria transmission intensity

The use of convenience sampling for monitoring malaria transmission is an attractive alternative to conventional surveillance methods involving intensive community based sampling or passively collected routine health facility data of variable quality. Chapter 5 demonstrated that both health facility and primary school surveys were able to provide reasonable estimates of malaria transmission in the community and reflects the spatial heterogeneity in transmission intensity across the study area. Importantly, the convenience sampling approach showed strong concordance with the community estimates when spatial overlap of location of residence was achieved highlighting the importance of accounting for the microepidemiological nature of malaria transmission when interpreting results. Routine health facility data was also collected as part of a passive case detection study. Although this data was able to rank facilities according to transmission intensity, data were based on confirmed malaria by microscopy

and/or RDT, which is not the case in a large proportion of facilities and required constant supervision to ensure data quality. Therefore, in this setting, the use of periodic surveys in health facilities or primary schools where robust diagnostic practices are enforced (134, 212, 263) provides an operationally attractive alternative to monitor malaria transmission and could be used to identify areas with increased burden while minimizing potential bias.

Detecting local-scale hotspots of malaria transmission

In chapter 4, hotspots were detected at the local level using data collected during an intensive community-based survey. Both model-based geostatistics (MBG) and the more commonly used SatScan approach were used to detect local level heterogeneity and models were informed using both sero- and pcr-positivity. Although obtaining a consistent delineation of hotspot boundaries was difficult, both cluster detection methods and malaria metrics were able to identify the general areas with higher burden and therefore offer insight into where interventions should be prioritized. The number of points used to inform the spatial models also had an impact on where the hotspot boundaries were drawn. Similar to convention statistics, large sample sizes resulted in more precise estimates of hotspot delineation. A better understanding of local level transmission dynamics would be useful to provide a gold standard with which to assess the error in terms of where hotspot boundaries are drawn by each cluster detection methods and the sample size or distribution of points needed to ensure robust results. (26) The malaria metrics tested are measuring different aspects of malaria transmission, which is likely to be one of the factors causing the differences observed in the resulting hotspots.

As discussed in chapter 4, the delineation of hotspots for targeting interventions to areas with increased burden is ultimately based on an operational decision. The choice of metric and statistical models will depend on many factors including the information available, technical capacity, and desired outcome. However, these results show that the results obtained must be interpreted appropriately and uncertainties acknowledged. The work on defining hotspots of transmission highlights the difficulties in translating the theory of targeting interventions to

malaria foci to practice and underlies some of the gaps that should be explored to further refine and operationalize such an approach. Furthermore, additional insight would also help identify what granularity will be good enough to achieve sustainable reductions in transmission, in other words, be good enough for decision making.

Convenience sampling to identify hotspots of malaria transmission

The use of convenience sampling was tested as a potential tool to identify hotspots of malaria transmission in the community for subsequent targeting (chapter 6). Those testing positive during the convenience sample, particularly school surveys, were associated with residing in a hotspot of malaria exposure identified based on a community survey. The use of convenience sampling to target hotspots of malaria has several advantages over the conventional approaches. Firstly, using convenience sampling has the potential to provide information with a greater spatial granularity. Information can be obtained either at the household level if spatial coordinates of case residences could be obtained (appendix 1.1) or alternatively to the catchment area of the institution, which tend to be smaller than district boundaries. (126, 285) Also, a reliance on cross-sectional studies minimizes the bias and subjectivity associated with focusing solely on suspected malaria cases (as discussed in section 5.3.3). These data suggest that as well as being a reliable metric for malaria surveillance, convenience sampling, and particularly primary school surveys, could provide an operationally attractive outlet to identify hotspots of malaria transmission in the community.

Targeting Hotspots of Malaria

Finally, if foci of transmission can be identified, it is likely that the distribution of parasite carriers within that hotspot will also be heterogeneous with some compounds being more at risk than others. (127) As discussed in section 6.3, current strategies that target interventions to parasite carriers rely on test and treat or mass treat campaigns, which either under- or over-target the parasite reservoir. Therefore, ascertaining the proportion of parasite carriers within known hotspots of infection that can be identified using an operationally tractable approach could further refine a targeted strategy and provide a benchmark with which to compare the efficacy of a convenience sampling approach. Chapter 6

presents an approach whereby a sentinel population consisting of individuals most likely to have infections detectable by RDT being tested, and if any individual was positive for malaria, it was assumed that subpatent and asymptomatic infections in the household were present. This approach managed to identify the majority of parasite carriers, however the best approach still missed over one-fifth of infections. The sensitivity of the focal testing approach was associated with the baseline parasite prevalence suggesting that this approach may be a viable and operationally attractive method in some settings.

7.3 Future research directions

That heterogeneity of transmission exists at all spatial scales is increasingly acknowledged and supports the focus on a hotspot-targeted strategy for malaria control and elimination. Although a better understanding of what spatial scale is ideal for a hotspot targeted approach in different settings is needed, the practicality of integrating such a strategy into a malaria control program will be contingent on three (not mutually exclusive) factors: that the approach should be operationally viable; that targeting hotspots will be more cost effective than a uniform strategy; and that there is an impact on malaria transmission, ideally in the hotspots as well as the broader community.

Firstly, for any intervention to be sustainable it must be integrated into the local malaria control program and be feasible given existing financial, technical, and logistical capacity. (222) This study identified two operationally attractive strategies that were able to measure transmission intensity in the community and identify hotspots of infection. The use of school and health facility surveys are appropriate for capturing the malaria burden given the epidemiological profile in the Kenyan highlands and probably in many other African settings, but in specific settings different populations may be more appropriate and acceptable to the community. For example, mobile populations may be better indicators of transmission potential in areas primarily affected by imported malaria or targeting markets or mining camps where cases are associated with certain occupations. (39, 286) Furthermore, the acceptability of such a program may also vary by population. For example, in some cultures it may be unacceptable to take biological

specimens in settings outside of a health care institution or at all. Therefore, it is essential to identify possible outlets to capture the malaria burden, adapt strategies to the local context, and integrate the communities into the decision making process.

Secondly, using a targeted approach will only become viable if the costs associated with identifying hotspots are lower than the money saved compared to conducting a universal campaign. The use of convenience sampling for monitoring malaria transmission has been shown to be more cost effective than community based surveys (232) and tends to provide more robust estimates compared to routinely collected malaria data (see section 5.3.1). (178, 223) However, more work is needed to further refine the use of operational approaches to target hotspots of transmission, for example whether case-aggregation over time or the use of a more re-active approach including a pre-specified buffer area around each case achieves better coverage of hotspots. Once strategies can be further refined, more robust cost-benefit analysis can be conducted to determine if the additional expenses associated with this approach compensates for the costs saved by targeting interventions to a subset of the population.

Thirdly, the focus on transmission hotspots is becoming an increasingly popular theory for application in malaria control and elimination programs. (121, 127) However, there is still no empirical evidence on the impact that such a targeted approach has on malaria transmission. The findings of this research and other work suggest that heterogeneity in malaria transmission exists and can be detected. (148, 287) Nevertheless, there are still several gaps in understanding the transmission dynamics at the local level and how much these foci actually contribute to malaria transmission. Understanding the local level transmission would facilitate refinement of methods to both define these, and ensure that impacts can be measured effectively. More concrete, longitudinal data are needed to identify such trends and provide a gold standard with which to measure boundary delineation and effectiveness of interventions against.

Lastly, the optimum malaria metric to inform cluster detection models merits further investigation. Ideally, mapping R_C at spatial resolutions relevant to local malaria control programs would be possible. Deciding where to intervene would be simple as would then target everywhere where R_C is greater than one. However, measuring R_O or approximations such as EIR or SCR are population based metrics and will inevitably introduce spatial bias in estimates. As mentioned above (section 7.2), cluster detection of seropositivity and PCR infection resulted in some consistency: both metrics were able to identify similar areas, presumably the most significant ones. Despite this consistency, there were always clusters identified by one approach that were missed by the other metric. As discussed, these metrics are inherently measuring different components of the malaria transmission cycle and provide different pieces of information. Obtaining a better understanding of how the different metrics relate would facilitate identifying the optimum metric or combination of metrics for hotspot detection. For example, is it necessary to employ both metrics for a combined measure for hotspot detection and how this compares with relying on the less precise diagnostic tools such as microscopy or RDT. Alternatively, employing an age-adjusted approach as has been employed by the global mapping projects, (9, 109) may provide a more accurate picture of hotspots. A pragmatic approach to a hotspot-targeted strategy would be to take any evidence of infection or exposure as a trigger for intervention. However, to refine such a strategy, a better understanding of what is driving these differences and whether they matter in terms of employing a hotspot targeted approach would be important. Once a gold-standard metric for defining hotspots can be identified, if possible, assessing the bias introduced with use of alternative, more operational practical diagnostic tools would be possible.

If a hotspot targeted approach were to be employed it will be important to identify what proportion of the parasite reservoir within such areas must be targeted to ensure a sustainable reduction in transmission. Community uptake of malaria control interventions is not uniform. It has been found that some individuals will consistently and correctly participate whereas others will never engage. (122) Therefore, identifying what threshold coverage must be achieved, and what the ideal package of interventions is, to ensure a sustainable impact on transmission

will be useful to provide a benchmark to strive towards and would facilitate monitoring and evaluation of such strategies. Identifying this threshold, which may vary in different populations, would also help gauge the efficacy of employing convenience sampling to target hotspots. Even though this research suggests that convenience sampling is able to identify hotspots for targeting, if the coverage achieved despite the use of buffer zones or other method is not sufficient, then alternative strategies would be needed.

In conclusion, the overall objective of this study was to determine if operationally attractive approaches for the identification of hotspots of malaria transmission in the western Kenyan highlands are possible and can provide viable alternatives to a community based survey approach. This thesis provides the first rigorous examination of defining hotspots of malaria transmission and the potential role for integrating operationally attractive approaches to both malaria surveillance and for targeting hotspots of malaria transmission. The findings show that the distribution of malaria in the Kenyan highlands is highly heterogeneous and that operational strategies can provide a sensitive method to monitor malaria transmission and to identify hotspots for subsequent targeting. However, there remains a need to further understand the role of hotspots in malaria transmission and how these can best be measured and ensuring that correct inferences are made from the available data.

Chapter 8 - References

8.1 Hotspot Detection Paper (P1) References (section 4.2)

1. WHO, 2013. World Malaria Report 2013. Program MC, ed. Geneva, Switzerland: World Health Organization.
2. WHO, 2014. World Malaira Report 2014. Organization WH, ed. Geneva, Switzerland: World Health Organization.
3. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW, 2004. The global distribution and population at risk of malaria: past, present, and future. The Lancet Infectious Diseases 4: 327-336.
4. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, Bousema T, Drakeley CJ, Ferguson NM, Basanez MG, Ghani AC, 2010. Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS Med 7.
5. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, Ghani A, Drakeley C, Gosling R, 2012. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS Med 9: e1001165.
6. Clements ACA, Reid HL, Kelly GC, Hay SI, 2013. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? The Lancet Infectious Diseases 13: 709-718.
7. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, Otieno S, Carneiro I, Cox J, Msuya E, Kleinschmidt I, Maxwell C, Greenwood B, Riley E, Sauerwein R, Chandramohan D, Gosling R, 2010. Identification of hot spots of malaria transmission for targeted malaria control. J Infect Dis 201: 1764-74.
8. Kleinschmidt I, Sharp BL, Clarke GPY, Curtis B, Fraser C, 2001. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. Am J Epidemiol 153: 1213-1221.
9. Nourein AB, Abass MA, Nugud AH, El Hassan I, Snow RW, Noor AM, 2011. Identifying residual foci of Plasmodium falciparum infections for malaria elimination: the urban context of Khartoum, Sudan. PLoS One 6: e16948.

10. Bejon P, Williams TN, Liljander A, Noor AM, Wambua J, Ogada E, Olotu A, Osier FH, Hay SI, Farnert A, Marsh K, 2010. Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. PLoS Med 7: e1000304.
11. Bousema T, Stevenson J, Baidjoe A, Stresman G, Griffin JT, Kleinschmidt I, Remarque EJ, Vulule J, Bayoh N, Laserson K, Desai M, Sauerwein R, Drakeley C, Cox J, 2013. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. Trials 14: 36.
12. Bejon P, Turner L, Lavstsen T, Cham G, Olotu A, Drakeley CJ, Lievens M, Vekemans J, Savarese B, Lusingu J, von Seidlein L, Bull PC, Marsh K, Theander TG, 2011. Serological evidence of discrete spatial clusters of Plasmodium falciparum parasites. PLoS One 6: e21711.
13. Mosha JF, Sturrock HJ, Greenwood B, Sutherland CJ, Gadalla NB, Atwal S, Hemelaar S, Brown JM, Drakeley C, Kibiki G, Bousema T, Chandramohan D, Gosling RD, 2014. Hot spot or not: a comparison of spatial statistical methods to predict prospective malaria infections. Malar J 13: 53.
14. Mosha JF, Sturrock HJ, Greenhouse B, Greenwood B, Sutherland CJ, Gadalla N, Atwal S, Drakeley C, Kibiki G, Bousema T, Chandramohan D, Gosling R, 2013. Epidemiology of subpatent Plasmodium falciparum infection: implications for detection of hotspots with imperfect diagnostics. Malar J 12: 221.
15. Diggle PJ, Tawn JA, 1998. Model-based geostatistics. Appl. Statist 47: 299-350.
16. Diggle PJ, Thomson MC, Christensen OF, Rowlingson B, Obsomer V, Gardon J, Wanji S, Takougang I, Enyong P, Kamgno J, Remme JH, Boussinesq M, Molyneux DH, 2007. Spatial modelling and the prediction of Loa loa risk: decision making under uncertainty. Annals of Tropical Medicine and Parasitology 101: 499-509.
17. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, Manh BH, Elyazar IR, Brooker S, Smith DL, Moyeed RA, Snow RW, 2009. A world malaria map: Plasmodium falciparum endemicity in 2007. PLoS Med 6: e1000048.

18. Elyazar IR, Gething PW, Patil AP, Rogayah H, Kusriastuti R, Wismarini DM, Tarmizi SN, Baird JK, Hay SI, 2011. Plasmodium falciparum malaria endemicity in Indonesia in 2010. PLoS One 6: e21315.
19. Stevenson JC, Stresman GH, Gitonga CW, Gillig J, Owaga C, Marube E, Odongo W, Okoth A, China P, Oriango R, Brooker SJ, Bousema T, Drakeley C, Cox J, 2013. Reliability of school surveys in estimating geographic variation in malaria transmission in the western Kenyan highlands. PLoS One 8: e77641.
20. Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, Thaithong S, Brown KN, 1993. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. Molecular and Biochemical Parasitology 61: 315-320.
21. Baidjoe A, Stone W, Ploemen I, Shagari S, Grignard L, Osoti V, Makori E, Stevenson J, Kariuki S, Sutherland C, Sauerwein R, Cox J, Drakeley C, Bousema T, 2013. Combined DNA extraction and antibody elution from filter papers for the assessment of malaria transmission intensity in epidemiological studies. Malar J 12: 272.
22. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SL, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya WM, Lemnge MM, Cox J, Reyburn H, Riley EM, 2005. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci U S A 102: 5108-13.
23. Corran P, Coleman P, Riley E, Drakeley C, 2007. Serology: a robust indicator of malaria transmission intensity? Trends Parasitol 23: 575-82.
24. Garbrecht J, Martz LW, 1997. The assignment of drainage direction over flat surfaces in rater digital elevation models. J Hydrology 193: 204-213.
25. Gleyzer A, Denisyuk M, Rimmer A, Salingar Y, 2004. A fast recursive GIS algorithm for computing Strahler stream order in braided and nonbraided networks. Journal of the American Water Resources Association 40: 937-946.
26. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, Hagan P, Hii JLK, Ndhlovu PD, Quinnell RJ, Watts CH, Chandiwana SK, Anderson RM, 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.

27. Zhang Z, Huang Y, 2012. A Linear Regression Framework for the Receiver Operating Characteristic (ROC) Curve Analysis. J Biom Biostat 3.
28. Bousema T, Youssef RM, Cook J, Cox J, Alegana VA, Amran J, Noor AM, Snow RW, Drakeley C, 2010. Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. Emerg Infect Dis 16: 392-9.
29. Cook J, Kleinschmidt I, Schwabe C, Nseng G, Bousema T, Corran PH, Riley EM, Drakeley CJ, 2011. Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, equatorial Guinea. PLoS One 6: e25137.
30. Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I, Diallo M, Diawara S, Ouattara A, Diakite M, Doumbo OK, 2006. Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. BMC Public Health 6: 286.
31. Bejon P, Williams TN, Nyundo C, Hay SI, Benz D, Gething PW, Otiende M, Peshu J, Bashraheil M, Greenhouse B, Bousema T, Bauni E, Marsh K, Smith DL, Borrmann S, 2014. A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. Elife 3: e02130.
32. Hay SI, Smith DL, Snow RW, 2008. Measuring malaria endemicity from intense to interrupted transmission. The Lancet Infectious Diseases 8: 369-378.
33. Tusting LS, Bousema T, Smith DL, Drakeley C, 2014. Measuring changes in Plasmodium falciparum transmission: Precision, accuracy and costs of metrics. Adv Parasitol 84: 151-208.
34. Smith DL, McKenzie FE, Snow RW, Hay SI, 2007. Revisiting the basic reproductive number for malaria and its implications for malaria control. PLoS Biol 5: e42.
35. Pullan RL, Sturrock HJ, Soares Magalhaes RJ, Clements AC, Brooker SJ, 2012. Spatial parasite ecology and epidemiology: a review of methods and applications. Parasitology 139: 1870-87.
36. Aamodt G, Samuelsen SO, Skrondal A, 2006. A simulation study of three methods for detecting disease clusters. Int J Health Geogr 5: 15.

37. Rothman KJ, 1990. A sobering start for the cluster busters' conferenc. Am J Epidemiol 132: S6-S13.
38. Jackson MC, Huang L, Luo J, Hachey M, Feuer E, 2009. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. Int J Health Geogr 8: 55.

8.2 Schools vs. Community Transmission Paper (P2) References (section 5.2.2)

1. Hay SI, Smith DL, Snow RW. (2008) Measuring malaria endemicity from intense to interrupted transmission. Lancet Infect Dis 8:369-78.
2. Snow RW, Amratia P, Kabaria CW, Noor AM, Marsh K. (2012) The changing limits and incidence of malaria in Africa: 1939-2009. Adv Parasitol 78:169-262.
3. Cox J, Hay SI, Abeku TA, Checchi F, Snow RW. (2007) The uncertain burden of *Plasmodium falciparum* epidemics in Africa. Trends Parasitol 23(4):142-8.
4. Satoguina J, Walther B, Drakeley C, Nwakanma D, Oriero EC, et al. (2009) Comparison of surveillance methods applied to a situation of low malaria prevalence at rural sites in The Gambia and Guinea Bissau. Malar J 8:274.
5. Greenhouse B, Ho B, Hubbard A, Njama-Meya D, Narum DL, et al. (2011) Antibodies to *Plasmodium falciparum* antigens predict a higher risk of malaria but protection from symptoms once parasitemic. J Infect Dis 204(1):19-26.
6. O'Meara WP, Mwangi TW, Williams TN, McKenzie FE, Snow RW, et al. (2008) Relationship Between Exposure, Clinical Malaria, and Age in an Area of Changing Transmission Intensity. Am J Trop Med Hyg 79(2):185-91.
7. Kendjo E, Agbenyega T, Bojang K, Newton CRJC, Bouyou-Akotet M, et al. (2013) Mortality patterns and site heterogeneity of severe malaria in African children. PloS One 8(3):e58686.
8. Beadle C, McElroy PD, Oster CN, Beier JC, Oloo AJ, et al. (1995) Impact of transmission intensity and age on *Plasmodium falciparum* density and associated fever: Implications for malaria vaccine trial design. J Infect Dis 172:1047-54.

9. Doolan DL, Dobano C, Baird JK. (2009) Acquired immunity to malaria. Clin Microbiol Rev 22(1):13-36,
10. Noor AM, Kirui VC, Brooker SJ, Snow RW. (2009) The use of insecticide treated nets by age: implications for universal coverage in Africa. BMC Public Health 9:369.
11. Brooker S, Kolaczinski JH, Gitonga CW, Noor AM, Snow RW. (2009) The use of schools for malaria surveillance and programme evaluation in Africa. Malar J 8:231.
12. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, et al. (2005) Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci USA 102(14):5108-13.
13. Corran P, Coleman P, Riley E, Drakeley C. (2007) Serology: a robust indicator of malaria transmission intensity? Trends Parasitol 23(12):575-82.
14. Reyburn HM, Mbatia R, Drakeley, C. Bruce J, Carneiro I, et al. (2005) Association of transmission intensity and age with clinical manifestations and case fatality of severe *Plasmodium falciparum* malaria. JAMA 293(12):1461-70.
15. Gitonga CW, Karanja PN, Kihara J, Mwanje M, Juma E, et al. (2010) Implementing school malaria surveys in Kenya: Towards a national surveillance system. Malar J 9:306.
16. Gitonga CW, Kihara JH, Njenga SM, Awuondo K, Noor AM, et al. (2012) Use of rapid diagnostic tests in malaria school surveys in Kenya: Does their under-performance matter for planning malaria control? Am J Trop Med Hyg 87(6):1004-11.
17. Stuckey EM, Stevenson JC, Cooke MK, Owaga C, Marube E, et al. (2012) Simulation of malaria epidemiology and control in the highlands of Western Kenya. Malar J 11:357.
18. Blackwelder WC. (1982) "Proving the Null Hypothesis" in Clinical Trials. Control Clin Trials 3:345-53.

19. Ministry of Public Health and Sanitation. (2010) National Guidelines for HIV Testing and Counselling in Kenya - 2010. In: Programme NAaSC, editor. Nairobi, Kenya: Ministry of Public Health and Sanitation.
20. Cook J, Reid H, Iavro J, Kuwahata M, Taleo G et al. (2010) Using serological measures to monitor changes in malaria transmission in Vanuatu. Malar J 9:169.
21. Corran PH, Cook J, Lynch C, Leendertse H, Manjurano A, et al. (2008) Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. Malar J 7:195.
22. Youden WJ. (1950) Index for rating diagnostic tests. Cancer 3(1):32-5.
23. Altman DG, Bland JM. (1983) Measurement in medicine: The analysis of method comparison studies. The Statistician 32(3):307-17.
24. Lin LIK. (1989) A concordance correlation coefficient to evaluate reproducibility. Biometrics 45(1):255-68.
25. Bradley EL, Blackwood LG. (1989) Comparing paired data: A simultaneous test for means and variances. Am Stat 43(4):234-5.
26. Ashton RA, Kefyalew T, Tesfaye G, Pullan RL, Yadeta D, et al. (2011) School-based surveys of malaria in Oromia Regional State, Ethiopia: A rapid survey method for malaria in low transmission settings. Malar J 10(1):25.
27. Bejon P, Turner L, Lavstsen T, Cham G, Olotu A, et al. (2011) Serological evidence of discrete spatial clusters of *Plasmodium falciparum* parasites. PloS One 6(6):e21711.
28. Greenwood BM. (1989) The microepidemiology of malaria and its importance to malaria control. Trans R Soc Trop Med Hyg 83(Suppl):25-9.
29. Sexton CJ, Costenbader EC, Vinh DT, Chen PL, Hoang TV, et al. (2012) Correlation of prospective and cross-sectional measures of HIV type 1 incidence in a higher-risk cohort in Ho Chi Minh City, Vietnam. AIDS Res Hum Retroviruses 28(8):866-73.
30. White MT, Griffin JT, Drakeley CJ, Ghani AC. (2010) Heterogeneity in malaria exposure and vaccine response: implications for the interpretation of vaccine efficacy trials. Malar J 9:82.

31. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, et al. (1997) Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proc Natl Acad Sci U S A 94:338-42.
32. Icks A, Haastert B, Glaeske G, Stumpf U, Windolf J, et al. (2012) Correction factor for the analysis of the hip fracture incidence--differences between age, sex, region, and calendar year. Cent Eur J Med 124(11-12):391-4.
33. Guyatt HL, Brooker S, Donnelly CA. (1999) Can prevalence of infection in school-aged children be used as an index for assessing community prevalence? Parasitology 118 (Pt 3):257-68.
34. Lucas AM, Mbiti IM. (2012) Does free primary education narrow gender differences in schooling? Evidence from Kenya. J Afr Econ 21(5):691-722.
35. Government of Kenya. (2006) Kenya Integrated Household Budget Survey (KIHBS) 2005/06. Revised ed. Nairobi, Kenya: Government of Kenya.
36. Aitchison J. (1982) The statistical analysis of compositional data. J R Stat Soc Series B Stat Methodol 44(2):139-77.

8.3 Health Facility Transmission Intensity Paper (P3) References (section 5.3.3)

1. Sturrock HJ, Hsiang MS, Cohen JM, D.L. S, Greenhouse B, Bousema T, Gosling R, 2013. Targeting asymptomatic malaria infections: Active surveillance in control and elimination. PLoS Med 10: e1001467.
2. The malERA Consultative Group on Monitoring E, and Surveillance, 2011. A research agenda for malaria eradication: Monitoring, evaluation, and surveillance. PLoS Med 8: e1000400.
3. Tusting LS, Bousema T, Smith DL, Drakeley C, 2014. Measuring changes in Plasmodium falciparum transmission: Precision, accuracy and costs of metrics. Adv Parasitol 84: 151-208.
4. Khosa E, Kuonza LR, Kruger P, Maimela E, 2013. Towards the elimination of malaria in South Africa: a review of surveillance data in Mutale Municipality, Limpopo Province, 2005 to 2010. Malar J 12: 7.
5. Juma E, Zurovac D, 2011. Changes in health workers' malaria diagnosis and treatment practices in Kenya. Malar J 10: 1.

6. Olotu A, Fegan G, Williams TN, Sasi P, Ogada E, Bauni E, Wambua J, Marsh K, Borrmann S, Bejon P, 2010. Defining clinical malaria: the specificity and incidence of endpoints from active and passive surveillance of children in rural Kenya. PLoS One 5: e15569.
7. Afrane Y, Zhou G, Githeko A, Yan G, 2013. Utility of health facility-based malaria data for malaria surveillance. PLoS One 8: e54305.
8. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, Otieno S, Carneiro I, Cox J, Msuya E, Kleinschmidt I, Maxwell C, Greenwood B, Riley E, Sauerwein R, Chandramohan D, Gosling R, 2010. Identification of hot spots of malaria transmission for targeted malaria control. J Infect Dis 201: 1764-74.
9. Oduro AR, Bojang KA, Conway DJ, Corrah T, Greenwood BM, Schellenberg D, 2011. Health centre surveys as a potential tool for monitoring malaria epidemiology by area and over time. PLoS One 6: e26305.
10. Macedo de Oliveira A, Mutemba R, Morgan J, Streat E, Roberts J, Menon M, Mabunda S, 2011. Prevalence of malaria among patients attending public health facilities in Maputo City, Mozambique. Am J Trop Med Hyg 85: 1002-7.
11. Okell LC, Bousema T, Griffin JT, Ouedraogo AL, Ghani AC, Drakeley CJ, 2012. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. Nat Commun 3: 1237.
12. Okell LC, Ghani AC, Lyons E, Drakeley CJ, 2009. Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. J Infect Dis 200: 1509-17.
13. Lindblade KA, Steinhardt L, Samuels A, Kachur SP, Slutsker L, 2013. The silent threat: Asymptomatic parasitemia and malaria transmission. Expert Rev Anti Infect ther 11: 623-639.
14. Nguyen HV, van den Eede P, van Overmeir C, Thang ND, Hung le X, D'Alessandro U, Erhart A, 2012. Marked age-dependent prevalence of symptomatic and patent infections and complexity of distribution of human Plasmodium species in central Vietnam. Am J Trop Med Hyg 87: 989-95.
15. Baliraine FN, Afrane YA, Amenya DA, Bonizzoni M, Menge DM, Zhou G, Zhong D, Vardo-Zalik AM, Githeko AK, Yan G, 2009. High prevalence of

asymptomatic plasmodium falciparum infections in a highland area of western Kenya: a cohort study. J Infect Dis 200: 66-74.

16. Harris I, Sharrock WW, Bain LM, Gray KA, Bobogare A, Boaz L, Lilley K, Krause D, Vallely A, Johnson ML, Gatton ML, Shanks GD, Cheng Q, 2010. A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. Malar J 9: 254.
17. Karl S, Gurarie D, Zimmerman PA, King CH, St Pierre TG, Davis TM, 2011. A sub-microscopic gametocyte reservoir can sustain malaria transmission. PLoS One 6: e20805.
18. Bousema T, Stevenson J, Baidjoe A, Stresman G, Griffin JT, Kleinschmidt I, Remarque EJ, Vulule J, Bayoh N, Laserson K, Desai M, Sauerwein R, Drakeley C, Cox J, 2013. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. Trials 14: 36.
19. Sanitation MoPHa, 2011. 2010 Kenya Malaria Indicator Survey. Control DoM, ed. Nairobi, Kenya: Ministry of Public Health and Sanitation.
20. Stevenson JC, Stresman GH, Gitonga CW, Gillig J, Owaga C, Marube E, Odongo W, Okoth A, China P, Oriango R, Brooker SJ, Bousema T, Drakeley C, Cox J, 2013. Reliability of school surveys in estimating geographic variation in malaria transmission in the western Kenyan highlands. PLoS One 8: e77641.
21. Stevenson J, St. Laurent B, Lobo NF, Cooke MK, Kahindi SC, Oriango R, Harback RE, Cox J, Drakeley C, 2012. Novel vectors of malaria parasites in the western highlands of Kenya. Emerg Infect Dis 18: 1547-1549.
22. WHO, 2012. Malaria rapid diagnostic test performance: Results of WHO product testing of malaria RDTs: Round 4 (2012). Diseases SPfRaTiT, ed. Switzerland: World Health Organization.
23. Sanitation MoPHa, 2010. National Guidelines for HIV Testing and Counselling in Kenya. Programme NAaSC, ed. Nairobi, Kenya: Ministry of Public Health and Sanitation.
24. Baidjoe A, Stone W, Ploemen I, Shagari S, Grignard L, Osoti V, Makori E, Stevenson J, Kariuki S, Sutherland C, Sauerwein R, Cox J, Drakeley C,

Bousema T, 2013. Combined DNA extraction and antibody elution from filter papers for the assessment of malaria transmission intensity in epidemiological studies. Malar J 12: 272.

25. Corran P, Coleman P, Riley E, Drakeley C, 2007. Serology: a robust indicator of malaria transmission intensity? Trends Parasitol 23: 575-82.
26. Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, Thaithong S, Brown KN, 1993. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. Molecular and Biochemical Parasitology 61: 315-320.
27. Koepfli C, Ross A, Kiniboro B, Smith TA, Zimmerman PA, Siba P, Mueller I, Felger I, 2011. Multiplicity and diversity of Plasmodium vivax infections in a highly endemic region in Papua New Guinea. PLoS Negl Trop Dis 5: e1424.
28. Koepfli C, Schoepflin S, Bretscher M, Lin E, Kiniboro B, Zimmerman PA, Siba P, Smith TA, Mueller I, Felger I, 2011. How much remains undetected? Probability of molecular detection of human Plasmodia in the field. PLoS One 6: e19010.
29. Liljander A, Wiklund L, Falk N, Kweku M, Martensson A, Felger I, Farnert A, 2009. Optimization and validation of multi-coloured capillary electrophoresis for genotyping of Plasmodium falciparum merozoite surface proteins (msp1 and 2). Malar J 8: 78.
30. McMorrow ML, Aidoo M, Kachur SP, 2011. Malaria rapid diagnostic tests in elimination settings - can they find the last parasite? Clin Microbiol Infect 17: 1624-1631.
31. Stresman G, Stevenson J, Owaga C, Marube E, Anyango C, Drakeley C, Bousema T, Cox J, 2014. Validation of three geolocation strategies for health facility attendees for research and public health surveillance in a rural setting in western Kenya. Epidemiology and Infection.
32. Barry AE, Schultz L, Senn N, Nale J, Kiniboro B, Siba PM, Mueller I, Reeder JC, 2013. High levels of genetic diversity of Plasmodium falciparum populations in Papua New Guinea despite variable infection prevalence. Am J Trop Med Hyg 88: 718-25.

33. Fox E, Strickland GT, Sarwar M, Shamim M, Zafar-Latif A, Khaliq AA, 1987. Reliable assessment of malaria prevalence through village clinics. Trans R Soc Trop Med Hyg 81: 115-117.
34. Cotter C, Sturrock HJW, Hsiang MS, Liu J, Phillips AA, Hwang J, Gueye CS, Fullman N, Gosling RD, Feachem RGA, 2013. The changing epidemiology of malaria elimination: new strategies for new challenges. The Lancet 382: 900-911.
35. Stresman GH, 2010. Beyond temperature and precipitation: ecological risk factors that modify malaria transmission. Acta Trop 116: 167-72.
36. Stresman GH, Kamanga A, Moono P, Hamapumbu H, Mharakurwa S, Kobayashi T, Moss WJ, Shiff C, 2010. A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. Malar J 9: 265.
37. WHO, 2012. Disease surveilance for malaria control: An operational manual. Project GM, ed. Switzerland: World Health Organization.
38. Mosha JF, Sturrock HJ, Greenhouse B, Greenwood B, Sutherland CJ, Gadalla N, Atwal S, Drakeley C, Kibiki G, Bousema T, Chandramohan D, Gosling R, 2013. Epidemiology of subpatent Plasmodium falciparum infection: implications for detection of hotspots with imperfect diagnostics. Malar J 12: 221.
39. Beshir KB, Sutherland CJ, Sawa P, Drakeley CJ, Okell L, Mweresa CK, Omar SA, Shekalaghe SA, Kaur H, Ndaro A, Chilongola J, Schallig HD, Sauerwein RW, Hallett RL, Bousema T, 2013. Residual Plasmodium falciparum parasitemia in Kenyan children after artemisinin-combination therapy is associated with increased transmission to mosquitoes and parasite recurrence. J Infect Dis 208: 2017-24.
40. Nassir E, Abdel-Muhsin AM, Suliaman S, Kenyon F, Kheir A, Geha H, Ferguson HM, Walliker D, Babiker HA, 2005. Impact of genetic complexity on longevity and gametocytogenesis of Plasmodium falciparum during the dry and transmission-free season of eastern Sudan. Int J Parasitol 35: 49-55.
41. Iriemenam NC, Shah M, Gatei W, van Eijk AM, Ayisi J, Kariuki S, V. EJ, Owino SO, Lal AA, Omosun YO, Otieno K, Desai M, ter Kuile FO, Nahlen B, Moore J, Hamel MJ, Ouma P, Slutsker L, Shi YP, 2012. Temporal trends of

sulphadoxine-pyrimethamine (SP) drug-resistance molecular markers in Plasmodium falciparum parasites from pregnant women in western Kenya. Malar J 11: 134.

42. Zhong D, Afrane Y, Githeko A, Cui L, Menge DM, Yan G, 2008. Molecular epidemiology of drug-resistant malaria in western Kenya highlands. BMC Infect Dis 8: 105.
43. Alves FP, R. DR, Menezes MJ, Krieger H, da Silva LHP, Camargo EP, 2002. High prevalence of asymptomatic Plasmodium vivax and Plasmodium falciparum infections in native amazonian populations. Am J Trop Med Hyg 66: 641-648.

8.4 Focal Mass Drug Administration Paper References (P4)

1. Cotter C, Sturrock HJW, Hsiang MS, et al. The changing epidemiology of malaria elimination: new strategies for new challenges. Lancet **2013**; 382:900-11.
2. Woolhouse MEJ, Dye C, Etard J-F, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proc Natl Acad Sci U S A **1997**; 94:338-42.
3. Baliraine FN, Afrane YA, Amenya DA, et al. High prevalence of asymptomatic plasmodium falciparum infections in a highland area of western Kenya: a cohort study. J Infect Dis **2009**; 200:66-74.
4. Okell LC, Bousema T, Griffin JT, Ouedraogo AL, Ghani AC, Drakeley CJ. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. Nat Commun **2012**; 3:1237.
5. Diarra A, Nebie I, Tiono A, et al. Seasonal performance of a malaria rapid diagnosis test at community health clinics in a malaria-hyperendemic region of Burkina Faso. Parasit Vectors **2012**; 5:103.
6. Harris I, Sharrock WW, Bain LM, et al. A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. Malar J **2010**; 9:254.

7. Cook J, Xu W, Msellem M, et al. Mass screening and treatment using a Falciparum-specific rapid diagnostic test did not reduce malaria incidence in Zanzibar. J Infect Dis **2014**.
8. Nsobya SL, Parikh S, Kironde F, et al. Molecular evaluation of the natural history of asymptomatic parasitemia in Ugandan children. J Infect Dis **2004**; 189:2220-6.
9. Bousema T, Griffin JT, Sauerwein RW, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS Med **2012**; 9:e1001165.
10. Lindblade KA, Steinhardt L, Samuels A, Kachur SP, Slutsker L. The silent threat: Asymptomatic parasitemia and malaria transmission. Expert Rev Anti Infect Ther **2013**; 11:623-39.
11. malERA, TDR, WHO. A research agenda for malaria eradication: diagnoses and diagnostics. PLoS Med **2011**; 8:e1000396.
12. Tietje K, Hawkins K, Clerk C, et al. The essential role of infection-detection technologies for malaria elimination and eradication. Trends Parasitol **2014**; 30:259-66.
13. Drakeley CJ, Corran PH, Coleman PG, et al. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci U S A **2005**; 102:5108-13.
14. Pinkevych M, Petravic J, Berecsky S, Rooth I, Farnert A, Davenport MP. Understanding the relationship between parasite growth rate and multiplicity of infection with the malaria parasite Plasmodium falciparum. J Infect Dis **2014**.
15. Proietti C, Verra F, Bretscher MT, et al. Influence of infection on malaria-specific antibody dynamics in a cohort exposed to intense malaria transmission in northern Uganda. Parasite Immunol **2013**; 35:164-73.
16. Nguyen HV, van den Eede P, van Overmeir C, et al. Marked age-dependent prevalence of symptomatic and patent infections and complexity of distribution of human Plasmodium species in central Vietnam. Am J Trop Med Hyg **2012**; 87:989-95.

17. Sutcliffe CG, Kobayashi T, Hamapumbu H, et al. Reduced risk of malaria parasitemia following household screening and treatment: a cross-sectional and longitudinal cohort study. PloS One **2012**; 7:e31396.
18. von Seidlein L. The failure of screening and treating as a malaria elimination strategy. PLoS Med **2014**; 11:e1001595.
19. Hamainza B, Moonga H, Sikaala CH, et al. Monitoring, characterization and control of chronic, symptomatic malaria infections in rural Zambia through monthly household visits by paid community health workers. Malar J **2014**; 13:128.
20. Poirot E, Skarbinski J, Sinclair D, Kachur SP, Slutsker L, Hwang J. Mass drug administration for malaria Cochrane Database Syst Rev **2013**; 12.
21. von Seidlein L, Greenwood BM. Mass administrations of antimalarial drugs. Trends Parasitol **2003**; 19:452-60.
22. Maude RJ, Socheat D, Nguon C, et al. Optimising strategies for Plasmodium falciparum malaria elimination in Cambodia: primaquine, mass drug administration and artemisinin resistance. PloS One **2012**; 7:e37166.
23. Bejon P, Williams TN, Liljander A, et al. Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. PLoS Med **2010**; 7:e1000304.
24. Littrell M, Sow GD, Ngom A, et al. Case investigation and reactive case detection for malaria elimination in northern Senegal. Malar J **2013**; 12.
25. Stresman GH, Kamanga A, Moono P, et al. A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. Malar J **2010**; 9:265.
26. Bousema T, Stevenson J, Baidjoe A, et al. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. Trials **2013**; 14:36.
27. Stevenson JC, Stresman GH, Gitonga CW, et al. Reliability of school surveys in estimating geographic variation in malaria transmission in the western Kenyan highlands. PloS One **2013**; 8:e77641.
28. Sanitation MoPHa. National Guidelines for HIV Testing and Counselling in Kenya. In: Programme NAaSC, ed. 2nd ed. Nairobi, Kenya: Ministry of Public Health and Sanitation, **2010**.

29. Stresman GH, Stevenson JC, Ngwu N, et al. High Levels of Asymptomatic and Subpatent Plasmodium falciparum Parasite Carriage at Health Facilities in an Area of Heterogeneous Malaria Transmission Intensity in the Kenyan Highlands. Am J Trop Med Hyg **2014**.
30. Baidjoe A, Stone W, Ploemen I, et al. Combined DNA extraction and antibody elution from filter papers for the assessment of malaria transmission intensity in epidemiological studies. Malar J **2013**; 12:272.
31. Snounou G, Viriyakosol S, Zhu XP, et al. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. Mol Biochem Parasitol **1993**; 61:315-20.
32. Mueller I, Schoepflin S, Smith T, et al. Force of infection is key to understanding the epidemiology of Plasmodium falciparum malaria in Papua New Guinean children. Proc Natl Acad Sci U S A **2012**; 109:10030-5.
33. Schneider P, Schoone G, Schallig HD, et al. Quantification of Plasmodium falciparum gametocytes in differential stage of development by quantitative nucleic acid sequence-based amplification. Mol Biochem Parasitol **2004**; 137:35-41.
34. Armstrong Schellenberg JR, Mrisho M, Manzi F, et al. Health and survival of young children in southern Tanzania. BMC Public Health **2008**; 8:194.
35. Hsiang MS, Hwang J, Tao AR, et al. Mass drug administration for the control and elimiation of Plasmodium vivax malaria: An ecological study from Jiangsu province, China. Malar J **2013**; 12.
36. Bejon P, Williams TN, Nyundo C, et al. A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. eLife **2014**; 3:e02130.
37. Searle KM, Shields T, Hamapumbu H, et al. Efficiency of household reactive case detection for malaria in rural Southern Zambia: simulations based on cross-sectional surveys from two epidemiological settings. PloS One **2013**; 8:e70972.
38. Bousema T, Okell L, Felger I, Drakeley C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. Nat Rev Microbiol **2014**; 12:833-40.

39. Tiono AB, Ouedraogo A, Ogutu B, et al. A controlled, parallel, cluster-randomized trial of community-wide screening and treatment of asymptomatic carriers of Plasmodium falciparum in Burkina Faso. Malar J **2013**; 12:79.
40. Greenhouse B, Ho B, Hubbard A, et al. Antibodies to Plasmodium falciparum antigens predict a higher risk of malaria but protection from symptoms once parasitemic. J Infect Dis **2011**; 204:19-26.
41. Bejon P, Warimwe G, Mackintosh CL, et al. Analysis of immunity to febrile malaria in children that distinguishes immunity from lack of exposure. Infect Immun **2009**; 77:1917-23.
42. Beshir KB, Sutherland CJ, Sawa P, et al. Residual Plasmodium falciparum parasitemia in Kenyan children after artemisinin-combination therapy is associated with increased transmission to mosquitoes and parasite recurrence. J Infect Dis **2013**; 208:2017-24.
43. Bell DR, Wilson DW, Martin LB. False-positive results of a Plasmodium falciparum Histidine-Rich protein 2-detecting malaria rapid diagnostic test due to high sensitivity in a community with fluctuating low parasite density. Am J Trop Med Hyg **2005**; 73:199-203.

8.5 Main Text References

1. McKenzie FE, Smith DL, O'Meara WP, Riley EM. Strain theory of malaria: The first 50 years. Advances in parasitology. 2008;66:1-46.
2. Snow RW, Amratia P, Kabaria CW, Noor AM, Marsh K. The changing limits and incidence of malaria in Africa: 1939-2009. Advances in parasitology. 2012;78:169-262.
3. Bousema T, Drakeley C. Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. Clinical microbiology reviews. 2011;24(2):377-410.
4. Doolan DL, Dobano C, Baird JK. Acquired immunity to malaria. Clinical microbiology reviews. 2009;22(1):13-36.
5. WHO. World Malaria Report 2013. In: Program MC, editor. Geneva, Switzerland: World Health Organization; 2013.

6. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, et al. A world malaria map: Plasmodium falciparum endemicity in 2007. PLoS medicine. 2009;6(3):e1000048.
7. Murray CJL, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, et al. Global malaria mortality between 1980 and 2010: a systematic analysis. The Lancet. 2012;379(9814):413-31.
8. WHO. World Malaira Report 2014. In: Organization WH, editor. Geneva, Switzerland: World Health Organization; 2014.
9. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, et al. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malaria journal. 2011;10:378.
10. Tatem AJ, Campiz N, Gething PW, Snow RW, Linard C. The effects of spatial population dataset choice on estimates of population at risk of disease. Population health metrics. 2011;9:4.
11. O'Meara WP, Collins WE, McKenzie FE. Parasite prevalence: A static measure of dynamic infections. The American journal of tropical medicine and hygiene. 2007;77(2):246-9.
12. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM. Malaria. The Lancet. 2013.
13. Hay SI, Smith DL, Snow RW. Measuring malaria endemicity from intense to interrupted transmission. The Lancet Infectious Diseases. 2008;8(6):369-78.
14. Noor AM, Clements AC, Gething PW, Moloney G, Borle M, Shewchuk T, et al. Spatial prediction of Plasmodium falciparum prevalence in Somalia. Malaria journal. 2008;7:159.
15. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. The Lancet Infectious Diseases. 2004;4(6):327-36.
16. Stresman GH. Beyond temperature and precipitation: ecological risk factors that modify malaria transmission. Acta tropica. 2010;116(3):167-72.
17. Smith DL, Dushoff J, McKenzie FE. The risk of a mosquito-borne infection in a heterogeneous environment. PLoS biology. 2004;2(11):e368.

18. Weiss DJ, Bhatt S, Mappin B, Van Boeckel TP, Smith DL, Hay SI, et al. Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000-2012: A high-resolution spatiotemporal prediction. Malaria journal. 2014;13:171.
19. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA, et al. A long neglected world malaria map: Plasmodium vivax endemicity in 2010. PLoS neglected tropical diseases. 2012;6(9):e1814.
20. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature. 2005;434(7030):214-7.
21. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of Plasmodium falciparum malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. The Lancet. 2014;383(9930):1739-47.
22. Gething PW, Battle KE, Bhatt S, Smith DL, Eisele TP, Cibulskis RE, et al. Declining malaria in Africa: improving the measurement of progress. Malaria journal. 2014;13(39).
23. Ochola LA, Ayieko C, Kisia L, Magak NG, Shabani E, Ouma C, et al. Changes in antigen-specific cytokine and chemokine responses to Plasmodium falciparum antigens in a highland area of Kenya after a prolonged absence of malaria exposure. Infection and immunity. 2014;82(9):3775-82.
24. John CC, McHugh MM, Moormann AM, Sumba PO, Ofulla AV. Low prevalence of Plasmodium falciparum infection among asymptomatic individuals in a highland area of Kenya. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2005;99(10):780-6.
25. Park AW, Magori K, White BA, Stallknecht DE. When more transmission equals less disease: reconciling the disconnect between disease hotspots and parasite transmission. PloS one. 2013;8(4):e61501.
26. Pullan RL, Sturrock HJ, Soares Magalhaes RJ, Clements AC, Brooker SJ. Spatial parasite ecology and epidemiology: a review of methods and applications. Parasitology. 2012;139(14):1870-87.

27. Lefevre T, Vantaux A, Dabire KR, Mouline K, Cohuet A. Non-genetic determinants of mosquito competence for malaria parasites. PLoS pathogens. 2013;9(6):e1003365.
28. Clark TD, Greenhouse B, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, Staedke SG, et al. Factors determining the heterogeneity of malaria incidence in children in Kampala, Uganda. The Journal of infectious diseases. 2008;198(3):393-400.
29. Stevenson J, St. Laurent B, Lobo NF, Cooke MK, Kahindi SC, Oriango R, et al. Novel vectors of malaria parasites in the western highlands of Kenya. Emerging infectious diseases. 2012;18(9):1547-9.
30. Chiyaka C, Tatem AJ, Cohen JM, Gething PW, Johnston G, Gosling R, et al. The stability of malaria elimination. Science. 2013;339:909-10.
31. Okara RM, Sinka ME, Minakawa N, Mbogo CM, Hay SI, Snow RW. Distribution of the main malaria vectors in Kenya. Malaria journal. 2010;9:69.
32. Ernst KC, Lindblade KA, Koech D, Sumba PO, Kuwuor DO, John CC, et al. Environmental, socio-demographic and behavioural determinants of malaria risk in the western Kenyan highlands: a case-control study. Tropical medicine & international health : TM & IH. 2009;14(10):1258-65.
33. Tusting LS, Willey B, Lucas H, Thompson J, Kafy HT, Smith R, et al. Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis. The Lancet. 2013;382(9896):963-72.
34. Cohen JM, Ernst KC, Lindblade KA, Vulule JM, John CC, Wilson ML. Topography-derived wetness indices are associated with household-level malaria risk in two communities in the western Kenyan highlands. Malaria journal. 2008;7:40.
35. Lwetoijera DW, Kiware SS, Mageni Z, Dongus S, Harris C, Devine GJ, et al. A need for better housing to further reduce indoor malaria transmission in areas with high bednet coverage. Parasites & vectors. 2013;6(57).
36. Ye Y, Hoshen M, Louis V, Seraphin S, Traore I, Sauerborn R. Housing conditions and Plasmodium falciparum infection: protective effect of iron-sheet roofed houses. Malaria journal. 2006;5:8.

37. West PA, Protopopoff N, Rowland M, Cumming E, Rand A, Drakeley C, et al. Malaria risk factors in North West Tanzania: the effect of spraying, nets and wealth. PloS one. 2013;8(6):e65787.
38. Rehman AM, Coleman M, Schwabe C, Baltazar G, Matias A, Gomes IR, et al. How much does malaria vector control quality matter: the epidemiological impact of holed nets and inadequate indoor residual spraying. PloS one. 2011;6(4):e19205.
39. Cunha MG, Silva ES, Sepulveda N, Costa SPT, Saboia TC, Guerreiro JF, et al. Serologically defined variations in Malaria endemicity in Para State, Brazil. PloS one. 2014;9(11):e113357.
40. Churcher TS, Cohen JM, Novotny J, Ntshalintshali N, Kunene S, Cauchemez S. Measuring the path toward malaria elimination. Science. 2014;344(6189):1230-2.
41. Satitvipawee P, Wongkhang W, Pattanasin S, Hoithong P, Bhumiratana A. Predictors of malaria-association with rubber plantations in Thailand. BMC public health. 2012;12(1115).
42. Verhulst NO, Qiu YT, Beijleveld H, Maliepaard C, Knights D, Schulz S, et al. Composition of human skin microbiota affects attractiveness to malaria mosquitoes. PloS one. 2011;6(12):e28991.
43. Batista EPA, Costa EFM, Silva AA. Anopheles darlingi (Diptera: Culicidae) displays increased attractiveness to infected individuals with Plasmodium vivax gametocytes. Parasites & vectors. 2014;7(251).
44. De Moraes CM, Stanczyk NM, Betz HS, Pulido H, Sim DG, Read AF, et al. Malaria-induced changes in host odors enhance mosquito attraction. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(30):11079-84.
45. Smallegange RC, van Gemert GJ, van de Vegte-Bolmer M, Gezan S, Takken W, Sauerwein RW, et al. Malaria infected mosquitoes express enhanced attraction to human odor. PloS one. 2013;8(5):e63602.
46. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of malaria in Africa. PLoS medicine. 2005;2(12):e340.
47. McCall MB, Hopman J, Daou M, Maiga B, Dara V, Ploemen I, et al. Early interferon-gamma response against Plasmodium falciparum correlates with

interethnic differences in susceptibility to parasitemia between sympatric Fulani and Dogon in Mali. The Journal of infectious diseases. 2010;201(1):142-52.

48. Bejon P, Warimwe G, Mackintosh CL, Mackinnon MJ, Kinyanjui SM, Musyoki JN, et al. Analysis of immunity to febrile malaria in children that distinguishes immunity from lack of exposure. Infection and immunity. 2009;77(5):1917-23.
49. Greenhouse B, Ho B, Hubbard A, Njama-Meya D, Narum DL, Lanar DE, et al. Antibodies to Plasmodium falciparum antigens predict a higher risk of malaria but protection from symptoms once parasitemic. The Journal of infectious diseases. 2011;204(1):19-26.
50. Keh CE, Jha AR, Nzarubara B, Lanar DE, Dutta S, Theisen M, et al. Associations between antibodies to a panel of Plasmodium falciparum specific antigens and response to sub-optimal antimalarial therapy in Kampala, Uganda. PloS one. 2012;7(12):e52571.
51. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? Trends in parasitology. 2007;23(12):575-82.
52. Liljander A, Bejon P, Mwacharo J, Kai O, Ogada E, Peshu N, et al. Clearance of asymptomatic P. falciparum Infections Interacts with the number of clones to predict the risk of subsequent malaria in Kenyan children. PloS one. 2011;6(2):e16940.
53. Bousema T, Okell L, Felger I, Drakeley C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. Nature reviews Microbiology. 2014;12(12):833-40.
54. Okell LC, Ghani AC, Lyons E, Drakeley CJ. Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. The Journal of infectious diseases. 2009;200(10):1509-17.
55. Tran TM, Li S, Doumbo S, Doumtabe D, Huang CY, Dia S, et al. An intensive longitudinal cohort study of Malian children and adults reveals no evidence of acquired immunity to Plasmodium falciparum infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2013;57(1):40-7.

56. Filipe JA, Riley EM, Drakeley CJ, Sutherland CJ, Ghani AC. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. PLoS computational biology. 2007;3(12):e255.
57. Goncalves BP, Huang CY, Morrison R, Holte S, Kabyemela E, Prevots DR, et al. Parasite burden and severity of malaria in Tanzanian children. The New England journal of medicine. 2014;370(19):1799-808.
58. Okell LC, Bousema T, Griffin JT, Ouedraogo AL, Ghani AC, Drakeley CJ. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. Nature communications. 2012;3:1237.
59. Tiono AB, Ouedraogo A, Ogutu B, Diarra A, Coulibaly S, Gansane A, et al. A controlled, parallel, cluster-randomized trial of community-wide screening and treatment of asymptomatic carriers of Plasmodium falciparum in Burkina Faso. Malaria journal. 2013;12:79.
60. Ouedraogo AL, Bousema T, Schneider P, de Vlas SJ, Ilboudo-Sanogo E, Cuzin-Ouattara N, et al. Substantial contribution of submicroscopical Plasmodium falciparum gametocyte carriage to the infectious reservoir in an area of seasonal transmission. PloS one. 2009;4(12):e8410.
61. Churcher TS, Bousema T, Walker M, Drakeley C, Schneider P, Ouedraogo AL, et al. Predicting mosquito infection from Plasmodium falciparum gametocyte density and estimating the reservoir of infection. eLife. 2013;2:e00626.
62. Wilson AL, Chen-Hussey V, Logan JG, Lindsay SW. Are topical insect repellents effective against malaria in endemic populations? A systematic review and meta-analysis. Malaria journal. 2014;13(446).
63. Killeen GF. Characterizing, controlling and eliminating residual malaria transmission. Malaria journal. 2014;13(330).
64. Pluess B, Tanser FC, Lengeler C, Sharp BL. Indoor residual spraying for preventing malaria. Cochrane Database Syst Rev. 2010(4):CD006657.
65. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, et al. Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS medicine. 2010;7(8).

66. ter Kuile FO, Terlouw DJ, Phillips-Howard PA, Hawley WA, Friedman JF, Kolczak MS, et al. Impact of permethrin-treated bed nets on malaria and all-cause morbidity in young children in an area of intense perennial malaria transmission in western Kenya: cross-sectional survey. The American journal of tropical medicine and hygiene. 2003;68(Suppl 4):100-7.
67. Gitonga CW, Edwards T, Karanja PN, Noor AM, Snow RW, Brooker SJ. Plasmodium infection, anaemia and mosquito net use among school children across different settings in Kenya. Tropical medicine & international health : TM & IH. 2012;17(7):858-70.
68. West PA, Protopopoff N, Wright A, Kivaju Z, Tigererwa R, Mosha FW, et al. Indoor residual spraying in combination with insecticide-treated nets compared to insecticide-treated nets alone for protection against malaria: A cluster randomised trial in Tanzania. PLoS medicine. 2014;11(4):e1001630.
69. Wong J, Hamel MJ, Drakeley CJ, Kariuki S, Shi YP, Lal AA, et al. Serological markers for monitoring historical changes in malaria transmission intensity in a highly endemic region of Western Kenya, 1994-2009. Malaria journal. 2014;13(1):451.
70. Hamel MJ, Otieno P, Bayoh N, Kariuki S, Were V, Marwanga D, et al. The combination of indoor residual spraying and insecticide-treated nets provides added protection against malaria compared with insecticide-treated nets alone. The American journal of tropical medicine and hygiene. 2011;85(6):1080-6.
71. Fillinger U, Sonye G, Killeen GF, Knols BGJ, Becker N. The practical importance of permanent and semipermanent habitats for controlling aquatic stages of Anopheles gambiae sensu lato mosquitoes: Operational observations from a rural town in western Kenya. Tropical medicine & international health : TM & IH. 2004;9(12):1274-89.
72. Fillinger U, Ndenga B, Githeko A, Lindsay SW. Integrated malaria vector control with microbial larvicides and insecticide-treated nets in western Kenya: a controlled trial. Bulletin of the World Health Organization. 2009;87(9):655-65.
73. Smith DL, Cohen JM, Chiyaka C, Johnston G, Gething PW, Gosling R, et al. A sticky situation: the unexpected stability of malaria elimination.

Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2013;368(1623):20120145.

74. Clark TD, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, Greenhouse B, Staedke SG, et al. Incidence of malaria and efficacy of combination antimalarial therapies over 4 years in an urban cohort of Ugandan children. PloS one. 2010;5(7):e11759.
75. WHO. Guidelines for the treatment of malaria. 2nd ed. Geneva, Switzerland: World Health Organization; 2010.
76. WHO. Malaria case management: Operations manual. In: Program GM, editor. Switzerland: World Health Organization; 2009.
77. Pigott DM, Atun R, Moyes CL, Hay SI, Gething PW. Funding for malaria control 2006-2010: A comprehensive global assessment. Malaria journal. 2012;11(245).
78. Smith DL, Hay SI, Noor AM, Snow RW. Predicting changing malaria risk after expanded insecticide-treated net coverage in Africa. Trends in parasitology. 2009;25(11):511-6.
79. Stern DI, Gething PW, Kabaria CW, Temperley WH, Noor AM, Okiro EA, et al. Temperature and malaria trends in highland East Africa. PloS one. 2011;6(9):e24524.
80. O'Meara WP, Mangeni JN, Steketee RW, Greenwood B. Changes in the burden of malaria in sub-Saharan Africa. Lancet Infect DIs. 2010;10:545-55.
81. Feachem RGA, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, et al. Shrinking the malaria map: progress and prospects. The Lancet. 2010;376(9752):1566-78.
82. Jagannathan P, Muhindo MK, Kakuru A, Arinaitwe E, Greenhouse B, Tappero J, et al. Increasing incidence of malaria in children despite insecticide-treated bed nets and prompt anti-malarial therapy in Tororo, Uganda. Malaria journal. 2012;11:435.
83. Tatem AJ, Gething PW, Smith DL, Hay SI. Urbanization and the global malaria recession. Malaria journal. 2013;12(133).
84. Smith DL, McKenzie FE, Snow RW, Hay SI. Revisiting the basic reproductive number for malaria and its implications for malaria control. PLoS biology. 2007;5(3):e42.

85. Ouedraogo AL, Bastiaens GJ, Tiono AB, Guelbeogo WM, Kobylinski KC, Ouedraogo A, et al. Efficacy and safety of the mosquitocidal drug ivermectin to prevent malaria transmission after treatment: A double-blind, randomized, clinical trial. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2014.
86. Slater HC, Walker PG, Bousema T, Okell LC, Ghani AC. The Potential Impact of Adding Ivermectin to a Mass Treatment Intervention to Reduce Malaria Transmission: a Modelling Study. The Journal of infectious diseases. 2014.
87. Bousema T, Sutherland CJ, Churcher TS, Mulder B, Gouagna LC, Riley EM, et al. Human immune responses that reduce the transmission of Plasmodium falciparum in African populations. International journal for parasitology. 2011;41(3-4):293-300.
88. Nunes JK, Woods C, Carter T, Raphael T, Morin MJ, Diallo D, et al. Development of a transmission-blocking malaria vaccine: progress, challenges, and the path forward. Vaccine. 2014;32(43):5531-9.
89. Pollitt LC, Churcher TS, Dawes EJ, Khan SM, Sajid M, Basanez MG, et al. Costs of crowding for the transmission of malaria parasites. Evolutionary applications. 2013;6(4):617-29.
90. Ouedraogo AL, de Vlas SJ, Nebie I, Ilboudo-Sanogo E, Bousema JT, Ouattara AS, et al. Seasonal patterns of Plasmodium falciparum gametocyte prevalence and density in a rural population of Burkina Faso. Acta tropica. 2008;105(1):28-34.
91. Sawa P, Shekalaghe SA, Drakeley CJ, Sutherland CJ, Mweresa CK, Baidjoe AY, et al. Malaria transmission after artemether-lumefantrine and dihydroartemisinin-piperaquine: a randomized trial. The Journal of infectious diseases. 2013;207(11):1637-45.
92. Bousema T, Okell L, Shekalaghe S, Griffin JT, Omar S, Sawa P, et al. Revisiting the circulation time of Plasmodium falciparum gametocytes: molecular detection methods to estimate the duration of gametocyte carriage and the effect of gametocytocidal drugs. Malaria journal. 2010;9:136.
93. Okell LC, Drakeley CJ, Ghani AC, Bousema T, Sutherland CJ. Reduction of transmission from malaria patients by artemisinin combination therapies: a pooled analysis of six randomized trials. Malaria journal. 2008;7:125.

94. Okell LC, Drakeley CJ, Bousema T, Whitty CJ, Ghani AC. Modelling the impact of artemisinin combination therapy and long-acting treatments on malaria transmission intensity. PLoS medicine. 2008;5(11):e226.
95. Cotter C, Sturrock HJW, Hsiang MS, Liu J, Phillips AA, Hwang J, et al. The changing epidemiology of malaria elimination: new strategies for new challenges. The Lancet. 2013;382(9895):900-11.
96. Zhang Q, Lai S, Zheng C, Zhang H, Zhou S, Hu W, et al. The epidemiology of Plasmodium vivax and Plasmodium falciparum malaria in China, 2004 - 2012: from intensitified control to elimination. Malaria journal. 2014;13(419).
97. Sturrock HJ, Hsiang MS, Cohen JM, D.L. S, Greenhouse B, Bousema T, et al. Targeting asymptomatic malaria infections: Active surveillance in control and elimination. PLoS medicine. 2013;10(6):e1001467.
98. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the impact of human mobility on malaria. Science. 2012;338(6104):267-70.
99. O'Meara WP, Mwangi TW, Williams TN, McKenzie FE, Snow RW, Marsh K. Relationship between exposure, clinical malaria, and age in an area of changing transmission intensity. The American journal of tropical medicine and hygiene. 2008;79(2):185-91.
100. O'Meara WP, Bejon P, Mwangi TW, Okiro EA, Preshu N, Snow RW, et al. Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. Lancet. 2008;372:1555-62.
101. Roca-Feltrer A, Carneiro I, Smith L, Schellenberg JR, Greenwood B, Schellenberg D. The age patterns of severe malaria syndromes in sub-Saharan Africa across a range of transmission intensities and seasonality settings. Malaria journal. 2010;9:282.
102. Carneiro I, Roca-Feltrer A, Griffin JT, Smith L, Tanner M, Schellenberg JA, et al. Age-patterns of malaria vary with severity, transmission intensity and seasonality in sub-Saharan Africa: A systematic review and pooled analysis. PloS one. 2010;5(2):e8988.

103. Ceesay SJ, Casals-Pascual C, Erskine J, Anya SE, Duah NO, Fulford AJ, et al. Changes in malaria indices bewteen 1999 and 2007 in The Gambia: A retrospective analysis. Lancet. 2008;372:1545-54.
104. Reyburn H, Mbatia R, Drakeley C, Bruce J, Carneiro I, Olomi R, et al. Association of transmission intensity and age with clinical manifestations and case fatality of severe Plasmodium falciparum malaria. JAMA. 2005;293(12):1461-70.
105. Mutanda AL, Cheruiyot P, Hodges JS, Ayodo G, Odero W, John CC. Sensitivity of fever for diangosis of clinical malaria in a Kenyan ara of unstable, low malaria transmission. Malaria journal. 2014;13(163).
106. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS medicine. 2012;9(1):e1001165.
107. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proceedings of the National Academy of Sciences of the United States of America. 1997;94:338-42.
108. Thomson MC, Connor SJ, D'Alessandro U, Rowlingson B, Diggle PJ, Cresswell M, et al. Predicting malaria infection in Gambian children from satellite data and betnet use surveys: The importance of spatial correlation in the interpretation of results. The American journal of tropical medicine and hygiene. 1999;61(1):2-8.
109. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the Plasmodium falciparum parasite rate. Malaria journal. 2007;6:131.
110. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, et al. Identification of hot spots of malaria transmission for targeted malaria control. The Journal of infectious diseases. 2010;201(11):1764-74.
111. Greenwood B. The microepidemiology of malaria and its importance to malaria control. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1989;83(Suppl):25-9.
112. Wanzirah H, Tusting LS, Arinaitwe E, Katureebe A, Maxwell K, Rek J, et al. Mind the gap: house structure and the risk of malaria in Uganda. PloS one. 2015;10(1):e0117396.

113. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. Bull WHO. 2000;78:1401-11.
114. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, Hay SI, et al. The limits and intensity of Plasmodium falciparum transmission: Implications for malaria control and elimination worldwide. PLoS medicine. 2008;5(2):e38.
115. Cook J, Kleinschmidt I, Schwabe C, Nseng G, Bousema T, Corran PH, et al. Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, equatorial Guinea. PloS one. 2011;6(9):e25137.
116. Johnston GL, Smith DL, Fidock DA. Malaria's missing number: calculating the human component of R0 by a within-host mechanistic model of Plasmodium falciparum infection and transmission. PLoS computational biology. 2013;9(4):e1003025.
117. Smith DL, Drakeley CJ, Chiyaka C, Hay SI. A quantitative analysis of transmission efficiency versus intensity for malaria. Nature communications. 2010;1:108.
118. Tatem AJ, Smith DL, Gething PW, Kabaria CW, Snow RW, Hay SI. Ranking of elimination feasibility between malaria-endemic countries. The Lancet. 2010;376(9752):1579-91.
119. Smith DL, Dushoff J, Snow RW, Hay SI. The entomological inoculation rate and Plasmodium falciparum infection in African children. Nature. 2005;438(7067):492-5.
120. Cook J, Xu W, Msellem M, Vonk M, Bergstrom B, Gosling R, et al. Mass screening and treatment using a Falciparum-specific rapid diagnostic test did not reduce malaria incidence in Zanzibar. The Journal of infectious diseases. 2014.
121. Sturrock HJ, Novotny JM, Kunene S, Dlamini S, Zulu Z, Cohen JM, et al. Reactive case detection for malaria elimination: real-life experience from an ongoing program in Swaziland. PloS one. 2013;8(5):e63830.
122. Lutambi AM, Chitnis N, Briet OJ, Smith TA, Penny MA. Clustering of vector control interventions has important consequences for their effectiveness: a modelling study. PloS one. 2014;9(5):e97065.

123. Simon C, Moakofhi K, Mosweunyane T, Jibril HB, Nkomo B, Motlaleng M, et al. Malaria control in Botswana, 2008-2012: The path towards elimination. Malaria journal. 2013;12(458).
124. Asale A, Getachew Y, Hailesilassie W, Speybroeck N, Duchateau L, Yewhalaw D. Evaluation of the efficacy of DDT indoor residual spraying and long-lasting insecticidal nets against insecticide resistant populations of Anopheles arabiensis Patton (Diptera: Culicidae) from Ethiopia using experimental huts. Parasites & vectors. 2014;7:131.
125. Minzi O, Maige S, Sasi P, Ngasala B. Adherence to artemether-lumefantrine drug combination: A rural community experience six years after change of malaria treatment policy in Tanzania. Malaria journal. 2014;13(267).
126. Noor AM, Amin AA, Gething PW, Atkinson PM, Hay SI, Snow RW. Modelling distances travelled to government health services in Kenya. Tropical medicine & international health : TM & IH. 2006;11(2):188-96.
127. Bejon P, Williams TN, Nyundo C, Hay SI, Benz D, Gething PW, et al. A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. eLife. 2014;3:e02130.
128. Mwangi TW, Fegan G, Williams TN, Kinyanjui SM, Snow RW, Marsh K. Evidence for over-dispersion in the distribution of clinical malaria episodes in children. PloS one. 2008;3(5):e2196.
129. Bejon P, Williams TN, Liljander A, Noor AM, Wambua J, Ogada E, et al. Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. PLoS medicine. 2010;7(7):e1000304.
130. Jaenisch T, Sullivan DJ, Dutta A, Deb S, Ramsan M, Othman MK, et al. Malaria incidence and prevalence on Pemba island before the onset of the successful control intervention on the Zanzibar archipelago. Malaria journal. 2010;9:32.
131. Clarke KC, McLafferty SL, Tempalski BJ. On epidemiology and geographic information systems: A review and discussion of future directions. EID. 1996;2(2):85-92.
132. Diggle PJ, Tawn JA. Model-based geostatistics. Appl Statist. 1998;47(3):299-350.

133. Bi Y, Hu W, Yang H, Zhou XN, Yu W, Guo Y, et al. Spatial patterns of malaria reported deaths in Yunnan Province, China. The American journal of tropical medicine and hygiene. 2013;88(3):526-35.
134. Oduro AR, Bojang KA, Conway DJ, Corrah T, Greenwood BM, Schellenberg D. Health centre surveys as a potential tool for monitoring malaria epidemiology by area and over time. PloS one. 2011;6(11):e26305.
135. Nourein AB, Abass MA, Nugud AH, El Hassan I, Snow RW, Noor AM. Identifying residual foci of Plasmodium falciparum infections for malaria elimination: the urban context of Khartoum, Sudan. PloS one. 2011;6(2):e16948.
136. Clements ACA, Reid HL, Kelly GC, Hay SI. Further shrinking the malaria map: How can geospatial science help to achieve malaria elimination? The Lancet infectious diseases. 2013;13(8):709-18.
137. Gething PW, Noor AM, Gikandi PW, Hay SI, Nixon MS, Snow RW, et al. Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data. Geographical analysis. 2008;40(2):167-88.
138. Kleinschmidt I, Bagayoko M, Clarke GPY, Craig M, Le Sueur D. A spatial statistical approach to malaria mapping. International journal of epidemiology. 2000;29(2):355-61.
139. Kleinschmidt I, Sharp BL, Clarke GPY, Curtis B, Fraser C. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in KwaZulu Natal, South Africa. American journal of epidemiology. 2001;153(12):1213-21.
140. Clennon JA, Kamanga A, Musapa M, Shiff C, Glass GE. Identifying malaria vector breeding habitats with remote sensing data and terrain-based landscape indices in Zambia. International journal of health geographics. 2010;9:58.
141. Hay SI, Snow RW, Rogers DJ. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1998;92:12-20.

142. Giorgi E, Sesay SS, Terlouw DJ, Diggle PJ. Combining data from multiple spatially references prevalence surveys using generalized linear geostatistical models. Submitted.
143. Chagneau P, Mortier F, Picard N, Bacro JN. A hierarchical bayesian model for spatial prediction of multivariate non-gaussian random fields. Biometrics. 2011;67(1):97-105.
144. Gething PW, Patil AP, Hay SI. Quantifying aggregated uncertainty in Plasmodium falciparum malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. PLoS computational biology. 2010;6(4):e1000724.
145. Hay SI, Snow RW. The malaria Atlas Project: developing global maps of malaria risk. PLoS medicine. 2006;3(12):e473.
146. Diggle PJ, Thomson MC, Christensen OF, Rowlingson B, Obsomer V, Gardon J, et al. Spatial modelling and the prediction of Loa loa risk: decision making under uncertainty. Annals of tropical medicine and parasitology. 2007;101(6):499-509.
147. Jackson MC, Huang L, Luo J, Hachey M, Feuer E. Comparison of tests for spatial heterogeneity on data with global clustering patterns and outliers. International journal of health geographics. 2009;8:55.
148. Mosha JF, Sturrock HJ, Greenwood B, Sutherland CJ, Gadalla NB, Atwal S, et al. Hot spot or not: A comparison of spatial statistical methods to predict prospective malaria infections. Malaria journal. 2014;13(1):53.
149. Rogerson PA. Statistical methods for the detection of spatial clustering in case-control data. Stat Med. 2006;25(5):811-23.
150. Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. International journal of health geographics. 2006;5:15.
151. Song C, Kulldorff M. Power evaluation of disease clustering tests. International journal of health geographics. 2003;2(9).
152. Bousema T, Youssef RM, Cook J, Cox J, Alegana VA, Amran J, et al. Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. Emerging infectious diseases. 2010;16(3):392-9.

153. Gething P, Atkinson P, Noor A, Gikandi P, Hay S, Nixon M. A local space-time kriging approach applied to a national outpatient malaria dataset. Computers & geosciences. 2007;33(10):1337-50.
154. Kobayashi T, Chishimba S, Shields T, Hamapumbu H, Mharakurwa S, Thuma PE, et al. Temporal and spatial patterns of serologic responses to Plasmodium falciparum antigens in a region of declining malaria transmission in southern Zambia. Malaria journal. 2012;11:438.
155. Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I, et al. Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. BMC public health. 2006;6:286.
156. Coleman M, Coleman M, Mabuza AM, Kok G, Coetzee M, Durrheim DN. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. Malaria journal. 2009;8:68.
157. Hsiang MS, Hwang J, Kunene S, Drakeley C, Kandula D, Novotny J, et al. Surveillance for malaria elimination in Swaziland: a national cross-sectional study using pooled PCR and serology. PloS one. 2012;7(1):e29550.
158. WHO. Disease surveillance for malaria elimination: An operational manual. In: Program GM, editor. Switzerland: World Health Organization; 2012.
159. WHO. Disease surveilance for malaria control: An operational manual. In: Project GM, editor. Switzerland: World Health Organization; 2012.
160. Protopopoff N, Van Bortel W, Marcotty T, Van Herp M, Maes P, Baza D, et al. Spatial targeted vector control in the highlands of Burundi and its impact on malaria transmission. Malaria journal. 2007;6:158.
161. Okebe J, Affara M, Correa S, Muhammad AK, Nwakanma D, Drakeley C, et al. School-based countrywide seroprevalence survey reveals spatial heterogeneity in malaria transmission in the gambia. PloS one. 2014;9(10):e110926.
162. Bousema T, Kreuels B, Gosling R. Adjusting for heterogeneity of malaria transmission in longitudinal studies. The Journal of infectious diseases. 2011;204(1):1-3.
163. Smith DL, Perkins TA, Reiner RC, Jr., Barker CM, Niu T, Chaves LF, et al. Recasting the theory of mosquito-borne pathogen transmission dynamics

and control. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2014;108(4):185-97.

164. Oduro AR, Conway DJ, Schellenberg D, Satoguina J, Greenwood BM, Bojang KA. Seroepidemiological and parasitological evaluation of the heterogeneity of malaria infection in the Gambia. Malaria journal. 2013;12(222).
165. Hsiang MS, Hwang J, Tao AR, Liu Y, Bennett A, Shanks GD, et al. Mass drug administration for the control and elimiation of Plasmodium vivax malaria: An ecological study from Jiangsu province, China. Malaria journal. 2013;12(383).
166. Stresman GH, Kamanga A, Moono P, Hamapumbu H, Mharakurwa S, Kobayashi T, et al. A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. Malaria journal. 2010;9:265.
167. Cohen JM, Dlamini S, Novotny JM, Kandula D, Kunene S, Tatem AJ. Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. Malaria journal. 2013;12(61).
168. Sturrock HJ, Picon D, Sabasio A, Oguttu D, Robinson E, Lado M, et al. Integrated mapping of neglected tropical diseases: epidemiological findings and control implications for northern Bahr-el-Ghazal State, Southern Sudan. PLoS neglected tropical diseases. 2009;3(10):e537.
169. Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in Plasmodium falciparum transmission: Precision, accuracy and costs of metrics. Advances in parasitology. 2014;84:151-208.
170. malERA, TDR, WHO. A research agenda for malaria eradication: diagnoses and diagnostics. PLoS medicine. 2011;8(1):e1000396.
171. Gimnig JE, Walker ED, Otieno P, Kosgei J, Olang G, Ombok M, et al. Incidence of malaria among mosquito collectors conducting human landing catches in western Kenya. The American journal of tropical medicine and hygiene. 2013;88(2):301-8.
172. Kilama M, Smith DL, R. H, Kigozi R, Yeka A, Lavoy G, et al. Estimating the annual entomological inoculation rate for Plasmodium falciparum transmitted by Anopheles gambiae s.l. using three sampling methods in three sites in Uganda. Malaria journal. 2014;13(111).

173. WHO. Malaria Microscopy Quality Assurance Manuel Version 1. Geneva, Switzerland: World Health Organization; 2009.
174. McMorrow ML, Aidoo M, Kachur SP. Malaria rapid diagnostic tests in elimination settings - can they find the last parasite? Clin Microbiol Infect. 2011;17:1624-31.
175. Manjurano A, Okell L, Lukindo T, Reyburn H, Olomi R, Roper C, et al. Association of sub-microscopic malaria parasite carriage with transmission intensity in north-eastern Tanzania. Malaria journal. 2011;10:370.
176. Alexander N, Schellenberg D, Ngasala B, Petzold M, Drakeley C, Sutherland C. Assessing agreement between malaria slide density readings. Malaria journal. 2010;9:4.
177. Koepfli C, Schoepflin S, Bretscher M, Lin E, Kiniboro B, Zimmerman PA, et al. How much remains undetected? Probability of molecular detection of human Plasmodia in the field. PloS one. 2011;6(4):e19010.
178. Wafula R, Sang E, Cheruiyot O, Aboto A, Menya D, O'Meara WP. High sensitivity and specificity of clinical microscopy in rural health facilities in western Kenya under an external quality assurance program. The American journal of tropical medicine and hygiene. 2014;91(3):481-5.
179. Patil AP, Okiro EA, Gething PW, Guerra CA, Sharma SK, Snow RW, et al. Defining the relationship between Plasmodium falciparum parasite rate and clinical disease: statistical models for disease burden estimation. Malaria journal. 2009;8:186.
180. Zhao J, Lama M, Korenromp E, Aylward P, Shargie E, Filler S, et al. Adoption of rapid diagnsotic tests for the diagnosis of malaria, a preliminary analysis of the global fund program data, 2005 to 2010. PloS one. 2012;7(8):e43549.
181. Barber BE, William T, Grigg MJ, Piera K, Yeo TW, Anstey NM. Evaluation of the sensitivity of a pLDH-based and an aldolase-based rapid diagnostic test for diagnosis of uncomplicated and severe malaria caused by PCR-confirmed Plasmodium knowlesi, Plasmodium falciparum, and Plasmodium vivax. Journal of clinical microbiology. 2013.
182. WHO. Malaria rapid diagnostic test performance: Results of WHO product testing of malaria RDTs: Round 4 (2012). In: Diseases SPfRaTiT, editor. Switzerland: World Health Organization; 2012.

183. Masaninga F, Sekeseke-Chinyama M, Malambo T, Moonga H, Babaniyi O, Counihan H, et al. Finding parasites and finding challenges; Improved diagnostic access and trends in reported malaria and anti-malarial drug use in Livingstone district, Zambia. Malaria journal. 2012;11(341).
184. Thiam S, Thwing J, Diallo I, Fall FB, Diouf MB, Perry R, et al. Scale-up of home-based management of malaria based on rapid diagnostic tests and artemisin-based combination therapy in a resource-poor country: Results in Senegal. Malaria journal. 2012;11(334).
185. Aydin-Schmidt B, Mubi M, Morris U, Petzold M, Ngasala BE, Premji Z, et al. Usefulness of Plasmodium falciparum-specific rapid diagnositc tests for assessment of parasite clearance and detection of recurrent infections after artemisinin-based combination therapy. Malaria journal. 2013;12:349.
186. Kyabayinze DJ, Tibenderana JK, Odong GW, Rwakimari JB, Counihan H. Operational accuracy and comparative persistent antigenicity of HRP2 rapid diagnostic tests for Plasmodium falciparum malaria in a hyperendemic region of Uganda. Malaria journal. 2008;7:221.
187. Shakely D, Elfving K, Aydin-Schmidt B, Msellem MI, Morris U, Omar R, et al. The usefulness of rapid diagnostic tests in the new context of low malaria transmission in Zanzibar. PloS one. 2013;8(9):e72912.
188. Baltzell KA, Shakely D, Hsiang M, Kemere J, Ali AS, Bjorkman A, et al. Prevalence of PCR detectable malaria infection among febrile patients with a negative Plasmodium falciparum specific rapid diagnostic test in Zanzibar. The American journal of tropical medicine and hygiene. 2013;88(2):289-91.
189. Mosha JF, Conteh L, Tediosi F, Gesase S, Bruce J, Chandramohan D, et al. Cost implications of improving malaria diagnosis: findings from north-eastern Tanzania. PloS one. 2010;5(1):e8707.
190. Kyabayinze DJ, Asiimwe C, Nakanjako D, Nabakooza J, Counihan H, Tibenderana JK. Use of RDTs to improve malaria diagnosis and fever case management at primary health care facilities in Uganda. Malaria journal. 2010;9:200.
191. Kamuliwo M, Chanda E, Haque U, Mwanza-Ingwe M, Sikaala C, Katebe-Sakala C, et al. The changing burden of malaria and association with vector

control interventions in Zambia using district-level surveillance data, 2006-2011. Malaria journal. 2013;12(437).

192. Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, et al. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. Molecular and biochemical parasitology. 1993;61:315-20.
193. Steenkeste N, Incardona S, Chy S, Duval L, Ekala MT, Lim P, et al. Towards high-throughput molecular detection of Plasmodium: new approaches and molecular markers. Malaria journal. 2009;8:86.
194. Singh B, Cox-Singh J, Miller AO, Abdullah MS, Snounou G, Rahman HA. Detection of malaria in Malaysia by nested polymerase chain reaction amplification of dried blood spots on filter paper. 1996.
195. Steenkeste N, Rogers WO, Okell L, Jeanne I, Incardona S, Duval L, et al. Sub-microscopic malaria cases and mixed malaria infection in a remote area of high malaria endemicity in Rattanakiri province, Cambodia: Implication for malaria elimination. Malaria journal. 2010;9:108.
196. Oriero EC, Jacobs J, Van Geertruyden JP, Nwakanma D, D'Alessandro U. Molecular-based isothermal tests for field diagnosis of malaria and their potential contribution to malaria elimination. The Journal of antimicrobial chemotherapy. 2014.
197. Polley SD, Gonzalez IJ, Mohamed D, Daly R, Bowers K, Watson J, et al. Clinical evaluation of a loop-mediated amplification kit for diagnosis of imported malaria. The Journal of infectious diseases. 2013;208(4):637-44.
198. Patel JC, Oberstaller J, Xayavong M, Narayanan J, DeBarry JD, Srinivasamoorthy G, et al. Real-time loop-mediated isothermal amplification (RealAmp) for the species-specific identification of Plasmodium vivax. PloS one. 2013;8(1):e54986.
199. Polley SD, Mori Y, Watson J, Perkins MD, Gonzalez IJ, Notomi T, et al. Mitochondrial DNA targets increase sensitivity of malaria detection using loop-mediated isothermal amplification. Journal of clinical microbiology. 2010;48(8):2866-71.
200. Patel JC, Lucchi NW, Srivastava P, Lin JT, Sug-Aram R, Aruncharus S, et al. Field evaluation of a real-time fluorescence loop-mediated isothermal

amplification assay, RealAmp, for the diagnosis of malaria in Thailand and India. The Journal of infectious diseases. 2014;210(8):1180-7.

201. Faucher JF, Aubouy A, Beheton T, Makoutode P, Abiou G, Doritchamou J, et al. What would PCR assessment change in the management of fevers in a malaria endemic area? A school-based study in Benin in children with and without fever. Malaria journal. 2010;9:224.
202. Proietti C, Verra F, Bretscher MT, Stone W, Kanoi BN, Balikagala B, et al. Influence of infection on malaria-specific antibody dynamics in a cohort exposed to intense malaria transmission in northern Uganda. Parasite immunology. 2013;35(5-6):164-73.
203. Drakeley C, Cook J. Chapter 5 Potential Contribution of Sero‐Epidemiological Analysis for Monitoring Malaria Control and Elimination: Historical and Current Perspectives. 2009;69:299-352.
204. Corran PH, Cook J, Lynch C, Leendertse H, Manjurano A, Griffin J, et al. Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. Malaria journal. 2008;7:195.
205. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SL, Carneiro I, et al. 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. 2005;102(14):5108-13.
206. Shekalaghe S, Alifrangis M, Mwanziva C, Enevold A, Mwakalinga S, Mkali H, et al. Low density parasitaemia, red blood cell polymorphisms and Plasmodium falciparum specific immune responses in a low endemic area in northern Tanzania. BMC infectious diseases. 2009;9:69.
207. Rolfes MA, McCarra M, Magak NG, Ernst KC, Dent AE, Lindblade KA, et al. Development of clinical immunity to malaria in highland areas of low and unstable transmission. The American journal of tropical medicine and hygiene. 2012;87(5):806-12.
208. Noor AM, ElMardi KA, Abdelgader TM, Patil AP, Amine AA, Bakhiet S, et al. Malaria risk mapping for control in the republic of Sudan. The American journal of tropical medicine and hygiene. 2012;87(6):1012-21.

209. Samadoulougou S, Maheu-Giroux M, Kirakoya-Samadoulougou F, De Keukeleire M, Castro MC, Robert A. Multilevel and geo-statistical modeling of malaria risk in children of Burkina Faso. Parasites & vectors. 2014;7(350).
210. Moss WJ, Hamapumbu H, Kobayashi T, Shields T, Kamanga A, Clennon J, et al. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: A cross-sectional and longitudinal community survey. Malaria journal. 2011;10:163.
211. Alemu K, Worku A, Berhane Y. Malaria infection has spatial, temporal, and spatiotemporal heterogeneity in unstable malaria transmission areas in northwest Ethiopia. PloS one. 2013;8(11):e79966.
212. Ashton RA, Kefyalew T, Tesfaye G, Pullan RL, Yadeta D, Reithinger R, et al. School-based surveys of malaria in Oromia Regional State, Ethiopia: a rapid survey method for malaria in low transmission settings. Malaria journal. 2011;10(1):25.
213. Bejon P, Turner L, Lavstsen T, Cham G, Olotu A, Drakeley CJ, et al. Serological evidence of discrete spatial clusters of Plasmodium falciparum parasites. PloS one. 2011;6(6):e21711.
214. Zachariah R, Harries AD, Ishikawa N, Rieder HL, Bissell K, Laserson K, et al. Operational research in low-income countries: What, why, and how? Lancet infectious diseases. 2009;9:711-7.
215. Zachariah R, Ford N, Maher D, Bissell K, Van den Bergh R, van den Boogaard W, et al. Is operational research delivering the goods? The journey to success in low-income countries. The Lancet infectious diseases. 2012;12(5):415-21.
216. Moonen B, Cohen JM, Tatem AJ, Cohen J, Hay SI, Sabot O, et al. A framework for assessing the feasibility of malaria elimination. Malaria journal. 2010;9:322.
217. Cao J, Sturrock HJ, Cotter C, Zhou S, Zhou H, Liu Y, et al. Communicating and monitoring surveillance and response activities for malaria elimination: China's "1-3-7" strategy. PLoS medicine. 2014;11(5):e1001642.
218. Guerin PJ, Olliaro P, Nosten F, Druilhe P, Laxminarayan R, Binka F, et al. Malaria: current status of control, diagnosis, treatment, and a proposed

agenda for research and development. The Lancet infectious diseases. 2002;2(9):564-73.

219. Zhou SS, Rietveld AE, Velarde-Rodriguez M, Ramsay AR, Zhang SS, Zhou XN, et al. Operational research on malaria control and elimination: a review of projects published between 2008 and 2013. Malaria journal. 2014;13(1):473.
220. Breman JG, Brandling-Bennett AD. The challenge of malaria eradication in the twenty-first century: research linked to operations is the key. Vaccine. 2011;29 Suppl 4:D97-103.
221. malERA, TDR, WHO. A research agenda for malaria eradication: health systems and operational research. PLoS medicine. 2011;8(1):e1000397.
222. Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, Smith DL, et al. Operational strategies to achieve and maintain malaria elimination. The Lancet. 2010;376(9752):1592-603.
223. Afrane Y, Zhou G, Githeko A, Yan G. Utility of health facility-based malaria data for malaria surveillance. PloS one. 2013;8(2):e54305.
224. Cibulskis R, Bell D, Christophel E-M, Hii J, Delacollette C, Bakkyaita N, et al. Estimating trends in the burden of malaria at country level. The American journal of tropical medicine and hygiene. 2007;77(Suppl 6):1330137.
225. Malaria RB. Malaria Indicator Survey: Basic Documentation for Survey Design and Implementation 2013 [Jan 2015]. Available from: http://www.malariasurveys.org.
226. Chirombo J, Lowe R, Kazembe L. Using structured additive regression models to estimate risk factors of malaria: analysis of 2010 Malawi malaria indicator survey data. PloS one. 2014;9(7):e101116.
227. Eyobo MB, Awur AC, Wani G, Julla AI, Remijo CD, Sebit B, et al. Malaria indicator survey 2009, South Sudan: Baseline results at household level. Malaria journal. 2014;13:45.
228. Cox J, Sovannaroth S, Soley LD, Ngor P, Mellor S, Roca-Feltrer A. Novel approaches to risk stratification to support malaria elimination: An example from Cambodia. Malaria journal. 2014;13(371).
229. Satoguina J, Walther B, Drakeley C, Nwakanma D, Oriero EC, Correa S, et al. Comparison of surveillance methods applied to a situation of low malaria

prevalence at rural sites in The Gambia and Guinea Bissau. Malaria journal. 2009;8:274.

230. malERA, TDR, WHO. A research agenda for malaria eradication: Monitoring, evaluation, and surveillance. PLoS medicine. 2011;8(1):e1000400.
231. Griffin JT, Ferguson NM, Ghani AC. Estimates of the changing age-burden of Plasmodium falciparum malaria disease in sub-Saharan Africa. Nature communications. 2014;5:3136.
232. Brooker S, Kolaczinski JH, Gitonga CW, Noor AM, Snow RW. The use of schools for malaria surveillance and programme evaluation in Africa. Malaria journal. 2009;8:231.
233. Proietti C, Pettinato DD, Kanoi BN, Ntege E, Crisanti A, Riley EM, et al. Continuing intense malaria transmission in northern Uganda. The American journal of tropical medicine and hygiene. 2011;84(5):830-7.
234. Shekalaghe SA, Drakeley C, van den Bosch S, ter Braak R, van den Bijllaardt W, Mwanziva C, et al. A cluster-randomized trial of mass drug administration with a gametocytocidal drug combination to interrupt malaria transmission in a low endemic area in Tanzania. Malaria journal. 2011;10:247.
235. Sutcliffe CG, Kobayashi T, Hamapumbu H, Shields T, Mharakurwa S, Thuma PE, et al. Reduced risk of malaria parasitemia following household screening and treatment: a cross-sectional and longitudinal cohort study. PloS one. 2012;7(2):e31396.
236. Koita K, Novotny J, Kunene S, Zulu Z, Ntshalintshali N, Gandhi M, et al. Targeting imported malaria through social networks: a potential strategy for malaria elimination in Swaziland. Malaria journal. 2013;12(219).
237. Badu K, Afrane YA, Larbi J, Stewart VA, Waitumbi J, Angov E, et al. Marked variation in MSP-119 antibody responses to malaria in western Kenyan highlands. BMC infectious diseases. 2012;12:50.
238. Alemu A, Muluye D, Mihret M, Adugna M, Gebeyaw M. Ten year trend analysis of malaria prevalence in Kola District, North Gondar, northwest Ethiopia. Parasites & vectors. 2012;5(173).

239. Bousema T, Stevenson J, Baidjoe A, Stresman G, Griffin JT, Kleinschmidt I, et al. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. Trials. 2013;14:36.
240. WHO, TDR. Evaluation of rapid diagnostic tests: Malaria. Nature Reviews: Microbiology. 2006;4:S34-S8.
241. Elliott SR, Fowkes FJ, Richards JS, Reiling L, Drew DR, Beeson JG. Research priorities for the development and implementation of serological tools for malaria surveillance. F1000prime reports. 2014;6:100.
242. Greenwood B, Bojang K, Tagbor H, Pagnoni F. Combining community case management and intermittent preventive treatment for malaria. Trends in parasitology. 2011;27(11):477-80.
243. Nankabirwa J, Brooker SJ, Clarke SE, Fernando D, Gitonga CW, Schellenberg D, et al. Malaria in school-age children in Africa: an increasingly important challenge. Tropical medicine & international health : TM & IH. 2014;19(11):1294-309.
244. Halliday KE, Okello G, Turner EL, Njagi K, McHaro C, Kengo J, et al. Impact of intermittent screening and treatment for malaria among school children in Kenya: a cluster randomised trial. PLoS medicine. 2014;11(1):e1001594.
245. Ndiaye Y, Ndiaye JLA, Cisse B, Blanas D, Bassene J, Manga IA, et al. Community case management in malaria: review and perspectives after four years of operational experience in Saraya district, south-east Senegal. Malaria journal. 2013;12(240).
246. Takem EN, Affara M, Amambua-Ngwa A, Okebe J, Ceesay SJ, Jawara M, et al. Detecting Foci of Malaria Transmission with School Surveys: A Pilot Study in the Gambia. PloS one. 2013;8(6):e67108.
247. Halliday KE, Karanja P, Turner EL, Okello G, Njagi K, Dubeck MM, et al. Plasmodium falciparum, anaemia and cognitive and educational performance among school children in an area of moderate malaria transmission: baseline results of a cluster randomized trial on the coast of Kenya. Tropical medicine & international health : TM & IH. 2012;17(5):532-49.
248. Brooker S, Okello G, Njagi K, Dubeck MM, Halliday KE, Inyega H, et al. Improving educational achievement and anaemia of school children: design

of a cluster randomised trial of school-based malaria prevention and enhanced literacy instruction in Kenya. Trials. 2010;11:93.

249. Stewart L, Gosling R, Griffin J, Gesase S, Campo J, Hashim R, et al. Rapid assessment of malaria transmission using age-specific sero-conversion rates. PloS one. 2009;4(6):e6083.
250. Rowe AK, Kachur SP, Yoon SS, Lynch M, Slutsker L, Steketee RW. Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa. Malaria journal. 2009;8:209.
251. Juma E, Zurovac D. Changes in health workers' malaria diagnosis and treatment practices in Kenya. Malaria journal. 2011;10:1.
252. Gimbel S, Micek M, Lambdin B, Lara J, Karagianis M, Cuembelo F, et al. An assessment of routine primary care health information system data quality in Sofala Province, Mozambique. Population health metrics. 2011;9:12.
253. Peters DH, Garg A, Bloom G, Walker DG, Brieger WR, Rahman MH. Poverty and access to health care in developing countries. Annals of the New York Academy of Sciences. 2008;1136:161-71.
254. Sumba PO, Wong SL, Kanzaria HK, Johnson KA, John CC. Malaria treatment-seeking behaviour and recovery from malaria in a highland area of Kenya. Malaria journal. 2008;7:245.
255. Rutherford M, Dockerty JD, Jasseh M, Howie SRC, Herbison P, Jeffries DJ, et al. Access to health care and mortality of children under 5 years of age in the Gambia: a case-control study. Bulletin of the World Health Organization. 2009;87(3):216-25.
256. O'Meara WP, Noor A, Gatakaa H, Tsofa B, McKenzie FE, Marsh K. The impact of primary health care on malaria morbidity--defining access by disease burden. Tropical medicine & international health : TM & IH. 2009;14(1):29-35.
257. Noor AM, Zurovac D, Hay SI, Ochola SA, Snow RW. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. Tropical medicine & international health : TM & IH. 2003;8(10):917-26.
258. Cibulskis RE, Pujari S, Otten MW. Do estimates of intervention coverage obtained from children at immunization clinics provide a reasonable

approximation to population values? The Journal of infectious diseases. 2012;205 Suppl 1:S91-102.

259. Stuckey EM, Stevenson J, Cooke MK, Owaga C, Marube E, Oando G, et al. Simulation of malaria epidemiology and control in the highlands of western Kenya. Malaria journal. 2013;11(357).
260. Sanitation MoPHa. 2010 Kenya Malaria Indicator Survey. In: Control DoM, editor. Nairobi, Kenya: Ministry of Public Health and Sanitation; 2011.
261. Hay SI, Noor AM, Simba M, Busolo M, Guyatt HL, Ochola SA, et al. Clinical epidemiology of malaria in the highlands of western Kenya. Emerging infectious diseases. 2002;8(6):543-8.
262. Omwami EM, Omwami RK. Public investment and the goal of providing universal access to primary education by 2015 in Kenya. International Journal of Educational Development. 2010;30(3):243-53.
263. Gitonga CW, Karanja PN, Kihara J, Mwanje M, Juma E, Snow RW, et al. Implementing school malaria surveys in Kenya: towards a national surveillance system. Malaria journal. 2010;9:306.
264. Noor AM, Alegana VA, Gething PW, Snow RW. A spatial national health facility database for public health sector planning in Kenya in 2008. International journal of health geographics. 2009;8:13.
265. Sanitation MoPHa. National Guidelines for HIV Testing and Counselling in Kenya. In: Programme NAaSC, editor. 2nd ed. Nairobi, Kenya: Ministry of Public Health and Sanitation; 2010.
266. Wakaba M, Mbindyo P, Ochieng J, Kiriinya R, Todd J, Waudo A, et al. The public sector nursing workforce in Kenya: a county-level analysis. Human resources for health. 2014;12:6.
267. Wesolowski A, Stresman G, Eagle N, Stevenson J, Owaga C, Marube E, et al. Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. Nature Scientific Reports. 2014;4(5678).
268. Alemu K, Worku A, Berhane Y, Kumie A. Spatiotemporal clusters of malaria cases at village level, northwest Ethiopia. Malaria journal. 2014;13:223.

269. Atkinson JA, Johnson ML, Wijesinghe R, Bobogare A, Losi L, O'Sullivan M, et al. Operational research to inform a sub-national surveillance intervention for malaria elimination in Solomon Islands. Malaria journal. 2012;11:101.
270. Chu BK, Deming M, Biritwum NK, Bougma WR, Dorkenoo AM, El-Setouhy M, et al. Transmission assessment surveys (TAS) to define endpoints for lymphatic filariasis mass drug administration: a multicenter evaluation. PLoS neglected tropical diseases. 2013;7(12):e2584.
271. Leslie T, Mikhail A, Mayan I, Anwar M, Bakhtash S, Nader M, et al. Overdiagnosis and mistreatment of malaria among febrile patients at primary healthcare level in Afghanistan: observational study. Bmj. 2012;345:e4389.
272. Kamanga A, Moono P, Stresman G, Mharakurwa S, Shiff C. Rural health centres, communities and malaria case detection in Zambia using mobile telephones: a means to detect potential reservoirs of infection in unstable transmission conditions. Malaria journal. 2010;9:96.
273. Altman DG, Bland JM. Measurement in Medicine: The analysis of method comparison studies. Journal of the Royal Statistica Society Series D (The Statistician). 1983;32(3):307-17.
274. Gething PW, Noor AM, Gikandi PW, Ogara EA, Hay SI, Nixon MS, et al. Improving imperfect data from health management information systems in Africa using space-time geostatistics. PLoS medicine. 2006;3(6):e271.
275. Mwanziva C, Shekalaghe S, Ndaro A, Mengerink B, Megiroo S, Mosha F, et al. Overuse of artemisinin-combination therapy in Mto wa Mbu (river of mosquitoes), an area misinterpreted as high endemic for malaria. Malaria journal. 2008;7:232.
276. Johansson EW, Gething PW, Hildenwall H, Mappin B, Petzold M, Peterson SS, et al. Diagnostic testing of pediatric fevers- Meta-analysis of 13 national surveys assessing influences of malaria endemicicity and source of care on test uptake for frebrile children under five years. PloS one. 2014;9(4):e95483.
277. Okiro EA, Alegana VA, Noor AM, Snow RW. Changing malaria intervention coverage, transmission and hospitalization in Kenya. Malaria journal. 2010;9:285.

278. Baidjoe A, Stone W, Ploemen I, Shagari S, Grignard L, Osoti V, et al. Combined DNA extraction and antibody elution from filter papers for the assessment of malaria transmission intensity in epidemiological studies. Malaria journal. 2013;12:272.
279. Toure OA, Kone PL, Coulibaly MAA, Ako BAA, Gbessi EA, Coulibaly B, et al. Coverage and efficacy of intermittent preventive treatment with sulphadoxine pyrimethamine against malaria in pregnancy in Cote d'Ivoire five years after its implementation. Parasites & vectors. 2014;7(495).
280. MacPherson P, Choko AT, Webb EL, Thindwa D, Squire SB, Sambakunsi R, et al. Development and validation of a global positioning system-based "map book" system for categorizing cluster residency status of community members living in high-density urban slums in Blantyre, Malawi. American journal of epidemiology. 2013;177(10):1143-7.
281. Hamainza B, Moonga H, Sikaala CH, Kamuliwo M, Bennett A, Eisele TP, et al. Monitoring, characterization and control of chronic, symptomatic malaria infections in rural Zambia through monthly household visits by paid community health workers. Malaria journal. 2014;13(1):128.
282. Cook J, Reid H, Iavro J, Kuwahata M, Taleo G, Clements A, et al. Using serological measures to monitor changes in malaria transmission in Vanuatu. Malaria journal. 2010;9:169.
283. Cook J, Speybroeck N, Sochanta T, Somony H, Sokny M, Claes F, et al. Sero-epidemiological evaluation of changes in Plasmodium falciparum and Plasmodium vivax transmission patterns over the rainy season in Cambodia. Malaria journal. 2012;11:86.
284. Drake TL, Okello G, Njagi K, Halliday KE, Jukes M, Mangham L, et al. Cost analysis of school-based intermittent screening and treatment of malaria in Kenya. Malaria journal. 2011;10:273.
285. Sturrock HJ, Cohen JM, Keil P, Tatem AJ, Le Menach A, Ntshalintshali NE, et al. Fine-scale malaria risk mapping from routine aggregated case data. Malaria journal. 2014;13(1):421.
286. Tatem AJ, Huang Z, Narib C, Kumar U, Kandula D, Pindolia DK, et al. Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. Malaria journal. 2014;13(52).

287. Mosha JF, Sturrock HJ, Greenhouse B, Greenwood B, Sutherland CJ, Gadalla N, et al. Epidemiology of subpatent Plasmodium falciparum infection: implications for detection of hotspots with imperfect diagnostics. Malaria journal. 2013;12:221.

Appendix 1 – Other Publications

Appendix 1.1 – Geolocation

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk
Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Epidemiology and Infection		
When was the work published	22/03/2014		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**CC-BY**	Was the work subject to academic pier review?	**Yes**

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	
Please list the paper's authors in the intended authorship order.	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Conceived the study, collected the data, conducted all database management and analysis. Prepared the manuscript.

Student Signature: ____  **_____ Date: ___17/02/2015_____**

Supervisor Signature: _____  **_____ Date: ___17/02/2015_____**

Title: Validation of three geolocation strategies for health facility attendees for research and public health surveillance in a rural setting in western Kenya

Authors: G. H. Stresman*[1], J. C. Stevenson[2,3,4], C. Owaga[3], E. Marube[3], C. Anyango[3], C. Drakeley[1], T. Bousema[1,5], J. Cox[2]

Affiliations:

[1] Department of Immunology & Infection, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London UK

[2] Department of Disease Control, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London UK

[3] Kenya Medical Research Institute, Centre for Global Health Research, Kisumu Kenya

[4] Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

[5] Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands

*Corresponding Author

Reference: Stresman et al., (2014) Epidemiology and Infection; 149(9): 1978-89.

Summary

Understanding the spatial distribution of disease is critical for effective disease control. Where formal address networks do not exist, tracking spatial patterns of clinical disease is difficult. Geolocation strategies were tested at rural health facilities in western Kenya. Methods included geocoding residence by head of compound, participatory mapping and recording the self-reported nearest landmark. Geocoding was able to locate 72.9% (95% CI: 67.7-77.6) of individuals to within 250 m of the true compound location. The participatory mapping exercise was able to correctly locate 82.0% of compounds (95% CI: 78.9-84.8) to a 2 km x 2.5 km area with a 500 m buffer. The self-reported nearest landmark was able to locate 78.1% (95% CI: 73.8-82.1) of compounds to the correct catchment area. These strategies tested provide options for quickly obtaining spatial information on individuals presenting at health facilities.

INTRODUCTION

Many infectious diseases show microepidemiological geographical variation. Outbreaks of (emerging) infectious diseases may be geographically confined or start in small pockets that later give rise to larger outbreaks [1-4]. For endemic infectious diseases with stable disease transmission, considerable geographical heterogeneity in the intensity of transmission has been described [2, 5-8]. Geographical variation for both epidemic and endemic infectious disease occurrence has important public health consequences. Identifying regions with higher disease burden can facilitate cost-effective prioritization of control efforts [9-11]. Within regions, identifying areas of persistent and intense transmission may prevent outbreaks of disease that spread from these areas and support disease elimination strategies when overall disease occurrence has declined [2, 12, 13]. To allow spatial targeting of disease control efforts, attributing a geographic location to each disease occurrence is ideal, and the minimum number required for accurate monitoring is likely to be disease specific [9, 14, 15].

Given adequate address information, automated geocoding software packages can generate accurate spatial coordinate data for a large proportion of individuals [16, 17], thereby providing a basis for the spatial analysis of disease transmission [18-20]. In circumstances where formal address data are unavailable or privacy concerns limit the use of precise spatial locations, other approaches have been used to obtain geographical information on incident cases. Catchment areas of, for example, community pharmacies or general practitioners have been used for describing spatial patterns in disease occurrence [6, 15, 20-22]. In areas with well-developed public health infrastructure, catchment areas tend to be well defined and sufficiently small to allow a meaningful attribution of localities to clinical cases based on the facility they attended [20, 22]. Geolocation approaches are likely to have less utility for resource poor settings where formal address systems are commonly unavailable and where health facility catchment areas are relatively large and poorly defined [5, 23, 24]. Alternative approaches to geolocation strategies are needed in such settings.

Two of the most commonly used geolocation strategies for rural resource-poor environments are distributing compound ID cards after an enumeration exercise or actively visiting compounds and geolocating the area of residence for individuals of interest [25]. Although these methods provide accurate spatial information, they are not operationally attractive outside research settings [10, 21, 25]. Approaches that can be implemented without the need for house-to-house visits would facilitate the incorporation of spatial information into routine data collection and public health planning at the local level. If this can be done with sufficient precision it would support the identification of local-level disease heterogeneity [5, 18, 25].

Here, we examine the accuracy and precision of three approaches to geolocate health facility attendees in a rural area of western Kenya: geocoding on name of head of compound, participatory mapping using satellite imagery, and attributing participants to the catchment area of the self-reported nearest landmark.

METHODS

Study site

The study was conducted in a rural area of Rachuonyo South district, Nyanza Province in the western Kenyan highlands that spans approximately 300 km^2. There is one main road that runs through the area and the landscape consists of rolling hills and several large rivers (figure 1). The population mostly comprises people from the Luo ethnic group whose main occupation is subsistence agriculture. Compounds typically comprise extended families living in proximity to their fields or in multi-unit structures in the few, more urban, market centres [26].

Five rural health facilities were identified whose catchments overlapped with community-based cross-sectional surveys being carried out (figure 1) [27]. Cross-sectional malaria surveys in the health facilities were conducted in October 2011 and in July 2012 to coincide with the bimodal seasonal peaks in malaria transmission. Four of the five health facilities were sampled during both surveys. One facility was replaced for the second year to maximize overlap with the ongoing community work. All patients and accompanying individuals attending the

outpatient clinic were recruited for the survey. A questionnaire was administered to all consenting participants to obtain information on malaria indicators and their area of residence, as described below. Tracing individual compounds from health facility attendees is a laborious and costly exercise because of the large catchment areas and inaccessible terrain and could therefore not be completed for all attendees. For operational reasons, following the facility survey, 30% of participants were randomly selected and traced to their compounds, to validate the geolocation strategies being tested. Compounds were mapped using a GPS receiver.

Geolocation strategies

Method 1: Geocoding. A system of geocoding was developed to match 'postal addresses' to an existing spatial database. In this setting in rural Kenya, compounds are known by the name of the compound head, usually the patriarch of the family. Individuals have three names, two given and one family name. Names of the compound head were collected as part of the questionnaire at the facility. Names were matched to an existing database of names of compound heads with associated spatial coordinates collected as part of a large cross-sectional survey in the area. This community survey sampled approximately one third of the population [27]. As not all compounds were sampled during the community survey, the names of the three nearest neighbors were also collected at the facility to increase the probability of finding a match. This method would be useful in areas that have existing and updated registries with accompanying spatial information and could easily be applied to all scales, depending on the availability of baseline data.

Analysis was restricted to those compounds located in the area of the community survey. Names from the two databases were matched using Levenshtein's distance algorithm [28] for string matching using Stata (version 12.1; StataCorp, TX, USA). Possible matches, where the matching probability was ≥80%, were checked manually. Matches were discarded if: (a) there was more than one compound head with the same name in either database; (b) if only one of the three names was recorded; or (c) if all three names were provided but at least one of the names did

not match. This process was repeated for the names of the nearest neighbors. All likely matches were plotted in ArcGIS (version 10.1; ESRI, CA, USA) and the distance between the actual geolocated compound and the matched compound from the community survey was calculated. Compounds from the health facility survey were considered successfully located if they were less than 250 m from the corresponding compound in the community survey. This resolution was a pragmatic choice as it was deemed an acceptable balance between accuracy and spatial resolution, as this area would only likely comprise 2 or 3 compounds.

Method 2: Participatory mapping. The second method assessed was participatory mapping, and was similar to the recently published 'map-book' exercise [25] and involved producing poster-sized, high-resolution satellite images (Quickbird; Digital Globe, CO, USA) of each facility catchment area (Figure 2). Locations of health facilities, schools, markets and other key landmarks were labeled on the image and a reference grid consisting of 2 km by 2.5 km 'blocks' was superimposed on the area [27]. Each block comprised 20 'cells', each measuring 500 x 500 m. Each block/cell combination was given a unique numeric identifier. The system (including size of polygon) was selected because it was familiar to the field workers and would provide them a better frame of reference for facilitating the exercise. As part of the participant questionnaire, the interviewer would explain the main features of the satellite map and with the participant, would attempt to locate the residence on the map and record the corresponding cell identifier. Due to the spatial resolution required to locate compounds, this approach is most applicable to local scale but could be scaled up if satellite imagery was indexed into a book-format instead of a poster.

Locations of participants followed to their compounds were plotted in ArcGIS and were classified as correctly located based on the participatory mapping exercise if they fell within the reported cell. To account for the likely edge effect with compounds located just outside a grid cell being considered incorrect, the proportion of compounds correctly identified within 500 m (one cell) or 1000 m (two cells) surrounding the reported block/cell was also calculated. The distance

between the edge of the cell/buffer and the incorrectly located compounds was calculated in ArcGIS to determine the mean error associated with the approach.

Method 3: Nearest self-reported landmarks. The final method tested was to see if participants resided in the catchment of self-reported nearest landmarks. This approach is the most flexible and could be easily applied at all scales, given a database of the relevant landmark with accompanying spatial information is available. At the health facility, each participant was asked to name the nearest health facility, primary school, market and church to their compound. Combinations of responses were also assessed using overlapping catchment areas to increase the precision of the approach. Locations of compounds were plotted using ArcGIS and a compound was considered to be correctly located if it fell within the catchment area or intersecting catchment areas that matched the response provided at the facility.

Catchment areas for each type of landmark were estimated based on both Euclidian- (straight-line) and cost-distance [29, 30]. There were some missing coordinates for certain reported schools. Therefore, analysis was restricted to participants who reported residing near the schools with known coordinates. Euclidian distances were calculated using the ArcGIS Euclidian distance tool in the spatial analyst package to delineate catchment areas for for both health facilities (figure 3A) and primary schools (figure 3B).

A cost-distance function to account for factors that may either impede or facilitate travel was also used to delineate landmark catchment areas. Given the gently undulating topography of the study area, it was assumed that ease and speed of travel between compounds and relevant landmarks is determined either by the presence of roads (facilitating travel) or by the presence of rivers (impeding travel). Roads and rivers in the study area were digitized using high-resolution Quickbird satellite imagery multispectral imagery at 2.8 m resolution sharpened with a 60 cm resolution panchromatic image. Roads were classified into four categories: (1) tarred roads where the likely maximum speed is 80 km/hr; (2) roads that are not tarred but vehicles travel a likely maximum speed of 40 km/hr;

(3) roads that are not tarred but accessible to a vehicle or motorbike with likely maximum speeds of 20 km/hr; (4) paths not likely traversed by a vehicle but by motorbike with likely maximum speeds of 10 km/hr. For all other surfaces, including walking paths or fields, a maximum speed of 5 km/hr was assumed [23]. Rivers were classified as barriers to movement except where they were intersected by a road or path. The cost-distance models for both health facilities (figure 3C) and primary schools (figure 3D) were created using IDRISI software (Clark Labs, MA, USA) and imported into ArcGIS for analysis.

The mean error for both methods was calculated as the distance between the border of the catchment and the location of the incorrectly located compound. The distance between each compound and the centroid of each polygon could have been calculated. However, due to the irregular shape of many of the polygons, the distance to the centroid would not be an accurate reflection of the error rate in this approach as points that are far away from the centroid but located to the correct catchment area would generate a large error rate and be misleading.

Ethical considerations. This study was approved by the ethics committees of the London School of Hygiene and Tropical Medicine (LSHTM 5956) and the Kenya Medical Research Institute (SSC 1589). Individual informed consent was sought from all participants of the health facility survey by signature or thumbprint accompanied with the signature of an independent witness. As defined in the Kenya national guidelines, participants below 18 years of age who were pregnant, married, or a parent were considered "mature minors" and consented for themselves [31].

Data analysis. The proportion of study participants whose compounds were correctly located using each geolocation strategy of all participants that provided responses for each method and corresponding binomial 95% confidence interval was calculated. Mean error of each method was determined by calculating the distance between the actual location of the compound and edge of the identified area. Plotting the proportions for each approach against the mean area identified

the optimum strategy: strategies located in the top left corner of the plot signified high precision and accuracy.

RESULTS

Across both surveys, 3034 people were enrolled of which 830 (27%) were able to be traced back to their compounds and included in the analysis. Those that could not be traced were mainly due to running out of time and inaccurate information provided at the facility. The participants that could not be traced were evenly distributed between years and facilities.

Method 1: Geocoding

Of the geolocated participants, 519 lived within the area of the community cross-sectional malaria survey and could be used for geolocation. Of the 328 matched compounds, 56% were successfully located using the head of compound. Of the participants that were matched, 72.9% were correctly located to within 250 m (95% CI: 67.7-77.6, median distance of 36.2 m). Possible reasons for why more people were not correctly matched may include people not being familiar with the full names of their neighbors or reporting different heads of compound for the same compound (e.g. the grandfather vs. the father of the family). The median distance from the true location to the matched compound of those that were incorrectly matched was 4440.9 m (IQR: 1610.1-8591.4 m).

Method 2: Participatory Mapping

Using the participatory mapping approach, 64.9% (95% CI: 61.2-68.4) of 695 participants who attempted the mapping exercise were successfully located to the appropriate 2 x 2.5 km block (table 1). When a 500 m buffer in all directions around the block was included, the proportion correctly located improved to 82% (95%CI: 78.9-84.8) at the block and from 12.4% (95%CI: 10.0-15.0) to 57.1% (95%CI: 53.3-60.8) at the cell level.

However, 135 (16.3%) participants did not participate in the mapping exercise. Reasons for refusal were not recorded, but there were no differences in sex or age distributions in the populations who did and did not participate in the exercise. Of

those willing to locate their residence, 61.5% were female compared to 58.9% in the not willing group (p=0.6). Similarly, the mean age in the adult populations in those not willing to locate their residence was slightly higher at 37.9 compared to 35.3 years in those that did attempt the exercise, although the difference was not significant (p=0.3).

For compounds that were incorrectly located, the median distance to the correct block was 489 m (IQR: 229-1036 m), 1036 m (IQR: 737-1737), and 1737 m (IQR: 1179-2728) for the block only, +500 m buffer, and +1000 m buffer, respectively. The median distance of compounds incorrectly located from the identified cells was 539 m (IQR: 236-1095 m), 1055 m (IQR: 737-1644) including a 500 m buffer, and 1588 m (IQR: 1200-2180 m) including a 1000 m buffer. Also, the proportion of people that were correctly identified to a specific block or cell significantly varied per facility (block only p=0.007, +500 m p=0.003, +1000 m p<0.0001).

Method 3: Nearest self-reported landmarks

Analysis of self-reported nearest landmarks indicated that responses for nearest market tended to predominantly consider relatively large markets, rather than smaller, local markets. In addition there was too much variability in responses concerning the nearest church, the majority of which were small establishments whose spatial coordinates had not been recorded, to conduct meaningful analysis. For these reasons only data relating to the nearest health facility and primary school were retained.

Overall, the nearest health facility and primary school were reported correctly 84.9% (95% CI: 82.2-87.2) and 73.4% (95% CI: 68.8-77.7) of the time, respectively, based on straight-line distance (median distance 1486 m, IQR: 1008 – 2241 m). The use of the self-reported nearest primary school was able to locate 82.0% (95% CI: 78.1-85.8) of participants' compounds to the correct Euclidian distance catchment area (mean area of 6.7 km^2) (table 2) with a median distance to the self-reported nearest school of 878 m (IQR: 522 – 1234 m). The self-reported nearest health facility was able to locate 78.1% (95% CI: 73.8-82.1) of compounds to an area of 12.3 km^2. When the combination of responses was tested,

the mean area reduced to 1.7 km^2 and 48.7% (95% CI: 43.6-53.6) of participants' compounds were correctly located.

Next, 77.1% (95% CI: 74.1-80.0) and 78.1% (95% CI: 73.8-82.1) of participants were located to the correct health facility and school catchments, respectively using the cost-distance catchment area. The combined responses were able to locate individuals based on the combination of responses with 72.4% (95% CI: 67.8-76.8) of compounds successfully located to a mean area of 3.7 km^2 (table 2).

Of those individuals who did not reside in the catchment area of the reported nearest landmark, the mean distance away from the edge of the catchment area was 1252 m (IQR: 261 – 1899 m) for catchments based on Euclidian distance and 496 m (IQR: 174 – 605 m) using the cost-distance model.

Optimal geolocation approach

Although not directly comparable due to the different scales, the results across all strategies showed a logarithmic relationship between mean catchment area and proportion of compounds correctly identified (figure 4). Points that are located in the top left corner represent the optimal combination of low mean area (high precision) and a high proportion of people correctly located using that strategy (high accuracy). The results of this analysis suggest that using the location of the nearest primary school as well as the participatory mapping with buffer was the most promising methods to geolocate rural health facility attendees in this rural study setting.

DISCUSSION

A simple and operationally feasible way to identify the spatial occurrence of disease in rural areas where homes have no formalized address would be an extremely useful tool and could easily be employed as an operationally attractive approach to spatial disease surveillance in a wide range of settings around the world. A recent study has been conducted in Blantyre, Malawi in an urban setting [25] however, our study is, to our knowledge, the first attempt to examine different methods to geolocate health facility attendees in a rural area and to gauge their

precision. Although strategies are not directly comparable due to the different spatial scales, the current study showed that there are options available to obtain spatial information in areas where no formal postal network exists. Results have shown that it was possible to correctly locate close to 80% of participant compounds using either a participatory mapping exercise (to 2 x 2.5 km blocks with buffer) or by using information about the nearest primary school. This is similar to the level of detection of most geocoding strategies when applied in developed countries, although the spatial resolution is not as good [17, 32]. In this study, methods based on name-matching or participatory mapping to the 500 x 500 m cell level proved to be less accurate, but are capable of greater spatial precision.

The ideal geolocation approach in a rural setting will ultimately depend on the information available, the objectives, whether it be monitoring for epidemics or planning for disease control interventions, and the required spatial precision/accuracy. The geocoding approach requires that an accurate and up to date list of names of compound heads is available, which is unlikely to be the case outside areas of active community-based research. The geocoding approach also relies on names recorded being complete and recorded consistently; a difficult task in busy facilities. There may also be challenges in obtaining correct information from people who may want to remain anonymous. Also, a systematic bias is inevitable as compounds whose head has a common name or is the head of multiple compounds will never be matched unless other variables are also considered. However, in areas where a complete database is available, through land registries for example, or if overall accuracy is less important, geocoding could provide a useful geolocation approach.

The participatory mapping exercise also has notable limitations. It requires that a map of the study area be available and that there are personnel familiar with the area capable of interpreting satellite imagery. Key features must be identifiable on the map to help orient readers. Although the age difference here was not significant, younger generations may also be more map literate than older generations. High-resolution satellite imagery can be expensive to acquire, up to

several thousand US dollars [25] however, free imagery with good resolution is becoming more widely available for even remote areas in rural and low-income settings and a similar exercise could be conducted using web-based platforms as is increasingly being utilized for disaster response [33-35]. Also, depending on the size of the area of interest, it may be possible to create a schematic map of the area using local knowledge [10].

To facilitate participatory mapping, a grid was superimposed onto the study area, leading to an edge effect whereby if a person was located just outside of the block/cell they would be classified incorrectly even though the error margin could be only a few meters. Edge effect will always be an important limitation that must be accounted for in any application of this methodology particularly when the focus is on locating residences at a precise spatial resolution. However, despite this limitation, this research has provided important insight into how the edge effect can be minimized and sensitivity increased by the addition of buffer zones. Other approaches could have been used including a hexagonal grid or larger clusters as was used in the study in Blantyre's urban slum area [25]. These approaches will likely reduce, but not completely eliminate the edge effect. Also, in this study, there was a significant difference in the proportion of people correctly located at each health facility and not every participant was willing to complete the exercise. This suggests that the familiarity of the interviewers with the area, their ability to read and explain the maps to local populations, and the time they have or choose to dedicate may be important determinants for success.

The use of the nearest landmark approach requires that the location of the feature in question (e.g. church, school) be known. This could be done by visiting and mapping each site using a GPS receiver, or sites could be located on a map by someone familiar with the area. National databases of the locations of such landmarks are becoming more common and therefore this limitation may be less relevant, however to be useful, databases must be up to date and include all government, faith based, and private facilities. In this study, people only correctly located the nearest landmark around 80% of the time and the accuracy of this approach was dependent on the definition of catchment area used. The reporting

bias may be due to factors such as spatial perceptions of 'closeness', the density of that type of landmark in the area, or reporting known or highly frequented landmarks rather than those that are closer. Other possible landmarks that could be used include nearest chief or assistant chief, nearest shop, or nearest local transport point. In terms of defining catchment areas, both methods produced similar results [36]. The analysis using the cost-distance catchment areas showed a lower error rate based on the distance from the edge of the catchment area suggesting that this approach may be more robust. However, the utility of this approach is limited to areas with digitized travel networks, access to the required software, and the expertise to create the cost-distance surface is required.

The goals of the geolocation exercise will influence the optimum strategy. Firstly, the ideal scale will depend on the spatial pattern of the disease and the size of the area of interest [5]. For example, if the objective was to identify foci of infections of a highly heterogeneous disease such as malaria in a low endemic or epidemic setting [7, 9, 20] then achieving higher precision would be essential. Conversely, if the distribution of sexually transmitted infections was being studied, less precision may be acceptable or even necessary to guarantee anonymity [20]. Secondly, the ideal strategy will depend on the purpose of geolocating cases. If it is for programmatic use such as passive public health surveillance, or to establish disease distribution at a regional or national level, then using the nearest health facility, with a larger mean catchment area may be sufficient. However, if greater precision and accuracy were required, for identification of foci for disease elimination or identifying where to implement control, for example, then knowing the exact boundaries of the catchment area or having a comprehensive postal network that can be geocoded to a high precision would be essential.

There were some limitations to this study. Firstly, it was only feasible to trace 27% of participants to their compounds. Although this provided a large sample, it is possible that if we could have traced all individuals, the results and the conclusions on the applicability of the techniques tested may have been different. However, as the sample was a random selection, the impact on the results is expected to be minimal. Similarly, spatial coordinates were only available for the government-run

primary schools in the area, thereby restricting the sample to those residing near these schools. The limited number of school locations that were available as well as the lack of covariates such as size or perception of academic rigor to include as part of delineation of catchment areas likely influenced the size of catchment areas as calculated by both approaches. However, although altered catchment area boundaries would impact both the precision and accuracy of the results, this is not likely to have a significant impact of the results.

Spatial monitoring of health facility data has strengthened public health programmes in developed countries and facilitates conducting research with passively collected data [6, 37]. However, the ability to efficiently geolocate individuals residing in areas where no formal address network exists or where the settlement pattern is not conducive to matching individuals to specific localities is currently lacking, particularly in areas around the world where infectious disease transmission persists [5, 38]. The geolocation strategies tested as part of this research exemplify alternative options for obtaining spatial information from health facility patients in a setting that is typical for much of rural sub-Saharan Africa and other parts of the world. Easily collected spatial information can supplement both passive and active disease surveillance to detect foci of transmission, enables the detection of outbreaks in a timely manner, and facilitates tracking of how disease spreads through the population over time [37, 39, 40]. If validated in other parts of the world, these results indicate that recording the nearest primary school or implementation of a participatory mapping exercise at rural health facilities offer potential strategies to facilitate spatial analysis of disease dynamics. Further research is needed to demonstrate their utility in a range of settings and their operational viability before formal testing in a broader operational context.

ACKNOWLEDGEMENTS

We would like to acknowledge the community of Rachuonyo South District, the Medical Officers of Health and the staff at the participating rural health facilities and schools for their support and cooperation. We also thank our partners at the Kenya Medical Research Institute (KEMRI)/US Centers for Disease Control and

Prevention, Kisumu, Kenya and the LSHTM Malaria Centre for their guidance and support. This article has been approved by the Director of KEMRI.

FUNDING

This work was supported by the Bill and Melinda Gates Foundation, under the Malaria Transmission Consortium [grant number 45114]. CD is supported by the Wellcome Trust [grant number 091924]. TB is supported by a Grant Challenge Grant from the Bill and Melinda Gates Foundation [grant number OPP1024438].

CONFLICTS OF INTEREST

All authors have no conflicts of interest to report

AUTHOR CONTRIBUTIONS

Conceived and designed the study: GHS JCS TB CD JC. Field Data Collection: GHS EM CO CA. Data Analysis: GHS JC. Wrote the manuscript: GHS JCS TB CD JC. All authors reviewed and approved of the paper.

REFERENCES

1 Schimmer B, et al. The use of a geographic information system to identify a dairy goat farm as the most likely source of an urban Q-fever outbreak. *BMC Infectious Diseases* 2010; **10**: 69.

2 Gatto M, et al. Generalized reproduction numbers and the prediction of patterns in waterborne disease. *Proceedings of the National Academy of Sciences* 2012; **109**: 19703-19708

3 Breban R, Riou J, Fontanet A. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *The Lancet* 2013; **382**: 694-699.

4 Cauchemez S, et al. Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: application to the emergence of US swine origin influenza A H3N2v virus. *PLoS Medicine* 2013; **10**: e1001399

5 Noor AM, et al. A spatial national health facility database for public health sector planning in Kenya in 2008. *International Journal of Health Geography* 2009; **8**: 13.

[6] Sturrock HJ, et al. Targeting asymptomatic malaria infections: Active surveillance in control and elimination. *PLoS Medicine* 2013; **10**: e1001467.

[7] Woolhouse MEJ, et al. Heterogeneities in the Transmission of Infectious Agents: Implications for the Design of Control Programs. *Proceedings of the National Academy of Sciences* 1997; **94**: 338-342.

[8] Zhang ZJ, et al. Identification of high-risk regions for schistosomiasis in the Guichi region of China: an adaptive kernel density estimation-based approach. *Parasitology* 2013; **140**: 868-875.

[9] Bousema T, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Medicine* 2012; **9**: e1001165.

[10] Dongus S, et al. Participatory mapping of target areas to enable operational larval source management to suppress malaria vector mosquitoes in Dar es Salaam, Tanzania. *International Journal of Health Geography* 2007; **6**: 37.

[11] Noor AM, et al. Malaria risk mapping for control in the republic of Sudan. *American Journal of Tropical Medicine and Hygiene* 2012; **87**: 1012-1021.

[12] Trevelyan B, Smallman-Raynor M, Cliff AD. The Spatial Dynamics of Poliomyelitis in the United States: From Epidemic Emergence to Vaccine-Induced Retreat, 1910-1971. *Annals of the Association of American Geographers* 2005; **95**: 269-293.

[13] Moss WJ, et al. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community survey. *Malaria Journal* 2011; **10**: 163.

[14] Curtis AB, et al. Using GIS and secondary data to target diabetes-related public health efforts. *Public Health Reports* 2013; **128**: 212-220.

[15] Han D, et al. Assessing bias associated with geocoding of historical residence in epidemiology research. *Geospatial Health* 2013; **7**: 369-374.

[16] Dearwent SM, Jacobs RR, Halbert JB. Locational Uncertainty in Georeferencing Public Health Datasets. *Journal of Exposure Analysis and Environmental Epidemiology* 2001; **11**: 329-334.

[17] Kumar S, Liu M, Hwang SA. A multifaceted comparison of ArcGIS and MapMarker for automated geocoding. *Geospatial Health* 2012; **7**: 145-151.

[18] Carvalho RM, Nascimento LF. Spatial distribution of dengue in the city of Cruzeiro, Sao Paulo State, Brazil: use of geoprocessing tools. *Revista Instituto de Medicina Tropical de Sao Paulo* 2012; **54**: 261-266.

[19] Noor AM, et al. Defining equity in physical access to clinical services using geographical information systems as part of malaria planning and monitoring in Kenya. *Tropical Medicine and International Health* 2003; **8**: 917-926.

[20] Owusu-Edusei K, Jr., Doshi SR. Assessing spatial gaps in sexually transmissible infection services and morbidity: an illustration with Texas county-level data from 2007. *Sexual Health* 2012; **9**: 334-340.

[21] Lash RR, Carroll DS, Hughes CM et al. Effects of georeferencing effort on mapping monkeypox case distributions and transmission risk. *International Journal of Health Geography* 2012; 11: 23.

[22] Florentinus SR, et al. Linking community pharmacy dispensing data to prescribing data of general practitioners. *BMC Medical Informatics and Decision Making* 2006; **6**: 18.

[23] Noor AM, et al. Modelling distances travelled to government health services in Kenya. *Tropical Medicine and International Health* 2006; **11**: 188-196.

[24] Francis D, et al. Health facility-based malaria surveillance: the effects of age, area of residence and diagnostics on test positivity rates. *Malaria Journal* 2012; **11**: 229.

[25] Macpherson P, et al. Development and Validation of a Global Positioning System-based "Map Book" System for Categorizing Cluster Residency Status of Community Members Living in High-Density Urban Slums in Blantyre, Malawi. *American Journal of Epidemiology* 2013; **177**: 1143-1147.

[26] Stuckey EM, et al. Simulation of malaria epidemiology and control in the highlands of Western Kenya. *Malaria Journal* 2012; **11**: 357.

[27] Bousema T, et al. The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. *Trials* 2013; **14**: 36.

[28] Levenshtein VI. Binary codes capable of correcting deletion, insertions, and reversals. *Soviet Physics Doklady* 1966; **10**: 707-710.

[29] Apparicio P, et al. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geography* 2008; **7**: 7.

[30] Delamater PL, et al. Measuring geographic access to health care: Raster and network-based methods. *International Journal of Health Geography* 2012; **11**: 15.

[31] Sanitation MoPHa. National Guidelines for HIV Testing and Counselling in Kenya - 2010. In: Programme NAaSC, editor. Nairobi, Kenya: Ministry of Public Health and Sanitation; 2010.

[32] Zimmerman DL. Estimating the intensity of a spatial point process from locations coarsened by incomplete geocoding. *Biometrics* 2008; **64**: 262-270.

[33] ESRI. ArGIS Services Directory: World Imagery (Map Server). 2013 [cited 2013 04-Jul]; Available from: http://services.arcgisonline.com/ArcGIS/rest/services/World_Imagery/MapServer.

[34] Laituri M, Kodrich K. On Line Disaster Response Community: People as Sensors of High Magnitude Disasters Using Internet GIS. *Sensors* 2008; **8**: 3037-3055.

[35] Voigt S, et al. Satellite image analysis for disaster and crisis-management support. *IEEE Transactions on Geoscience and Remote Sensing* 2007; **45**: 1520-1528.

[36] Phibbs CS, Luft HS. Correlation of travel time on roads versus straight line distance. *Medical Care Research and Review* 1995; **52**: 532-542.

[37] Stresman GH, et al. A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. *Malaria Journal* 2010; **9**: 265.

[38] Rowe AK, et al. Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa. *Malaria Journal* 2009; **8**: 209.

[39] de Souza Gomes EC, et al. Schistosomiasis transmission and environmental change: a spatio-temporal analysis in Porto de Galinhas, Pernambuco--Brazil. *International Journal of Health Geography* 2012; **11**: 51.

[40] Mosha JF, et al. Epidemiology of subpatent Plasmodium falciparum infection: implications for detection of hotspots with imperfect diagnostics. *Malaria Journal* 2013; **12**: 221.

FIGURES

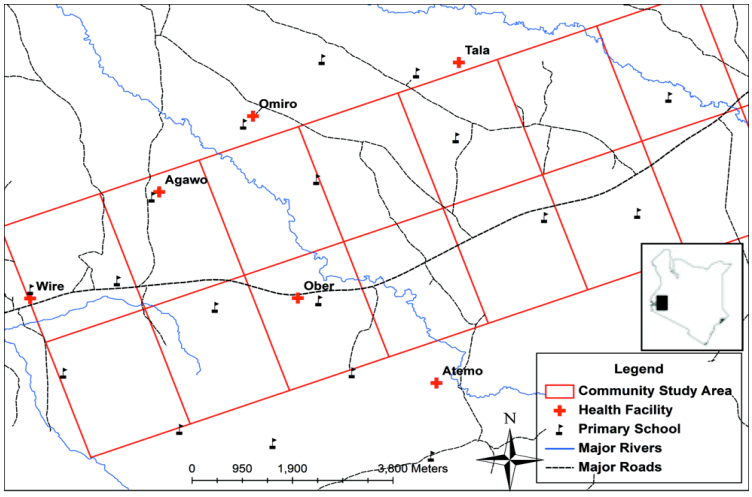

Figure 1: Map of the study area, Rachuonyo South, Kenya (2011-2012), showing the main roads (dashed lines), rivers (solid lines), location of schools (flags) and health facilities (crosses).

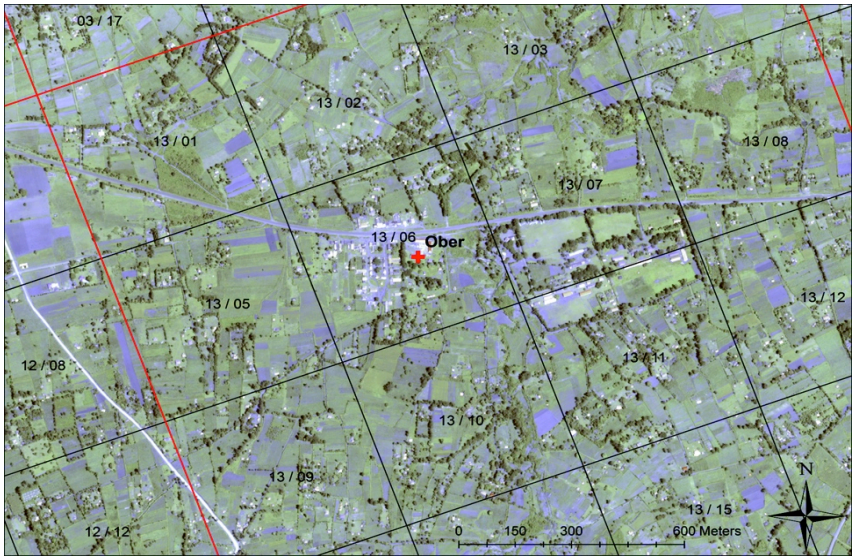

Figure 2: Participatory mapping example showing the grid of block and cells that were overlain on satellite imagery. The red lines outline the block and block numbers are shown. The cells are outlined by the black lines within each block and are counted from 1 to 20 starting with the upper left corner and counting from left to right (ie. 13/01 to 13/20).

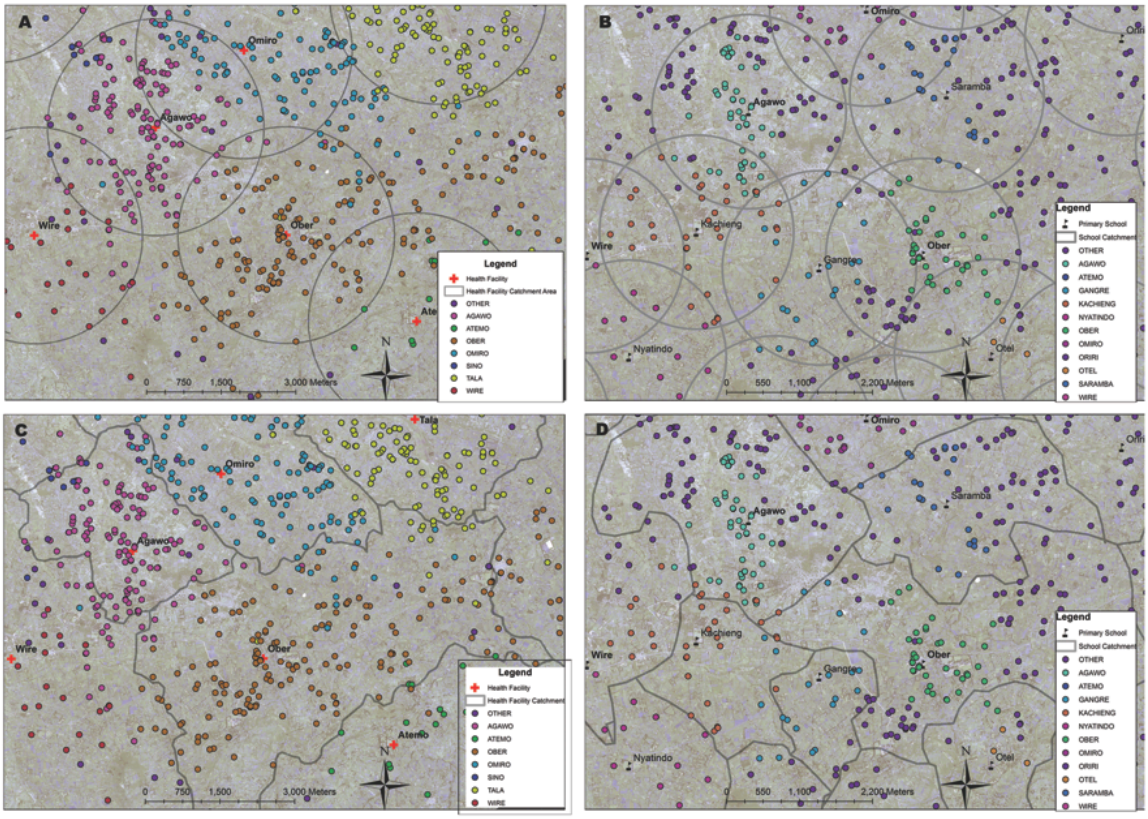

Figure 3: Examples of the catchment areas and the spatial distribution of responses for self reported nearest landmark for the Euclidian and cost-distance models, South Rachuonyo, Kenya, 2011-2012: A) Health facility catchment based on Euclidian distance model; B) Primary school catchment based on Euclidian distance model; C) Health facility catchment area based on cost-distance model; D) School catchment area based on cost-distance model.

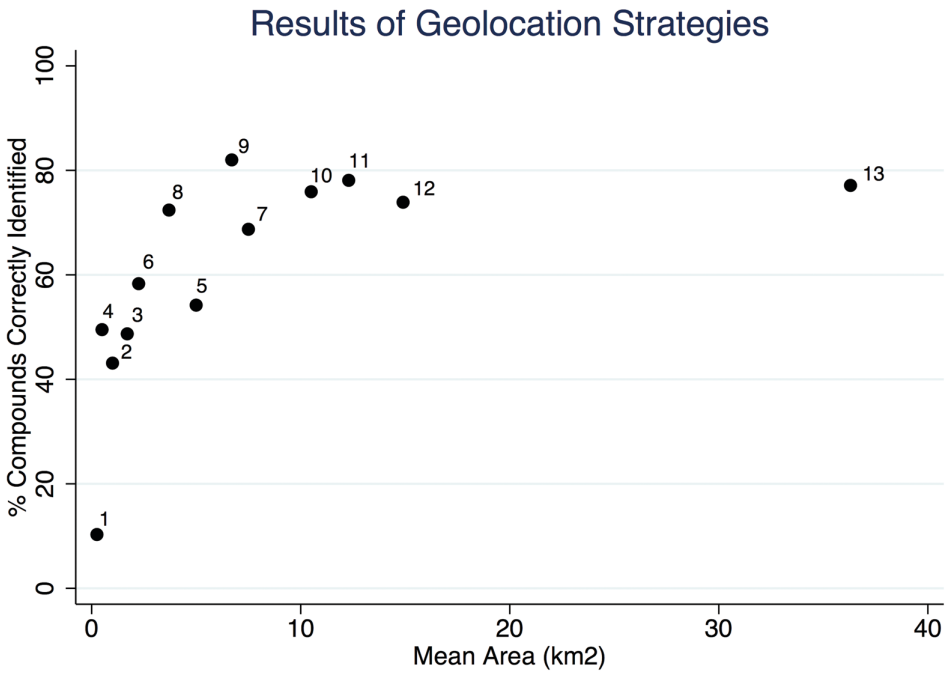

Figure 4: Scatter plot showing the summarized results of all geolocation strategies tested with the precision (mean area) of the approach plotted against the accuracy (% of compounds correctly located): 1-Cell [participatory mapping (PM)]; 2-Cell (+500 m)[PM]; 3-Combined Health Facility (HF) & Primary School (PS) (Euclidian distance - ED)[Nearest Landmark (NL)]; 4-Geocoding; 5-Block [PM]; 6-Cell (+1000 m) [PM]; 7-Block (+500 m)[PM]; 8-Combined HF & PS (cost-distance - CD)[NL]; 9-PS (ED)[NL]; 10-Block (+1000 m)[PM]; 11-PS (CD)[NL]; 12-HF (ED)[NL]; 13-HF (CD)[NL].

TABLES

Table 1: Results of Participatory Mapping Exercise, Rachuonyo South, Kenya, 2011-2012									
	Block/Cell Only			+ 500 m buffer			+ 1000 m buffer		
	Mean Area (km^2)	% Correct	95% CI	Mean Area (km^2)	% Correct	95% CI	Mean Area (km^2)	% Correct	95% CI
Block	5	64.9	61.2-68.4	7.5	82.0	78.9-84.8	10.5	90.6	88.2-92.7
Cell	0.25	12.4	10.0-15.0	1	57.1	53.3-60.8	2.25	77.1	73.8-80.2

Table 2: Results of self-reported nearest landmarks as a geolocation strategy, Rachuonyo South, Kenya, 2011-2012						
	Euclidian Distance			Cost Distance		
	Mean Area (km^2)	% Correct	95% CI	Mean Area (km^2)	% Correct	95% CI
Health Facility	14.9	73.9	70.7, 76.8	36.3	77.1	74.1, 80.0
Primary School	6.7	82.0	78.1, 85.8	12.3	78.1	73.8, 82.1
HF & Sch	1.7	48.7	43.6, 53.6	3.7	72.4	67.8, 76.8

Appendix 1.2 – Quantifying travel behavior

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Scientific Reports		
When was the work published	14/07/2014		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**CC-BY**	Was the work subject to academic pier review?	**Yes**

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	
Please list the paper's authors in the intended authorship order.	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Suported data analysis and contributed to preparation of the manuscript.

Student Signature: _____  **_____ Date: ___17/02/2015_____**

Supervisor Signature: ____  **_____ Date: ___17/02/2015_____**

Title: Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones

Authors: Amy Wesolowski[1]*, Gillian Stresman[2], Nathan Eagle[3,4] Jennifer Stevenson[5,6,7],Chrispin Owaga[6], Elizabeth Marube[6], Teun Bousema[2,8], Christopher Drakeley[2], Jonathan Cox[5], and Caroline O. Buckee[4,9]
* Corresponding author

Affiliations:
[1] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA.
[2] Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London, UK.
[3] Department of Computer Science, Northeastern University, Boston, MA, USA.
[4] Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA.
[5] Department of Disease Control, London School of Hygiene and Tropical Medicine, London, UK.
[6] Centre for Global Health Research, Kenya Medical Research Institute/Centers for Disease Control and Prevention, Kisumu, Kenya.
[7] Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
[8] Radboud University Nijmegen, Nijmegen, 6525 GA, Netherlands.
[9] Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, MA, USA.

Reference: Wesolowski et al., (2014) Scientific Reports, 4:5678.

Abstract
Human travel impacts the spread of infectious diseases across spatial and temporal scales, with broad implications for the biological and social sciences. Individual data on travel patterns have been difficult to obtain, particularly in low-income countries. Travel survey data provide detailed demographic information, but sample sizes are often small and travel histories are hard to validate. Mobile

phone records can provide vast quantities of spatio-temporal travel data but vary in spatial resolution and explicitly do not include individual information in order to protect the privacy of subscribers. Here we compare and contrast both sources of data over the same time period in a rural area of Kenya. Although both data sets are able to quantify broad travel patterns and distinguish regional differences in travel, each provides different insights that can be combined to form a more detailed picture of travel in low-income settings to understand the spread of infectious diseases.

Introduction

Improvements in transportation infrastructure and increasing human mobility are enabling unprecedented connectivity between populations at both local and global scales, allowing for the rapid dissemination of pathogens [1-6]. Humans are able to introduce diseases into immunologically naïve populations through direct transmission or by introducing them into the environment [1, 7-9], and travel plays a critical role in the spatial spread of influenza, polio, cholera, and dengue, as well as in the spatial spread of drug resistance among pathogens such as malaria [2, 3, 5, 10-15]. Quantifying population travel dynamics is difficult, however, particularly in low-income countries where individual level data sets that include information about travel behavior are difficult to obtain and collect.

Traditionally, travel history questions from household surveys or from census data have provided the most comprehensive source of travel information [16]. During these surveys, which often include data on variables such as age, sex, income, household structure, health status, or ethnicity, for example, individuals are asked questions about their movement patterns. Surveys therefore provide insights into the demographic biases and motivations underlying movement patterns. However, these data sets often only sample a small subset of the population and may be subject to recall bias. Moreover, these questions are typically nested in larger surveys with disparate objectives that may impact their generalizability and oversample individuals of interest to the larger survey objective, for instance they may be part of country wide Malaria Indicator Surveys, questions asked during hospitalization, or household budget surveys [17-21]. The most common source of

travel survey data in Africa is records from a national or micro-census, but these typically address only long-term changes in residence [22, 23].

In contrast, anonymized mobile phone usage data have recently been shown to provide a valuable source of information on regular movement patterns on various spatial scales [14, 15, 24-26]. Call detail records (CDRs) store locational information for each subscriber when they make a call or send a SMS (Short Message Service), providing a detailed temporal and spatial picture of often millions of people. Due to privacy concerns and pre-paid plans, individual socio-demographic data about subscribers are unavailable to researchers. Analysis from previous work has shown that mobile phone ownership is biased towards wealthy, urban males, despite remarkable levels of ownership across all income brackets in Kenya, for example [27]. Furthermore, phone sharing practices may hinder the use of mobile phone data to accurately capture individual level inferences about movement patterns [27]. Nevertheless, we have shown that these biases can be corrected for and are unlikely to impact the routes and relative volumes of travel between most populations [28].

We have previously quantified intra-national travel patterns from nearly 15 million mobile phone subscribers in Kenya on a range of spatial and temporal scales using mobile phone data, with a particular emphasis on the role of travel in the importation of malaria parasites across the country [15]. The volumes and direction of travel varied seasonally, and depended on both the origin and destination locations, with a large amount of travel occurring to and from the capital city, Nairobi. Here we compare a subset of these data with information from a detailed survey about travel from the same time and place, collected during cross-sectional surveys of 2,650 individuals in two districts in western Kenya. The travel survey was conducted as part of a study to characterize patterns of malaria transmission and risk factors for infection in an area of low malaria endemicity. We describe patterns of travel, highlight the differences and strengths in each data set and discuss how the data sets can be used in conjunction to enhance their utility.

Results

Travel history results from community surveys Travel data was collected as part of a malariometric survey conducted in February 2009 and covered 2,650 (0.13% of the population in the survey sites N=19,744) individuals in two districts: Kisii Central (formerly part of Kisii district) and Rachuonyo South (formerly part of Rachuonyo) (see Figure 1). Despite being predominantly rural, these districts have the relatively high population densities (707 (Kisii) and 705 (Rachuonyo) individuals/km^2, total populations: 457,105 and 307,126 individuals from the national census) that characterize the regions surrounding Lake Victoria. The individuals included in the study are from the rural parts of both districts. Kisii is primarily made up of the Kisii ethnic group whereas Rachuonyo is primarily made up of the Luo ethic group (see Figure 1). Travel surveys provided general socio-demographic information (see Table 1) that was used to analyze travel patterns stratified by age, gender, and other covariates.

One of the most striking findings in the travel survey was that the vast majority of people (90%, N= 2,388) reported that they had not made an overnight trip to another district within the last 3 months (see Table S1). More individuals within households in Rachuonyo reported traveling more often than those in Kisii (see Figure 2). When individuals did travel, they reported spending the majority of their time in neighboring districts or those including a major city, predominantly Nairobi (Figure 3, Table S2). The primary motivations for travel were either visiting family or friends (54%, N = 105) or attending a funeral (17%, N = 46) (see Tables S3-S4). Of those who traveled, most reported taking only a single trip that had most often occurred less than four weeks ago (64%, N = 125 see Table S5-S6, see Figure 4). Of adults (aged 15 or older) who have traveled, men were slightly more likely to have taken an overnight trip (males: 13% = 70/525, females 11% = 83/657, χ^2= 2.3889, p = 0.6646). The destinations for travel were primarily the same for both men and women, although men reported that they traveled to Nairobi more often than women (16% = 11/70 versus 7% = 6/83, χ^2=29, p <0.001) (see Table S7). Children (under 15 years of age) were less likely to travel than adults (3% (42/1318) of children had taken an overnight trip).

Less than half (47%, N = 366) of households reported having a mobile phone. Mobile phone ownership (on a household level) was positively correlated with the likelihood of reporting having traveled (see Table 2). The percentage of households where at least one person reported traveling was 60% (83/138) in households with a mobile phone versus 40% (55/138) (χ^2=10.72, p=0.001) for those without a mobile phone.

Mobile phone data analysis. We analyzed CDRs using methods previously described [15] (see Materials and Methods), identifying 34,861 subscribers (4.6% of the total population in these districts assuming each subscriber is an individual) in the region (see Materials and Methods). Briefly, cell tower locations were assigned to districts, demarcated by political boundaries. Using a daily time series of tower locations over the course of the data set, subscribers whose most used mobile phone tower was within 3km, the typical service range, of the study site were considered (see Materials and Methods). During the three-month study time period corresponding to the travel survey, movement between districts was quantified.

In contrast to the travel survey, we inferred from the CDRs that the vast majority of mobile phone subscribers had spent at least one night outside Kisii and Rachuonyo districts during the time frame of the survey (61% from Kisii, 95% from Rachuonyo, in total 27,668 subscribers, see Table 3). As observed in the survey data, subscribers from Rachuonyo traveled more than those from Kisii, possibly related to the geographic distribution of the Luo ethnic group. We excluded travel between Kisii and Rachuonyo because many cell towers lie on the border between the two districts, making it difficult to separate travelers within this sub-region. Half of subscribers traveled for at least 2 days away from Kisii and Rachuonyo to other districts (36% from Kisii, 63% from Rachuonyo, 17,560 subscribers, see Table 3). Thus, we estimate that between 17,560 (two days or more) and 27,668 (one night or more) subscribers traveled to other districts during the study time frame. Including travel lasting at least one night, subscribers took a total of 13,860 trips. These trips were often short with 65% lasting less than three days (see Table S8).

Comparing travel between data sources. Given the wide divergence in terms of the magnitude of travel between the two data sets, we calculated an adjustment to compare the two data sets (see Table S9). The survey sites had a collective population of 19,744 individuals when accounting for the total enumerated population for the areas that represented the survey clusters. Using the survey data, we estimated that between 2,500 and 11,500 mobile phone subscribers were located in the study site at the time, with the range determined by the estimated number of subscribers per household (see Supplementary Information and Table 4). This value is up to one order of magnitude less than the number of mobile phone subscriber IDs we have included in the analysis, indicating that i) we may be capturing subscribers who reside in neighboring areas in our CDR analysis, ii) individuals own multiple mobile phones or SIM cards, and/or iii) estimates from the two data sources are extremely different.

We cannot address this last option, but it seems unlikely that on average each individual owned 5 SIM cards. Furthermore, even if we assume that all mobile phone subscribers were adult men, since they represent the most mobile demographic group, at most 12% of men reported traveling away from their home district in the survey. This would correspond to 200 to 1,800 mobile phone subscribers within the study site traveling (see Supplementary Information, Tables S10-S11). Since this value is orders of magnitude less than measured number of trips by mobile phone subscribers (approximately 28,000, see Table 4), the two sources of data remain markedly different in their estimates of the number of travelers, although both were able to identify the main districts where people travel.

We next compared the percentage of individuals taking between one and 60 trips from each data set. In general, individuals from the survey data took a fewer number of trips than the mobile phone data would suggest, although individuals traveled more frequently in the survey data (see Figure 4). Possible reasons for the discrepancies between the two data sets include recall bias or misreporting in the travel surveys, differences in the populations represented in each data set, and

mobile phone sharing practices. We hypothesize that the first is highly likely, and although the last two are possible, they cannot account for the entirety of the difference [9, 29]. It is likely, therefore, that all three of these contribute to varying degrees and actual travel falls somewhere between the two estimates.

Impact of travel estimates on predictions about malaria exchange. One of the most important reasons to quantify human mobility is in the assessment of the spread of disease in the region, including malaria. Previously, we quantified malaria (*Plasmodium falciparium*) importation within Kenya using mobile phone data [15] and spatial *P. falciparium* ($PfPR_{2\text{-}10}$) prevalence data from the Malaria Atlas Project [30]. Using a simplified metric that does not require as detailed data as in [15], we used a measure of malaria exchange (as opposed to malaria importation) that utilizes population-weighted travel as well as prevalence data (see Materials and Methods and Supplementary Information) [31]. In particular, this metric does not require information on the duration of travel since it is unavailable in the survey data. This measure describes the estimated exchange of malaria parasites adjusted based on the prevalence data between two locations. It almost certainly overestimates the impact of travel, since we use the higher parasite rate found in children age 2-10 years old, but illustrates the possible range of importation of parasites to and from the region.

For travelers from Kisii and Rachuonyo, the mobile phone data produces total malaria exchange estimates an order of magnitude greater than the survey data, in this case comparing the total number of travelers from both data sets (see Tables 5, S12). Both data sets predict that the amount of malaria being brought (?) coming(?) into Rachuonyo is much greater than into Kisii, and were both able to identify the major routes. Mobile phone data predicted that malaria exchange occurs between nearly all districts. However, the community survey data suggest that malaria parasites are likely to predominantly come from a few districts (see Supplementary Information). These findings have important implications for targeted surveillance in the region, since the overall volume and locations contributing to malaria exchange may be a more important consideration for control programs than travel surveys would indicate. We propose that while travel

surveys provide important information about motivations for travel, the type of people who are traveling, and identify the main travel destinations, they are also likely to under-estimate the volume and range of mobility (see Supplementary Information).

Discussion

Overall, the community survey provided a snapshot of travel behavior for 2,650 individuals. The volume of travel reported from the surveys was considerably lower than that captured by mobile phone data. It is possible that mobile phone subscribers were simply not captured by the survey, since working age men are often absent during community surveys. Other possible reasons for under-reporting of travel include recall problems of interviewed individuals; details about trips taken may be forgotten or when trips were taken not accurately reported. Lack of knowledge of, or recent changes in administrative boundaries may also result in underreporting of travel. Surveys are challenging to conduct on a large scale and it is not feasible to sample the majority of residents within even small geographic areas. Cross-sectional surveys can only collect travel data for each individual at one point in time and therefore do not provide a dynamic picture of overall movement patterns. For example, Nyamira district was once part of Kisii district and this may have caused confusion in the travel survey that would not be observed in the mobile phone data.

Mobile phone data enables researchers to estimate travel patterns for a large sample of the population over time, but can only provide an estimate of travel for mobile phone subscribers and is limited by mobile phone tower density. Community surveys are able to compliment mobile phone data by approximating travel patterns of non-subscribers. Here we used anonymized CDRs where every subscriber is assigned a unique ID. Subscriber IDs may not reflect individuals due to phone sharing and/or multiple SIM card ownership [27]. Subscribers also represent a biased sample of the general population, with ownership more prevalent among more educated, urban, males [27]. However, based on the results from the travel survey, it appears that those households that do not own a mobile

phone are also less likely to travel, so bias of estimates due to skewed mobile phone ownership may not be as large as previously thought.

Interestingly, it appears that in this setting ethnicity influences travel behavior. From both data sets, we observed that those living in Rachuonyo travel more than those in Kisii (see Figure 2). Rachuonyo is predominantly Luo whereas Kisii is predominately Kisii [32]. The large geographic coverage of the Luo ethnic group (see Figure 1) may go some way to explain this. The main reasons for travel given during the surveys were to visit family and friends or attend a funeral, both of which are more likely to have strong ethnic influences. However, at present we can only suggest this as a possible explanation. Aside from ethnicity, road access and travel times to other districts may also impact travel and we suggest that this should be investigated in future work.

Quantifying human travel patterns can have broad applications in epidemiology, particularly the spatial spread of infectious diseases. Being able to accurately parameterize movement patterns will be invaluable in identifying areas that are at risk of re- or continued importation of disease, which has major implications for control and elimination programs. Here we compared travel survey questions with mobile phone data over the same time period in western Kenya. We found that the survey data produces lower estimates of travel, although it did provide demographic information about travelers and motivations for travel. Mobile phone data can give a refined, spatio-temporal description of travel patterns, although it lacks information about subscribers, is often difficult to obtain, and as more providers become available such comprehensive estimates as presented here become even more challenging to achieve. In the case of malaria exchange via travel within these districts, although the volume of exchange differs by data source, both surveys were able to identify the some areas where the majority of exchange is likely to originate. In conjunction, these two data sources can be used to form a quantitative and qualitative description of travel within rural Kenya.

Methods

Travel survey data. A malariometric community survey was conducted using a cluster design in the highland districts Kisii Central and Rachuonyo South (referred to as Kisii and Rachuonyo in this paper), Nyanza province, western Kenya. For the survey, 23 enumeration areas (EA) (administrative areas with approximately 100 households or 500 residents) were randomly selected. Each EA was enumerated and mapped and 12-15 households were randomly selected for the survey.

The cross sectional surveys took place during February 2009. During this survey, individual informed consent was sought from all residents of the compound above the age of 6 months by signature or thumbprint accompanied by the signature of an independent witness. Consent for children under the age of 18 was provided by a parent/guardian and children between 14 and 17 years also provided written assent by signature or thumbprint accompanied by the signature of an independent witness. Individuals between 15 and 18 years of age who were pregnant, married, or a parent were considered "mature minors" according to national policy and were able to consent for themselves. The household was interviewed to assess household wealth indices and use of anti-malarial measures. All consenting individuals above the age of 6 months were tested for malaria and anemia. Individuals in both surveys were asked basic travel questions about themselves and their children, although the questions varied by survey (see Table 6). In the survey, individuals were asked if they had made any overnight trips to another district, the total number of overnight trips, when they came back from their journey, and the reason for traveling.

Mobile phone data. Call data records (CDR) from June 2008 till June 2009 for 14,816,521 subscribers within Kenya were obtained from all months except for February 2009. For each entry in the CDR, the sender, receiver, date, and location of the call (or SMS) was recorded by the leading mobile phone provider. In total, subscribers sent and received approximately 12 billion calls and SMS geolocated at one of 12,502 mobile phone towers. For each subscriber, we approximated their daily location based on the location of the mobile phone tower that serviced the

majority of their calls (or SMS) or the tower that serviced their most recent call (or SMS) if no call was made. For this analysis, we aggregated tower locations to districts based on the location of the mobile phone tower. We only considered subscriber IDs where the majority of their calls were serviced by mobile phone towers within the service area (3 km) and the district of each study site to conservatively only consider travel by subscribers whose primary mobile phone tower location was in the study site. At the time of data collection, this was the standard service area for mobile phone towers. This data was then restricted to include the sets of subscribers that overlap with the area of the community survey, one set in Kisii and two in Rachuonyo (see Figure 1). In total, we considered the data generated from 16,196 (based at 6 mobile phone towers) and 18,665 (9 mobile phone towers) subscribers in Kisii and Rachuonyo respectively (see Table 3).

We only considered travel that crossed district boundaries outside of the study area and not local movement within the study site (i.e. no travel between Kisii and Rachuonyo). Although the study site spans a district border, climate and topography are similar and we wanted to assess the extent of travel to areas where disease transmission would be markedly different. Also, there were a number of mobile phone towers along the borders of these districts making differentiating travel between the two locations more difficult. To match the time period of the survey, we only considered travel that occurred between the start of November 2008 till the end of January 2009. The mobile phone data describe the movements to the entire country of approximately 35,000 subscribers primarily call from one of 15 mobile phone towers. No other demographic information is available from cell phone data.

Comparing between the two sources of travel data. To compare between the mobile phone and survey data, we estimated the number of subscribers using the survey data and calculated a range for the number of trips taken by these subscribers. To estimate the number of mobile phone subscribers in the study area from the survey data, we used the number of individuals in the study area (~20,000), number of households (776), percentage of households with a mobile phone (47%

reported in the survey), as well as the average number of individuals per sleeping structure (3.7). We did not know the number of subscribers per household but assumed a range of 1-4 subscribers per household to produce a range of ~2,500-11,500 mobile phone subscribers in the study area from the survey data.

To estimate the number of subscribers who have traveled using only the survey data, we considered a range of the percentage of subscribers who have traveled. At the low end, 8% of individuals have reported traveling results in 200-920 subscribers who have traveled. At the high end, adult males living with a household with a mobile phone were the demographic group with the highest percentage of travelers (16%). This value would imply that between 400-1,800 subscribers have traveled. As reported in the results section, these estimates are at least an order of magnitude lower than the measured values from the mobile phone data.

Quantifying malaria exchange. To further compare both data sets, we quantified a malaria (P. falciparum) exchange metric using each set of travel data along with malaria endemicity data. Spatially explicit quantitative malaria endemicity estimates were obtained from the Malaria Atlas Project [31]. *P. falciparum* malaria endemicity data were obtained from the MAP (www.map.ox.ac.uk/) as measured by the parasite rate in the 2-10 age group ($PfPR_{2-10}$) [30]. This measure is an overestimate on the parasite rate since we are quantifying travel by adults, who generally have lower rates of parasite carriage. We use prevalence in children to avoid complex adjustments for patterns of prevalence by age, which vary with transmission intensity and are not straightforward to measure since many semi-immune individuals have sub-patent infections. Our estimates therefore represent an upper limit, and are intended to reflect the potential range and extent of spatial spread of malaria.

We calculated population rescaled travel from Rachuonyo and Kisii to other districts using the mobile phone and census data. For the mobile phone data, the population in each district's coverage area was the number of subscribers (18,665 and 16,196 in Rachuonyo and Kisii) whereas in the survey data it was the total

number of individuals surveyed (1,297 and 1,352 in Rachuonyo and Kisii). From the survey data, we separated individuals by their district study site and considered travel to other districts.

In previous work, we utilized the mobile phone data to quantify the role of travel for malaria importation within Kenya [15]. However, due to the coarseness of the travel survey data and inability to describe the duration and exact destinations for all trips reported in the survey data, we choose to use a simplified malaria-travel metric that describes malaria exchange between locations [31] (see Table 4, Supplementary Information for further discussion). This metric, *Pfm*, is based on travel between individuals from the study sites (i) to all other districts (j) is defined as:

$$Pfm_{i,j} = \frac{PfPR_i * PfPR_j}{PfPR_i + PfPR_j} m_{i,j}$$

where $m_{i,j}$ is the population weighted travel to other districts.

Statistical analysis. The proportion of people traveling to another district was calculated for both datasets and summary values compared. Data from the travel survey data were analyzed to estimate the conditional probabilities of travel outside the district to provide insight on the demographics of travelers. Statistical and spatial analyses were carried out using R statistical analysis software (R v3.0.1, The R Foundation for Statistical Computing).

Geographic analysis. Mapping shown in Figures 1-3 was carried out by one of the co-authors using ArcGIS v10.1.

Ethical considerations. The community surveys were conducted and approved by the ethical committees of the London School of Hygiene and Tropical Medicine (LSHTM) and the Kenya Medical Research Institute (KEMRI) under protocol number SSC1802. Call data record were provided by the leading mobile phone provider to one of the co-authors of the paper. All received records were anonymized and could not be linked to individual users. The de-identified mobile

phone records analysis was approved as not human subjects researchers by Harvard University IRB.

Acknowledgments

We are grateful to the community of Kisii and Rachuonyo for their cooperation. We also thank the support of our partners in the Kenya Medical Research Institute/ US Centers for Disease Control and Prevention, Kisumu, Kenya.

This project was funded by the Bill and Melinda Gates Foundation, under the Malaria Transmission Consortium, grant no. 45114. APW was supported by the National Science Foundation Graduate Research Fellowship program (#0750271) and the Department of Engineering and Public Policy at Carnegie Mellon University. COB was supported by the Models of Infectious Disease Agent Study program (cooperative agreement 1U54GM088558). *The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health. Mobile phone data were provided by an anonymous service provider in Kenya and is not available for distribution.* The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

This article has been approved by the Director of the Kenya Medical Research Institute.

Author contribution statement

A.W. designed the research, performed the analysis of data, and wrote the manuscript. C.D., J.C and J.S designed and implemented the field travel surveys. A.W., G.S., J.S., N.E., T.B., E.M., C.O., C.D., J.C, and C.O.B. all contributed to designing the research and manuscript preparation. All authors reviewed the manuscript.

Competing financial interests

All authors have no competing financial interests to declare.

References

1. Balcan, D., et al., *Multiscale mobility networks and the spatial spreading of infectious diseases.* Proc Natl Acad Sci U S A, 2009. 106(51): p. 21484-9.

2. Ferguson, N.M., et al., Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature, 2005. 437(7056): p. 209-14.
3. Longini, I.M., Jr., et al., *Containing pandemic influenza at the source.* Science, 2005. 309(5737): p. 1083-7.
4. Morens, D.M., G.K. Folkers, and A.S. Fauci, *The challenge of emerging and re-emerging infectious diseases.* Nature, 2004. 430(6996): p. 242-9.
5. Kilpatrick, A.M., et al., *Predicting the global spread of H5N1 avian influenza.* Proc Natl Acad Sci U S A, 2006. 103(51): p. 19368-73.
6. Funk, S., M. Salathe, and V.A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: a review. J R Soc Interface, 2010. 7(50): p. 1247-56.
7. Prothero, R.M., *Disease and mobility: a neglected factor in epidemiology.* Int J Epidemiol, 1977. 6(3): p. 259-67.
8. Meloni, S., et al., Modeling human mobility responses to the large-scale spreading of infectious diseases. Sci Rep, 2011. 1: p. 62.
9. Stoddard, S.T., et al., The role of human movement in the transmission of vector-borne pathogens. PLoS Negl Trop Dis, 2009. 3(7): p. e481.
10. Kimman, T.G. and H. Boot, The polio eradication effort has been a great success--let's finish it and replace it with something even better. Lancet Infect Dis, 2006. 6(10): p. 675-8.
11. Kyle, J.L. and E. Harris, *Global spread and persistence of dengue.* Annu Rev Microbiol, 2008. 62: p. 71-92.
12. Bajardi, P., et al., Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. PLoS One, 2011. 6(1): p. e16591.
13. Gaudart, J., et al., Spatio-temporal dynamics of cholera during the first year of the epidemic in Haiti. PLoS Negl Trop Dis, 2013. 7(4): p. e2145.
14. Le Menach, A., et al., Travel risk, malaria importation and malaria transmission in Zanzibar. Sci Rep, 2011. 1: p. 93.
15. Wesolowski, A., et al., *Quantifying the impact of human mobility on malaria.* Science, 2012. 338(6104): p. 267-70.
16. Santos, A., McGuckin, N., Nakamoto, H. Y., Gray, D., Liss, S. , *Summary of travel trends: 2009 national household travel survey.* National Technical Information Service: Trends in travel behavior 2011.

17. Buliung, R.N., Remmel, T. K, Open source, spatial analysis, and activity-travel behaviour research: capabilities of the aspace package. J. Geogr Syst 2008. 10: p. 191-216
18. Schlich, R., Axhausen, K. W. , Habitual travel behaviour: evidence from a six-week travel diary. Transp., 2003. 30: p. 13-36.
19. Noor, A.M., et al., Establishing the extent of malaria transmission and challenges facing pre-elimination in the Republic of Djibouti. BMC Infect Dis, 2011. 11: p. 121.
20. Craig, M.H., et al., Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa: part II. The impact of non-climatic factors. Trop Med Int Health, 2004. 9(12): p. 1258-66.
21. Service, G.S., Ghana Statistical Service: Ghana Living Standards Survey, 2008.
22. Wesolowski, A., et al., The use of census migration data to approximate human movement patterns across temporal scales. PLoS One, 2013. 8(1): p. e52971.
23. Pindolia, D.K., et al., Human movement data for malaria control and elimination strategic planning. Malar J, 2012. 11(1): p. 205.
24. Gonzalez, M.C., C.A. Hidalgo, and A.L. Barabasi, *Understanding individual human mobility patterns.* Nature, 2008. 453(7196): p. 779-82.
25. Song, C., et al., *Limits of predictability in human mobility.* Science, 2010. 327(5968): p. 1018-21.
26. Simini, F., et al., *A universal model for mobility and migration patterns.* Nature, 2012. 484(7392): p. 96-100.
27. Wesolowski, A., et al., Heterogeneous mobile phone ownership and usage patterns in Kenya. PLoS One, 2012. 7(4): p. e35319.
28. Wesolowski, A., et al., The impact of biases in mobile phone ownership on estimates of human mobility. J R Soc Interface, 2013. 10(81): p. 20120986.
29. Clarke, M., Dix, M., Jones, P. , *Error and uncertainty in travel surveys.* Transp., 1981. 10: p. 105-126.
30. Noor, A.M., et al., *The risks of malaria infection in Kenya in 2009.* BMC Infect Dis, 2009. 9: p. 180.
31. Tatem, A.J. and D.L. Smith, International population movements and regional Plasmodium falciparum malaria elimination strategies. Proc Natl Acad Sci U S A, 2010. 107(27): p. 12222-7.

32. Morgan, W.T.W., The ethnic geography of Kenya on the eve of independence: the 1962 census. Erdkunde 2000. 54: p. 76-87

FIGURES

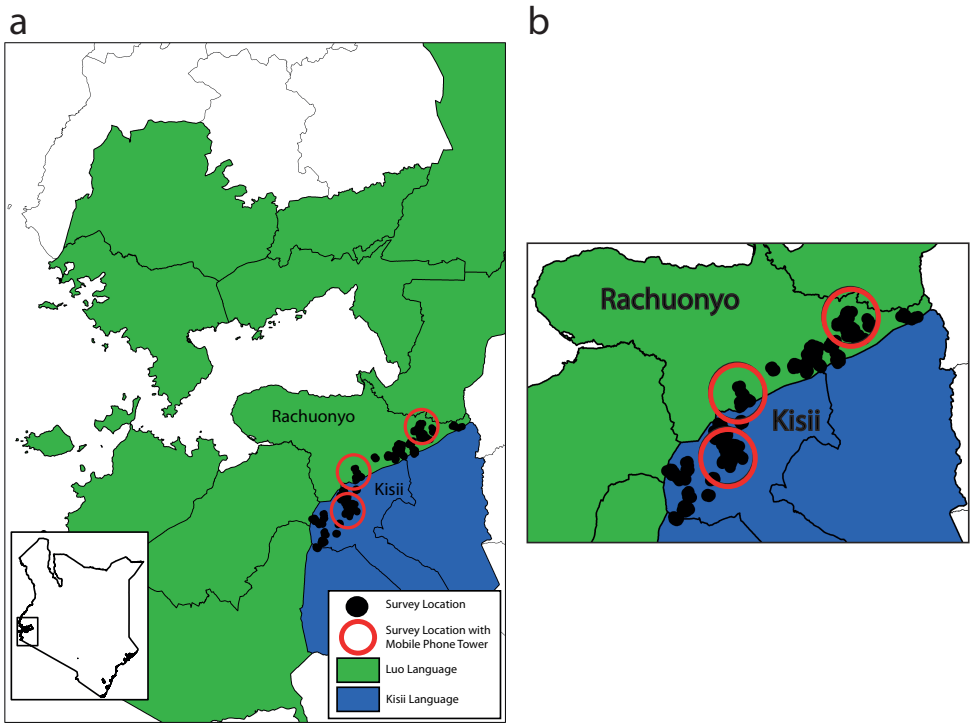

Figure 1: The household survey locations within the study site in western Kenya.

A) Surveys were taken at households within western Kenya (larger map is highlighted in the inset, created using ArcGIS v10.1) with their locations mapped as black points. Households within 3 km of a mobile phone tower are outlined in red. Areas are colored by their dominant language with DhoLuo (Luo language) in green and Kisii in blue. In Rachuonyo district, the dominant language is DhoLuo whereas it is Kisii in Kisii district. B) A zoomed image of the study site along with mobile phone towers.

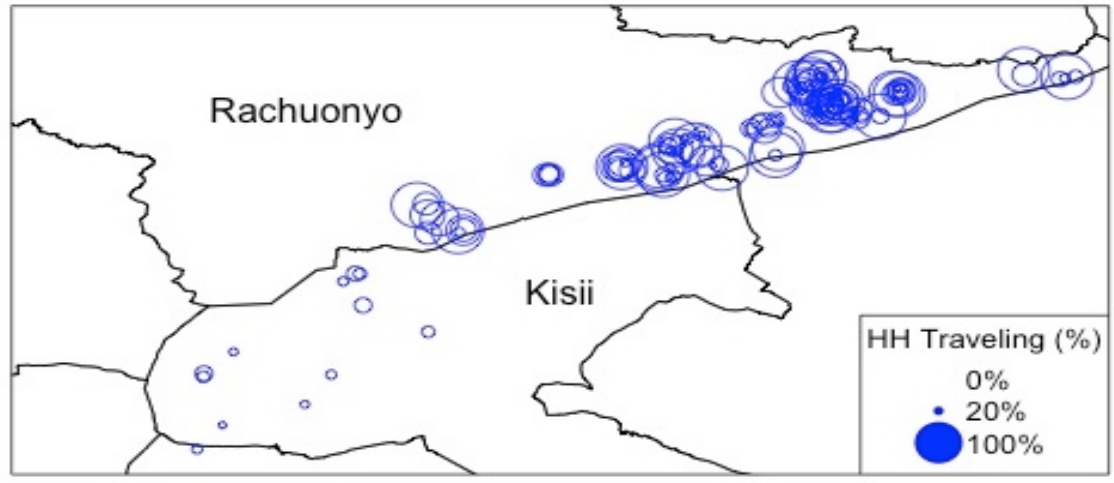

Figure 2: The percentage of individuals within a household who reported traveling. From the travel survey data, the percentage of individuals per household who reported traveling was quantified . Households within Rachuonyo traveled much more than those in Kisii (t=-7.401, df = 410.141, p-value < 0.001). This map was created using ArcGIS v10.1.

a **b**

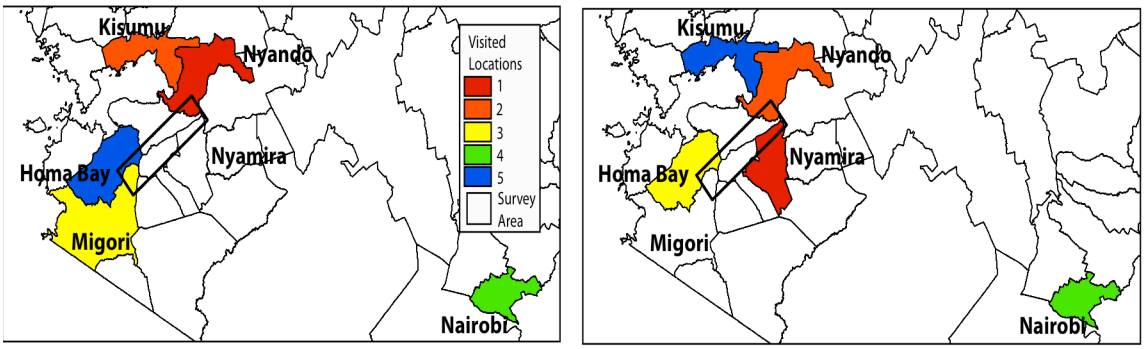

Figure 3: The locations of the most commonly visited districts. From the A) survey data and B) mobile phone data, the five most commonly visited districts are colored by their rank with the survey area outlined in black. The most common districts visited were Nyamira, Nyando, Homa Bay, Nairobi, Kisumu, and Migori, also primarily nearby districts and those including major population centers (Kisumu and Nairobi) (in descending order). This did vary slightly between Kisii and Rachuonyo. Amongst subscribers in Kisii the districts most commonly visited were: Nyamira, Nairobi, Gucha, and Migori whereas those in Rachuonyo commonly visited Nyamira, Nyando, Homa Bay, and Kisumu. The map was created using ArcGIS v10.1.

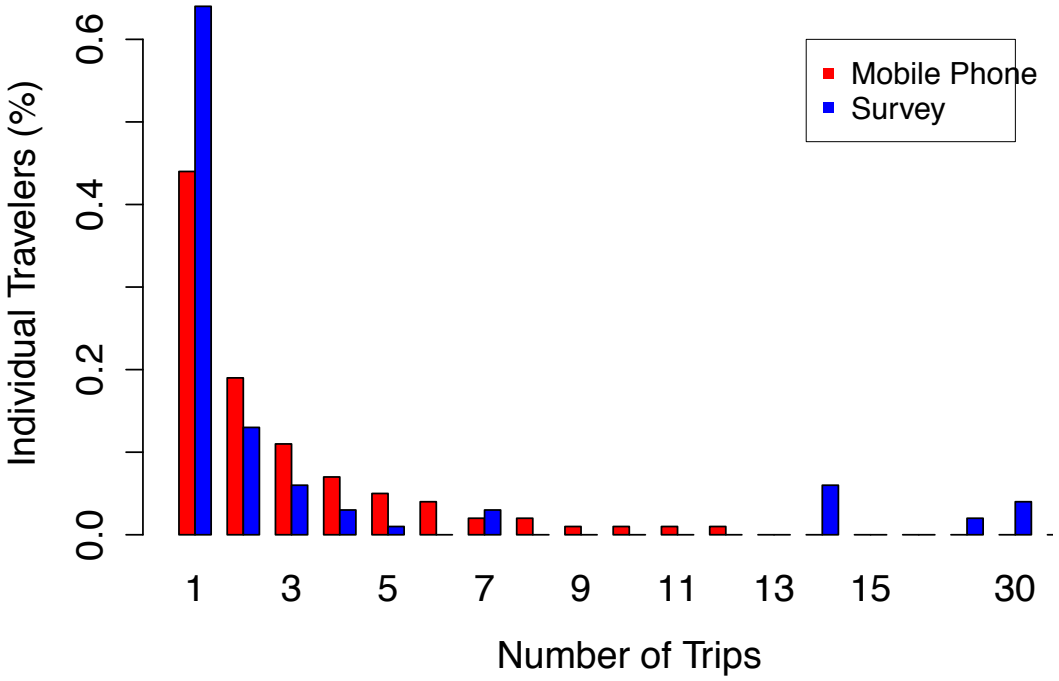

Figure 4: The number of trips taken by individuals from each data source. The distribution of the number of trips (between 1 – 60 trips) taken by individuals who traveled from the mobile phone data (red) and the survey data (blue) is shown.. In both surveys, individuals rarely reported taking more than one trip, whereas in the mobile phone data multiple trips were measured from a substantial number of subscribers (KS statistic: 0.7947, p = 0.0005).

TABLES

Table 1: Basic survey descriptive statistics. All percentages (sample size) do not necessarily add to 1 if the survey respondent did not answer the question.

	Travel Survey
Number of individuals	2650
% Male	46 (1222)
% Female	53 (1398)
% Adults (15+)	45 (1194)
% Children (0-14)	65 (1456)
Survey Dates	February, 2009
Number of households	776
Average household size	3.7

Table 2: The percentage of adults within a household who travel versus the percentage of those households who own a mobile phone. For households where 0-100% of the adults in the HH have traveled, the percentage of those HH who own a mobile phone. In general, the households where a higher percentage of adults have traveled are more likely to own a mobile phone (t=2.6441, df = 699.45, p-value = 0.0084).

Percentage of adults within a HH who have traveled	**Percentage of those HH who own a mobile phone**
0	45 (277)
20	75 (3)
25	100 (2)
33	50 (5)
50	64 (37)
67	100 (3)
100	56 (29)

Table 3: The basic travel statistics from the mobile phone data.

	Kisii	**Rachuonyo**
Number of mobile phone towers within study site	6	9
Number of subscribers	16,196	18,665
Number of travelers, trips lasting at least 1 day	61% (N=9,880)	95% (N=17,732)
Number of travelers, trips lasting at least 2 days	36% (N=5,830)	63% (N=11,759)

Table 4: A comparison between the two data sets. For both data sets, the type of travel data available and scale (spatial and population) available. In general, the survey data is able to provide a coarser picture of travel, although refined socio-demographic data about travelers. The mobile phone data can only provide estimates on subscriber travel and is not able to provide any socio-demographic information about travelers. In order to compare between both data sets, we estimated the number of subscribers and the number of subscribers who have traveled from the survey data (see Materials and Methods). In comparison to the actual values quantified using the mobile phone data, the survey data produces estimates an order of magnitude less than the observed quantities.

	Survey Data	**Mobile Phone Data**
Number of trips taken by individuals	Yes	Yes, for subscribers
Primary travel destination	Yes, district level	Yes, mobile phone tower
All destinations visited during traveling	No, only the primary destination	Yes, mobile phone tower
Duration of travel	No	Yes
Socio-demographic information about travelers	Yes	No
Spatial scale	District level	Mobile phone tower (~3km)
	Estimated Value – Survey Data	**Actual Value – Mobile Phone Data**
Total Population of Survey Site	19,744	
Number of Subscribers	2,500—11,500	35,000
Number of Subscribers who Travel	200—1,800	28,000

Table 5: The population weighted malaria-travel metric estimated importation from the survey data and mobile phone data in both study areas.

Data Source, Location	Population Weighted Malaria-Travel Metric	Top destination district for malaria-travel metric
Survey, Kisii	0.00014	Butere/Mumias
Survey, Rachuonyo	0.015	Nyando
Mobile Phone Data, Kisii	0.055	Nyamira
Mobile Phone Data, Rachuonyo	0.25	Nyamira

Table 6: A brief outline of the travel questions asked in the travel survey.

Travel Question
Have you made any overnight trips to another district in the last 3 months?
How many overnight trips have you made in the last 3 months?
Where did you spend the majority of time during this trip?
When did you get back?
What was your reason for traveling?

Appendix 1.3 – Intervention trial protocol

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk
Registry
T: +44(0)2072994646
F: +44(0)207299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS

SECTION A – Student Details

Student	**Gillian Stresman**
Principal Supervisor	**Dr. Teun Bousema**
Thesis Title	**Operational Strategies for the Identification and Targeting of Hotspots of Malaria Transmission**

If the Research Paper has previously been published please complete section B , if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Trials		
When was the work published	2013		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	CC-BY	Was the work subject to academic pier review?	**Yes**

***If yes, please attach evidence of retention, if no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.**

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published	
Please list the paper's authors in the intended authorship order.	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	Design of intervention package and preparation and conduct of surveys. Contributed to manuscript prepatation.

Student Signature: _____ _____ Date: ___17/02/2015_____

Supervisor Signature: ____ _____ Date: ___17/02/2015_____

Title: Study Protocol: A cluster-randomized trial to determine the impact of hotspot-targeted interventions on malaria transmission

Authors: Teun Bousema[1,2*§], Jennifer Stevenson[3*], Amrish Baidjoe[2], Gillian Stresman[1], Jamie T. Griffin[4], Immo Kleinschmidt[5], Edmond J. Remarque[6], John Vulule[7], Nabie Bayoh[7], Kayla Laserson[7,8], Meghna Desai[7,8], Robert Sauerwein[2], Chris Drakeley[1], Jonathan Cox[3]

[*]Contributed equally; [§]Corresponding author

Affiliations:

[1]Department of Immunology & Infection; Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2]Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands

[3]Department of Disease Control; Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[4]MRC Centre for Outbreak Analysis & Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

[5]MRC Tropical Epidemiology Group, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, United Kingdom..

[6]Department of Parasitology, Biomedical Primate Research Centre, Rijswijk, The Netherlands

[7]Kenya Medical Research Institute, Centre for Global Health Research, Kisumu, Kenya

[8]Centers for Disease Control and Prevention, Division of Parasitic Diseases, Atlanta, USA

Reference: Bousema et al. (2013). Trials, 14:36.

Abstract

Background: Malaria transmission is highly heterogeneous in most settings resulting in the formation of recognizable malaria hotspots. Targeting these

hotspots may represent a highly efficacious way to control or eliminate malaria if they fuel malaria transmission to the wider community.

Design: Hotspots of malaria will be determined based on spatial patterns in aged-adjusted prevalence and density of antibodies against malaria antigens apical membrane antigen-1 and merozoite surface protein-1. The community effect of interventions targeted at these hotspots will be determined. The intervention will comprise larviciding, focal screening and treatment of the human population, distribution of long-lasting insecticide-treated nets and indoor residual spraying. The impact of the intervention will be determined inside and up to 500 m outside the targeted hotspots by PCR-based parasite prevalence in cross-sectional surveys, malaria morbidity by passive case detection in selected facilities and entomological monitoring of larval and adult *Anopheles* populations.

Discussion: This study aims to provide direct evidence for a community effect of hotspot-targeted interventions. The trial is powered to detect large effects on malaria transmission in the context of ongoing malaria interventions. Follow-up studies will be needed to determine the effect of individual components of the interventions and the cost-effectiveness of a hotspot targeted approach where savings in the number of compounds that need to receive interventions should outweigh the costs for hotspot-detection.

Trial registration: **NCT01575613**

Introduction

The transmission of infectious agents is highly heterogeneous in space and time. For many infectious diseases, a small number of human hosts are most frequently or most heavily infected while the majority of a local population is much less affected [1-4]. In malaria this heterogeneity of disease transmission often results in variation in malaria incidence within small areas [5-10]. In some settings the non-random distribution of malaria incidence between households appears to conform to the ''20/80 rule'' [2], whereby approximately 20% of a host population contributes 80% of the cases of an infectious organism [5,9]. The factors underlying the micro-epidemiology of malaria are not fully understood but include variation in distance to the nearest mosquito breeding site [5-9,11], wind direction[12], house construction features [6,8,9,13,14], human behavioural [7,8,13] and genetic factors [7,8,15].

Heterogeneity in malaria transmission has implications for malaria control. Individuals who are bitten most often are most likely to be infected and can amplify transmission by infecting a large number of mosquitoes with malaria parasites. Estimates of the basic reproductive number (R_0), a central concept in infectious disease epidemiology defined as the average number of secondary cases arising in a susceptible population as a result of a single human case over the course of their infection, are sensitive to assumptions of heterogeneous mosquito exposure. R_0 may increase up to fourfold as a consequence of heterogeneous mosquito exposure [2,4,16].

The large influence of heterogeneous exposure on malaria transmission also suggests that interventions targeting areas of comparatively high exposure can be highly effective. Woolhouse and colleagues suggested that, depending on the costs of identifying hotspots of transmission, treating the core 20% might be preferable to non-targeted interventions on economic grounds [2]. If hotspots fuel transmission to a wider geographical region, community protection may be achieved by targeting those individuals that are most important for disease transmission. This hotspot targeted approach will only be (cost) effective if the assumption that hotspots fuel transmission in surrounding areas is correct – and then only if such hotspots can be reliably detected [4]. Several approaches to identify hotspots of malaria transmission have been proposed in recent years. Incidence of clinical malaria is a frequently used indicator of hotspots of malaria transmission [8-10] but is affected by a differential acquisition of protective immune responses inside and outside hotspots [17,18]. Geographical clustering of asymptomatic parasite carriage may be a more stable indicator of hotspots of transmission [10] and in areas of moderate or low endemicity hotspots might be most readily detected using serological markers of malaria exposure [9,10,19-22]. In an area of moderate endemicity in Tanzania, serological data have been used to identify clinically- and entomologically-confirmed hotspots of malaria transmission with 96% sensitivity and 82% specificity [9].

This manuscript describes a methodological approach to identifying hotspots of malaria transmission and a protocol for the evaluation of a hotspot targeted intervention. The aim of this intervention study is to determine whether the simultaneous roll-out of interventions in hotspots of malaria transmission has a community-wide effect that extends beyond the hotspot boundaries and results in local elimination of malaria.

Defining the intervention clusters

Study area

The study will be conducted in highland fringe localities (1400-1600 m altitude) in Rachuonyo South District, Western Kenya (34.75-34.95°E, 0.41-0.52°S). The predominant ethnicity in Rachuonyo is Luo. Local residents depend upon farming, cattle and goat herding for subsistence. Compounds comprise an average of 2 houses (IQR 1-3) and are distributed broadly across a rolling landscape intersected with small streams and rivers. The main malaria vectors in the area are *Anopheles gambiae s.s.*, *An. arabiensis*, and *An. funestus*. Malaria transmission is seasonal, with two seasonal peaks in malaria cases reflecting the bimodal rainfall pattern, with the heaviest rainfall typically occurring between April and June and a smaller peak between October and November each year. Most malaria is caused by *Plasmodium falciparum*. Community cross-sectional surveys conducted in 2010 indicated parasite prevalence averaging 14.8% in the general population but ranging between 0% and 51.5%. School surveys carried out in primary schools in the same year indicated an average parasite prevalence of 25.8% in 7-18 year olds (range for individual schools 0-71.4%). Insecticide Treated Nets (ITNs) have been promoted by the Ministry of Public Health and Sanitation for many years and distribution campaigns have taken place through antenatal and child health clinics, reaching a coverage for under 5s of 82.7%, as determined in surveys in 2010 (unpublished data). In addition, community-wide mass distribution of ITNs was undertaken by the DOMC in 2011. Indoor Residual Spraying (IRS) was first carried out in Rachuonyo South in mid-2008 with financial support of the US President's Malaria Initiative. Reported house coverage with IRS in Rachuonyo South was estimated at 70.3% in 2009 and 74.3% in 2010.

Sampling strategy to identify hotspots of transmission

We will select a 5x20 km (100 km^2) area within which results from recent community and school malaria surveys suggest highly heterogeneous malaria exposure. The study area will be divided in to 400 cells of 500x500 m that are further subdivided in four sub-cells of 250x250 m.

All structures in the area have been geo-located using contemporaneous high-resolution satellite data [Quickbird; DigitalGlobe Services, Inc., Denver, Colorado] that were acquired and processed using standard digital image processing techniques [ENVI 4.8, Exelis Visual Information Solutions, McLean, VA USA]. Pan-sharpened colour images were then imported into a geographic information system [ArcGIS 9.2; Environmental Systems Research Institute, Redlands, USA] and all structures were digitized manually giving a total of 8,632 structures with a median of 45 (interquartile range 35-52) per 500x500 m cell. We aim to obtain measurements from ≥50 individuals per 500x500 m cell since estimates of sero-conversion rates from fewer than 50 observations from all age-groups combined are likely to be unreliable [9]. To maximize the discriminative power of serological markers of exposure, we will sample individuals in pre-defined age strata (≤5 years; 6-11 years; 12-15 years; 16-25 years and >25 years). For logistical reasons, our unit of sampling will be the compound.

To limit the chances of two selected structures belonging to the same compound an iterative sampling approach will be used that involved randomly selecting a "seed" structure and then removing all closely neighbouring structures (within 50 m) from the sample universe before proceeding to select a second structure. This process will be repeated until all possible "non-neighbouring" structures have been selected. From the resulting list of eligible structures a stratified sample of 16 structures will be chosen from each 500m x 500m cell. To ensure maximum geographical coverage, at least one compound will be selected from each 250x250 m sub-cell while the number of compounds selected from each of the sub-cells will be weighted by the structure density in these sub-cells.

All other structures in which people sleep and which are associated with each selected compound will be included. The target number of 50 observations per 500x500 m cell is chosen irrespective of the population density of the cells.

Data collection and measurements to identify hotspots of transmission

Enumeration

For planning purposes the field area will be sub-divided into 20 blocks of 5x4 cells (i.e. 2.5x2 km in size). Teams will be provided with a printed overview map of the block they are working in (figure 1), as well as detailed high resolution maps incorporating the QuickBird satellite data for each 500x500 m cell. Each team will also be provided with a handheld global positioning system (GPS) receiver [Garmin 62*S*; Garmin International, Inc., Olathe, KS, USA] that has been pre-loaded with the selected compound positions, track locations and cell boundaries. An enumeration team, comprising one field worker, a reporter and a local guide, will visit selected compounds to explain the study procedures, enumerate inhabitants, collect information on house characteristics and inform residents that the survey team will visit later that day. In situations where none of the structures within a selected compound corresponds with a residential building, the selected compound will be replaced with the nearest visible inhabited compound. The location of this replacement will be recorded on the satellite images, mapped using the GPS and recorded on the enumeration forms.

All compounds where at least one adult (>20 years) and one child (<15 years) are permanent residents (defined as sleeping in the structure) qualify for enrolment. If the head of the compound agrees to participate, the geographical coordinates of the main house of the compound will be recorded and compound and individual house codes will be written on the doors of all sleeping structures with a permanent marker. The names and ages of all compound members will be recorded on study forms and information on compound and house characteristics, including structure type, ITN coverage, and IRS history, will be collected using a pre-coded questionnaire (Programmed in Visual Basic, Visual CE v11.0) on a Personal Digital Assistant (HP Ipaq 210, Windows Mobile 6.1). A personal study

identification card will be issued to each individual, which has to be shown to the sampling team when they visit later that same day.

The field workers will carry a checklist to record the cumulative number of selected individuals for each age category. The order in which compounds are visited will be randomly selected based on a computer-generated list. After completing a compound, the enumeration team continues to the next compound until at least 10 compounds have been enumerated. If the checklist indicates that age targets are not met at this point, they will continue visiting compounds according to the list until each age target is met.

Sampling

After enumeration, participating compounds will be visited by a sampling team consisting of two fieldworkers trained in interviewing and sampling techniques. Sampling teams will be provided with relevant maps, compound lists, enumeration forms and ID cards in advance. Compounds will be identified by codes marked on doors at the point of enumeration; compound occupants will be asked to present their identification card for formal confirmation. Informed consenting will be conducted and the name, gender, age, residency history, travel history, ITN use and sleeping times of each compound member will be recorded. The temperature of each compound member will be measured by auxiliary thermometer. For all febrile individuals (>37.2 °C), a rapid diagnostic test [RDT; Paracheck®, Orchid BiomedicalSystems, India] detecting *P. falciparum*-specific histidine rich protein-2 will be performed. For all individuals surveyed, a single finger prick sample will be taken for haemoglobin (Hb) measurement using a HemoCue photometer [HemoCue 201+, Angelholm, Sweden] and three droplets transferred onto a filter paper [3MM Whatman, Maidstone, UK] for serum and DNA collection. After transfer to a field laboratory, filter papers will be dried overnight and stored in plastic bags with silica gel. Once a week, samples will be transported to the KEMRI/CDC laboratory in Kisumu and stored at -20 °C until further processing. All individuals with an Hb ≤11 g/dL will be given hematenics; individuals with an Hb ≤ 6g/dL will be accompanied to a nearby health centre for further care. Febrile individuals who are found parasitaemic by RDT will be given artemether-

lumefantrine [AL, Coartem®, Novartis, Switzerland]; women of child bearing age who are RDT positive will be assessed for pregnancy and offered a pregnancy test if deemed appropriate. Febrile children below 6 months of age and women who are suspected or found to be pregnant or are unwilling to be tested will be transported to the nearest health facility for a full assessment and treatment.

Malaria parasite prevalence

A combined extraction of DNA and elution of antibodies will be performed on the samples collected. Two discs with a diameter of 2.5 mm will be cut from the centre of a single filter paper bloodspot using a hole-puncher and will be eluted in deep well plates with addition of 1120 μL of a 0.5% saponin/phosphate buffered saline solution [Sigma Aldrich]. DNA will be extracted using the protocol described by Plowe [23]; parasites will be detected by nested PCR [24,25].

Serological markers of malaria exposure

Total immunoglobulin G (IgG) antibodies against *P. falciparum* apical membrane antigen (AMA-1) and merozoite surface protein 1 ($MSP\text{-}1_{19}$) will be detected by ELISA using standard methodology[26,27]. Three serological outcome measures will be used to determine spatial patterns in malaria exposure: i) the combined antibody prevalence i.e seropositive for AMA-1 and/or $MSP\text{-}1_{19}$; ii) the age-adjusted $\log_{10}$-transformed optical density (OD) [21,28]; iii) the age-dependent sero-conversion rate (SCR) for combined AMA-1, $MSP\text{-}1_{19}$ antibody prevalence [21,26].

Definition of hotspots

SaTScan software[29] will be used for the detection of spatial clustering in antibody prevalence (Bernouilli model) and log10-transformed age-adjusted OD values (normal probability model). Circular and elliptic shaped windows [29,30] will be used to systematically scan the study area as a whole and segments of the study area using a 2x4 km rolling window. Hotspots will be allowed to be <1 km in radius and include <25% of the population of each window scanned. Scanning of segments of the study area will be done to improve the sensitivity of the scan to detect local hotspots. Local hotspots may not be detected when scanning the area as a whole since altitude differences in the study area result in variations in average levels of

transmission intensity. A hotspot will be defined as an area for which there is strong evidence ($p<0.05$) that the observed prevalence and/or density of combined AMA-1 and MSP-1_{19} antimalarial antibodies is higher than expected values. Expected values are based on average values for the area as a whole and for the 2x4 km rolling window.

Since malaria antibodies are relatively long-lived and may indicate current as well as past malaria exposure, parasite prevalence inside and outside hotspots of malaria exposure will be determined by PCR to confirm ongoing transmission in serologically defined hotspots.

Selection of hotspots and evaluation areas

Since habitation in the study area is fairly evenly distributed, with every 500x500 m cell having six or more residential structures, clusters are unlikely to be isolated geographically. To minimise the influence of neighbouring hotspots on malaria transmission in selected intervention or control hotspots, we will select hotspots for which there are no other hotspots detected within 1 km in any direction from the hotspot boundary. The hotspot targeted intervention will be evaluated in the area surrounding the hotspot (evaluation zones). The evaluation zone will comprise the area surrounding the hotspot up to 500 m from the hotspot boundary in each direction.

Components of the intervention

Intervention clusters

Four interventions will be rolled out in the period preceding the long rainy season: larviciding, focal screening and treatment (FSAT), LLIN distribution and indoor residual spraying (IRS). The details of interventions, and their timing have been agreed upon in collaboration with the Kenyan Division of Malaria Control (DOMC) of the Kenyan Ministry of Public Health and Sanitation (MOPHS). Ten per cent of households will be visited 1-2 weeks after the intervention to assess any short-term side-effects of the FSAT, LLINs and IRS.

Larviciding

All permanent aquatic mosquito habitats in hotspots will be mapped using handheld GPS receivers during the dry season. In the period preceding the long rainy season (April), and throughout the long rainy season (until September) all stagnant water bodies (permanent and temporary) inside hotspots will be treated on a weekly basis with water-dispersible granule formulations of the commercial strains of *Bacillus thuringiensis var. israelensis* (Bti), VectoBac, that will be provided by Valent BioSciences Corp., IL. Larviciding will be carried out using previously published protocols [31]; the entire hotspot area will be examined for water bodies on a weekly basis, all of which will be included in the intervention. Spot-checks for surviving anopheline larvae and pupae will be done on a weekly basis.

Focal screen and treatment (FSAT)

All compounds in hotspots will be visited and the temperature of each individual will be determined. All individuals aged 6 months -15 years regardless of temperature and all older individuals who are febrile (tympanic temperature ≥37.5 °C) will be tested for malaria parasites using HRP-2 and pLDH based RDT (First Response®, Premier Medical Corporation Ltd.,India). If one or more individuals are found to be RDT positive the entire compound will receive a curative dose of AL with the exception of pregnant women and children below 6 months of age. Because of the different times at which treatment is initiated, only the first treatment dose will be supervised by community health workers and given with fatty food (>1.5g fat) to facilitate absorption. The second daily dose will be taken without direct supervision but advice on taking the treatment with food will be given; all empty blisters will be collected by community health workers after treatment has been completed to monitor adherence.

Long-lasting insecticide treated nets

All compounds in hotspots will receive one LLIN per two house members. LLINs (Permanet® 3.0) were donated by Vestergaard Frandsen. House members will be given verbal information and leaflets on proper use of nets and study personnel will assist in hanging the LLINs within houses. Correct usage and retention of study nets will be assessed by questionnaire 6 weeks after distribution.

Indoor residual spraying

Routine annual indoor residual spraying (IRS) with Deltamethrin or lambda cyhalothrin (ICON) is undertaken at six-monthly intervals in all structures where people are sleeping. The IRS campaigns are jointly funded by the Government of Kenya and the US President's Malaria Initiative, and implemented by the Research Triangle Institute (RTI) with the DOMC and District Health Management Teams. For this study IRS will continue as normal except that implementation will be scheduled prior to the rains and start of the malaria transmission season (March-April) in intervention hotspots.

Control clusters

Control clusters will receive the routine malaria control measures which for 2012 will be the annual IRS programme as detailed above and continued case management at health facilities. The IRS is scheduled to take place in April- May 2012. No LLIN distribution campaigns are planned for 2012.

Design of the randomized evaluation

Sensitization and recruitment

Prior to the implementation of the interventions, meetings with district administrative and health representatives in the selected areas will be organised. Community meetings will be held with local chiefs, community elders and opinion leaders, school representatives and church leaders. All compound in the selected intervention areas will be visited prior to the intervention; the procedures of the interventions and evaluation procedures will be explained to all compound members present. ID cards will be distributed that will be used for identification purposes during compound visits and for identification of compound members who visit health facilities in the area.

Randomization

Hotspots with their surrounding evaluation areas, will be randomized to the intervention or control arm using computer generated tables. No stratification by parasite prevalence or altitude will be undertaken.

Hypotheses and Outcomes

Hypotheses

Hotspot targeted interventions with larviciding, LLINs, IRS and FSAT will reduce malaria transmission inside and outside hotspots of malaria transmission.

The community effect of hotspot targeted interventions, defined as the impact on parasite prevalence in the evaluation zone surrounding the hotspot, is a function of distance to the hotspot boundary.

Primary and secondary outcome measures

The primary outcome measure is:

Parasite prevalence by PCR in the evaluation zone surrounding malaria hotspots in intervention and control clusters

Secondary outcome measures are:

Parasite prevalence by PCR in the evaluation zone surrounding malaria hotspots in relation to distance to the boundary of hotspots in intervention and control clusters

Indoor and outdoor *Anopheles* mosquito densities inside and outside hotspots of malaria transmission in intervention and control clusters

The presence of *Anopheles* larvae in mosquito breeding sites in malaria hotspots in intervention and control clusters

The number of malaria cases reporting at health facilities, coming from intervention and control clusters

Reported side effects and acceptability of FSAT, LLINs and IRS

Evaluation

Cross-sectional surveys

Three cross-sectional surveys will be conducted: at baseline prior to the interventions, during the peak transmission season, and at the end of the peak

transmission season. For each cross-sectional survey, 25 compounds that are located inside hotspots, 25 compounds that are located <250 m from the hotspot boundary and 25 compounds that are located 250-500 m from the hotspot boundary will be randomly selected. This strategy is expected to give ≥ 100 individual observations from each of these three areas. To minimize confounding by neighbouring hotspots, an exclusion buffer will be incorporated in the selection of compounds, ensuring a minimum distance of ≥500 m from neighbouring hotspots.

Study teams will visit selected compounds and, subject to obtaining informed consent, collect information from inhabitants of all houses that belong to that compound using PDAs. For individuals older than 6 months, tympanic temperature will be measured and a finger prick blood sample (~300 μL) will be collected for assessment of haemoglobin concentration and for collection of nucleic acids and serum on Whatman 3MM filter paper [Maidstone, UK]. Whole blood will be collected in BD K2EDTA microtainers [BD Becton, Dickinson and Company, UK] in selected clusters for more detailed molecular analyses. A RDT will be used to determine malaria infection for all febrile individuals. Those with a positive RDT will receive AL and/or will be referred to a health centre for further care.

Passive case detection

A passive case detection system will be introduced in government and mission health facilities to monitor individuals presenting with malaria. Facilities will be selected to cover intervention and control clusters. For this, the catchment areas of health facilities in the area have been determined. Individuals from intervention and control clusters will be asked to present a household card whenever visiting a health facility. This household card will be linked to geo-located compounds. For individuals who present without a household card, other information that allows geo-location will be collected, such as nearest school. Tympanic temperature will be measured, and an RDT used to determine parasite carriage for each individual with measured or reported fever.

Entomological monitoring

In a subset of the control and intervention clusters, larval and adult mosquito abundance will be monitored. All permanent breeding sites in hotspots will be mapped and from a random selection of 15 sites per hotspot the presence or absence of early and late stage anopheline larvae and pupae will be assessed using a 250 ml mosquito dipper . Five dips will be done in sites smaller than 5 m^2; 10 dips in sites larger than 5 m^2. This will be carried out at two-weekly intervals. Adult collections of anophelines will be carried out twice per month in 36 randomly selected houses in each cluster selected in cross-sectional surveys.. Twelve houses will be selected within the hotspots, of which 4 will be sampled by pyrethrum spray catch (PSC), 4 for indoor light trap collections and 4 for outdoor light trap collections. Outside the hotspot 24 houses will be randomly selected of which 8 will be sampled by PSC, 8 for indoor and 8 for outdoor light traps.

PSC will be carried out indoors according to standard WHO protocols [32]. CDC miniature light traps [Model 512; John W. Hock Company, Gainesville, Florida, US] will be used following previously published procedures to sample mosquitoes indoors [33] and outdoors [34]. The effective range of CDC light traps for outdoor mosquito sampling has been estimated as 5 m [35]. Accordingly, outdoor sampling will take place 15 m from selected houses to prevent inhabitants acting as unshielded bait. All traps will be set at 1830 hours and collected at 0630 hours. On eight randomly selected light traps indoors and/or outdoors, a collection bottle rotator will be fitted (Model 1512, John W. Hock Company, Gainesville, Florida, US) which allows collection cups to rotate every 2 hours to estimate vector abundance at intervals throughout the night. Vector abundance, parity rates and the proportion of anopheline females unfed, fed, gravid, and infected will be determined for each species [36] and compared between the two study arms.

Statistical considerations

Sample size

All available malaria simulation models indicate that malaria transmission in the area surrounding intervention hotspots will decrease considerably because malaria transmission is effectively interrupted in those compounds that seed

transmission to a larger geographical area [2,16,37]. However, there are no published studies that quantify the impact of hotspot-targeted interventions. We estimated the predicted impact of targeted interventions in our study area using one of the leading individual-based simulation models [37] using human, entomological and parasitological characteristics collected at our sites in Kenya. We modelled three scenarios in situations with a pre-intervention parasite prevalence in the human population of 10-20%: i) no additional interventions; ii) targeted distribution of long lasting insecticide treated nets (LLINs), reaching 90% of the population in hotspots and iii) targeted LLINs and targeted effective IRS reaching 90% of the population in hotspots (Figure 2). The impact of larviciding is currently insufficiently parameterized to be included in the model [37].

Our simulations show that targeted interventions can interrupt transmission completely, both inside and outside hotspots of malaria transmission, reducing overall parasite prevalence to <5%, in a manner that appears sustainable in the following years (see figure 2). These predictions have to be interpreted with caution since i) the simulation model has not been prospectively tested; ii) there is no published evidence that quantifies the impact of hotspot targeted interventions on transmission intensity in the wider community; and iii) the intensity of transmission will be highly variable between hotspots in our study area. There is insufficient evidence on which to base power calculations for a cluster-randomized trial; however, these simulations can give an indication of the size of the effect of the planned interventions. The primary outcome measure is parasite prevalence in the evaluation zone. A previous study on the community benefits of insecticide treated nets in Asembo Bay, western Kenya, indicated that an indirect beneficial effect on malaria transmission is most pronounced within 500 m from the intervention area [38]. We used this finding to define our evaluation zone surrounding the hotspots. Assuming a sample of 200 randomly selected individuals in the evaluation zone of each cluster, a coefficient of variation of true proportions between clusters within each group (k) of 0.4 and mean parasite prevalence of 15% and ≤ 5% in the control and intervention clusters respectively, would require 5 clusters per study arm for 80% power and 5 significance (α). This

power calculation is based on a comparison between arms and the assumption that parasite prevalence will remain unaltered in the control arm.

To estimate the impact of the interventions on the hotspots themselves a sample size of 100 individuals in each of 5 of clusters (hotspots) per study arm, will be required to detect a similar difference between intervention and control clusters (≤ 5% versus 15% mean prevalence), assuming k=0.5, 80% power and 5% significance.

Data analysis

The primary analysis will be based on intention to treat whereby all evaluation areas are included in the analysis, regardless of the level of coverage. The main outcome measure, parasite prevalence, will be analysed as binary variable. For the primary study outcome, we will compare parasite prevalence in the evaluation zones of intervention and control clusters using a logistic multilevel generalised linear model using Stata version 12 [Stata Corporation, Texas, US] to account for clustering per compound and random effects to account for differences between study clusters. For secondary study outcomes, we will relate parasite prevalence to distance to the hotspot boundary in meters and in bins of 100 m; this analysis will be done for each of the clusters separately by Generalized Estimating Equations (GEE), adjusting for correlations between observations from the same compound. Indoor and outdoor *Anopheles* densities will be compared between study arms using GEE models and Poisson or Negative Binomial distributions. The proportion of productive breeding sites will be compared between intervention and control hotspots by GEE models, adjusting for correlations between observations from the same clusters.

Ethics considerations

Ethics approval

The study proposal received ethics approval from Scientific Steering Committee (SSC), the Ethical Review Committee (ERC) of the Kenya Medical Research Institute (KEMRI) Nairobi under proposal number SSC 2163, the London School of Hygiene

& Tropical Medicine ethics committee (#6111), and from Centers for Disease Control and Prevention (with exempt status).

Informed consent

IRS is to be conducted as part of the routine district-wide malaria control programme. Consent will be obtained verbally at the compound by community health workers and spray operators recruited by the Ministry of Public Health and Sanitation as is consistent with their operating procedures. Ahead of targeted distribution of LLINs, informed, written consent will be sought at the house level from the head of the household or representative in the presence of an independent witness. Larviciding will be done after consulting with and receiving approval from the DOMC, the Kenyan Pest Control Product Board (PCPB), the district administrative, fisheries and health teams and after community meetings. Verbal consent will be sought from owners of or persons responsible for any privately owned permanent breeding sites in the intervention areas (e.g. fish ponds). Since most mosquito breeding sites are not restricted to particular households, consent at household level is not practical and approval from the community, DOMC and PCPB is considered adequate.

Before FSAT and cross-sectional surveys, informed written consent will be sought from individuals; and/or their parents/guardians, and confirmed by an independent witness. Assent forms will be signed by children between the ages of 13 and 17 years and by their parents/guardians. Each assent form will be accompanied by a consent form signed by the parent/guardian. All consent and assent forms will be countersigned by the staff member obtaining consent and a copy will be left at the households.

Trial oversight

Ethical and safety aspects of the study are overseen by an independent monitor. No data safety and monitoring board (DSMB) will be installed. IRS and LLINs form part of routine malaria control in Kenya and will be undertaken in collaboration with the DOMC. Larviciding with Bti has been undertaken previously in neighbouring districts and has previously been shown to pose no health risk [39].

The proposed form of FSAT where household members of parasite carriers are treated regardless of their parasite status by microscopy is not part of the current malaria strategy of the Kenyan DOMC although screening and treatment of asymptomatic parasite carriers is recommended[40]. Our FSAT approach is based on the assumption of a high proportion of submicroscopic infections among asymptomatic individuals [41], especially among household members of microscopy positive individuals [42]. The drug used throughout the study is the first line antimalarial treatment in most of East Africa, including Kenya.

Discussion

Targeting interventions to hotspots of malaria transmission is frequently mentioned as a cost-effective approach for malaria control and elimination [2,4,5,43] although direct evidence for a community effect of hotspot-targeted interventions is currently unavailable. The present study aims to determine this effect in a cluster-randomized intervention trial.

Valuable information on how to quantify community effects of malaria control interventions comes from trials with ITNs [44]. Mortality rates[45], incidence of severe malaria [46], incidence of uncomplicated malaria[38,46], anaemia[38] and high density parasitaemia [38] have been shown to be reduced in compounds without ITNs that were in close proximity of compounds with ITNs. Hawley and colleagues found that individuals living in control villages within 300 m of ITN villages in Kenya experienced a level of protection similar to that experienced by individuals living in ITN villages and that this was plausibly due to area-wide effects on vector densities and sporozoite positive mosquitoes [38]. Despite similarities, hotspot targeted interventions may differ considerably from untargeted ITN campaigns in their community impact. Mathematical simulation models suggest that the impact of hotspot targeted interventions may be much larger than that of community-wide ITN distributions and may lead to local malaria elimination [4]. In line with this, our trial is powered to detect large effects on malaria transmission. However, two of the major assumptions underlying the optimistic model outcomes are incompletely understood. Firstly, the stability of hotspots is central to ensure sustainable community effects. Hotspots of (asymptomatic) parasite carriage are

generally assumed to be stable [4,10]. However, a recent report that wind direction in relation to breeding site location may be a key element in determining the location of hotspots [12], suggests that local environmental factors may also influence the spatial stability of hotspots. We believe that our approach to define hotspots serologically may be less susceptible to (short-term) variations in wind direction or other ecological factors since it effectively bases hotspot-detection on immunological markers of cumulative malaria exposure [26]. Secondly, a community effect of hotspot-targeted interventions strongly depends on mosquito mixing patterns. Mosquito mixing patterns are unlikely to be homogeneous. Reported site-fidelity where mosquitoes are likely to return to the same compounds [47,48] remains to be confirmed but could considerably reduce the community effect of hotspot-targeted interventions. The most informative measure of mixing patterns may be an approach where parasite populations are tracked in human populations, inside and outside hotspots of malaria transmission.

Research on the impact of community interventions where 'herd coverage' is required to ensure effectiveness raises a number of practical issues. Similar to mass drug administration campaigns, high community coverage [49,50] is required in our study to reduce R_0 to values below 1. Our intervention is further challenged by a dependence on community participation in control measures that are only rolled out in a selected proportion of this community. Gaining community trust is essential to the study's success and we expect good participation rates after our lengthy sensitization process and strong involvement of community leaders and local workers in all aspects of the study preparation, intervention and evaluation.

Even with excellent participation rates, the nature of our intervention will remain susceptible to contamination from neighbouring hotspots. An ideal study setting would comprise a large number of geographically isolated clusters, each being an independent focus of malaria transmission, with within these clusters clearly defined hotspots [4]. Our real-life setting falls short of this ideal scenario. The continuous inhabitation in the area makes it unlikely that clusters are geographically isolated. We aim to minimize contamination from non-targeted malaria hotspots by incorporating an exclusion zone in our selection of eligible

hotspots and in the selection of compounds in the evaluation phase. We nevertheless expect that there will be residual contamination that will be reflected in a spatial component in the effect of hotspot targeted interventions: the level of contamination will be highest in areas furthest away from the targeted hotspot and nearest to untargeted hotspots. Similarly, the effect of the intervention within the targeted hotspots may be largest in those compounds that are most remote from the nearest untargeted compound. Mathematical simulation models of are expected to be valuable as an integral part of the evaluation of our intervention to assess the plausibility that a change in transmission intensity can be attributed to the intervention.

The current study is not designed to determine the effect of individual interventions. While simulations suggest that targeted interventions with LLINs and IRS will be sufficient to eliminate malaria locally [4], we chose a relatively comprehensive package of malaria control measures incorporating a wide variety of available interventions, targeting both the mosquito vector and the malaria parasite in humans. If findings from the current study prove promising, a next step will be to determine the optimum package of tools for hotspot-targeted interventions across a range of settings. This package will differ between different settings. Larviciding, for example, will be most beneficial in settings where breeding sites are discrete and well-defined [51-53] while the effects of IRS and ITNs will be affected by insecticide resistance, amongst other factors [44]. Importantly, follow-up studies should determine the cost-effectiveness of the hotspot approach to assess whether savings in the number of compounds that need to be targeted for conventional vector control in the absence of hotspot treatment outweigh the costs for hotspot-detection and coordination of hotspot interventions.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

Trial design: TB, JS, IK, RS, CD, JC; design of intervention packages JS, AB, GS, JV, NB, KL, MS, JC; design of analytical plan: TB, IK, JTG, KL, MD, RS, CD, JC; preparation and

conduction of surveys: JS, AB, GS; contribution of reagents EJR, contributed to manuscript preparation TB, JS, AB, GS, IK, JTG, EJR, JV, NB, MS, RS, CD, JC. The author(s) declare that they have no competing interests.

Acknowledgements

We thank project staff, the community of Kabondo and Kasipul, Rachuonyo South, and KEMRI/CDC Kisumu. This project is funded by the Bill and Melinda Gates Foundation, under the Malaria Transmission Consortium, Grant No.45114 and the Grand Challenge Grant No. OPP1024438. This manuscript has been approved by the Director of the Kenya Medical Research Institute.

References

1 Manning, S. D., Woolhouse, M. E. & Ndamba, J. Geographic compatibility of the freshwater snail *Bulinus globosus* and schistosomes from the Zimbabwe highveld. *Int J Parasitol* **25**, 37-42, doi:0020-7519(94)00097-8 [pii] (1995).

2 Woolhouse, M. E. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc Natl Acad Sci U S A* **94**, 338-342 (1997).

3 Criscione, C. D. *et al.* Landscape genetics reveals focal transmission of a human macroparasite. *PLoS Negl Trop Dis* **4**, e665, doi:10.1371/journal.pntd.0000665 (2010).

4 Bousema, T. *et al.* Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med* **9**, e1001165, doi:10.1371/journal.pmed.1001165 (2012).

5 Carter, R., Mendis, K. N. & Roberts, D. Spatial targeting of interventions against malaria. *Bull World Health Organ* **78**, 1401-1411 (2000).

6 Oesterholt, M. J. *et al.* Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania. *Malar.J.* **5**, 98 (2006).

7 Clark, T. D. *et al.* Factors determining the heterogeneity of malaria incidence in children in Kampala, Uganda. *J Infect Dis* **198**, 393-400 (2008).

8 Kreuels, B. *et al.* Spatial variation of malaria incidence in young children from a geographically homogeneous area with high endemicity. *J Infect Dis* **197**, 85-93 (2008).

9 Bousema, T. *et al.* Identification of hot spots of malaria transmission for targeted malaria control. *J Infect Dis* **201**, 1764-1774, doi:10.1086/652456 (2010).

10 Bejon, P. *et al.* Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med* **7**, e1000304, doi:10.1371/journal.pmed.1000304 (2010).

11 Ghebreyesus, T. A. *et al.* Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey. *BMJ* **319**, 663-666 (1999).

12 Midega, J. T. *et al.* Wind direction and proximity to larval sites determines malaria risk in Kilifi District in Kenya. *Nat Commun* **3**, 674, doi:ncomms1672 [pii] 10.1038/ncomms1672 (2012).

13 Ghebreyesus, T. A. *et al.* Household risk factors for malaria among children in the Ethiopian highlands. *Trans.R.Soc.Trop.Med.Hyg.* **94**, 17-21 (2000).

14 Gamage-Mendis, A. C. *et al.* Clustering of malaria infections within an endemic population: risk of malaria associated with the type of housing construction. *Am.J.Trop.Med.Hyg.* **45**, 77-85 (1991).

15 Mackinnon, M. J., Mwangi, T. W., Snow, R. W., Marsh, K. & Williams, T. N. Heritability of malaria in Africa. *PLoS Med* **2**, e340, doi:05-PLME-RA-0236R1 [pii] 10.1371/journal.pmed.0020340 (2005).

16 Smith, D. L., McKenzie, F. E., Snow, R. W. & Hay, S. I. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol.* **5**, e42 (2007).

17 Bousema, T., Kreuels, B. & Gosling, R. Adjusting for heterogeneity of malaria transmission in longitudinal studies. *J Infect Dis* **204**, 1-3, doi:jir225 [pii] 10.1093/infdis/jir225 (2011).

18 Clarke, S. E. *et al.* Risk of malaria attacks in Gambian children is greater away from malaria vector breeding sites. *Trans R Soc Trop Med Hyg* **96**, 499-506 (2002).

19 Cook, J. *et al.* Serological markers identify heterogeneity of effectiveness of malaria control interventions on Bioko island, Equatorial Guinea. *PLoS One* (2011).

20 Bejon, P. *et al.* Serological evidence of discrete spatial clusters of Plasmodium falciparum parasites. *PLoS One* **6**, e21711, doi:10.1371/journal.pone.0021711 PONE-D-11-04349 [pii] (2011).

21 Bousema, T. *et al.* Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008. *Emerg Infect Dis* **16**, 392-399 (2010).

22 Stone, W. *et al.* IgG Responses to Anopheles gambiae Salivary Antigen gSG6 Detect Variation in Exposure to Malaria Vectors and Disease Risk. *PLoS One* **7**, e40170, doi:10.1371/journal.pone.0040170 PONE-D-12-09375 [pii] (2012).

23 Plowe, C. V., Djimde, A., Bouare, M., Doumbo, O. & Wellems, T. E. Pyrimethamine and proguanil resistance-conferring mutations in Plasmodium falciparum dihydrofolate reductase: polymerase chain reaction methods for surveillance in Africa. *Am J Trop Med Hyg* **52**, 565-568 (1995).

24 Steenkeste, N. *et al.* Towards high-throughput molecular detection of Plasmodium: new approaches and molecular markers. *Malar J* **8**, 86, doi:1475-2875-8-86 [pii] 10.1186/1475-2875-8-86 (2009).

25 Hsiang, M. S. *et al.* PCR-based pooling of dried blood spots for detection of malaria parasites: optimization and application to a cohort of Ugandan children. *J Clin Microbiol* **48**, 3539-3543, doi:JCM.00522-10 [pii] 10.1128/JCM.00522-10 (2010).

26 Drakeley, C. J. *et al.* Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc.Natl.Acad.Sci.U.S.A* **102**, 5108-5113 (2005).

27 Corran, P. H. *et al.* Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. *Malaria Journal* **7** 195 (2008).

28 Wilson, S. *et al.* Age-adjusted *Plasmodium falciparum* antibody levels in school-aged children are a stable marker of microgeographical variations in exposure to Plasmodium infection. *BMC.Infect.Dis.* **7**, 67 (2007).

29 SatScan, T. http://www.satscan.org/.

30 Kulldorff, M., Huang, L., Pickle, L. & Duczmal, L. An elliptic spatial scan statistic. *Stat Med* **25**, 3929-3943, doi:10.1002/sim.2490 (2006).

31 Fillinger, U. *et al.* A tool box for operational mosquito larval control: preliminary results and early lessons from the Urban Malaria Control Programme in Dar es Salaam, Tanzania. *Malar J* **7**, 20, doi:1475-2875-7-20 [pii]
10.1186/1475-2875-7-20 (2008).

32 World Health, O. Manual on practical entomology in malaria. Part II., (World Health Organization, Geneva, 1975).

33 Mboera, L. E., Kihonda, J., Braks, M. A. & Knols, B. G. Influence of centers for disease control light trap position, relative to a human-baited bed net, on catches of Anopheles gambiae and Culex quinquefasciatus in Tanzania. *American Journal of Tropical Medicine and Hygiene* **59**, 595-596 (1998).

34 Govella, N. J., Chaki, P. P., Mpangile, J. M. & Killeen, G. F. Monitoring mosquitoes in urban Dar es Salaam: evaluation of resting boxes, window exit traps, CDC light traps, Ifakara tent traps and human landing catches. *Parasit Vectors* **4**, 40, doi:1756-3305-4-40 [pii]
10.1186/1756-3305-4-40 (2011).

35 Odetoyinbo, J. A. Preliminary investigation on the use of a light-trap for sampling malaria vectors in the Gambia. *Bull World Health Organ* **40**, 547-560 (1969).

36 Wirtz, R. A., Burkot, T. R., Graves, P. M. & Andre, R. G. Field evaluation of enzyme-linked immunosorbent assays for Plasmodium falciparum and Plasmodium vivax sporozoites in mosquitoes (Diptera: Culicidae) from Papua New Guinea. *J.Med.Entomol.* **24**, 433-437 (1987).

37 Griffin, J. T. *et al.* Strategies towards Plasmodium falciparum malaria elimination in Africa using currently available tools. *PLoS Med* **6**, e20179 (2010).

38 Hawley, W. A. *et al.* Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. *Am.J.Trop.Med.Hyg.* **68**, 121-127 (2003).

39 World Health, O. Report of the seventh WHOPES working group meeting. Review of: VectoBac WG, Permanet, Gokilaht-S 5EC. WHO/CDS/WHOPES/2004.8. (2004).

40 DOMC. National Malaria Strategy 2009-2017: Towards a malaria-free Kenya. (Division of Malaria Control; Ministry of Public Health and Sanitation, Nairobi, 2009).

41 Okell, L. C., Ghani, A. C., Lyons, E. & Drakeley, C. J. Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. *J Infect Dis* **200**, 1509-1517, doi:10.1086/644781 (2009).

42 Stresman, G. H. *et al.* A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia. *Malar J* **9**, 265, doi:1475-2875-9-265 [pii] 10.1186/1475-2875-9-265 (2010).

43 Carter, R. Spatial simulation of malaria transmission and its control by malaria transmission blocking vaccination. *Int J Parasitol* **32**, 1617-1624 (2002).

44 Killeen, G. F. *et al.* The importance of considering community-level effects when selecting insecticidal malaria vector products. *Parasit Vectors* **4**, 160, doi:1756-3305-4-160 [pii] 10.1186/1756-3305-4-160 (2011).

45 Binka, F. N., Indome, F. & Smith, T. Impact of spatial distribution of permethrin-impregnated bed nets on child mortality in rural northern Ghana. *Am.J.Trop.Med.Hyg.* **59**, 80-85 (1998).

46 Howard, S. C. *et al.* Evidence for a mass community effect of insecticide-treated bednets on the incidence of malaria on the Kenyan coast. *Trans.R.Soc.Trop.Med.Hyg.* **94**, 357-360 (2000).

47 McCall, P. J., Mosha, F. W., Njunwa, K. J. & Sherlock, K. Evidence for memorized site-fidelity in Anopheles arabiensis. *Trans.R.Soc.Trop.Med.Hyg.* **95**, 587-590 (2001).

48 Charlwood, J. D., Graves, P. M. & Marshall, T. F. Evidence for a 'memorized' home range in Anopheles farauti females from Papua New Guinea. *Med Vet Entomol* **2**, 101-108 (1988).

49 Shekalaghe, S. *et al.* A cluster-randomized trial of mass drug administration with a gametocytocidal drug combination to interrupt malaria transmission in a low endemic area in Tanzania. *In preparation* (2011).

50 Okell, L. C. *et al.* The potential contribution of mass treatment to the control of Plasmodium falciparum malaria. *PLoS One* **6**, e20179 (2011).

51 Fillinger, U., Ndenga, B., Githeko, A. & Lindsay, S. W. Integrated malaria vector control with microbial larvicides and insecticide-treated nets in western Kenya: a controlled trial. *Bull World Health Organ* **87**, 655-665, doi:S0042-96862009000900009 [pii] (2009).

52 Fillinger, U., Sonye, G., Killeen, G. F., Knols, B. G. & Becker, N. The practical importance of permanent and semipermanent habitats for controlling aquatic stages of Anopheles gambiae sensu lato mosquitoes: operational observations from a rural town in western Kenya. *Trop Med Int Health* **9**, 1274-1289, doi:TMI1335 [pii] 10.1111/j.1365-3156.2004.01335.x (2004).

53 Geissbuhler, Y. *et al.* Microbial larvicide application by a large-scale, community-based program reduces malaria infection prevalence in urban Dar es Salaam, Tanzania. *PLoS One* **4**, e5107, doi:10.1371/journal.pone.0005107 (2009).

FIGURES

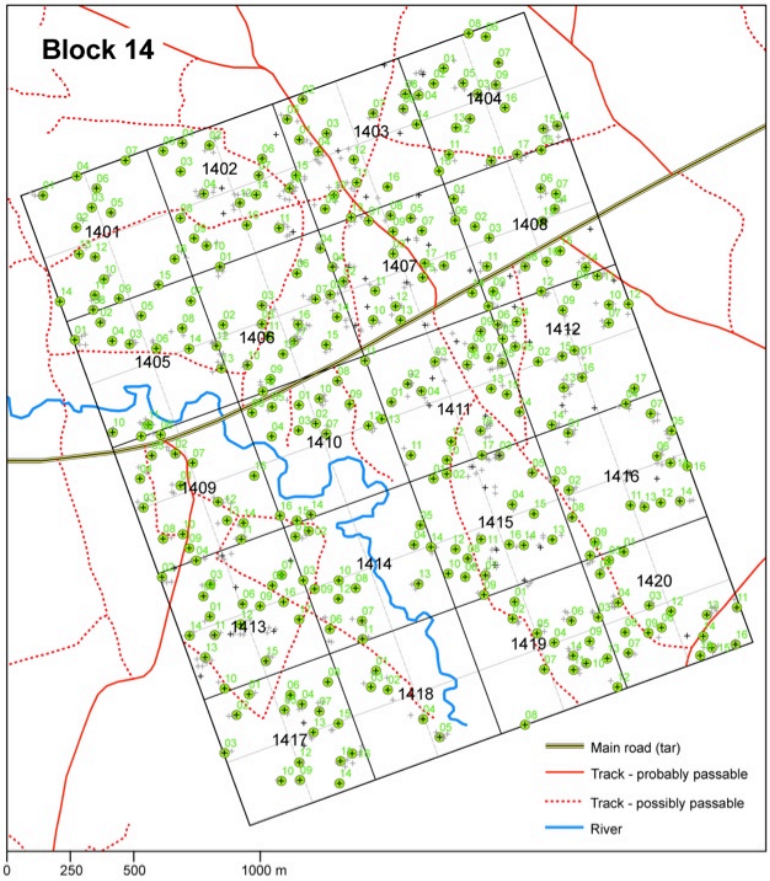

Figure 1 - Overview map of one block in the study area comprising 20 cells. A map of a part of a 2km x 2.5km part of the study area that comprises 20 500m x 500m cells and 80 sub-cells. Cell numbers are given in black bold letters; grey crosses indicate structures; green circles with black crosses indicate selected and numbered households. Rivers and roads are indicated in the map as given in the legend.

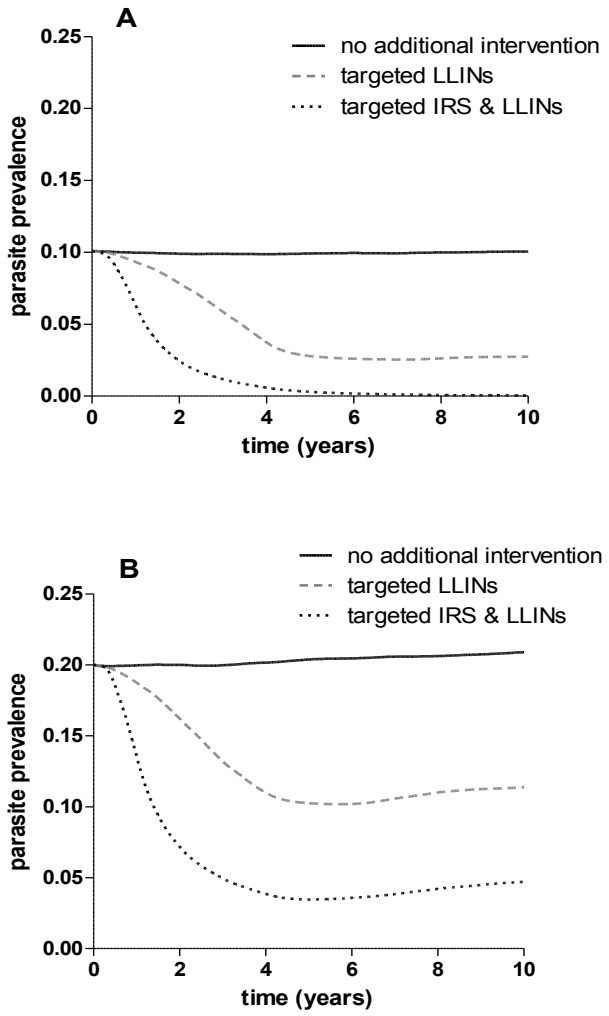

Figure 2 - Simulation of intervention outcome

The figure presents a simulation of hotspot targeted interventions in areas with a baseline parasite prevalence of 10% or 20%. ITN coverage is assumed to be 41% across all age groups (83% in under fives). Plotted is smoothed parasite prevalence in the total population as a function of time in years since the start of the intervention. No interventions (solid black line), hotspot-targeted increase in LLIN coverage to reach 90% effective coverage in hotspots (dashed grey line) and hotspot-targeted increase in LLIN coverage to reach 90% effective coverage in hotspots in combination with targeted IRS reaching 90% of households in hotspots (dashed black line).

Appendix 1.4 – Model Based Geostatistics: Statistical Methods

This document outlines the underlying statistical methodology of the "Methods" section in the paper. In the remainder of the document we will refer both to PCR and seroprevalence as "test". In Section 1 we describe the geostatistical models that were fitted to the PCR and seroprevalence data. In Section 2, we give some details on the metrics used for the sample size calculations.

---

1 Geostatistical analysis

1.1 Model

Let Y_i denote the number of positive counts of the test in the i-th compound, each associated with sampling locations x_i, for i=1, . . . , n. Conditionally on the realization of the random effect $T(x_i)$, the response variable Y_i follow a Binomial distribution with expected value $E[Y_i] = n_i p_i$ where n_i is the number of compound members and p_i is the probability of having a positive test. We use the canonical logit link function defined as

$$\log\left\{\frac{p_i}{1-p_i}\right\} = T(x_i) = d(x_i)^\top \beta + S(x_i) + Z_i,\ i = 1, \ldots, n, \tag{1}$$

where $d(x_i)$ is a vector spatial covariates with associated vector of regression coefficients β; $S(x_i)$ is a stationary isotropic zero-mean Gaussian process with variance σ^2 and correlation function $\rho(u) = \exp(-u/\Phi)$ with u being the distance between two compounds and scale parameter $\Phi > 0$; and Z_i are mutually independent Gaussian variables that are used to account for non-spatial variation within compounds.

The set of spatial covariates $d(x_i)$ used in the model were selected using ordinary logistic regression and that were significant at 5% confidence level; in table 1, these are reported indicating their includion in the PCR and seroprevalence models.

	Term	PCR	Seroprevalence
1	Intercept	Yes	Yes
2	Mean elevation	Yes	Yes
3	Maximum NDVI	Yes	No
4	Mean NDVI	Yes	No
5	Distance from closest fish pond	Yes	Yes
6	Tree cover	Yes	Yes
7	Maximum TWI	No	Yes
8	Mean TWI	No	Yes
9	Distance from the 3rd order stream	No	Yes
10	Distance from the 2nd order stream	No	Yes

Table 1: Identified spatial covariates using an ordinary logistic regression; third and fourth columns indicate their presence in the models for PCR and seroprevalence, respectively

1.2 Parameter estimation

We use the Monte Carlo maximum likelihood (MCML) method (Geyer & Thompson, 1992; Geyer, 1994, 1996, 1999) for estimation of the model parameters. This procedure uses conditional simulations of the random effect T given the data Y to obtain a computationally efficient approximation to the intractable likelihood function. More details on the analytical derivation of such an approximation can be found in Christensen (2004) and Giorgi et al. (2015).

1.3 Prediction

Now, consider the prediction of $T^*=(T(x_{n+1}), \ldots, T(x_{n+q}))^T$ at q additional prediction locations forming a regular grid at spacing 100 m over the entire surveyed area. All relevant explanatory variables, listed in Table 1, were also available at the prediction locations. We do not include the mutually independent random variables Z_i in 1 as part of our target for prediction, since, in our case, these are interpreted as non-spatial variation within compounds.

Using a Monte Carlo Markov chain algorithm proposed by Christensen et al. (2006), we obtain 10^4 samples from the distribution of T^* given Y by simulating 110000 samples and retaining every 10^{th} sample after a burn-in of 10^4 simulations. Let $t_{(1)}(x_{n+i}), \ldots, t_{(10^4)}(x_{n+i})$, denote the simulated samples for the i-th grid locations x_{n+i}. Predicted prevalences are obtained by transforming the sampled values $t_{(j)}(x_{n+i})$ to $p_{(j)}(x_{n+i}) = \exp\{t_{(j)}(x_{n-i})\}/(1+\exp\{t_{(j)}(x_{n+i})\})$ for $i=1, \ldots, q$ and $j=1, \ldots, 10^4$. We then summarize the resulting set of predicted prevalence surfaces with the following indices.

- Point-wise mean are obtained as $\sum_{j=1}^{10^4} p_{(j)}(x_{n+i})/10^4$, for $i = 1, \ldots, q$.
- Exceedance probabilities are obtained as $\sum_{j=1}^{10^4} I_{(c,1)}\{p_{(j)}(x_{n+i})\}/10^4$, for $i=1, \ldots, q$ where c is a pre-defined prevalence threshold ($c = 0.28$ for PCR and $c = 0.70$ for seroprevalence) and $I_{(c,1)}\{p\}$ is an indicator function that takes value 1 if $c < p < 1$ and 0 otherwise.

2 Computational details of the sample size calculations

Let A denote the surveyed region of interest in R^2, $\mu(x) = d(x)^T\beta$ the fixed effect part of the linear predictor in (1) and define $\hat{S}p(x)$ to be the kriging predictor of $S(x)$ obtained by a sample corresponding to (100 x p)% of the total population. The integrated mean-square error (IMSE) and the discrimination index (DI) are the defined as follows (Fanshawe & Diggle, 2013)

$$\text{IMSE} = \int_A E\left\{\left[\exp(\mu(x)+S(x)) - \exp(\mu(x)+\hat{S}_p(x))\right]^2\right\} dx, \quad (2)$$

$$\text{DI} = -\int_A \left\{P\left(\mu(x)+\hat{S}_p(x) > l\right) - 0.5\right\}^2 dx, \quad (3)$$

where $E\{\bullet\}$ is the expected value with respect to the distribution of $S(x)$ and $l = \log(c/(1 - c))$, with c given prevalence threshold as specified in Section 1.3. The IMSE index in (2) quantifies the overall mean-square error in A of the odds ratio spatial predictor. In (3), DI measures how well the design of a given sample size discriminates hotspots; under ideal circumstances all the predictive probabilities would be either 0 or 1.

In order to compute the intractable integrals in (2) and (3), we impose the spatially continuous process $S(x)$ to be piecewise constant over a regular grid ($x_1, \ldots$

. , x_N) in A at spacing 220 m. For a given proportion p of the total population, we then compute the IMSE and DI metrics using the following Monte Carlo procedure.

1. Select randomly a set of locations corresponding to (100 x p)% of the total digitized structures aggregated to the compound level, with the distance between any two sampled locations no less than 20 m in order to guarantee a good spatial coverage in A.
2. Simulate 10^4 surfaces of $S(x)$ over the regular grid in A, setting the covariance parameters equal to the respective MCML estimates (see Section 1.2). Let now $S_{(i)}^{\top} = (S_{(i)}(\tilde{x}_1), \ldots, S_{(i)}(\tilde{x}_N))$ be the i-th simulated surface.
   i) For each of the randomly chosen locations obtained from 1, select $S_{(i)}(x)$ where x is the closest grid point, and add Gaussian noise Z, corresponding to non-spatial variation between compounds with variance equal to the respective MCML estimate.
   ii) Compute the kriging predictor for $S_{(i)}$, denoted by $S_{(i)}$, denoted by $\hat{S}_{(i)}^{\top} = (\hat{S}_{(i)}(\tilde{x}_1), \ldots, \hat{S}_{(i)}(\tilde{x}_N))$
   iii) Repeat i) and ii) for $I = 1, \ldots, 10^4$ and finally approximate (2) and (3) as

$$\text{IMSE} \approx \frac{1}{N \times 10^4} \sum_{i=1}^{10^4} \sum_{j=1}^{N} \left[\exp(\mu(\tilde{x}_j) + S_{(i)}(\tilde{x}_j)) - \exp(\mu(\tilde{x}_j) + \hat{S}_{(i)}(\tilde{x}_j)) \right]^2,$$

$$\text{DI} \approx -\frac{1}{N \times 10^4} \sum_{i=1}^{10^4} \sum_{j=1}^{N} \left\{ I\{\mu(\tilde{x}_j) + S_{(i)}(\tilde{x}_j) > l\} - 0.5 \right\}^2,$$

   Where $I(a > l)$ is an indicator function that takes value 1 if a > l and 0 otherwise.

References

CHRISTENSEN, O. F. (2004). Monte Carlo maximum likelihood in model-based geostatistics. *Journal of Computational and Graphical Statistics*. **3**, 702-718.

CHRISTENSEN, O. F., ROBERTS, G. O. & SKÖLD, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics.* **15**, 1-17.

FANSHAWE, T. & DIGGLE, P. J. (2013). Adaptive sampling design for spatio-temporal prediction. In *Spatio-temporal designs*, J. Mateu & W. G. Müller, eds. John Wiley and Sons, Ltd, pp. 249-268.

GEYER, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, Series B* **56**, 261-274.

GEYER, C. J. (1996). Estimation and optimization of functions. In *Markov Chain Monte Carlo in Practice*, W. Gilks, S. Richardson & D. Spiegelhalter, eds. London: Chapman and Hall, pp. 241-258.

GEYER, C. J. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry, Likelihood and Computations.* O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout, eds. Boca Raton, FL; Chapman and Hall/CRC, pp. 79-140.

GEYER, C. J. & THOMPSON, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, Series B* **54**, 657-699.

GIORGI, E., SESAY, S. S., TERLOUW, D. J. & DIGGLE, P. J. (2015). Combining data from multiple spatially referenced prevalence surveys using generalized linear geostatistical models. *Journal of the Royal Statistical Society, Series A*. In press.

Appendix 2 – Survey Questions

Appendix 2.1 – School Survey

Part I-Household Information-Fill out once for each house

(NB. Translations for questions in Kisii, DhoLuo and Swahili (K, L, S respectively) are given only for those questions directed to household members and not for observational questions or for answers as all interviewers speak English)

1. District_____
2. Division_____
3. Location_____
4. Enumeration Area_____
5. Village Name_____
6. Head of Household_____
7. For the head of household, what is the highest level of education completed?

   K: *Omonene bwe'nyomba eye ngayi asomete agaika*

   L: *Wuon odni osomo nyaka klas adi?*

   S: *Ni kiwango kipi cha juu zaidi cha masomo ambacho mwenye nyumba ame hitimu?*

   a. NONE
   b. PRIMARY
   c. SECONDARY
   d. HIGHER
8. What type of wall was used for the construction of this house?
   a. CLAY OR MUD
   b. BRICK OR STONE
   c. CEMENT/PLASTERED
   d. CEMENT/PAINTED
   e. OTHER
9. What type of roof was used for the construction of this house?
   a. GRASS THATCH
   b. IRON SHEET

c. TILES
d. OTHER

10. How many windows have glass?
    a. NONE
    b. SOME
    c. ALL
11. Are the eaves open or closed?
    A. OPEN
    b. CLOSED
    c. PARTIALLY OPEN
12. What is the main material of the floor?
    a. EARTH/SAND
    b. DUNG
    c. WOOD PLANKS
    d. PALM/BAMBOO
    e. PARQUET OR POLISHED WOOD
    f. VINYL OR ASPHALT STRIPS
    g. CERAMIC TILES
    h. CEMENT
    i. CARPET
    j. OTHER)_____
13. What animals are found in your compound?

    K: *Ntugo ki ere ase omochie oyo*

    L: *Jamni mage ma un godo e dalau ka?*

    S: *Wanyama wapi kati ya hawa wanapatikana kwenye ua la nyumba yako?*
    a. COWS: YES/NO
    b. GOATS: YES/NO
    c. SHEEP: YES/NO
14. What is the main type of fuel used by your family for cooking?

    K: *Inko ogotumeka koruga ase omochioyo botambe*

    L: *En ang'o ma jo odni tiyogo kuom chweko e tedo mapile?*

    S: *Unatumia aina gani ya nguvu au moto kupika?*
    a. ELECTRICITY OR GAS/

b. KEROSENE
c. CHARCOAL
d. FIREWOOD
e. DUNG
f. OTHERS (SPECIFY) _____

15. At any time in the past 12 months, has anyone sprayed the interior walls of your dwelling?
    K: *Ase emetienyi ikomi nebere yaetire, monto'nde onya gosiara chinyasi chiaime chienyomba yao eriogo riogoita chivmbu?*
    L: *E dweche 12 ma osekalo, bende osekir yath e kor odni gi iye?*
    S: *Katika jumla ya miezi kumi na miwili iliyopita, kuna yeyote amenyunyizia dawa kwenye kuta za nyumba yako?*
    a. YES
    b. NO
    c. DON'T KNOW
16. How many months ago was the house sprayed? (IF LESS THAN ONE MONTH, RECORD '00' MONTHS AGO)
    K: Ingaki ki yaetire korwa enyomba eye esiarerwa eriogo?
    L: Ma ne otimore dweche adi ma okalo?
    S: Ni miezi mingapi imepita tangu nyumba yako inyunyiziwe dawa?
    _____
17. Who sprayed the house?
    K: Ning''o osiarete eriogo
    L: Ng''ano mane okiro yath e odni?
    S: Ni nani aliyeinyunyizia dawa nyumba yako?
    a. GOVERNMENT WORKER/PROGRAMME
    b. PRIVATE COMPANY
    c. HOUSEHOLD MEMBER
    d. OTHER (specify)_____
    e. DON'T KNOW
18. Have any of the following been used in your house over the last week?

K: Kemo kiebi kianya gotumeka nyomba mwao ase amatuko atano nabere aetire?

L: Bende usetiyo gi achiel kuom magi e otka e juma ma okaloni?

S: Kuna chochote kifuatacho ambacho kimetumiwa kwenye nyumba yako wiki iliyopita?

a. MOSQUITO COILS?
b. INSECTICIDE SPRAYS (e.g. DOOM)?
c. REPELLENTS?

19. Does your household have any mosquito nets that can be used while sleeping?

K: Enyomba yao nebwate eneti ye chiumbu egotumeka ekero mokorara?

L: Bende odi ka nitiere e net ma itiyogo ka ji nindo?

S: Una neti ya kujifunikia wakati wa kulala ili kujizuia mbu?

a. YES
b. NO

Part II-Household Listing-Fill out questionnaire for each person who stayed in house previous night. For children, pose the questions to the primary caretaker

1. Name of person)_____
2. Is (NAME) male or female?
   a. MALE
   b. FEMALE
3. Does (NAME) usually live here?

   K: [X] noo amenyete aiga?

   L: Bende [X] odak ga ka?

   S: Mtu huyu [X] anaishi hapa?

   a. YES
   b. NO
4. Did (NAME) stay here last night?

   K: [X] Naraire aiga botuko bwaetete?

   L: Bende [X] ne onindo ka otieno mokalo?

   S: [X] Amelala hapa jana usiku?

a. YES
b. NO

5. What is the date of birth of (NAME)?)

K: [X] mwaka ki aiboire?

L: [X] nonyuol e higa ane?

S: Tarehe ya [X[ya kuzaliwa?

_____

6. What is [X]'s date of birth?

K: [X] mwaka ki aiboire?

L: [X] nonyuol e higa ane?

S: Tarehe ya [X[ya kuzaliwa?

_____

7. Has (NAME) been ill with a fever at any time in the last 2 weeks?

K: [X] Onya koigwa riberera ngaki ende ase chichuma ibere chiaetire?

L: Bende [X] osebedo gi del ma ore e jumbe ariyo ma osekalo?

S: Katika wiki mbili zilizopita, [X] amekuwa mgonjwa na kusikia maumivu na joto mwilini?

a. YES
b. NO
c. NOT SURE

8. Did [X] seek advice or treatment for the fever from any source?

K: [X] naetwe riogo rinde gose nachiete nyagitari koegwa eriogo rieriberera?

L: Bende ne omanyo ng''ado rieko, kata thieth kuom del maore kamor amora?

S: [X] alitafuta huduma ya afya au matibabu popote?

a. YES
b. NO

2. Where did you seek advice or treatment? (Check all that apply)

K: Ase ng''o aetwe obosemi gose kogwenigwa korwa?

L: Ng''ado rieko kata thieth ne omanyo kanye?

S: Alienda kutafuta wapi huduma au matibabu hayo?

a. GOVT. HOSPITAL
b. GOVT. HEALTH CENTER
c. GOVT. HEALTH POST

d. MOBILE CLINIC
e. FIELD WORKER
f. OTHER PUBLIC
g. PVT. HOSPITAL/CLINIC
h. PHARMACY
i. PRIVATE DOCTOR
j. MOBILE CLINIC
k. FIELD WORKER
l. OTHER PVT. MEDICAL)_____
m. SHOP
n. TRAD. PRACTITIONER
o. OTHER)_____

3. Has [X] had a fever in the last 24 hours?

   K: [X] otwarire riberera korwa igoro?

   L: Bende [X] osebedo gi del maore nyoro kata kawuono?

   S: [X] amekuwa na maumivu na mwili wake kuwa na joto masaa ishirini na manne yaliyopita?

   a. YES
   b. NO
   c. DON'T KNOW

4. Has (NAME) taken any drugs in the last 2 weeks? (Check all that apply)

   K: [X] onyure riogo rinde ase chichuma ibere chiaetire?

   L: Bende [X] osemwonyoe yath e jumbe ariyo ma osekalo?

   S: [X] ametumia dawa yoyote wiki mbili zilizopita?

   a. SP/FANSIDAR
   b. CHLOROQUINE
   c. AMODIAQUINE
   d. QUININE
   e. COARTEM
   f. OTHER ANTIMALARIAL (SPECIFY)_____
   g. ASPIRIN
   h. ACETAMINOPHEN/PARACETAMOL
   i. IBUPROFEN

j. OTHER (SPECIFY)_____

k. DON'T KNOW

We are now going to ask some questions about where [X] has lived in the past, and where he has travelled to in recent months. The reason we are asking these questions is to find out whether [X] might have been at risk of getting malaria in other places.'

K: Inkogenda tore koboria amaborio igoro ya'se [X] amenyire ase matukoaetire, nase atarete gochia ase omotienyi oyo. Etokoboria ribori eri erinde torore gose [X] nabase kobwatwa na malaria ase ensemo ende

L: Koro adwaro penjo kuom kama [X] osebedo ka odakie e thuolo ma okalo, to gi kama osedhie wuoth e dweche matin mokalo. Penjagi konyowa ng'eyo ka onyalo yudo tuo mar malaria Kuonde moko opogore gi ka.

S: Tutakuuliza maswali kuhusu pale [X] ameishi tena pasipo hapa na kule ametembelea miezi iliyopita ya karibuni. Tunataka kuweza kujua kama [X] angeweza kupata malaria pahali pengine

5. Has (X) always lived in (THIS DISTRICT)?

   K: Tatiga ensemo eye [X] omoniyre ensemo endo ase emetienyi etano nomo yaetire?

   L: Kopogore gi (this district) ka bende nitiere district mne ma [X] osedake moloyo dweche auchiel?

   S: Badala ya (X) ali ishi kwa wilaya ingine kwa zaridi ya miezi sita?

   a. YES (Skip to question #)

   b. NO

6. If NO,

   Which other district have did [X} live in most recently?

   K : Insemo ki ende [X] amenya bwango iga ase engaki entambe

   L: District mane ma osedakie mang'enye moloyo

   S: [X] ameishi katika wilaya gain tena hivi majuzi?

   _____

7. For how long did you live in (X)?

   K: Engaki engana inaki [X] eberete ase omochi oyo

   L: [X] osedak e districtni kuom thuolo maromo nade

S: Ameishi huko siku ngapi? (jina la wilaya)

_____

8. Has [X] lived in any other districts?
   K: [X] onya komenya insemo ya omochie onde goetania emetienyi etato nomo?
   L: Bende osedak e district moro amora kuom thuolo moloyo dweche auchiel?
   S: [X] ameishi katika wilaya zingine?
   a. YES (Repeat previous 2 questions)
   b. NO (Skip to question #23)
9. Has [X] travelled outside (THIS DISTRICT) in the last 3 months?
   K: [X] onyagotarera insemo ya mochie onde ase emetienyi etato yaetire arare aroro?
   L: Bende sagar moro oseyudi mininde oko kuom dweche adek mose kalo?
   S: [X] ametembelea wilaya nyingine (jina la wilaya) miezi mitatu iliyopita?
   a. YES
   b. NO (Skip to question #36)
   c. DON'T KNOW (Skip to question #36)
10. How many trips has [X] made outside (THIS DISTRICT) in the last 3 months?
    K: Chisabari irenga [X] agenda isiko ya omochie oyo chiokorara (erieta ria omochie) ase engaki yemetienyi etato
    L: Edweche a dek ma osekalo osedhi wuoth oko mar district ni di di?
    S: Ameenda huko (jina la wilaya) mara ngapi?

    _____
11. When did [X] come back from your most recent trip?
    K: Indi [X] airanete korwa esabari egendete bwango iga yokorara?
    L: Oduogo ka ang'o koa e wuoth magik mane odhi oko mar district ni?
    S: Alirudi lini kutoka safari yake ya mwisho?
    a. <2 WEEKS AGO
    b. 2-4 WEEKS AGO
    c. >4 WEEKS AGO
12. Which district did [X] spend most time in during that trip?

K: Imochie ki [X] aberete amatuko amange ase chisabari achire chiokorara?

L: E wuodhe ma ogik mane odhie oko mar districtni, ne odhinyo dak e district mane?

S: Ni wilaya ipi aliweza kukaa sana katika safari yake?

_____

13. What was the main reason for taking the trip?

K: Ngeto ki kiagerete x akagenda chisabari echi chiokorara?

L: Ang''o maduong'' mane omiyo odhi e wuodh ni?

S: Kwa nini alienda kwenye safari hii?

    a. WORK/LOOKING FOR WORK
    b. BUYING/SELLING AT MARKET
    c. ATTENDING SCHOOL OR UNIVERSITY
    d. VISITING RELATIVES OR FRIENDS
    e. OTHER (SPECIFY) _____

14. Are there any other persons who live in this house that we have not listed?

K: Monto onde nare orarete nyomba aiga totararika rieta riaye?

L: Bendo nitiere ji mamoko manindo e odini mapodi ok andiko nying gi?

S: Kuna watu wengine wanaoishi kwenye nyumba hii ambao hatuna majina yao?

    a. YES (Go back to QUESTION 1 and fill out this form for that person)
    b. NO

15. Are there any other people who may not be members of your family, such as domestic servants, lodgers or friends who usually live here?

K: Omogeni nare oraa otari monto oino buna omokori egasi omomenyi gose omosani oino orarete nainwe aiga?

L: Bende nitiere ji mamoko ma ok jodala ka mapile, kaka jotich, wede, kata osiepe ma mane onindo ka otieno manyoro?

S: Kuna watu wengine wanaoishi kwenye nyumba hii kama marafiki, wafanyi kazi au watu wanaolipia malazi yao ambao hatuna majina yao?

    a. YES (Go back to QUESTION 1 and fill out this form for that person)
    b. NO (Go to net roster).

Part III-Net roster-Fill out questionnaire for each net in the house.

1. Net Number_____

2. Observed
    a. YES
    b. NO
3. How long ago did you obtain the net?
   K: Indi enyomba eye yanyorete/yaetwe chineti chiechiumbu?
   L: Nedni ichako bedogo karango?
   S: Mliipata neti hii wakati mgani?

   _____
4. Observe the brand of net:
    a. PermaNet
    b. Olyset
    c. SupaNet
    d. SupaNet Extra
    e. Other_____
5. Since you got the mosquito net, was it ever soaked or dipped in a liquid to repel mosquitoes or bugs?
   K: koru onyora eneti ye chiumbu, yanyagosibigwa neriogo riamache riogoseria chiumbu ne chinsuri?
   L: Nyaka ichak bedo gi nedni, bende oselwoke kod yath mageng'o suna?
   S: Tangu ulipoinunua neti hii, ume wahi kuiweka kwenye maji yaliyo na dawa ya kuuwa au kuwafukuza mbu na wadudu?
    a. YES
    b. NO (Go to question #7)
    c. NOT SURE (Go to question #7)
6. How long ago was the net last soaked or dipped? (If less than 1 month, enter '00')
   K: Amatuko arenga aetire korwa eneti yaoo esibigwa neriogo riamache riogoita chiumbu?
   L: Olwoke gi yadh suna ndalo adi ma osekalo?
   S: Ni wakati mgani umepita tangu uiweke neti kwenye maji yaliyo na dawa

   _____
7. Was this net hanging last night?
   K: Eneti iyasungire obotuko bwaigoro?

L: Bende ne nedni oyar e otieno mapiny orugo kwauono?

S: Neti hii ilitumika jana usiku kujikinga na mbu?

a. YES
b. NO (Skip to last question)

8. List all people who slept under this net last night:
    a. Person 1_____
    b. Person 2_____
    c. Person 3_____

Appendix 2.2 – Health Facility Survey

SECTION 1 - PATIENT DATA

1. Select health facility:

   (drop down menu with

   a. Agawo
   b. Ober
   c. Omiro
   d. Othoro
   e. Tala)

2. Patient ID Number: HF22-_ _-_ _ _

   (-First two digits should be automatically populated using the first 2 letters of the Health Facility Name selected in Question 1 e.g. AG.

   -Last 3 digits should be an autogenerated number increasing from from 001 to 500 based on the order of sampling.)

3. Date of sample:__ __ ____
4. Sex:
5. Year of birth
6. Age:
7. Axillary Temperature

SECTION 2 - AREA OF RESIDENCE

We would like to ask you some questions about the compound where you slept last night.

L:Wadwaro penji penjo ewi dala kamani ninde nyoro gotieno

S:Tungetaka kukuuliza maswali kadhaa kuhusu boma ulimolala jana usiku

8. Name of head of compound (3 names): _____

   L:Nying wuon dala

   S:Jina la mwenye Boma

9. Name of head of house (3 names): _____

   L:Nying wuon ot

   S:Jina la mwenye nyumba

10. District: _____

    L:district

S:Wilaya

(drop down menu with district list)

11. Location: _____

    L:lokeson

    S:Lokesheni

12. EA Name: _____

    L: Gweng

    S:Kijiji

13. What is your clan? _____

    L: Un jo dhoot mane?

    S: Unatoka ukoo gani?

14. What is the nearest health facility to this compound?

    _____

    L: Kar thieth mane mani machiegni gi dalau ka?

    S: Kituo cha afiya gani iko karibu na boma lenu?

15. What is the nearest primary school to this compound?

    _____

    L: Sikul mane mani machiegni gi dalau ka?

    S: Shule gani iko karibu na boma lenu?

16. What is the nearest market to this compound? _____

    L: Chiro mane mani machiengi gi dalau ka

    S: Soko lipi liko karibu na boma lenu?

17. What is the nearest Church to this compound: _____

    L:Kar lamo machiegni

    S:Kanisa iliyo karibu

18. Names of head of compound for 3 nearest neighbors:

    L:Nyinge weg mieche adek mokiewo kodu.

    S;Majina ya wenye boma matatu majirani wakaribu kabisa

    a. _____
    b. _____
    c. _____

GEOLOCATING QUESTIONNAIRE

Show the patient/guardian the provided map of the health facility catchment area. Using the geolocation questions and key features that have been located on the map, help the patient/guardian locate where their compound is on the map as accurately as possible and record the results.

19. Map details
    d. Able to locate residence on map?
       Drop down – Yes/No
       If Yes, move to b and c. If No, skip to section 3
    e. Block number:
       Number pad– 2 digits
    f. Cell number
       Number pad – 2 digits

SECTION 3 - PATIENT INTERVIEW

20. Has [X] had a fever in the last 24 hours? YES/NO/DON'T KNOW
    L: Bende [X] osebedo gi del maore nyoro kata kawuono?
    S: [X] amekuwa na maumivu na mwili wake kuwa na joto masaa ishirini na manne yaliyopita?
21. Why is [X] visiting the clinic today?_____
    L: Ango' momiyo [X] obiro ei osiptal ni kawuono?
    S: Kwanini [X] ametembelea hospitali hii leo?
22. What symptoms does [X] have now?
    L: Gin ranyisi mage mag tuo ma [X] nigodo sani?
    S: Ni dalili zipi za ugonjwa [X] anazo sasa hivi?
    _____
    _____

23. Has (X) been ill with a fever at any time in the last 2 weeks?
    L: Bende [X] osebedo gi del ma ore e jumbe ariyo ma osekalo?
    S: Katika wiki mbili zilizopita, [X] amekuwa mgonjwa na kusikia maumivu na joto mwilini?
    a. YES

b. NO (SKIP TO 8)

c. DON'T KNOW

24. Has [X] had a fever in the last 24 hours?

    L: Bende [X] osebedo gi del maore nyoro kata kawuono?

    S: [X] amekuwa na maumivu na mwili wake kuwa na joto masaa ishirini na manne yaliyopita?

    a. YES
    b. NO
    c. DON'T KNOW

25. Did [X] seek advice or treatment for the fever from any source?

    L: Bende ne omanyo ng''ado rieko, kata thieth kuom del maore kamoro amora?

    S: [NAME] alitafuta huduma ya afya au matibabu popote?

    a. YES
    b. NO
    c. DON'T KNOW

26. Where did you seek advice or treatment? (Check all that apply)

    L: Ng''ado rieko kata thieth ne omanyo kanye?

    S: Alienda kutafuta wapi huduma au matibabu hayo?

    a. GOVT. HOSPITAL
    b. GOVT. HEALTH CENTER
    c. GOVT. HEALTH POST
    d. MOBILE CLINIC
    e. FIELD WORKER
    f. OTHER PUBLIC
    g. PVT. HOSPITAL/CLINIC
    h. PHARMACY
    i. PRIVATE DOCTOR
    j. MOBILE CLINIC
    k. FIELD WORKER
    l. OTHER PVT. MEDICAL_____
    m. SHOP
    n. TRAD. PRACTITIONER

o. OTHER_____

27. Has (X) taken any drugs in the last 2 weeks? (Check all that apply)

L: Bende [X] osemwonyoe yath e jumbe ariyo ma osekalo?

S: [X] ametumia dawa yoyote wiki mbili zilizopita?

a. SP/FANSIDAR
b. CHLOROQUINE
c. AMODIAQUINE
d. QUININE
e. COARTEM
f. OTHER ANTIMALARIAL (SPECIFY)_____
g. ASPIRIN
h. ACETAMINOPHEN/PARACETAMOL
i. IBUPROFEN
j. OTHER (SPECIFY)_____
k. NO
l. DON'T KNOW

28. Did [X] sleep under a net last night?

L: Be ng'ani noninde e bwo net gotieno manyoro?

S: Je huyu alila chini ya neti jana usiku?

a. YES
b. NO
c. DON'T KNOW

29. At any time in the past 12 months, has anyone with a back pack come and sprayed the interior walls of your dwelling with an insecticide to kill mosquitoes? YES/NO/DON'T KNOW

L: E dweche 12 ma osekalo, bende osekir yath e kor odni gi iye?

S: Katika jumla ya miezi kumi na miwili iliyopita, kuna yeyote amenyunyizia dawa kwenye kuta za nyumba yako?

We are now going to ask some questions about where [X] has lived in the past, and where he has travelled to in recent months. The reason we are asking these questions is to find out whether [X] might have been at risk of getting malaria in other places.'

L: Koro adwaro penjo kuom kama [X] osebedo ka odakie e thuolo ma okalo, to gi kama osedhie wuoth e dweche matin mokalo. Penjagi konyowa ng'eyo ka onyalo yudo tuo mar malaria Kuonde moko opogore gi ka.

S: Tutakuuliza maswali kuhusu pale [X] ameishi tena pasipo hapa na kule ametembelea miezi iliyopita ya karibuni. Tunataka kuweza kujua kama [X] angeweza kupata malaria pahali pengine

30. Apart from "EA Name" has (X) lived in any other EA for more than 6 months?

    L: Bende [X] osedak e EA moro ma opogore gi (Y) e thuolo mohingo dweche auchiel?

    S: Tofauti na hii "EA Name" [X] amewahi kuishi kwa EA ingine kwa muda wa inazidi miezi 6?

    a. YES (specify)
    b. NO
    c. DON'T KNOW

31. Has (X) made any overnight trips outside of "EA Name" in the last 3 months?

    L: Be [X] osedhi e gweng moro e dweche adek mokalo?

    S: [X] ametembelea kijiji kingine kwa kipidi cha miezi 3 zilizopita?

    a. YES (specify)
    b. NO (skip to results)
    c. DON'T KNOW (skip to results)

32. Final diagnosis_____

33. Treatment provided_____

Appendix 2.3 – Community Surveys and Focal Mass Drug Administration (FMDA)

(NB. Translations for questions in DhoLuo and Swahili (L, S respectively) are given only for those questions directed to household members and not for observational questions or for answers as all interviewers speak English)

Part I-Household Information-Fill out once for each house

1) House code _____

2) Head of Compound_____

3) Clan _____

4) Nearest Market _____

5) Nearest Primary School _____

6) Nearest Health Facility _____

7) What is the main occupation of the household head?

L: Tich mane ma wuon odni timo

S: Ni kazi ipi mwenye nyumba hufanya?

a) WORKS FOR PAY

b) INCOME FROM SPOUSE OR OTHER RELATIVE

c) UNPAID FAMILY BUSINESS

d) WORKS ON FAMILY FARM

e) UNEMPLOYED (ACTIVELY SEEKING WORK)

f) RETIRED

8) For the head of household, what is the highest level of education completed?

L: Wuon odni osomo nyaka klas adi?

S: Ni kiwango kipi cha juu zaidi cha masomo ambacho mwenye nyumba ame hitimu?

a) NONE

b) PRIMARY INCOMPLETE

c) PRIMARY COMPLETE

d) SECONDARY INCOMPLETE

e) SECONDARY COMPLETE

f) HIGHER INCOMPLETE

g) HIGHER COMPLETE

h) UNKNOWN

i) OTHER

9) How many windows have glass?

a) NONE

b) SOME

c) ALL

d) NO WINDOWS PRESENT

10) What is the main type of fuel used by your family for cooking?

L: En ang'o ma jo odni tiyogo kuom chweko e tedo mapile?

S: Unatumia aina gani ya nguvu au moto kupika?

a) ELECTRICITY OR GAS/

b) KEROSINE

c) CHARCOAL

d) FIREWOOD/STRAW

e) DON'T KNOW

f) OTHERS (SPECIFY) _____

11) What is the main source of drinking water in your house?

L: Ere kama jo odni goloe pi mar modho?

S: Uneyatoa wapi maki yako ya nyumbani ya kunywa?

a) PIPED WATER IN RESIDENCE

b) PUBLIC TAP/PUMP/PIPE

c) WELL IN OWN COMPOUND

d) PUBLIC WELL

e) RAIN WATER

f) RIVER/STREAM/SPRING OR OTHER SURFACE WATER

g) OTHER _____

h) DON'T KNOW

12) Observe/ask whether there are the following in this house

a) ELECTRICITY – YES/NO/DON'T KNOW

b) SOLAR POWER` – YES/NO/DON'T KNOW

c) TV – YES/NO/DON'T KNOW

d) MOBILE PHONE – YES/NO/DON'T KNOW

e) MOTORBIKE – YES/NO/DON'T KNOW

f) BICYCLE – YES/NO/DON'T KNOW

g) RADIO – YES/NO/DON'T KNOW

h) 2 SEATER – YES/NO/DON'T KNOW

i) CUSHIONS – YES/NO/DON'T KNOW

Part II-Person Questionnaire-Fill out questionnaire for each person who stayed in house previous night. For children, pose the questions to the primary caretaker

1) Name of person_____

2) Is (Name) listed on the house list?

a) Yes

b) No

3) Person ID number _____

4) Is (NAME) male or female?

a) MALE

b) FEMALE

5) Is (NAME) available for interview?

a) YES (If yes, skip to 7)

b) NO (If no, go to 6)

6) Will (NAME) be available for interview at another time?

a) YES, WHEN_____

b) NO

7) What is [NAME]'s date of birth?

L: [NAME] nonyuol e higa ane?

S: Tarehe ya [NAME] ya kuzaliwa?

a) YEAR_____

b) MONTH_____

8) Does (NAME) usually sleep here?

L: Bende [NAME] odak ga ka?

S: Mtu huyu [NAME] anaishi hapa?

a) YES

b) NO

9) Did (NAME) sleep here last night?

L: Bende [NAME] ne onindo ka otieno mokalo?

S: [NAME] Amelala hapa jana usiku?

a) YES

b) NO

10) Is (NAME) attending school?

L: Bende [NAME] dhi ga sikul?

S: Je [NAME] anaenda shule?

a) YES

b) NO

c) DON'T KNOW

11) Which school is (NAME) attending? (include whether primary or secondary)

L: Sikul mane ma [NAME] some?

S: Ni shule gani [NAME] anasoma?

_____

12) Which class is (NAME) in?

L: To [NAME] nie kilass adi?

S: [NAME] yuko darasa la ngapi

_____

13) What is the name of (NAME)'s class teacher?

L: Nying japuonj mar (NAME) en ango'?

S:Jina ya mwalimu wa (NAME) ni nani?

_____

14) Has (NAME) been ill with a fever at any time in the last 2 weeks?

L: Bende [NAME] osebedo gi del ma ore e jumbe ariyo ma osekalo?

S: Katika wiki mbili zilizopita, [NAME] amekuwa mgonjwa na kusikia maumivu na joto mwilini?

a) YES

b) NO (SKIP TO 16)

c) DON'T KNOW

15) Has [NAME] had a fever in the last 24 hours?

L: Bende [NAME] osebedo gi del maore nyoro kata kawuono?

S: [NAME] amekuwa na maumivu na mwili wake kuwa na joto masaa ishirini na manne yaliyopita?

a) YES

b) NO

c) DON'T KNOW

16) Has (NAME) taken any drugs in the last 2 weeks? (Check all that apply)

L: Bende [NAME] osemwonyoe yath e jumbe ariyo ma osekalo?

S: [NAME] ametumia dawa yoyote wiki mbili zilizopita?

a) SP/FANSIDAR

b) CHLOROQUINE

c) AMODIAQUINE

d) QUININE

e) COARTEM

f) OTHER ANTIMALARIAL (SPECIFY)_____

g) ASPIRIN

h) ACETAMINOPHEN/PARACETAMOL

i) IBUPROFEN

j) OTHER (SPECIFY)_____

k) DON'T KNOW

17) Did this person sleep under a net last night?

L: Be ng'ani noninde e bwo net gotieno manyoro?

S: Je huyu alila chini ya neti jana usiku?

a) YES

b) NO

c) DON'T KNOW

18) If NO, why not? (Check all that apply)

L: Ka da, nang'o?

S: Kama la kwanini?

a) IT IS TOO HOT UNDER THE NET

b) THERE IS NOT ENOUGH SPACE UNDER THE NET/I FEEL TOO

c) CLOSED IN

d) IT DOES NOT PROTECT AGAINST MOSQUITOES/INSECTS

e) NO MOSQUITOES AROUND

f) IT IS FOR ONLY CHILDREN/PREGNANT WOMEN

g) BEDNET USED BY PARENTS

h) BEDNET USED BY SIBLINGS

i) BEDNET BEING WASHED

j) BEDNET OLD

k) BEDNET KEPT FOR VISITORS

l) IT IS TOO EXPENSIVE/CANNOT AFFORD ENOUGH NETS FOR EVERYONE

m) IT IS NOT THE RAINY/MALARIA SEASON

n) CANNOT HANG IT OVER MY SLEEPING PLACE/SLEEPING

o) OUTSIDE

p) CHANGE MY SLEEPING PLACE TOO OFTEN

q) DO NOT KNOW

r) OTHER

We are now going to ask some questions about where [X] has travelled to in recent months. The reason we are asking these questions is to find out whether [X] might have been at risk of getting malaria in other places.'

L: Koro adwaro penjo kuom kama [X] osebedo ka odakie e thuolo ma okalo, to gi kama osedhie wuoth e dweche matin mokalo. Penjagi konyowa ng'eyo ka onyalo yudo tuo mar malaria Kuonde moko opogore gi ka.

S: Tutakuuliza maswali kuhusu pale [X] ameishi tena pasipo hapa na kule ametembelea miezi iliyopita ya karibuni. Tunataka kuweza kujua kama [X] angeweza kupata malaria pahali pengine

19) Has (NAME) made any overnight trips outside of "EA Name" within the last 3 months?

L: Be [NAME] osedhi e gweng moro e dweche adek mokalo?

S: [NAME] ametembelea kijiji kingine kwa kipidi cha miezi 3 zilizopita?

a) YES

b) NO (skip to 25)

c) DON'T KNOW

20) During the trips taken in the last 3 months outside "EA Name", which districts did (NAME) travel to? (check all that apply)

L;Ei dweche adek mokalo mane idhi e wuoth oko gi, ni dhi ei distrik mane?

S:Katika kipindi cha miezi tatu zilizopita ulipotembelea vijiji vingine, ulienda wilaya gani?

a) Suba, Homa Bay, Kuria or Migori

b) Trans Mara, Kisii, or Gucha

c) Nyamira, Bomet or Buret

d) Rachuonyo

e) Nyando, Kisumu, or Bondo

f) Kericho, Nandi, or Vihiga

g) Siaya, Butere, or Busia

h) Mumias or Kakamega

i) Bungoma or Teso

j) Other (Specify)

21) How many nights did (NAME) spend away the last three months?

L: [NAME] ne oduogo ka oa e woudhe manonindoe oko mogik karang'o

S: Ni usiku ngapi [NAME] amechukua akiwa kwa safari yake ya mwisho

Number_____ (Don't know=-1)

22) Has (NAME) made any overnight trips outside of "EA Name" in the last 1 month?

L: Be [NAME] osedhi e gweng moro e dweche 1 mokalo?

S: [NAME] ametembelea kijiji kingine kwa kipidi cha miezi 1 zilizopita?

a) YES

b) NO (skip to 25)

c) DON'T KNOW

23) Which districts did (NAME) travel to during the trip(s) outside "EA Name" during the last month? (check all that apply)

L;Ei dwe achiel mokalo ni, (NAME) ne odhi e wuoth ei distrik mage?

S:Katika mwezi moja ipitayo, (JINA) alitembelea wilaya gani?

a) Suba, Homa Bay, Kuria or Migori

b) Trans Mara, Kisii, or Gucha

c) Nyamira, Bomet or Buret

d) Rachuonyo

e) Nyando, Kisumu, or Bondo

f) Kericho, Nandi, or Vihiga

g) Siaya, Butere, or Busia

h) Mumias or Kakamega

i) Bungoma or Teso

j) Other (Specify)

24) How many nights did (NAME) spend away during the last month?

L: [NAME] ne oduogo ka oa e woudhe manonindoe oko mogik karang'o

S: Ni usiku ngapi [NAME] amechukua akiwa kwa safari yake ya mwisho

Number_____

(Don't know=-1)

If there are more people to interview that have been enumerated click yes to return to the start of the person questionnaire. If you have finished all of the people that have been listed ask the head of compound:

25) Are there any other people who may or may not be members of your family, such as domestic servants, lodgers or friends who usually sleep in this house?

L: Bende nitiere ji mamoko ma ok jodala ka mapile, kaka jotich, wede, kata osiepe ma nindo kae pile?

S: Kuna watu wengine wanaoishi kwenye nyumba hii kama marafiki, wafanyi kazi au watu wanaolipia malazi yao ambao wanalala hapa kila siku?

a) YES (Go back to person questionnaire start)

b) NO (Go to Results Section)