CHALLENGES TO MALARIA CONTROL IN THE DEMOCRATIC REPUBLIC OF CONGO AND BEYOND

Mark M. Janko

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geography.

Chapel Hill
2017

Approved by:

Michael Emch

Brian Reich

Conge Song

Aaron Moody

Steve Meshnick

ABSTRACT

Mark M. Janko: Challenges to Malaria Control in the Democratic Republic of Congo and Beyond
(Under the direction of Michael Emch)

Roughly 40% of the world's population lives in areas where they are at risk of malaria infection. In the last 15 years, the global health community has made considerable progress in reducing transmission. Despite this progress, a number of challenges to further reductions remain. This dissertation addresses three such challenges.

First, I focus on the ecology that serves as a backdrop to transmission, and focus on the role agriculture may play. In doing so, I attempt to understand how agriculture affects both mosquito behavior, as well as malaria risk in under-5 children in the Democratic Republic of Congo (DRC), a country with one of the world's highest malaria burdens. My findings from this work suggest that increasing exposure to agriculture is associated with increased indoor biting among *Anopheles gambiae* mosquitoes, which may be the mechanism driving the observed association between agriculture and increased malaria risk.

Second, I turn to address insecticide resistance, which may undermine the contributions that bed nets have in reducing transmission. One challenge in monitoring insecticide resistance is the difficulty in obtaining representative samples of mosquitoes. I make some progress in overcoming this limitation using population-based survey data collected from 2009-2016 in 21 countries across sub-Saharan Africa, and find that the effects of bed nets treated with different insecticides vary considerably, and that certain countries need to transition away from using certain insecticides.

Finally, I attempt to understand how malaria spreads. To do so, I leverage genetic data on the *Plasmodium falciaprum* malaria parasite from 28 neutral microsatellite markers drawn from malaria-infected children living in the DRC. I consider different population genetics tools to identify whether or not the malaria parasite population can be classified into smaller sub-populations, whether or not there is evidence of isoloation-by-distance, and if there appears to be gene flow between geographically and economically proximate regions. My results indicate that the malaria parasite population in DRC is best

characterized as single population with weak evidence of isolation-by-distance, with no strong evidence of gene flow or barriers to it. However, outliers were observed along DRC's border.

ACKNOWLEDGEMENTS

I am forever indebted to an uncountable number of people, who have helped me in an uncountable number of ways, from making sure that I had a beer (and Al's burger) in hand when I obviously needed one, to making sure that I had a dissertation in hand when I obviously needed one. First and foremost, I must give equal thanks to A-squad and α-squad, who have supported me in both endeavors throughout this long haul. Y'all know who you are, and I am grateful to have such incredible and supportive people in my life. Second, I must thank my committee. This dissertation would not have been possible without you. Special thanks go to Mike, who gave me space to figure out how I wanted to fit in this field, and for supporting my endeavor to pursue a masters degree in statistics at the same time as my PhD. I am also indebted to Steve, who ensured that this dissertation mattered. Indeed, my primary goal for pursuing a PhD was to develop skills that could be used to improve human well-being. You have been a tremendous mentor in helping me work towards that goal. I also want to thank Brian, who read through pages of my hack algebra, and who somehow managed to make Inverse Wisharts less scary (that is no easy task!). And thanks to Conghe and Aaron, geographers whose insights have made this dissertation better. Finally, I wish to thank my family, for all their love and support over the years, and for teaching me to work hard, because without being able to do that, none of what follows would have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDC	Centers for Disease Control and Prevention
DAG	Directed Acyclic Graph
DBS	Dried Blood Spot
DHS	Demographic and Health Survey
DRC	Democratic Republic of the Congo
FAO	Food and Agricultural Organization of the United Nations
IBD	Isolation-by-Distance
ITN	Insecticide-treated net
LLIN	Long-lasting Insecticide-treated net
MCMC	Markov Chain Monte Carlo
MIP	Molecular Inversion Probes
MIS	Malaria Indicator Survey
PCA	Principal Component Analysis
RDT	Rapid Diagnostic Test
UI	Uncertainty Interval
WHO	World Health Organization
WHOPES	World Health Organization Pesticide Evaluation Scheme

CHAPTER 1: INTRODUCTION

Bad air is not good

Malaria means bad air

Therefore it's not good

□

This dissertation is essentially about two things. First, this dissertation is about malaria, a disease that roughly 40% of the world's population is at risk of getting each year. Second, although less obvious, it is about how a geographer ought to study malaria or, for that matter, any other health outcome—that is, what can geographers offer that other disciplines cannot or do not? This is neither a new nor an idle question, but one asked 40 years ago in Melinda Meade's seminal 1977 paper that developed the theoretical framework that has motivated a large body of work in population health/medical geography ever since (Meade, 1977). These two things, of course, are not mutually exclusive, with the nature of malaria transmission being place-specific and spatially-varying, and geography being a discipline organized (in some way) around notions of place and space. The objective of this introduction, then, is to orient the reader towards both areas. With regard to malaria, this introduction will provide a broad overview of transmission, the disease burden, the efforts that have led to its reduction in the last 15 years, and current threats to that progress. Three of these threats make up the substantive focus of this dissertation, and I briefly mention each of these, leaving the details to chapters 2, 3, and 4, respectively. With regard to geography, this introduction will provide a brief overview of the theoretical framework for this dissertation, discuss how population health geographers have sought to answer scientific questions in this framework, the inherent limitations in these approaches, and then move to introduce a modeling framework that is more formally aligned with theory. Implementing this modeling strategy is largely the focus of Aim 1, but it has relevance to Aims 2 and 3 as well. I now turn to the substantive focus of this dissertation: malaria.

1.1 Malaria

1.1.1 The Disease

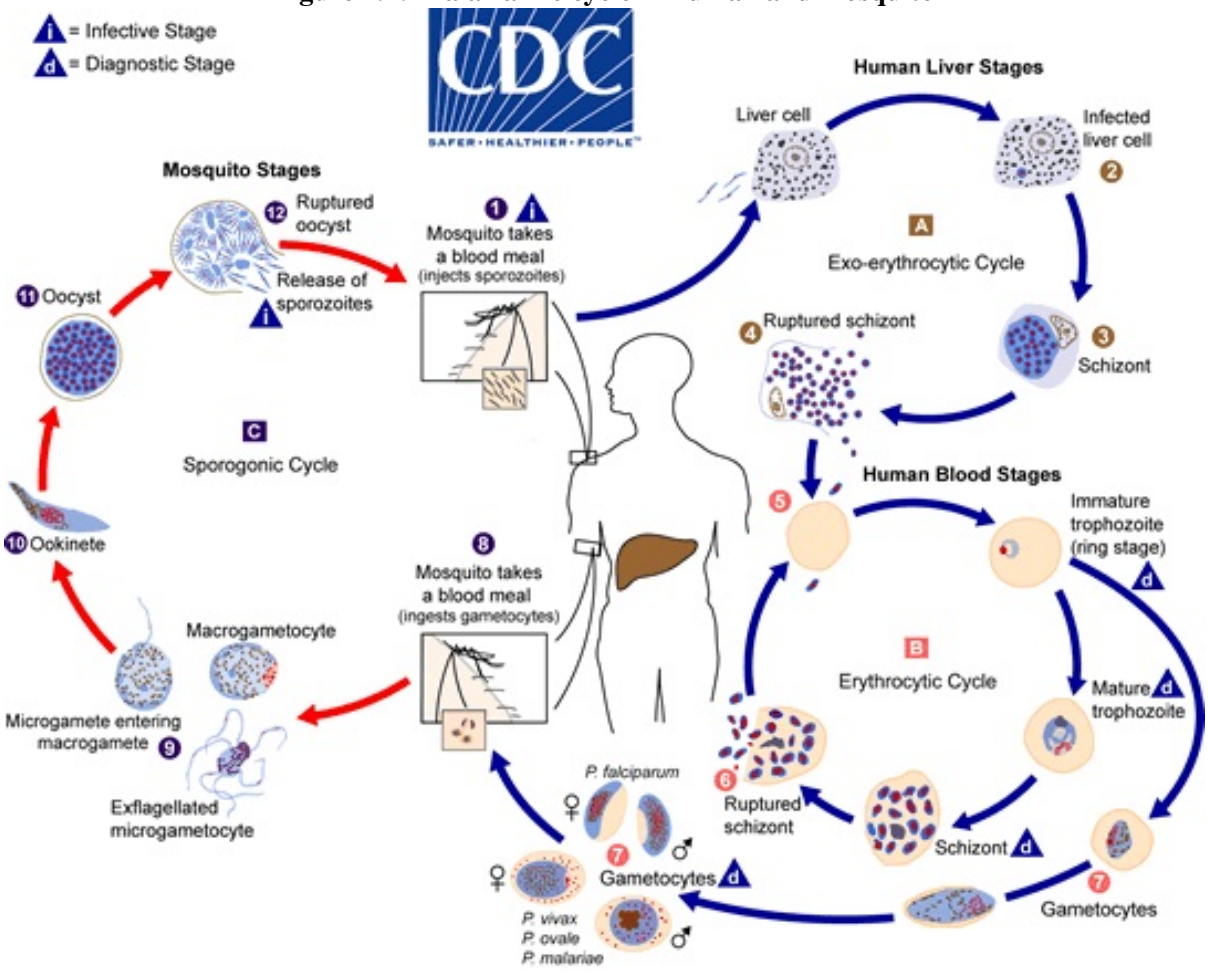
Malaria is a curable and preventable disease caused by an infection with one of five species of the *Plasmodium* parasite known to infect humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. The first two of these species are the most prevalent worldwide, but in Sub-Saharan Africa, *Plasmodium falciparum* is the predominant species. Worldwide, *P. falciparum* is also the deadliest. Following initial infection, and before any symptoms first appear, there is an incubation period that varies from 7 to 30 days, with *P. falciparum* infections tending to have shorter incubation times, and *P. malariae* longer times. Subsequent clinical symptoms in cases of uncomplicated malaria include fever, sweats, nausea and vomiting, chills, headaches, body aches, and general malaise (CDC, 2015). In more severe cases, such as those cases that go untreated or misdiagnosed, malaria infection may spread to the brain (causing cerebral malaria), lead to kidney failure or other major organ disease, severe anemia, respiratory distress, and death (CDC, 2015). Importantly, clinical symptoms may never emerge in an individual, meaning that infections can also be asymptomatic.

A malaria infection (of any species) is the result of a complex transmission cycle. At the most basic level, malaria is transmitted by an infectious bite from any one of 30-40 species of (female) *Anopheles* mosquito. Upon inoculation into a human host, the parasites multiply and grow, first in liver cells and then in red blood cells, eventually destroying these red blood cells and releasing merozoites ("offspring" parasites) that continue invading other red blood cells. During this blood stage of infection, gametocytes—the sexual stage of the parasite life cycle—circulate in the blood and are taken up by an *Anopheles* mosquito during a blood meal. Over the course of the next 10-18 days, these gametocytes develop into sporozoites through a series of intermediary steps, and then eventually enter the mosquito's salivary glands, where a further blood meal by the mosquito can transmit these sporozoites to another human, continuing the chain of transmission. Figure 1.1 below shows this life cycle in detail.

1.1.2 The Disease Burden and Malaria Control

Malaria is distributed across the tropical world, and the burden of disease is highest in sub-Saharan Africa (see Figure 1.2). For example, nine of the ten countries with the highest share of the global malaria

Figure 1.1: Malaria life cycle in human and mosquito

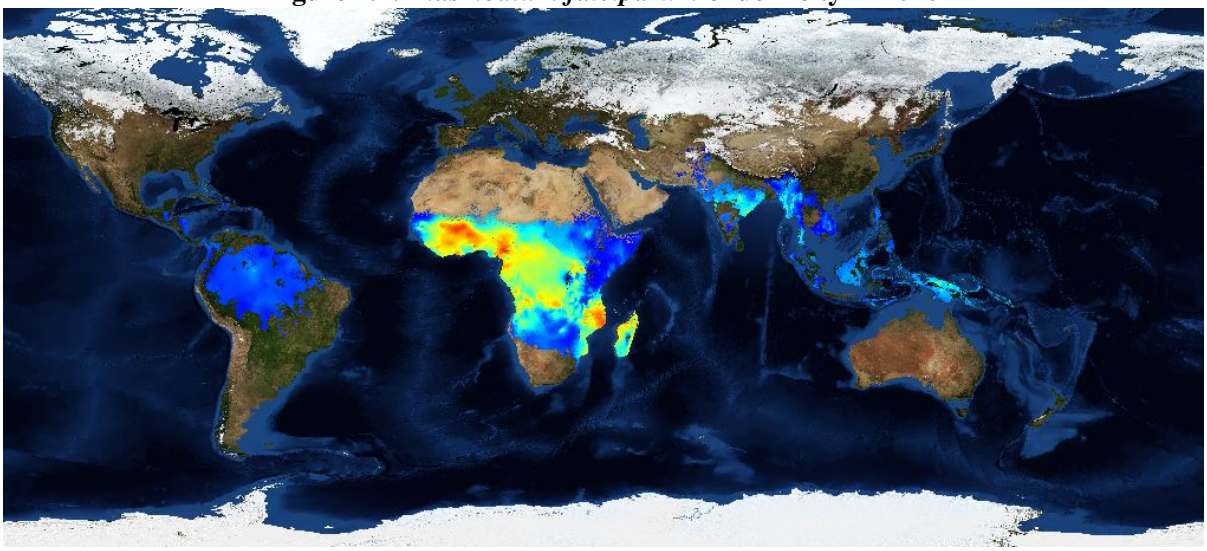


Source: CDC

burden (cases) in 2015 are in sub-Saharan Africa, while just five—Nigeria, the Democratic Republic of Congo (DRC), Uganda, Mozambique, and Cote d'Ivoire—accounted for 50% of the total burden (29% Nigeria, 9% DRC, and 4% in Uganda, Mozambique, and Cote d'Ivoire) (WHO et al., 2016). Malaria mortality is similarly concentrated in sub-Saharan Africa, with nine of the ten countries reporting the greatest share of deaths being in the region. Further, of the 429,000 (95% uncertainty interval [UI] $235,000 - 639,000$) estimated malaria deaths worldwide in 2015, five countries accounted for nearly 50% of them (Nigeria: 26%, DRC 10%, Mali 5%, Tanzania and Mozambique 4%) (WHO et al., 2016).

While 429,000 deaths represents stark evidence for the considerable work that remains to eliminate malaria, it nevertheless represents considerable progress towards that end. In the year 2000, for example, the WHO estimates there were more than twice as many malaria deaths worldwide (estimate 864,000;

Figure 1.2: *Plasmodium falciparum* endemicity in 2010



Source: Malaria Atlas Project

95% UI $655,000 - 1,087,000$)(WHO et al., 2016). This reduction is largely the result of a renewed commitment to eliminating malaria, starting with the 1998 launch of the Roll Back Malaria campaign, which set out an ambitious goal of halving the malaria burden by 2010 (WHO, 2005) by scaling up interventions with insecticide-treated bed nets (ITNs), indoor residual spraying (IRS), and prompt diagnosis and subsequent treatment using artemisinin-based combination therapies (ACTs).

This set of interventions essentially acts on two fronts. ITNs and IRS work to break the transmission cycle through controlling the vector population with insecticides that either kill a mosquito attempting to bite an individual sleeping under a net, or while resting on a sprayed surface. Additionally, ITNs provide a barrier that both prevents an infected mosquito from transmitting malaria to an uninfected human, or an uninfected mosquito from taking a blood meal from an infected human. Conversely, prompt diagnosis and treatment reduces the parasite population, such that an individual successfully treated will not transmit the parasite to a mosquito, thereby breaking the transmission cycle.

Although the scale-up of these interventions has been uneven across Africa, the reductions in transmission previously observed point to their success, and recent estimates suggest that they averted 663 million cases between 2000 and 2015 (95% UI $542 - 753$ million) (Bhatt et al., 2015). ITNs are estimated to have led to the greatest declines, responsible for an estimated 68% of the reductions (95% UI $62\% - 72\%$), followed by treatment with ACTs and IRS, with 22% (95% UI $17\% - 28\%$) and 10%

(95% UI $5\% - 14\%$), respectively (Bhatt et al., 2015). Importantly, this variability does not indicate that one intervention is more effective than another, but may reflect the different times and scales different interventions were deployed. Nevertheless, the impressive reductions in malaria transmission of the last 15 years will depend on the continued success of these interventions. This success, however, is by no means certain.

1.1.3 Challenges to Malaria Control

Despite the dramatic reductions in transmission as a result of large-scale interventions worldwide, a number of challenges may slow or reverse these important gains. In this dissertation, I focus on three.

Aim 1: Agriculture

The first of these challenges is the ecology that serves as the backdrop to transmission. Understanding the ecology of malaria and its vectors is an essential component of successful control (Ferguson et al., 2010). One of the most important factors influencing this ecology is environmental change, and in particular agriculture, which is of concern since over half of global population growth from now until 2050 is expected to occur in Africa, and UN projections suggest the population could double, from 1.2 billion in 2015 to 2.5 billion in 2050, with much of this growth occurring in rural areas (Bongaarts and Casterline, 2013; DESA, 2015; Jayne et al., 2014). Such growth places considerable demand on Africa's food supply, and governments are considering large-scale agricultural projects to meet this increased need (Jayne et al., 2014; Ijumba and Lindsay, 2001). This is particularly important in the DRC, a country with a high disease burden, and where the Food and Agriculture Organization of the United Nations (FAO) has made agricultural development a top priority (FAO, 2017).

Such agricultural development has the potential to undermine malaria control efforts, however, since expanding agriculture may produce habitats favored by *An. gambiae* mosquitoes, sub-Saharan Africa's most efficient malaria vector (Gimnig et al., 2001; Sinka et al., 2010). Work in this area is somewhat limited, with some studies focusing on agriculture and its effects on the vector population, while others have looked at agriculture and its effects on transmission. Few have attempted to investigate agriculture, the vector population, and transmission simultaneously. Further, many of these studies are conducted in a small number of sites, and may not be representative of the broader population, nor generalizable across ecological zones. They also frequently lack data on other factors relevant to transmission (e.g. bed net use), further limiting inferences on the agriculture-malaria relationship (Zhang et al., 2008).

In this study, I make some progress in addressing these limitations by using population-based survey data on under-5 children collected across the DRC, as well as entomological surveillance data collected contemporaneously across six sites representing DRC's different ecological zones. This aim, then, works to understand possible mechanisms by which exposure to agriculture among both mosquitoes and humans may give rise to increased (or decreased) risk of malaria.

Aim 2: Bed Net Effectiveness

The second challenge addressed in this dissertation relates insecticide resistance, and the ongoing need to monitor the effectiveness of bed nets. Indeed, as noted previously, insecticides play a vital role in bed nets, killing or repelling mosquitoes that come into contact with them, thereby helping to maintain protective efficacy after holes develop in nets through normal wear and tear (Darriet et al., 1984). However, resistance to pyrethroids, the only class of insecticides approved for use in bed nets, has been reported across sub-Saharan Africa (Ranson et al., 2011). Importantly, the epidemiological consequences of increasing levels of insecticide resistance are not clear, with nets failing to protect against malaria in some settings of high insecticide resistance, but dramatically reducing the odds of infection in others (Protopopoff et al., 2007, 2008; Henry et al., 2005).

As with agriculture, however, efforts to understand the potential effect of insecticide resistance have been limited to a small number of study sites, and as such results are not generalizable to the population. Moreover, surveillance efforts are challenged by the inability to representatively sample the vector population, meaning that molecular surveys cannot characterize the prevalence of knock-down resistance in the mosquito population. Molecular markers, however, are not able to detect all forms of resistance, such as behavioral resistance, which occurs when mosquito behavior changes to avoid contact with nets (Ranson et al., 2011; Russell et al., 2011). As a result, much more work is needed in this area.

Here too, I make some progress on addressing some of the limitations noted. In this aim, however, I expand the geographic scope beyond the DRC, and investigate the effects of bed nets treated with different insecticides using population-based survey data collected in 21 countries (including DRC) between 2009 and 2016. The rationale for doing so is simple: the burden of disease in Africa is disproportionately high, as are the consequences to malaria control if bed nets lose their efficacy there.

Aim 3: Diffusion

Finally, I conclude this dissertation by returning to the DRC to begin to explore a fundamental question: How does malaria move from place to place? At the most basic level, the spread of malaria

is driven by two processes. Over short distances, malaria spreads by way of mosquitoes that transmit the parasite among a local population. Over longer distances, however, malaria spreads through the movement of people. Understanding human movement and its role in the transmission of vector-borne pathogens has long been of interest in both geography and epidemiology (Prothero, 1977; Meade, 1977; Anderson et al., 1992; Stoddard et al., 2009). That said, understanding such a process poses certain challenges, chief among them the fact that we do not observe population movement in the data most commonly available to the malaria research community. Instead, inferences about population movement, and, more importantly, the spread of malaria, must be inferred from observations on the malaria parasite itself (Carrel et al., 2015; Patel et al., 2014). In this aim, therefore, I use genetic data obtained from malaria parasites infecting 496 children living in 262 communities across the DRC to begin to explore the structure of the parasite population, and hypothetical routes through which it may spread.

The public health rationale for doing so are several-fold, all of them related to informing malaria control. For example, concern is growing worldwide amidst reports of false-negative malaria rapid diagnostic tests (RDT) results—that is, malaria infections that go undetected by RDT. False-negative test results occur in malaria parasites that have deleted the Histidine Rich Protein 2 gene that codes for the target antigen of the most widely-used RDTs. A recent study conducted in the DRC found that approximately 6% of malaria infections in children under 5 years of age harbored this gene deletion, and that the parasite population exhibited considerable spatial and genetic structure (Parr et al., 2016). Importantly, the expansion of this parasite population threatens to undermine a cornerstone of malaria control, since RDTs represent the primary mode of malaria diagnosis, and their use is designed to ensure that only confirmed malaria cases are given antimalarial drugs. Similarly, antimalarial drug resistance has long posed a challenge to malaria control, and was attributed with the rise in prevalence across Africa in the 1980s and 1990s (WHO, 2005). Historically, drug resistant malaria has first arisen in Southeast Asia before spreading to Africa, where it first emerges in the East, followed by the west, suggesting a possible role for Central Africa in the spread (Wongsrichanalai et al., 2002; Taylor et al., 2013). For these reasons, understanding how the parasite population is structured, and exploring the dimensions of human movement and how it contributes to the spread of malaria, can help guide future malaria control efforts in the presence of emerging challenges to control.

Given this brief introduction to the substantive aims of this dissertation, I now turn to discuss the theoretical grounding that underlies much of my thinking regarding this work.

1.2 Population Health/Medical Geography

1.2.1 Introduction

As noted, while the substantive focus of this dissertation is malaria, this work is conducted with geography in mind. Of particular concern, moreover, are the theoretical underpinnings governing much work in population health/medical geography. Specifically, forty years ago, in 1977, Melinda Meade developed the 'Triangle of Human Ecology,' which has served as a core theoretical framework motivating work in the discipline ever since. She opens this work by noting: "Health professionals frequently wonder how medical geography differs from epidemiology, or what geographers do that health planners do not. These are not idle questions" (Meade, 1977). Writing 16 years later, Kearns notes that "Geographers have asked questions of who gets what, where, and why with respect to illness, which is also the central goal of epidemiology (Kearns, 1993). Indeed, this question remains relevant today, as tools commonly used by geographers are also used by public health professionals whose training is not necessarily rooted in geographic thought. One such tool, the hierarchical model, also known as a mixed model, random coefficient model, or multilevel model, is frequently deployed by population health/medical geographers, and serves as the primary vehicle for inference in two closely related fields, and what Arcaya and colleagues have described as two largely separate veins of inquiry: Neighborhoods and Health, and Spatial Epidemiology (Arcaya et al., 2012). Interest in the former largely lies in trying to identify associations between ecological exposures and individual health outcomes. Examples across a range of publication outlets show efforts to understand herd effects of a vaccine or other type of intervention (such as bednets) to prevent infectious diseases such as cholera or malaria (Ali et al., 2005; Messina et al., 2011; Perez-Heydrich et al., 2014), or to understand associations between the neighborhood food environment and an individual's diet (Morland et al., 2002), above and beyond the effects of individual characteristics. Conversely, spatial epidemiology is largely interested in one of three phenomena, depending on the type of data at hand. In the case of point-referenced data, spatial prediction is frequently a central goal, with perhaps the most common prediction goal being to estimate the prevalence of a disease across a study domain using a finite set of points where prevalence has been measured (Banerjee et al., 2014). In an areal data setting, the goal is typically smoothing, and proceeds by fitting a model with some variation of a conditionally autoregressive (CAR) prior to borrow information from neighboring areas to produce a

map that smooths over noise induced by data features such as small sample sizes within an area. In a public health setting, these models are frequently fit to learn about variability in disease risk over space, with a related aim to identify covariates associated with increased/decreased disease risk (Banerjee et al., 2014; de Araújo et al., 2013; Lawson, 2013). Finally, in the case of point-process data, the central goal is often to identify clustering, for example of disease cases around a point source (Banerjee et al., 2014).

These fields of inquiry are often divided theoretically. For example, Neighborhoods and Health research is generally organized around the principle that place (i.e. context) matters for health (Kearns, 1993), and that particular characteristics of a place (e.g. its composition) influence health outcomes. From a methodological perspective, place is largely defined as membership within a (often geographically-referenced) unit of some sort, be it a school, hospital, county, village, or city. Spatial Epidemiology is also interested in place, but with goals of prediction, smoothing, and clustering, it is the spatial arrangement of these places that tend to be of greater interest. In the broadest sense, then, Neighborhoods and Health tends to focus on estimating the effect a place's characteristics have on individuals within that place, while spatial epidemiology is concerned with spatial associations between places. Importantly, public health researchers are paying increasing attention to the need to consider both.

Earlier work in this area was largely aspatial, with hierarchical models fit to geographic data without geographic structure built into the model, with spatial methods being employed primarily for distance calculations and spatial aggregations, and formal spatial modeling largely being done on aggregated data (Auchincloss et al., 2012). The increase in software with the capacity to handle different correlation structures, however, has led to an increase in modeling that can address both concerns. Perhaps the most frequent (and often not explicitly mentioned) approach to incorporating both is to fit a hierarchical model separate from a spatial model. For example, Carrel and colleagues fit a hierarchical model aspatially to assess the relationship between rurality and HIV infection, but also provide a descriptive map of how HIV prevalence has changed over time using Bayesian kriging (Carrel et al., 2016). Work such as this is limited, however, in that the functional form of the two models are different, both in terms of the correlation structure for the random effects, as well as the covariate information included. In the case of the former, missing spatial random effects precludes learning about a spatial process after accounting for covariates. In the case of the latter, covariate information is often missing in the spatial model, limiting our ability to learn about important covariates, as well as the spatial process, since some of the spatial structure can likely be explained by conditioning on relevant covariate information.

Efforts to extend the basic modeling setting above have been done through comparing the two approaches—typically by way of information criteria such as BIC—in an effort to more formally understand contextual and spatial forces underlying health phenomena. Chaix and colleagues, for example, conducted two studies in which hierarchical models implemented aspatially were compared to spatial models on the aggregated data to learn about contextual and residual spatial effects on healthcare utilization and mental health (Chaix et al., 2005a,b). In the study on mental health, the authors also work in a geo-additive modeling framework with a smoothing term for longitude and latitude such that risk can be predicted across the study area, but acknowledge that this framework does not provide inference for the scale and range of the spatial process. Finally, other work is now emerging that identifies contextual (i.e. place) effects on health while also learning about an underlying spatial process. An example of this can be found in the work of Hajat and colleagues, who investigate associations between air pollution and individual and neighborhood level socioeconomic status (Hajat et al., 2013). In that study, the authors write down a hierarchical model to identify the contextual effect of SES on exposure to pollution, but also incorporate spatial correlation in the random effects.

While this brief review by no means covers the vast amount of recent work in both Neighborhoods and Health and Spatial Epidemiology, it does highlight a trend showing that model sophistication is increasing to correspond with the complexity of the data we have at hand, as well as the complexity of the spatial and contextual questions we wish to ask. Nevertheless, there are limitations and opportunities to overcome them. In all of the modeling scenarios addressed above, for example, spatial random effects were considered as part of an error process that persisted after accounting for covariate information, the effects of which are assumed to be invariant over space. Yet, as I will argue, an exposure's effect on a health outcome may vary across space as well, possibly due to missing covariate information. Furthermore, the methods used in Neighborhoods and Health and Spatial Epidemiology are built upon the same modeling framework—the hierarchical model—suggesting that the two veins of inquiry would benefit greatly through further integration of ideas. Given that Population Health/Medical geographers are core members of the Neighborhoods and Health and Spatial Epidemiology research communities, core theory from the discipline can help to integrate the ideas of space and place in a general and flexible modeling framework. Doing so represents an initial effort to develop formal inferential goals for Population Health/Medical geography, thereby extending the contribution geographers can make to

public health, as well as to the advancement of the closely related Neighborhoods and Health and Spatial Epidemiology subfields.

In the following sections, I hope to make progress in this front. In 1.2.2, I briefly outline Meade's Triangle of Human Ecology, and how it connects to ideas common in Neighborhoods and Health and Spatial Epidemiology. In Section 1.2.3, I then attempt to link Meade's theoretical framework to the effects of place and space on health. In this section, I pay particular attention to the role of unmeasured confounding, and discuss the theoretical and practical contribution that spatially modeling such confounding has in terms of moving us towards better understanding the processes promoting or preventing health outcomes. In section 1.2.4, I turn to an empirical example, where my focus is to understand the effect of two environmental risk factors on the probability of malaria infection among children under 5 years of age in the DRC. Section 1.2.5.

1.2.2 Triangle of Human Ecology

Meade's theory owes much of its intellectual origin to the work of Jacques May, who became the "father" of Population Health/Medical geography in the United States with a series of publications throughout the 1950s, beginning in 1950 with "Medical Geography: its methods and objectives," in which he acknowledged that "today we recognize that disease is a multiple phenomenon which occurs only if various factors coincide in time and space" (May, 1950). He describes the multiple phenomena in terms of causative agents—those agents such as *Plasmodium* parasites that actually cause disease—as well as other agents that are necessary for the causative agent to interact with a susceptible human subject, such as the *Anopheles* mosquito necessary for malaria transmission. While an infectious bite from an *Anopheles* mosquito thus represents the causal mechanism by which transmission occurs, he goes on to note geographical factors relevant for such occurrence. Geographical features are wide ranging, multiple, and disease-specific, though he broadly characterizes those most relevant for human disease as being related in some fashion to heat, light, and humidity.

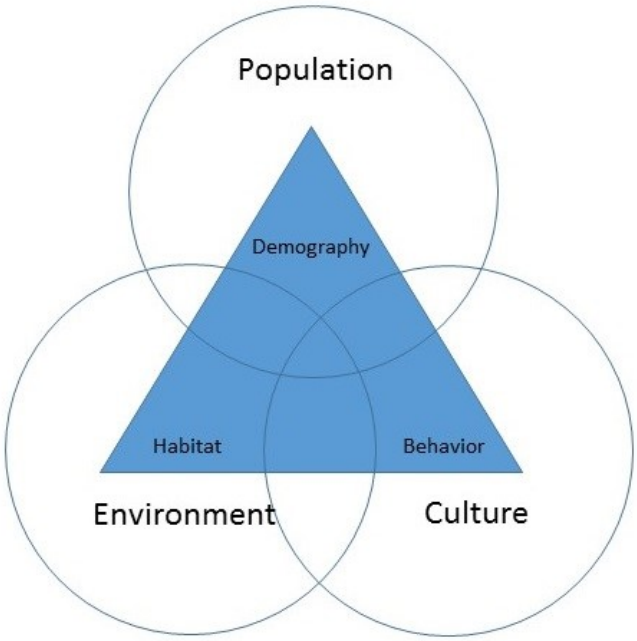
Working in a data landscape that is starved by today's standards, May followed this initial work with a series of maps which, for the first time, showed the global distributions of various diseases (Meade and Emch, 2010). Then, in 1959, he released *The Ecology of Human Disease*, the result of an intellectual evolution in which he came to see that cultural and environmental conditions produced or limited health and disease in different ways (May et al., 1959). In this body of work, the map is, not surprisingly,

a central object in population health/medical geographic thought. At the same time, however, May's thoughts of disease evolved from an initial observation while working as a physician in southeast Asia that his patients both experienced and responded to disease differently than did his patients back in Europe. Following in the French tradition of geography at the time, he began to see disease as being a function of population, environmental, and behavioral traits. In one immediately obvious sense, this work represented the groundwork for what would later be organized into Meade's theory. Such recognition has further importance in that it points to the need for population health geographers to contribute to understanding underlying diseases processes—i.e. understanding the 'multiple phenomena'—and not simply describe patterns. Indeed, as John Hunter noted: geographers are not to be "dot mappers," but have as one of their tasks the need to understand not just pattern, but process (Hunter et al., 1974).

Melinda Meade was a contemporary of Hunter (as well as a former student), and her work built substantially upon his and May's. In particular, her development of the 'triangle of human ecology' is largely a coalescence of both May's observations of multifactorial determinants of disease and Hunter's efforts to move the discipline to understand process. For example, in formalizing core dimensions of population health geography, she argues that by considering health as adaptability, this in turn implies that disease is the byproduct of the ways in which humans and their behavior, culture, socioeconomic context and demographic characteristics interact with the environment, be it natural or built. Such a conceptualization leads to the 'triangle of human ecology' represented in Figure 1.3 below. In this arrangement, we can see the broader aspects of human, population, and cultural features present, with disease-relevant features from each of these broader categories inside a triangle with differing intersections corresponding to different levels of risk.

One of the key contributions of Meade's theory that is perhaps overlooked, however, relates to spatial structure. Specifically, she writes that "behavior, the observable aspect of culture, usually has spatial expression" (Meade, 1977). Further, we can also easily imagine that other relevant features exhibit spatial expression. Environmental features are certainly correlated in space, and population characteristics can be as well. Importantly, such variability in these distributions and arguments as to their spatial structure provide theoretical underpinnings from within Population Health/Medical Geography's origins up to current motivations behind studies in the Neighborhoods and Health and Spatial Epidemiology fields. Indeed, May's observation that people experienced different health outcomes in different places can be thought of as an intellectual step towards thinking about the effects of a place on individual health

Figure 1.3: Triangle of Human Ecology



outcomes (the organizing principal in Neighborhoods and Health), while Meade's extension of this to consider spatial expression is an intellectual step towards thinking about associations between places (the organizing principal in Spatial Epidemiology). In this sense, then, space and place are integrally linked, implying that efforts to understand both would benefit from being addressed simultaneously.

1.2.3 Linking Meade, Space, Place, and Spatially Varying Coefficients

From a methodological perspective, handling space and place simultaneously within a single modeling framework requires working with hierarchical spatial models. As noted previously, the typical implementation involves fitting a model in which only the intercept is spatially varying. A fuller model would consist of models in which the effects of covariates are modeled as spatially varying as well, and that fitting such models represents a core contribution geographers can make within public health broadly, and within Neighborhoods and Health and Spatial Epidemiology specifically. Moreover, we can justify this modeling framework both on theoretical grounds, as well as from a more classical epidemiological perspective—that is, the perspective of causal inference.

One reason for this relates to the multiple relevant factors for disease that May describes, and in particular the practical challenge of gathering data on all of them. For example, in the case of malaria or other vector-borne disease, data on the vector is generally not collected, and environmental characteristics such as temperature and precipitation, important to many vector populations, are used in models of disease risk instead. However, the vector population is itself heterogeneous, with, for example, different species of *Anopheles* mosquitoes responsible for the transmission of malaria and exhibiting considerable spatial structure. These different species, moreover, have their own habitat preferences, and respond to the same environmental conditions in different ways. Thus, fitting a hierarchical model in which covariate effects are considered invariant over space will not account for any confounding induced by the unobserved and heterogeneous vector population.

Situations such as this can become particularly problematic if meaningful variation in effects is naively masked. Indeed, to take an extreme example, consider a study domain defined over a region. Now imagine that for people living in one half of this region, a given exposure is protective against a health outcome, while over the other half of the region the exposure increases risk. If we fit a hierarchical model in which we assume that the effect of the exposure on the outcome is invariant over space, we may discover, wrongly, that the exposure is not associated with the health outcome, and the conclusions we draw will be detrimental to half of the study population. In fact, we will have estimated a relationship that we assume holds everywhere, but in fact holds nowhere—an "everywhere is nowhere" effect, in essence.

To take a more concrete example, consider the case study included in the paper introducing spatially varying coefficient regression models (Gelfand et al., 2003). In that work, the authors set out to model the selling price of homes based on various features of each home, such as the square footage, number of bathrooms, and age. If the study region of interest is a large metropolitan area, however, each of these features will affect the selling price of a home in different ways. For example, older homes will have higher values in some areas than in others, meaning that the effect of age will vary over the region as a function of, for example, the desirability of the neighborhood. From another perspective, without information about the desirability of an area that a home is located in, regression models assuming constant effects across the study region will fail to adequately capture the data generating process. From a more theoretical, as well as practical, perspective, collecting such information is a challenge from the start, given the difficulty in defining such terms as desirability and developing a means to measure it. Indeed, conceptualizing place and features of it have long been challenges in Population Health/Medical

geography, with an ongoing need to re-examine the interplay between place and space (Kearns and Joseph, 1993). Thus, even when measurements are made, they almost certainly fail to capture the full dimension of the construct they aim to, and in this case spatially-varying effects may result as well.

While this is certainly a limitation that needs further work by population health scientists and others, I argue here that, even though we cannot perfectly measure characteristics of a place, or even observe them at all, we can still learn about key dimensions of them. Indeed, if we can imagine what component of the hypothesized data generating process is unobserved, such as a vector population, or only partially observed, such as behavior, we can then think through how it might induce a spatially varying coefficient in other parts of the data generating process. Further, we argue that learning about this unobserved process and its structure over space should be a *fundamental* goal of Population Health/Medical geography, as it delivers local inferences about place effects on health while simultaneously learning about how these place effects vary over space, thereby integrating both space and place into a rich and flexible modeling framework. To demonstrate this, we turn now to an empirical example. In the tradition of May, who paid particular attention to the importance of the role of heat and water to disease transmission, we take as our example the effects of temperature and precipitation on malaria risk.

1.2.4 Spatially Modeling the Effects of Temperature and Precipitation on Malaria Risk

As can be seen in Figure 1.2, malaria exhibits considerable spatial heterogeneity, a phenomenon which is a frequently described, but poorly understood (Bousema et al., 2010). It is thus naturally suited to study within Meade's framework, with risk factors that span behaviors such as bednet use, population characteristics such as age, and environmental factors such as altitude, temperature, and precipitation. As noted previously, the unobserved mosquito population is sensitive to these environmental conditions. For example, *An. arabiensis* and *An gambiae* have similar larval habitats, but different biting behaviors, with the former being relatively more zoophilic (Sinka et al., 2010). Given these similar habitats, environmental conditions will clearly affect both species. For example, precipitation will increase the number of available breeding pools, thereby leading to a possible decrease in competition between vectors, or to a more rapid expansion of one population such as *An. gambiae* over *An. arabiensis*. Further, while increased temperature is generally believed to increase development rates of the parasite, and thus favoring transmission, this is not necessarily the case (Paaijmans et al., 2011). Given these complexities, and given that we do not observe the vector population, its composition, or its behavior,

we turn to work in a spatially varying coefficient framework and model the effects of temperature and precipitation on malaria risk in 6,657 under-5 children sampled across 492 communities as part of the 2013-14 Demographic and Health Survey (DHS) administered across the Democratic Republic of Congo.

The details of these data and the model fit will be presented more fully in the next chapter, as well as in Appendix I, where the full conditional distributions for all model parameters are derived. Here, I simply note write down the basic model as follows:

$$Y = \mathbf{X}\beta + \mathbf{Z}\theta + \epsilon$$

Where $\mathbf{Y}$ is an $6,657 \times 1$ vector of binary responses indicating each child's malaria status, $\mathbf{X}$ is a $6,657 \times 3$ design matrix with an intercept and two covariates: the average temperature (measured in degrees Celcius) during the month of the survey, and total precipitation (in cm) the month prior to the survey. β is a 3×1 vector of regression coefficients linking covariates to response, $\mathbf{Z}$ is a $6,657 \times 1,476$ random effects design matrix that maps a spatial random effect at location $j, j = 1, \ldots, 492$ to individual $i, i = 1, \ldots, n$. The spatial random effects are thus represented by θ, a $1,476 \times 1$ vector. Finally, ϵ is white noise process assumed to follow a standard normal distribution.

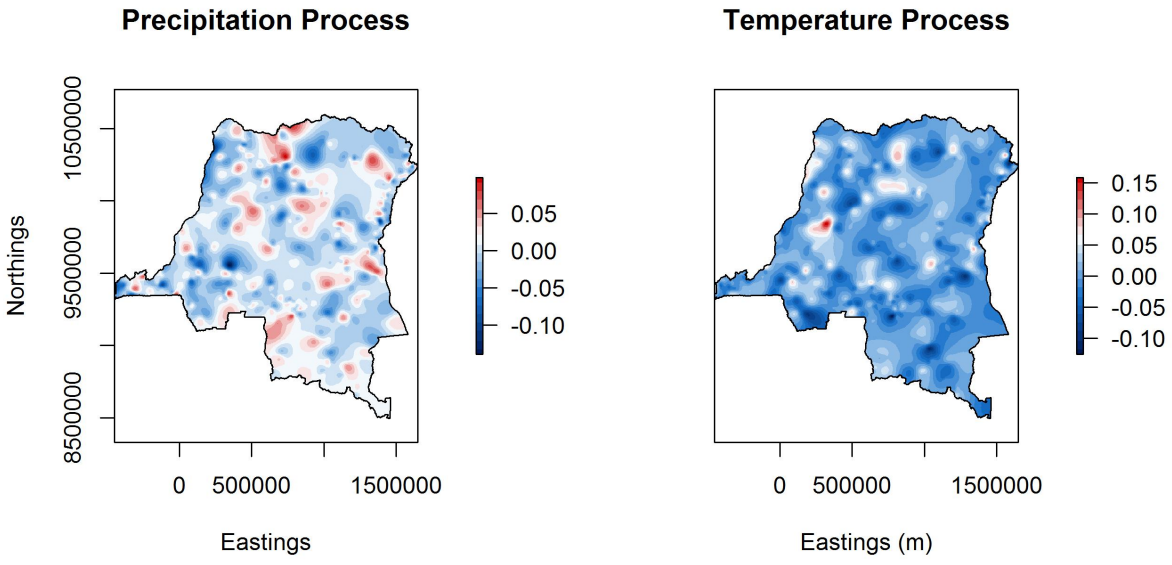
We adopt a probit specification and introduce latent normal variables for response $\mathbf{Y}$. Working in a Bayesian setting, we complete the model specification by assigning prior distributions to all unknown parameters in the model. Specifically, we assign a diffuse, zero-centered normal prior for the regression coefficients β. We model the spatial random effects as realizations from a zero-centered Gaussian process with separable covariance structure. We assign a low-precision Inverse Wishart prior for the spatial variance-covariance matrix, and use an exponential covariance structure with a uniform prior for the spatial range. We fit the model using MCMC and run the sampler for 120,000 iterations, discarding the first 20,000 as burn in and thinning the Markov chain to collect every tenth posterior sample. Inference for all parameters is thus based on 10,000 posterior draws.

Subsequent to model fitting, we then take 10,000 draws from the posterior predictive distribution to obtain estimates of the intercept and slope processes across the DRC. Doing so leads to the precipitation and temperature surfaces seen in Figure 1.4, which show considerable variability in the effects of each, especially in the precipitation process. The challenge, then, is to begin to develop hypotheses

about what may be driving the heterogenetity observed. One possibility is that precipitation measures are considerably noisy, and this may account for some of the variability observed. Beyond concerns over measurement, we may further hypothesize that other factors may contribute to this heterogeneity, such as different land cover types in the DRC. Indeed, as noted, temperature and precipitation are largely used as measures since the vector population is unaccounted for in the model. In particular, *An. gambiae* mosquitoes prefer transient pools of sunlit water, such as those created when land is cleared for agricultural development, thus, regions of higher risk may be explained through understanding the role of agriculture in transmission.

Still other reasons may explain this variability. In the model implemented here, for example, demographic and behavioral covariates are not included in the model, and their inclusion would almost certainly account for some of the spatial variability observed here. Thus, this initial model can be considered the start of a larger model-building effort. For example, additional models would include other covariates and consider different correlation structures to address different hypotheses about the underlying data generating process, with model comparisons proceeding via different measures of fit, such as DIC or out-of-sample based approaches.

Figure 1.4: Results from spatially varying coefficient regression model



1.2.5 Discussion

Ultimately, the message from this introduction is that malaria is a complex disease that exhibits considerable heterogeneity. This heterogeneity arises from multiple sources, including a vector population that is heterogeneous both in its species composition and behavior, which changes as a function of human interventions. Further, human behaviors are themselves heterogeneous, and environmental processes are as well. Thus, efforts attempting to understand a heterogeneous disease process fundamentally need to try to identify and understand the sources of such heterogeneity.

Population Health/Medical Geography, as delineated by Meade and her predecessors' work, essentially calls for the study of heterogeneity in its various forms. As it pertains to geography, then, the theme of this dissertation is that the contribution geographers can make to public health is through investigating variability. In this introductory example, and in the next chapter as well, spatially-varying coefficients are one way to accomplish this goal. But variability over space can be investigated aspatially, such as through analyses that stratify by geographical region, or through efforts to understand human movements. These latter two approaches represent the theoretical and methodological underpinnings of chapters 2 and 3, respectively. With this in mind, I now turn to the substantive aims of the dissertation.

CHAPTER 2: AGRICULTURE AND MALARIA RISK IN UNDER-5 CHILDREN

Understanding the ecology of malaria and its vectors is an essential component of successful control (Ferguson et al., 2010). One of the largest factors influencing this ecology is land use and anthropogenic land use change, such as agricultural development (Patz et al., 2004, 2000). Agriculture is of concern since over half of global population growth from now until 2050 is expected to occur in Africa, and UN projections suggest the population could double, from 1.2 billion in 2015 to 2.5 billion in 2050, with much of this growth occurring in rural areas (Bongaarts and Casterline, 2013; DESA, 2015; Jayne et al., 2014). Such growth places considerable demand on Africa's food supply, and governments are considering agricultural projects to meet this increased need (FAO, 2017; Boserup, 2005; Jayne et al., 2014; Ijumba and Lindsay, 2001). However, agricultural projects may reverse reductions in malaria transmission over the past decade, since expanding agriculture may produce habitats favored by *Anopheles gambiae* mosquitoes, sub-Saharan Africa's most efficient malaria vector. Specifically, *An. gambiae* mosquitoes prefer transient sunlit pools of water with little or no surrounding vegetation (Gimnig et al., 2001; Sinka et al., 2010).

Few studies, however, collect data on both vector populations and malaria incidence, relying instead on the relationship between the environment and transmission indicators. Findings from this work suggest that agricultural development is associated with changes in mosquito indoor resting density, human biting rates, sporozoite rates, entomological inoculation rates, larval abundance, reproduction rates, gonotropic cycles, and vector capacity (Afrane et al., 2006, 2005, 2008; Munga et al., 2006; Lyimo et al., 1992; Vittor et al., 2009; Ijumba et al., 2002). Such changes, however, do not necessarily increase malaria risk. For example, increased larval density is associated with longer mosquito development times, and different agricultural practices and crop types had varying effects on malaria risk (Lyimo et al., 1992; Ijumba and Lindsay, 2001; Zhang et al., 2008). Importantly, these studies were conducted in a small number of sites, are not representative of the broader population, nor generalizable across ecological zones. They also lack data on key factors governing transmission (e.g. bed net use), limiting inferences

on risk (Yang et al., 2008). Thus, given the diversity of vectors and of human ecosystems, much more work in this area is needed.

In this study, we examine the relationship between agriculture, the mosquito population, and malaria risk using a population-based survey of under-5 children living in the Democratic Republic of Congo (DRC), a large, ecologically diverse country containing 47% of Africa's potential agricultural land, and accounting for 10% of global malaria deaths in 2015 (Jayne et al., 2014; WHO et al., 2016). We also use contemporaneous entomological monitoring data collected over time across the DRC's ecological zones. Doing so allows us to consider possible mechanisms through which increases in agriculture may lead to a hypothesized increase in malaria risk. Findings from this study can therefore provide insights into a hypothesized but understudied driver of transmission.

2.1 Methods

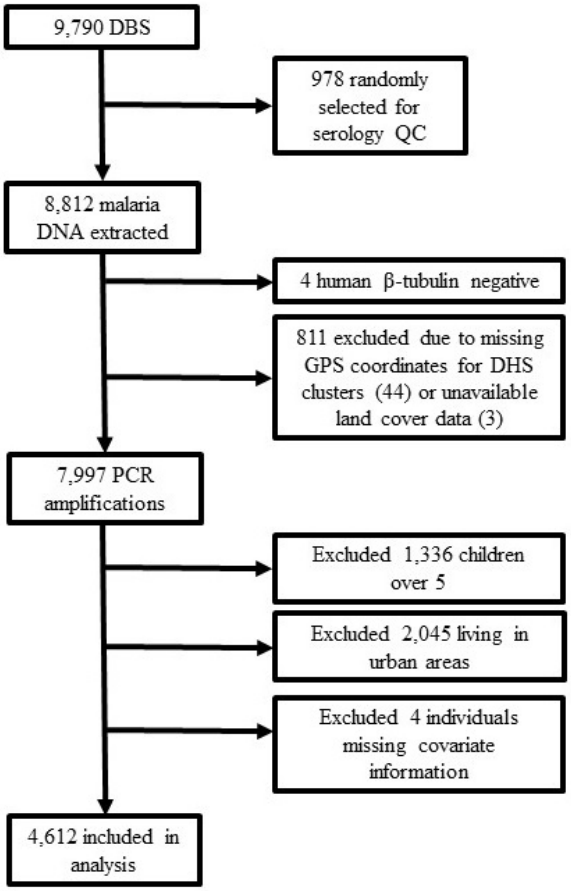
2.1.1 Study Population

The study population for this analysis consists of rural, under-5 children sampled as part of the 2013-14 DRC Demographic and Health Survey (DHS), a population-based cluster household survey. The sampling methods for the DHS are described elsewhere (Hancioglu and Arnold, 2013). The outcome for this study is each child's malaria status, as determined by polymerase chain reaction (PCR) analysis of dried blood spots (DBS) according to a previously published protocol (Singh et al., 1999; Taylor et al., 2011). A total of 8,808 DBS from children were available for this study. Spatial information and land cover data were unavailable for 44 and 3 DHS clusters, respectively, reducing the sample to 7,997. The DHS does not provide survey information on over-5 children, and these children were excluded. Finally, we include only children living in rural areas because agriculture in DRC is predominantly rural. This reduced the sample to 4,616 participants in 331 survey clusters, 4,612 of whom had no missing data. Figure 2.1 shows a flow diagram for the study.

The outcome measure for this study is each child's PCR-confirmed malaria status. DBS from the survey were shipped from DRC to the University of North Carolina at Chapel Hill (UNC) for PCR analysis. DNA was extracted from 9,790 children using Chelex, and PCR amplification for Plasmodium falciparum lactate dehydrogenase (DNA) was done according to a previously published protocol (Singh et al., 1999; Taylor et al., 2011). Of these, 978 blood spots were randomly selected for use in another

project. No corresponding survey data was available for 112 blood spots, leaving 8,700 blood spots for this study. Of these, 7,250 were samples from children under five years of age. These data were then merged to the DHS spatial data, which reduced the sample to 6,661 children under five (owing to the lack of spatial information for 44 survey clusters) with a PCR-confirmed result. Finally, given that interest here is in agriculture, a predominantly rural phenomenon, we include only those children living in rural areas, reducing the sample to 4,616 participants in 331 survey clusters, 4,612 of whom had no missing data on covariates of interest. Figure 2.5 shows a flow diagram for the study.

Figure 2.5: Study Flow Diagram



2.1.2 Exposure to Agriculture

Agricultural cover was derived using the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type data product (MCD12Q1), which provides yearly estimates of land cover at 500-meter resolution. In that dataset, two different classification schemes measure agricultural land cover,

one from the International Geosphere-Biosphere Programme (IGBP), the other from the University of Maryland (UMD). The IGBP measure includes two agricultural land cover classes, while the UMD measure includes one. We consider both classification schemes, and estimate the percent agricultural land cover within ten kilometers of each DHS cluster for each, then average the two to lessen the effect any extreme measures in one classification may have on inference. We chose a radius of ten kilometers since it represents the maximum flight distance of a female, human blood fed *An. gambiae* mosquito, representing the maximum extent over which human and mosquito populations interact (Kaufmann and Briegel, 2004).

2.1.3 Population, Behavioral, and Environmental Confounders

Population, behavioral, and environmental confounders are derived from the DHS and satellite remote sensing sources. Data extracted from the DHS include child age, sex, individual and community bed net use, altitude, and household construction materials—which represent both socioeconomic status and paths/barriers to mosquito entry. Individual bed net use was measured as use of a net treated with deltamethrin or alphacypermethrin, permethrin, or other kind of net. We consider net use in this manner owing to high levels of observed insecticide resistance to permethrin and remaining efficacy of deltamethrin and alphacypermethrin (AIRS, 2014; Levitz et al., 2017). Similarly, we calculate community bed net coverage according to the proportion of other respondents in the community sleeping under a deltamethrin- or alphacypermethrin-treated net, since these nets still possess a knockdown effect. Household wall construction was coded as natural (no walls or cane/palm/trunks), rudimentary (e.g. bamboo with mud), finished (cement), or other material according to the DHS. Roof construction was dichotomized as either finished (e.g. metal) or not, owing to small sample sizes in the rudimentary (27) and other (10) categories.

Precipitation and temperature were derived from multiple satellite platforms. We calculated the average temperature (in Celsius) the month the survey was conducted using the University of East Anglia's Climate Research Unit (CRU) TS3.23 data product, together with data from the National Centers for Environmental Prediction (NCEP) and the National Oceanic and Atmospheric Administration (NOAA). Precipitation was measured as the total rainfall (in centimeters) the month prior to the survey using Tropical Rainfall Monitoring Mission (TRMM) and CRU data. These measures were also calculated

within a ten-kilometer radius of each survey cluster and averaged. Our measures of temperature and precipitation are consistent with other work on malaria in DRC (Messina et al., 2011).

2.1.4 Entomological Monitoring Data

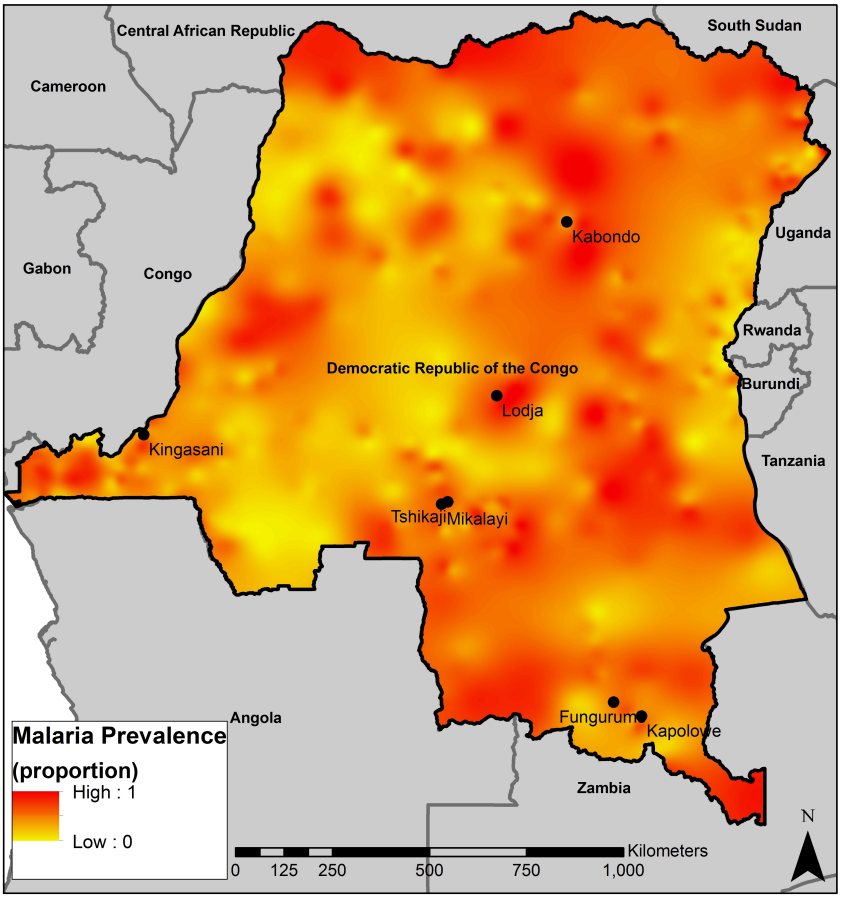
We consider the effect of agriculture on the vector population using entomological surveillance of *An. gambiae s.l.* (*An. gambiae*, hereafter), *An. paludis*, *An. moucheti s.l.* (*An. moucheti*, hereafter), and *An. funestus s.l.* (*An. funestus*, hereafter). In 2013, the Africa Indoor Residual Spraying (AIRS) Project conducted two rounds of mosquito surveillance (in August and November) across four sites chosen to represent equatorial, tropical, and mountainous ecological regions of the country. In 2014, three more sites were added, yielding seven total sites for 2014 surveillance, which occurred in February, April, and July. One of these sites was in an urban setting (Kinshasa), and was excluded since interest is in rural transmission (AIRS, 2014). Figure 2.6 maps each site together with background malaria prevalence in under-5 children. Mosquito collection occurred both indoors and outdoors using human landing catch (HLC), and we assume those caught indoors were intending to bite, and treat them as indoor biting mosquitoes.

In each site, eight households were chosen and HLC was performed in two of them each night for four nights by two mosquito collectors between 1800 and 0600 hours. One collector performed HLC indoors and the other outdoors. The two collectors switched places hourly to prevent mosquito attraction bias. Collectors were given malaria chemoprophylaxis. Mosquito species identification was done morphologically. Measuring agriculture, temperature, and precipitation around mosquito surveillance sites followed the same protocol as that used for the DHS survey.

2.1.5 Statistical Analyses

Probit regression models implemented in a hierarchical Bayesian setting were used to assess 1) the relationship between agriculture and mosquito indoor biting behavior, and 2) the relationship between agriculture and malaria risk. Three separate models were fit for indoor biting behavior among *An. gambiae*, *An. paludis*, and *An. funestus* mosquitoes. Insufficient numbers of *An. moucheti* mosquitoes (2) prevented modeling. All three models included a random intercept that varied independently across surveillance sites, and controlled for temperature, precipitation, and month of surveillance.

Figure 2.6: *Anopheles* surveillence sites and background malaria prevalence



Three models were also fit to assess the relationship between agriculture and malaria risk using DHS data. The three models addressed the survey sampling design, the unobserved vector population, and variability in crop types, the latter two representing sources of bias. The first model incorporates an independently varying random intercept to account for the correlation induced by the survey's cluster-sampling design. Such a model assumes unmeasured confounding exhibits no spatial structure. Given that the vector population is dependent on environmental conditions, which are spatially structured, we extend this model and incorporate spatial correlation in the intercept, thereby allowing for inference of unmeasured confounding across the DRC. Notably, both specifications assume no unmeasured confounding in the agriculture-malaria relationship. However, there may be variability in the effect due to different crop types, and different vectors may respond to agriculture in different ways. We therefore introduce a spatially varying coefficient process for the agriculture-malaria relationship (Gelfand et al., 2003).

All models are fit in a Bayesian setting. Continuous covariates (age, agriculture, temperature, precipitation, community bed net coverage) are first centered and scaled, such that regression coefficients represent effects per standard deviation increase in these variables. Regression coefficients are assigned standard normal prior distributions, while spatial structure is modeled using a Gaussian process with exponential covariance, consistent with other spatial models of malaria transmission (Hay and Snow, 2006; Hay et al., 2009). We implement the model using Markov chain Monte Carlo (MCMC) and run the sampler for 120,000 iterations, discarding the first 20,000 as burn in and thinning the Markov chain to collect every 10th posterior sample, such that final inferences are based on 10,000 posterior samples. Performance for models on malaria risk is assessed using Brier scores, area under the ROC curve, and DIC, with final inferences based on the best fitting model. Appendix I derives the full conditional distributions for all model parameters.

2.2 Results

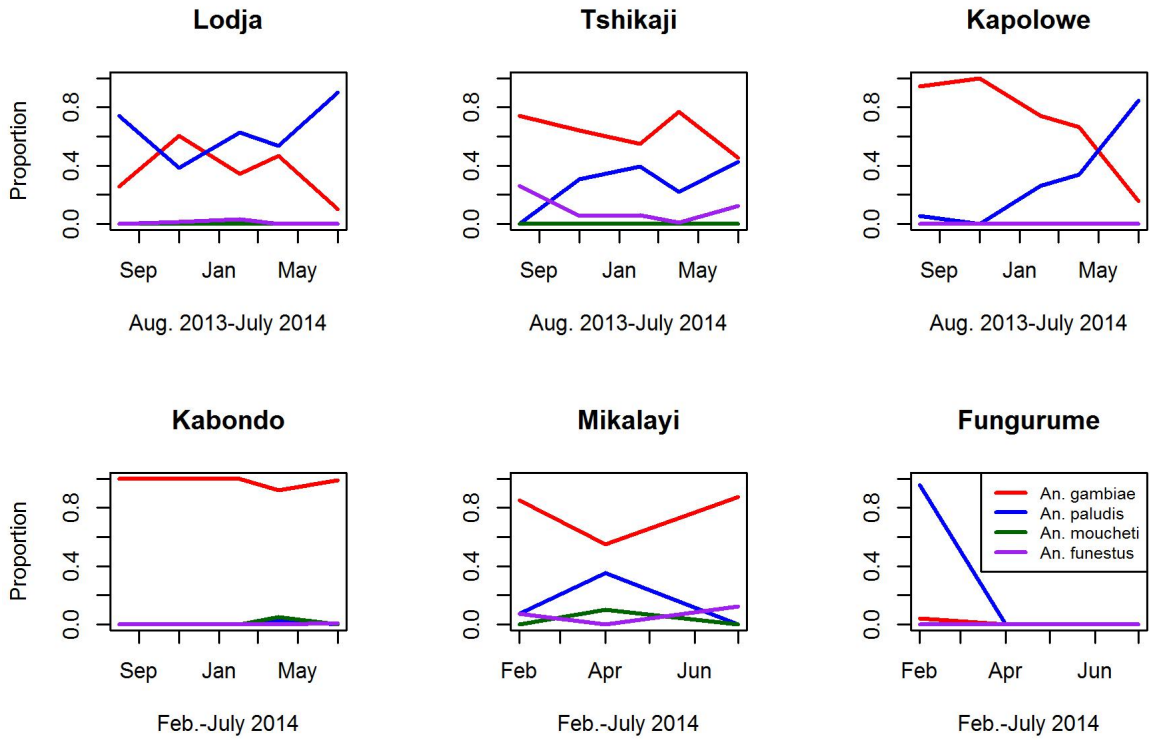
2.2.1 Effects of Agriculture on Mosquito Behavior

An. gambiae and *An. paludis* were the dominant vectors across all sites and over all time periods, making up 48% and 51% of 5,713 Anopheles caught by HLC, respectively. *An. funestus* and *An. moucheti* were relatively rare across all sites and over all collections. Further, relative abundance between *An. gambiae* and *An. paludis* varies in some sites, with relatively more *An. gambiae* from November through April, and more *An. paludis* during July and August. In the Kapolowe site, relative abundance appears unrelated to season, with *An. gambiae* abundance declining over the monitoring periods. Table 2.1 shows the relative abundance of each species caught over time, while Figure 2.3 presents these relative proportions over time at each site.

Table 2.1: Proportion of each *Anopheles* species by HLC collection period

Species	Collection Period					Total Caught
	Aug. 2013	Nov. 2013	Feb. 2014	Apr. 2014	Jul. 2014	
An. gambiae s.l.	0.44	0.79	0.58	0.61	0.17	2,744
An. funestus s.l.	0.03	0.01	0.02	0 (n=1)	0.01	71
An. moucheti s.l.	0	0	0	0 (n=2)	0	2
An. paludis	0.53	0.20	0.40	0.39	0.82	2,896
Total caught	455	606	2,143	895	1,614	5,713

Figure 2.7: Relative abundance of *Anopheles* mosquitoes by HLC by site



Agricultural coverage across all six sites ranged from 3.7% to 25.3% (mean=15%, sd=7%), while total precipitation ranged from 0 to 26.5 centimeters (mean=7.9, sd=94.8). Average temperature ranged from 19 to 26 degrees Celsius (mean=24, sd=2). Results from probit models on indoor biting behavior among *An. gambiae* mosquitoes suggest that increasing exposure to agriculture was associated with increased probability of indoor biting with high posterior probability (estimate: 0.22, 95% uncertainty interval [UI]: $-0.21 - 0.68$; Pr(estimate$>$ 0) = 0.85), controlling for available confounders. Given a 15% increase in agricultural cover, this estimate is associated with risk differences in indoor biting ranging from 0.14 (95% UI: $-0.19 - 0.30$) to 0.27 (95% UI: $-0.22 - 0.67$), depending on factors such as season (month of surveillance), temperature, and precipitation. Conversely, there was no indoor biting response to agriculture among *An. paludis* mosquitoes (estimate: -0.01, 95% UI: $-0.12 - 0.09$). Among *An. funestus*, increasing agriculture was associated with decreased indoor biting with high posterior probability (estimate: -0.13, 95% UI: $-0.37 - 0.09$; Pr(estimate$<$ 0) $= 0.88$). However, *An. funestus*

were not present in high abundance in any site. Table 2.2 presents full model results, with parameter estimates and 95% uncertainty intervals for all three species.

Table 2.2: Results of hierarchical probit models assessing the effect of agriculture on indoor biting behavior among *Anopheles* mosquitoes in DRC

	An. gambiae s.l.			*An. paludis*			*An. funestus s.l.*		
Variable	**estimate**	**2.5%**	**97.5%**	**estimate**	**2.5%**	**97.5%**	**estimate**	**2.5%**	**97.5%**
Intercept	-0.76	-1.08	-0.45	-1.30	-2.96	0.35	0.30	-1.58	2.16
Agriculture	0.22	-0.21	0.68	-0.01	-0.12	0.09	-0.13	-0.37	0.09
Precipitation	-0.17	-0.25	-0.10	0.05	0.03	0.08	0.02	-0.12	0.08
Temperature	-0.30	-0.40	-0.20	-0.02	-0.11	0.07	0.07	-0.12	0.26
August 2013					Reference				
November 2013	0.75	0.47	1.03	-0.83	-1.33	-0.32	-0.34	-2.11	1.46
February 2014	1.02	0.74	1.30	0.52	0.24	0.80	0.44	-0.79	1.67
April 2014	1.20	0.95	1.44	1.76	1.39	2.11	-0.75	-2.37	0.85
July 2014	0.59	0.39	0.78	0.85	0.59	1.12	-0.24	-1.32	0.80

2.2.2 Effects of Agriculture on Malaria Risk

Table 2.3 presents descriptive statistics on the agricultural exposure and potential confounders included in all models, together with their expected relationships on malaria risk. There was no difference in the percent of agricultural land cover surrounding a community by overall malaria prevalence in under-5 children, with malaria-infected and uninfected children exposed to 11.1% and 11.2% agricultural land cover on average. There were also no differences on the temperature and precipitation measures, although malaria prevalence was 7.5% higher among those living at altitudes under 1,000 meters. Other differences included age, with malaria-infected children being slightly older. Individual bed net use also varied, with a higher proportion of malaria-negative children sleeping under deltamethrin- or alphacypermethrin-treated nets, and a higher proportion of malaria-positive children sleeping under permethrin-treated nets. Further, malaria-negative children were exposed, on average, to higher levels of community bed net protection, while malaria-positive children tended to live in poorer quality housing.

Among the three models fit, the model with an independently varying intercept yielded the best fit to the data. Appendix II presents fit statistics. Further, while not the best fitting models, the spatial models can be suggestive of potential future areas of concern, and we also include a discussion in Appendix II. Table 2.4 presents results from the best-fitting model, including parameter estimates, 95% uncertainty intervals, and the posterior probability that the exposure increases malaria risk. Values near 1 indicate

Table 2.3: Descriptive statistics for variables included in hierarchical models

	Malaria-positive	Malaria-negative	Expected relationship
Individual-level variables			
Age (years) (mean)	2.9	2.6	+
Sex (% female)	48.2	50.1	+
Bed net use (% under net)			
Delta/Alphacypermethrin	37.3	48.3	-
Permethrin	9.3	6.8	0
Other	2.1	1.1	0
Community-level variables			
Household wall material (% living in HH)			
Natural	13.7	21.8	+
Rudimentary	76.8	66.2	+
Finished	8.8	9.8	-
Other	0.6	2.2	0
Household roof material (% living in HH)			
Natural	88.3	85.1	+
Rudimentary (n=27)	0.6	0.6	+
Finished	11.0	14.0	-
Other (n=10)	0.1	0.3	0
Community bed net use (%)	39.2	47.1	-
Altitude (% $>$ 1000m)	12.4	19.9	-
Precipitation (cm) (mean)	16.2	16.5	+
Temperature (C) (mean)	24.7	24.4	+
Agricultural land cover (%)	11.1	11.2	+

Notes: n=4,612. Due to a small number of children sleeping in households with rudimentary or other roofing material, these categories were collapsed into the reference category in model fitting. A + indicates that the variable of interest is expected to increase risk, while a - indicates decreased risk. A 0 indicates no expected effect on risk.

high probability of increased risk. Values near 0 indicate high probability of decreased risk, while values near 0.5 correspond to little or no effect. Increasing exposure to agriculture was associated with increased malaria risk with a high posterior probability (estimate: 0.07, 95% UI: $-0.04 - 0.17$, Pr(estimate>0) $= 0.89$). As before, a 15% increase in agricultural cover is associated with risk differences ranging from 0.00 ($-0.00 - 0.01$) to 0.03 ($-0.01 - 0.07$), depending on other risk factors such as bed nets treated with deltamethrin or alphacypermethrin, age, housing quality, and altitude.

To generate a sense of how large-scale agricultural expansion may affect children under 5 in the DRC, we plot the hypothetical probability of malaria infection for each child as a function of agriculture, with coverage ranging from 0% to 75% (the minimum and maximum observed values in the data). Figure 2.4 shows this plot, where each unique child is shaded and stratified according to their risk based on other covariates from the model. We further stratify these children according whether or not a child's risk fell into the lowest 25%, middle 50% (i.e. interquartile range), or highest 25% quantiles. Children at the extremes—i.e. those at very low (green) or very high (red) risk for malaria based on other risk

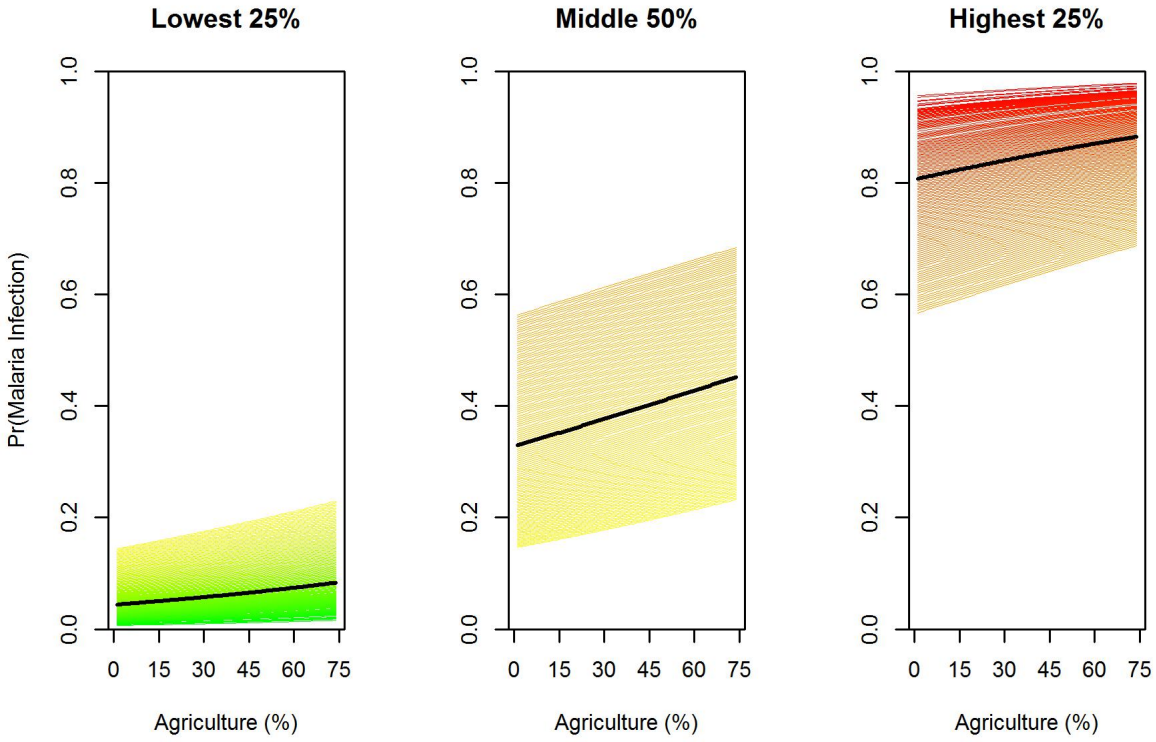
Table 2.4: Results for final hierarchical probit regression model on agriculture and malaria risk

Variable	Estimate	2.5%	97.5%	Pr(estimate>0)
Individual-level variables				
Intercept	-0.34	-0.51	-0.16	0
Age (years)	0.18	0.14	0.23	1
Female sex	0.03	-0.06	0.11	0.73
Bed net use				
Delta/Alphacypermethrin	-0.15	-0.25	-0.05	0
Permethrin	0.02	-0.17	0.21	0.58
Other	0.19	-0.18	0.56	0.84
Community-level variables				
Household wall material				
Natural		Reference		
Rudimentary	0.11	-0.04	0.27	0.92
Finished	0.05	-0.18	0.29	0.66
Other	-0.26	-0.77	0.26	0.17
Finished Household roof material	-0.12	-0.29	0.06	0.09
Community bed net use (%)	-0.21	-0.31	-0.12	0
Altitude (living $>$ 1000m)	-0.30	-0.70	0.11	0.07
Precipitation (z-score)	-0.07	-0.19	0.04	0.11
Temperature (z-score)	0.17	0.03	0.32	0.99
Agricultural land cover (z-score)	0.07	-0.04	0.17	0.89

Notes: Pr(estimate>0) indiates that the posterior probability that the regression coefficient is greater than 0, indicating posterior probability of increased risk. Values near or at 0 indicate that the effect is protective, while values at or near 1 indicate that the covariate is a risk factor. Values near 0.5 indicate no effect

factors—exhibit a small increase in risk due to large-scale agricultural expansion. For those whose risk is not at either extreme, however, dramatic increases in agriculture are accompanied by sizeable increases in malaria risk, as high as 0.13 $(-0.07 - 0.32)$, indicating increases in malaria risk due to potential large-scale agricultural development may be offset through simultaneous investments in housing quality, bed nets, and other interventions.

Figure 2.8: Hypothetical changes in malaria risk due to agriculture among children under 5 in DRC



Note: Shading indicates malaria risk based on other covariates, with green indicating low risk and red indicating high risk. For visualization, we further stratify based on each childs risk falling in the 25% (”Lowest 25%”), 25-75% (”Middle 50%”), and 75% (”Highest 25%”) quantiles. The black line represents the mean trend lines within each quantile.

2.3 Discussion

The agriculture-malaria relationship is complex, involving interactions between individuals, vectors, and the environment, all of which are highly diverse. To our knowledge, this is the first study attempting to understand the agriculture-malaria relationship using a population-based survey while incorporating available vector data from contemporaneous surveillance.

Results from entomological analyses suggest that increases in agriculture are associated with increases in indoor biting among *An. gambiae* mosquitoes, but not among *An. paludis*, and was associated with decreased indoor biting among *An. funestus*. Given the high abundance of *An. gambiae*, these results suggest that the agriculture-malaria relationship may be mediated through effects on indoor biting

among *An. gambiae*, and that while *An. funestus* showed a decreased probability of indoor biting with increasing agriculture, it only accounted for 1% of the mosquito population. Important seasonal patterns also existed among vectors, with the relative abundances of *An. paludis* and *An. gambiae* varying some sites, while indoor biting behavior among both species also varied, peaking in April 2014.

Results from analyses of DHS data indicate increased malaria risk with increasing agriculture, and that the relationship does not meaningfully vary over space due to confounding from the unobserved vector population or crop types. Further, our findings suggest that exposure to large-scale agricultural expansion will have a minimal effect on those at the most or least risk for infection based on other risk factors, but could have profound effects on those not at either of these extremes. Such an effect is of concern in the DRC, which has the largest proportion of potentially available cropland in sub-Saharan Africa, as well as one of the world's highest malaria burdens.

Considerable work remains to fully understand the relationship between agriculture and malaria risk in sub-Saharan Africa. Work on adult populations is limited, with one study in the DRC finding no effect (Messina et al., 2011). Additionally, the relationship between agriculture, temperature, and precipitation needs additional examination. In this study, we treat them as confounders, but they may also modify risk, and their roles are complex (Paaijmans et al., 2010, 2011; Krefis et al., 2011; Ageep et al., 2009; Hernández-Avila et al., 2006; Stresman, 2010; Paaijmans et al., 2008). That complexity is not fully captured here. It is also impossible to representatively sample the vector population, although our population was sampled in different ecological zones.

Work is also needed to understand the role of *An. paludis*, which has received virtually no attention in the malaria literature. Recent work to identify Africa's predominant malaria vectors predicted that *An. gambiae* was the dominant vector in DRC, which we do not dispute (Sinka et al., 2010). However, work from the 1990s suggested that *An. paludis* may be an important vector in the DRC, and given its observed presence and the country's high malaria burden, further work is needed to understand its role in transmission (Fontenille, 1999; Karch and Mouchet, 1992). Additional work is needed to try to understand how different types of agriculture may affect transmission.

In conclusion, this work provides the first evidence that increased exposure to agriculture increases malaria risk in children under 5 across rural and ecologically diverse settings, and may be due to increased indoor biting among *An. gambiae* mosquitoes. This is an area of growing concern for public health as transmission declines and as governments consider agricultural projects to respond to population growth,

projects that should be accompanied by additional malaria control measures (Ferguson et al., 2010; Pates and Curtis, 2005; Ijumba and Lindsay, 2001).

2.4 From Agriculture to ITN effectivness

This paper sought to characterize the effect of agriculture on the *Anopheles* mosquito population and on malaria risk. Doing so represented a contribution to efforts to understand the underlying ecology of malaria transmission, a prerequisite for malaria control (Ferguson et al., 2010). That said, another prerequisite to malaria control is that current interventions remain effective. Indeed, ITNs have been a mainstay of all malaria control programs worldwide. Yet, insecticide resistance poses a serious threat, and it has been observed across sub-Saharan Africa. This is of particular concern in light of our findings from this work, where increasing agricultural land cover was associated with increased probability that *An. gambiae* were caught biting indoors. If insecticide resistance renders certain bed nets ineffective, and if the ecological changes such as agricultural development lead the most pernicious malaria vector to bite indoors, then the loss of efficacy of ITNs due to insecticide resistance could be further compounded, and the transmission declines we have observe over the course of the last 15 years could be reversed.

Addressing this concern is the subject of chapter 3. Further, given the role of ITNs as a cornerstone of malaria control, we address this concern in 21 countries in sub-Saharan Africa.

CHAPTER 3: BED NET EFFECTIVENESS BY INSECTICIDE IN UNDER-5 CHILDREN ACROSS SUB-SAHARAN AFRICA

3.1 Introduction

Substantial investments in malaria control have greatly reduced malaria transmission since 2000, with interventions over this time period averting an estimated 663 million cases worldwide (95% uncertainty interval [UI]: 542 - 753) (Bhatt et al., 2015). Of the estimated cases averted, 68% (62% - 72%) have been attributed to use of insecticide treated nets (ITNs), a cornerstone of malaria control that have led to sustained reductions in transmission (Bhatt et al., 2015; Lindblade et al., 2006). ITNs possess two different mechanisms of action: they act as a physical barrier to prevent contact between mosquito and human, while insecticides provide an additional chemical barrier that can repel or kill susceptible mosquitoes on contact by targeting the voltage-gated sodium channel in the insect's nervous system (Darriet et al., 1984; Lengeler et al., 2004; Davies et al., 2007). This latter mechanism, moreover, helps maintain the net's protective efficacy after holes develop (Darriet et al., 1984). Thus, the continued success of ITNs largely depends on the continued effectiveness of insecticides, and the WHO has recommended all insecticide-treated nets (ITNs) in use be long-lasting insecticidal nets (LLIN), which are designed to retain insecticide activity for at least three years in field conditions, and are also of sturdier construction (WHO et al., 2007; Graham et al., 2005). Nevertheless, resistance to pyrethroids, the only class of insecticides approved for use in bed nets, has been reported across sub-Saharan Africa, the region with the highest malaria burden (Ranson et al., 2011; Kelly-Hope et al., 2008; Trape et al., 2011; Pinto et al., 2007; N'Guessan et al., 2007; Ranson et al., 2009; Hargreaves et al., 2000). Understanding the impact that such resistance may have on malaria control is therefore of critical importance.

Insecticide surveillance efforts have led to a growing body of work on the links between pyrethroid resistance, bed nets, and malaria transmission. This work has largely focused on identifying the molecular markers responsible for resistance, estimating their prevalence, and understanding their effect on malaria transmission. With regard to the former, two major mechanisms of insecticide resistance have been

identified: metabolic resistance and target-site resistance. The former is not as well understood, but is thought to result from the over expression of certain enzymes that sequester or detoxify the insecticide before it reaches the target site. Target-site resistance (also known as knock-down resistance, or *kdr*), conversely, is the result of point mutations in the insecticide's sodium channel target, and allow insects to withstand exposure to insecticides without being knocked down (Ranson et al., 2011). Prevalence of these molecular markers varies considerably over space and time, while additional evidence suggests that increasing intervention coverage is exerting selection pressure on the gene expression and mutation processes responsible for resistance, and prevalence of resistance often increases as a result (Ranson et al., 2011; Czeher et al., 2008; Stump et al., 2004; Vulule et al., 1999; Protopopoff et al., 2008; Mathias et al., 2011; Padonou et al., 2012; Yadouleton et al., 2010; Djogbénou et al., 2011).

The epidemiological consequences of such increases are not fully understood. For example, in the highland province of Karuzi, Burundi, interventions with pyrethroid-based IRS and LLINs reduced transmission intensity in children by 90%, despite increasing levels of insecticide resistance following the interventions (Protopopoff et al., 2008, 2007). Additionally, results from village randomized controlled trials in a region of Cote d'Ivoire with confirmed pyrethroid resistance among *An. gambiae* suggest that nets treated with lambda-cyhalothrin maintained a protective efficacy of 56% against malaria (Henry et al., 2005). Other work, however, suggests that the rise of insecticide resistance may lead to reverses in recent reductions in malaria transmission. Results from a longitudinal study in Dielmo, Senegal, for example, suggest that malaria incidence rebounded to near pre-intervention levels 27-30 months after the introduction of deltamethrin-treated LLINs, during which time the prevalence of insecticide resistance markers increased from 8% to 48%, and 37% of the *An. gambiae* population was resistant to deltamethrin (Trape et al., 2011). Additionally, trials conducted in northern and southern Benin among regular users of bed nets indicate that treated nets provide little or no protection in the presence of insecticide resistance, whereas in areas where the mosquito population was susceptible to insecticides, sleeping under a treated net reduced odds of malaria infection by 66% (Asidi et al., 2012). Another trial comparing different intervention strategies (targeted LLIN distribution to children and pregnant women; universal LLIN distribution; and combining either targeted or universal LLIN distributions with IRS) in areas of moderate pyrethroid resistance in southern Benin indicates that neither universal distributions, nor distributions combined with IRS, reduced malaria incidence or slow the emergence of *kdr* mutations (Corbel et al., 2012).

Monitoring insecticide resistance and developing countermeasures, such as when to change insecticides used in nets, presents challenges. For example, recent recommendations suggest yearly monitoring efforts take place at a minimum of 12 sentinel sites, with additional surveillance efforts in other sites and at higher frequency recommended if resources permit (Kelly-Hope et al., 2008). However, insecticide resistance is heterogeneous even across short distances (Ranson et al., 2009). In a surveillance study conducted in 30 sites across Benin, for example, mortality following permethrin exposure varied from 25% to 100%, as did the *kdr* allele frequencies (0 - 0.91) in *An. gambiae ss* mosquitoes (Djogbénou et al., 2011). One recommendation that addresses this limitation is to sample the local vector population before any insecticide-based intervention begins, such that the proper insecticide(s) can be used (Ranson et al., 2009). Logistically, this is likely impossible given the scope of mass bed net distribution campaigns. It is also impossible to draw a representative sample of the mosquito population, greatly limiting understanding resistance across an entire country and at the population level. Additionally, while molecular surveillance largely focuses on monitoring genetic mutations that confer resistance, resistance itself comes in multiple forms, including behavioral resistance, in which mosquitoes change their behavior to avoid contact with nets (Ranson et al., 2011; Mathias et al., 2011; Russell et al., 2011). This form of resistance is not well understood, and it is unknown whether or not behavioral traits have a genetic basis, meaning that, to date, molecular surveillance cannot detect behavioral changes (Ranson et al., 2011). Further, epidemiological studies on the links between insecticide resistance and transmission have also taken place in a limited number of sites, preventing generalizations to the broader population. Additional strategies are therefore needed to complement current entomological surveillance efforts.

Two additional strategies have been proposed by the World Health Organization Pesticide Evaluation Scheme (WHOPES), prospective, longitudinal studies and retrospective, population-based cross-sectional surveys, both of which can provide an evidence base regarding the durability and insecticide activity of nets under field conditions (WHO et al., 2011). An important advantage of using cross-sectional surveys, moreover, is that they are often conducted for other purposes, and thus do not require the additional resources associated with running a large longitudinal study. To our knowledge, however, large, population-based surveys have yet to be used to understand the effects of different insecticide-treated nets on malaria transmission. Therefore, in this study, our aim is to assess the effect bed nets treated with different insecticides on the probability of malaria infection in under-5 children across sub-Saharan Africa using national household surveys conducted between 2009 and 2016.

3.2 Methods

3.2.1 Study Design and Data Sources

We obtained data on malaria and bed net use in children under 5 years of age from publicly available Demographic and Health Surveys (DHS) and Malaria Indicator Surveys (MIS) conducted between 2009 and 2016. Briefly, DHSs are two-stage cluster household surveys designed to provide nationally and sub-nationally representative estimates across a number of public health domain areas among children under the age of 5, and women and men of reproductive age. MISs follow the same survey design, but are more limited in scope than a full DHS. All surveys collect extensive demographic and socioeconomic data from participants. Further details for these surveys are available elsewhere (Hancioglu and Arnold, 2013; DHS, 2017).

3.2.2 Outcome and Exposure Measures

The primary outcome in this study is the malaria status of each child, which was determined by a rapid diagnostic test (RDT) using a heel- or finger-prick blood sample. The primary exposure is whether or not a child slept under a bed net the previous night, coded according the insecticide used in each net. We ascertained insecticide status using the bed net brand reported in the surveys. Each child sleeping under a net was coded as sleeping under a net treated with deltamethrin, permethrin, alphacypermethrin—the three most common insecticides—or other type of net. The latter category was used in cases where no brand information was available, or in cases where the insecticide could not be determined from the brand.

3.2.3 Potential Confounders

Potential confounders included in this analysis were also derived from DHS and MIS surveys. They include: age in years, sex, whether or not the child lives in an urban or rural community, and wall and roof construction materials. Wall construction materials were coded as natural (e.g. no walls or cane/palm/trunks), rudimentary (e.g. wood with mud), or finished (e.g. cement/tin). Similarly, roof construction material was coded as natural (e.g. no roof or thatch/palm leaf), rudimentary (e.g. wood planks), or finished (e.g. metal or shingles). These categories represent different levels of barriers to

mosquito entry, with natural construction materials providing little or no barrier to entry, and finished materials providing the greatest barrier. Finally, we considered the possibility that community-level coverage with each type of net may influence individual malaria risk.

3.2.4 Statistical Methods

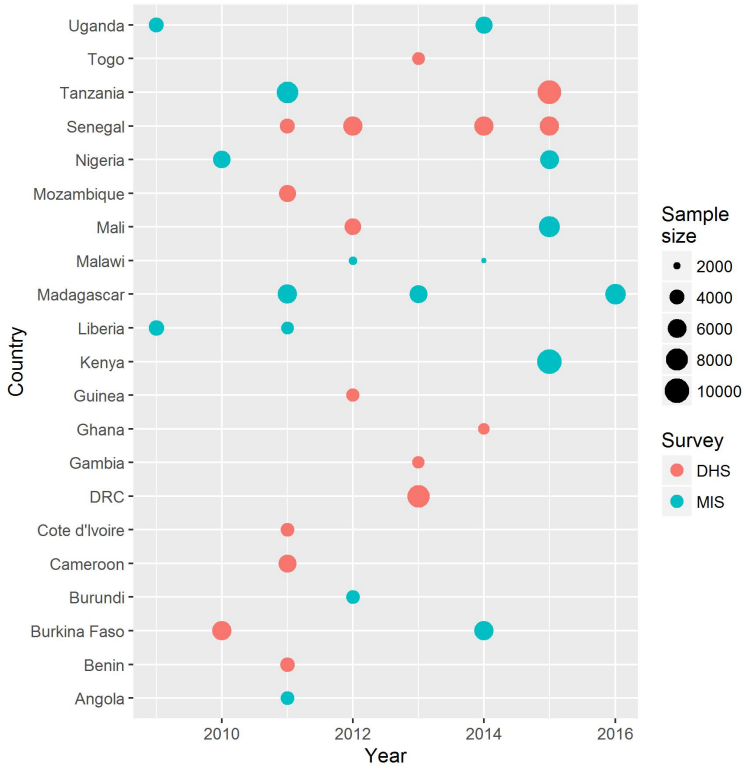
Our statistical analysis consisted of mixed effects logistic regression, and we stratify by country and year. All models incorporate survey sampling weights such that parameter estimates represent the expected effect in the under-5 population in each country and year in which a survey was conducted. We compared models including and excluding community-level coverage for each type of net using BIC. All analyses were done using R 3.3.1 and Stata 14.2.

3.3 Results

A total 168,118 children younger of 5 years of age from 33 DHS and MIS surveys conducted in 21 countries from 2009 to 2016 across sub-Saharan Africa were included in this analysis. Figure 3.9 shows a frequency plot of the surveys included in this study, together with sample sizes and type of survey.

A total of 92,698 (55%) children under 5 years of age slept under a bed net the previous night. Among them, 45,401 (49%) slept under a deltamethrin-treated net, with 31,041 (33.5%) and 4,983 (5.4%) sleeping under nets treated with permethrin and alphacypermethrin, respectively. A total of 11,273 (12.1%) children slept under a net with unknown brand or insecticide. Bed net insecticide use varied across countries and over time. Deltametherin-treated nets were the most common nets in use in 22 of the 33 surveys, while permethrin-treated nets were the most common in 9 surveys, although in some cases both nets were widely used (e.g. Kenya, Mali, Uganda). Alphacypermethrin was the most common insecticide used in the Malawi 2012 survey, although only 34% of nets were treated with this insecticide, while deltamethrin and permethrin each made up 27% of insecticide coverage. Among the nine countries with more than one survey, the primary insecticide in use changed across surveys in only two. In Burkina Faso, 51% of nets in use were treated with deltamethrin in 2010, decreasing to 25% in 2014, with a corresponding increase in coverage with permethrin-treated nets from 7% to 61%. Conversely, 50% of nets used in Madagascar in 2011 and 2013 were treated with permethrin, but in 2016, deltamethrin was the primary insecticide used, with coverage increasing from 14% to 82% between 2013 and 2016. In

Figure 3.9: Frequency plot of surveys included in the study

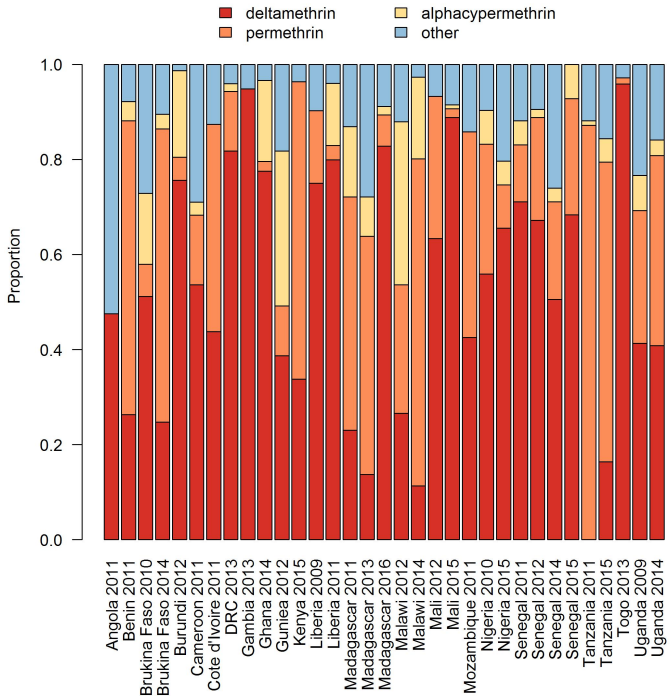


Malawi, coverage with both alphacypermethrin and deltamethrin declined from 2012 and 2014, while coverage with permethrin increased to 69%. Figure 3.10 shows the proportion of each type of insecticide in use by country and year.

Results from mixed effects logistic regression models indicate that excluding community-level bed net coverage (by insecticide) yielded better fit in 25 out of 33 models, and that the effects of sleeping under a net treated with a given insecticide did not meaningfully differ between the models considered. Table 3.5 shows the effects of sleeping under a bed net treated with different insecticides by country and year for both the full and reduced models, as well as the malaria prevalence at the time of the survey. Additionally, Figure 3.11 maps the odds ratio of each insecticide using the most recent survey, and therefore represents the current best estimate of the effectiveness of each insecticide at the population level.

The effect of using nets treated with different insecticides varied considerably across the different surveys and across different transmission contexts. In some countries, certain insecticides appear to confer little to no protection against infection. In Nigeria, for example, none of the three major insecticides

Figure 3.10: Proportion of nets treated with different insecticides by country and year



were associated with meaningful reductions in odds of infection, both in the 2010 and 2015 surveys, where malaria prevalence declined from 51.4% to 45.1%. Among those children sleeping under a deltamethrin-treated net in 2010, the odds of infection were only 6% lower than a child not sleeping under a bed net (OR 0.94, 95% UI 0.74 - 1.19). By 2015, this small effect had disappeared (OR 1.00, 95% UI 0.84 - 1.19). Among those sleeping under a net treated with permethrin, the odds of infection ranged from 0.90 (0.60 - 1.36) in 2010 to 1.28 (0.89 - 1.83) in 2015. For alphacypermethrin, a possible protective effect in 2010 (OR 0.79, 95% UI 0.36 - 1.72) disappeared in 2015 (OR 0.96, 95% UI 0.49 - 1.89), although considerable imprecision accompanies both estimates. Likewise, in Guinea, none of the three major insecticides in use were associated with reductions in odds of infection during the time of the survey, while in neighboring Liberia, the protective effect of deltamethrin—the primary insecticide in use in both the 2009 and 2011 surveys—appears to have disappeared from 2009 to 2011 (see Table 3.5).

Conversely, insecticides in other countries in West Africa had protective effects. In Cote d'Ivoire, which had an estimated malaria prevalence of 41.5% at the time of the survey in 2011, children sleeping under deltamethrin- and permethrin-treated nets had similar levels of protection (deltamethrin OR 0.73,

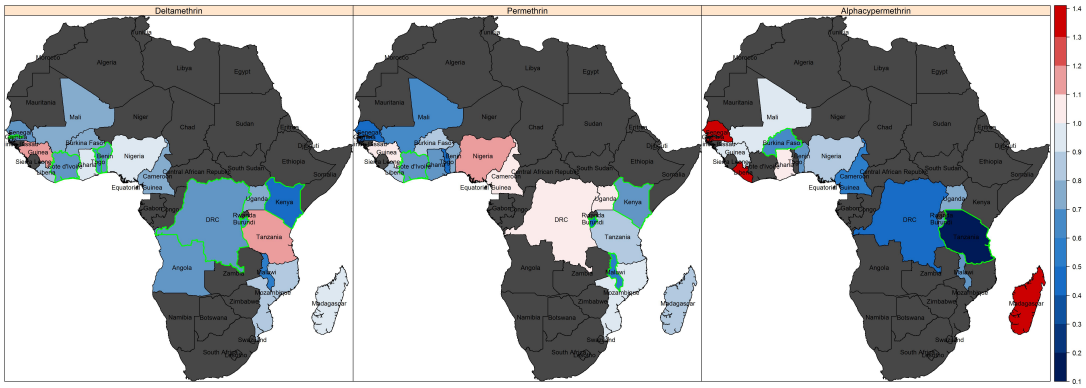
Table 3.5: Results from stratified logistic regression models

		Reduced Models			Full Models			
		Deltamethrin	Permethrin	Alphacypermethrin	Deltamethrin	Permethrin	Alphacypermethrin	
Country/Year	Malaria Prevalence (%)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	Favored Model (BIC)
Angola 2011	13.5	0.71 (0.47 - 1.11)	NA	NA	0.73 (0.47 - 1.12)	NA	NA	Reduced
Benin 2011	24.6	0.72 (0.51 - 0.99)	0.87 (0.68 - 1.12)	0.72 (0.38 - 1.38)	0.85 (0.60 - 1.20)	0.87 (0.68 - 1.11)	0.94 (0.49 - 1.81)	Full
Burkina Faso 2010	76.2	1.04 (0.87 - 1.25)	0.89 (0.61 - 1.29)	0.81 (0.61 - 1.09)	1.12 (0.92 - 1.35)	0.85 (0.58 - 1.24)	0.82 (0.61 - 1.10)	Reduced
Burkina Faso 2014	61.3	0.84 (0.68 - 1.05)	0.95 (0.76 - 1.18)	0.61 (0.38 - 0.96)	0.85 (0.68 - 1.06)	0.54 (0.43 - 0.67)	0.60 (0.38 - 0.94)	Reduced
Burundi 2012	12.8	0.67 (0.46 - 0.98)	0.40 (0.18 - 0.89)	0.79 (0.51 - 1.22)	0.75 (0.51 - 1.10)	0.40 (0.18 - 0.88)	0.80 (0.52 - 1.23)	Reduced
Cameroon 2011	30.0	0.82 (0.64 - 1.05)	1.13 (0.75 - 1.70)	0.58 (0.24 - 1.38)	0.80 (0.62 - 1.04)	1.08 (0.71 - 1.62)	0.56 (0.24 - 1.33)	Reduced
Cote d'Ivoire 2011	41.5	0.73 (0.57 - 0.94)	0.72 (0.54 - 0.95)	NA	0.64 (0.49 - 0.94)	0.74 (0.56 - 0.98)	NA	Full
DRC 2013	30.9	0.71 (0.59 - 0.85)	1.17 (0.78 - 1.76)	0.49 (0.18 - 1.33)	0.73 (0.61 - 0.88)	1.15 (0.75 - 1.75)	0.51 (0.19 - 1.39)	Reduced
Gambia 2013	2.6	0.49 (0.28 - 0.86)	NA	NA	0.45 (0.25 - 0.80)	NA	NA	Reduced
Ghana 2014	36.4	1.07 (0.83 - 1.39)	0.79 (0.38 - 1.64)	1.13 (0.71 - 1.81)	0.96 (0.74 - 1.24)	0.69 (0.33 - 1.46)	1.03 (0.64 - 1.66)	Reduced
Guinea 2012	46.8	1.24 (0.89 - 1.73)	1.13 (0.52 - 2.48)	1.02 (0.69 - 1.53)	1.11 (0.79 - 1.56)	1.07 (0.48 - 2.4)	1.02 (0.68 - 1.52)	Full
Kenya 2015	12.9	0.46 (0.35 - 0.59)	0.74 (0.60 - 0.92)	NA	0.46 (0.35 - 0.59)	0.76 (0.61 - 0.94)	NA	Full
Liberia 2009	36.6	0.87 (0.69 - 1.11)	0.70 (0.45 - 1.10)	NA	0.86 (0.68 -1.10)	0.69 (0.43 - 1.09)	NA	Full
Liberia 2011	44.7	0.97 (0.75 - 1.24)	0.92 (0.34 - 2.46)	1.39 (0.90 - 2.17)	0.94 (0.73 - 1.21)	0.99 (0.37 - 2.62)	1.47 (0.92 - 2.35)	Reduced
Madagascar 2011	8.7	0.85 (0.43 - 1.71)	0.61 (0.35 - 1.06)	0.93 (0.46 - 1.86)	0.77 (0.37 - 1.59)	0.54 (0.31 - 0.97)	0.71 (0.33 - 1.55)	Reduced
Madagascar 2013	10.0	1.05 (0.70 - 1.60)	1.03 (0.70 - 1.54)	1.06 (0.45 - 2.54)	0.89 (0.59 - 1.35)	0.82 (0.55 - 1.20)	0.90 (0.39 - 2.1)	Full
Madagascar 2016	5.1	1.01 (0.68 - 1.51)	0.96 (0.36 - 2.56)	2.76 (0.52 - 14.71)	0.87 (0.58 - 1.30)	0.88 (0.34 - 2.32)	2.56 (0.41-16.11)	Reduced
Mali 2012	47.0	0.92 (0.74 - 1.14)	1.08 (0.84 - 1.39)	NA	0.90 (0.72 - 1.11)	1.10 (0.85 - 1.42)	NA	Reduced
Mali 2015	23.4	0.85 (0.72 - 1.02)	0.65 (0.31 - 1.39)	1.07 (0.46 - 2.52)	0.85 (0.72 - 1.01)	0.67 (0.32 - 1.43)	1.09 (0.46 - 2.58)	Reduced
Mozambique 2011	38.2	0.95 (0.73 - 1.23)	1.02 (0.79 - 1.32)	NA	0.90 (0.69 - 1.18)	0.96 (0.74 - 1.24)	NA	Reduced
Malawi 2012	43.3	0.69 (0.40 - 1.17)	0.95 (0.64 - 1.43)	0.81 (0.56 - 1.18)	0.69 (0.40 - 1.18)	0.96 (0.65 - 1.44)	0.83 (0.57 - 1.21)	Reduced
Malawi 2014	37.2	0.58 (0.26 - 1.28)	0.50 (0.32 - 0.79)	0.71 (0.41 - 1.22)	0.58 (0.26 - 1.29)	0.48 (0.31 - 0.75)	0.76 (0.44 - 1.32)	Reduced
Nigeria 2010	51.4	0.94 (0.74 - 1.19)	0.90 (0.60 - 1.36)	0.79 (0.36 - 1.72)	0.94 (0.74 - 1.19)	0.87 (0.58 - 1.32)	0.82 (0.38 - 1.79)	Reduced
Nigeria 2015	45.1	1.00 (0.84 - 1.19)	1.28 (0.89 - 1.83)	0.96 (0.49 - 1.89)	0.98 (0.82 - 1.16)	1.23 (0.86 - 1.77)	1.00 (0.51 - 1.95)	Reduced
Senegal 2011	2.7	0.80 (0.50 - 1.30)	0.57 (0.20 - 1.62)	0.75 (0.17 - 3.42)	0.73 (0.44 - 1.21)	0.53 (0.17 - 1.63)	0.88 (0.22 - 3.48)	Full
Senegal 2012	3.3	0.57 (0.35 - 0.94)	0.53 (0.28 - 1.01)	0.65 (0.07 - 5.64)	0.55 (0.34 - 0.91)	0.70 (0.36 - 1.33)	1.00 (0.11 - 9.34)	Reduced
Senegal 2014	1.1	0.59 (0.25 - 1.39)	0.94 (0.34 - 2.55)	1.04 (0.20 - 5.51)	0.68 (0.25 - 1.84)	0.93 (0.29 - 2.94)	1.70 (0.73 - 3.97)	Reduced
Senegal 2015	0.6	0.74 (0.29 - 1.86)	0.44 (0.14 - 1.40)	1.83 (0.36 - 9.19)	0.63 (0.26 - 1.52)	0.45 (0.15 - 1.34)	0.81 (0.22 - 3.04)	Reduced
Tanzania 2011	9.4	NA	0.99 (0.76 - 1.29)	0.95 (0.28 - 3.2)	NA	0.95 (0.73 - 1.24)	0.93 (0.28 - 3.10)	Full
Tanzania 2015	14.4	1.22 (0.87 - 1.71)	1.00 (0.79 - 1.26)	0.19 (0.04 - 0.82)	1.07 (0.77 - 1.49)	0.88 (0.70 - 1.11)	0.24 (0.05 - 1.11)	Reduced
Togo 2013	38.2	0.70 (0.57 - 0.85)	0.44 (0.09 - 2.06)	NA	0.65 (0.54 - 0.80)	0.41 (0.09 - 1.90)	NA	Reduced
Uganda 2009	51.9	0.60 (0.45 - 0.80)	0.59 (0.40 - 0.86)	0.44 (0.26 - 0.75)	0.59 (0.45 - 0.79)	0.60 (0.41 - 0.88)	0.43 (0.25 - 0.73)	Reduced
Uganda 2014	29.9	0.86 (0.66 - 1.13)	1.12 (0.84 - 1.50)	0.84 (0.48 - 1.47)	0.86 (0.65 - 1.13)	1.05 (0.79 - 1.40)	0.86 (0.48 - 1.54)	Reduced

Notes: n = 168,118. All models adjust for child age (in years), sex, urban/rural status, and household wall and roof construction materials. The full model controls for community coverage with each type of net. NA indicates that the insecticide was either not in use, or present in such small numbers that there was no variability between net use and malaria outcome. All odds ratios are comparisons to a child not sleeping under a bed net of any kind.

95% UI 0.57 - 0.94; permethrin OR 0.72, 95% UI 0.54 - 0.95). In Benin in 2011, against a background malaria prevalence of 24.6%, deltamethrin was similarly protective (OR 0.72, 95% UI 0.51 - 0.99), although permethrin-treated nets were less so (OR 0.87, 95% UI 0.68 - 1.12).

Figure 3.11: Odds of malaria infection by insecticide use from each country's most recent survey



Green borders identify those countries in which a protective effect was observed at the 0.05 level.

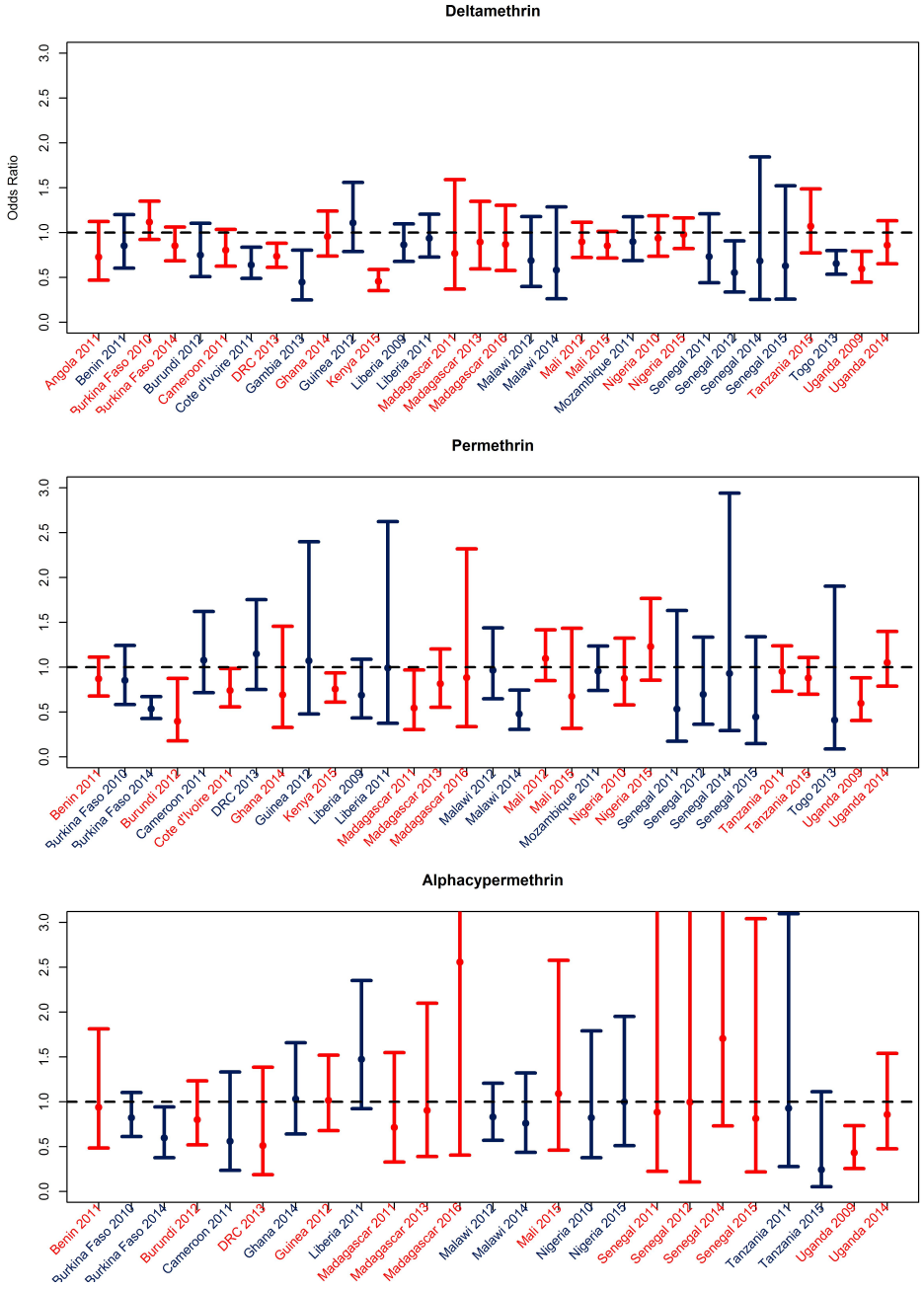
Similar patterns appear in Central and East Africa. In Uganda, all three insecticides appear to be losing efficacy based on the 2009 and 2014 surveys, particularly with regard to permethrin. This effect occurs against a backdrop of increased LLIN use among under-5 children, with LLIN coverage in this population going from 32% to 42% between surveys, and prevalence that decreased from 52% in 2009 to 30% in 2014. In neighboring Democratic Republic of Congo (DRC), where prevalence was 31% in 2013, deltamethrin-treated nets are associated with 29% lower odds of infection (OR 0.71, 95% UI 0.59 - 0.85), whereas permethrin-treated nets do not appear to confer any protective benefit (OR 1.17, 95% UI 0.78 - 1.76), while in nearby Burundi, both deltamethrin and permethrin are associated with considerably lower odds of infection (deltamethrin OR 0.67, 95% UI 0.46 - 0.98; permethrin OR 0.40, 95% UI 0.18 - 0.89).

More broadly, nets treated with different insecticides appear to have very limited efficacy across large parts of sub-Saharan Africa, although care must be taken in interpreting many these estimates, since in many surveys only a small number of individuals slept under a net treated with a given insecticide. This is particularly relevant for alphacypermethrin. In some cases, background malaria transmission levels can further contribute to this uncertainty. For example, Senegal's malaria prevalence was very low across all four surveys, ranging from 0.6% in 2015 to 3.3% in 2012, with a total of 61 and 254 cases, respectively. As such, imprecision for the estimates in Senegal likely results from very few cases of

malaria as well. The varying degrees of imprecision can be seen in Figure 3.12, which plots the odds of malaria infection according to insecticide use for all surveys, together with 95% uncertainty intervals.

Nevertheless, for a number of countries, there are clear indications as to which insecticides are associated with reduced odds of malaria infection, and which are not. Specifically, deltamethrin-treated nets appear to reduce odds of infection in Angola, Benin, Burkina Faso (2014 survey), Burundi, Cameroon, Cote dIvoire, DRC, Gambia, Kenya, Malawi, Mali (2015), and Togo. Conversely, in Ghana, Guinea, Madagascar, Mozambique, Nigeria, Tanzania, and Uganda (2014), there does not appear to be a protective effect. Similarly, permethrin appears to be ineffective across a number of countries, namely: Benin, Burkina Faso, Cameroon, DRC, Liberia, Mozambique, Nigeria, Tanzania, and Uganda (2014). In Burundi, Cote d'Ivoire, Kenya, and Malawi (2014), the nets were associated with 26 - 60% reductions in odds of infection. Finally, the role of alphacypermethrin-treated nets is much less clear owing to imprecision, although protective effects are evident in Burkina Faso, Tanzania, and Uganda (2009) (see Figure 3.12 and Table 3.5).

Figure 3.12: Odds of malaria infection by insecticide use



The alternating colors used here are to help facilitate visualization, with each change in color corresponding to a change in country.

3.4 Discussion

The spread of insecticide resistance threatens to undermine a pillar of malaria control, and in many parts of sub-Saharan Africa, countries likely need to switch insecticides. Determining which insecticide to use, and when to switch, requires a comprehensive understanding of insecticide resistance and its consequences on bed net effectiveness. Further, since bed net distribution decisions are made at the country level, it is important to understand the consequences of insecticide resistance country-wide and at the population level. However, current insecticide resistance monitoring efforts cannot be generalized to a country's population. Additionally, many of these monitoring efforts focus on molecular markers of resistance, which do not necessarily translate to reduced efficacy. They are also unable to address potential changes in mosquito behavior.

To address some of these limitations, we leveraged large, population-based surveys DHS and MIS surveys conducted across sub-Saharan Africa between 2009 and 2016. This approach has a number of strengths. First, DHS and MIS surveys provide detailed data on malaria status, bed net use, and bed net characteristics, together with important demographic and socioeconomic characteristics, for one of the populations at highest risk of malaria—children under 5 years of age. Second, the data are publicly available, such that researchers or government agencies can readily access the data to monitor bed net effectiveness in the presence of observed insecticide resistance. Further, while our analysis provides country-level estimates, the surveys are representative regionally within countries as well, facilitating a more detailed understanding within a given country if necessary.

Important limitations exist as well. First, bed net brand reporting was not consistent across all surveys. For example, the permanet brand was not reported in a manner that allowed us to distinguish between Permanet 2.0 and Permanet 3.0, the latter of which incorporates piperonyl butoxide (PBO), a synergistic compound that works to inhibit the metabolic enzyme that slows uptake of the insecticide in a mosquito (Moores and Bingham, 2005; Bingham et al., 2011). Additionally, bed net use was assessed based on whether or not a child slept under a net the night before, which is subject to reporting bias. Furthermore, considerable imprecision accompanied many of our estimates, owing to a small number of children sleeping under nets treated with certain insecticides, especially in low malaria prevalence settings (e.g. Senegal). We also lack data on the adult population in each of these countries, precluding

any inferences on the large majority of a country's residents. Finally, given the cross-sectional nature of these surveys, we cannot discount the role of residual confounding.

Despite these limitations, to our knowledge this study is the first to assess the effect of bed nets treated with different insecticides on malaria transmission using population-based data. Our results provide evidence against and in favor of different insecticides in different countries, evidence that can be used to complement ongoing entomological surveillance efforts within countries. In Nigeria, for example, a number of studies have documented target-site and metabolic resistance to permethrin and deltamethrin in a range of settings (Awolola et al., 2005, 2009; Kristan et al., 2003; Djouaka et al., 2008). These results are consistent with our findings here, in which neither net appeared protective against malaria. In a study conducted from May to September 2011 in a southern area of Cote d'Ivoire, high levels of resistance to both permethrin and deltamethrin were reported. Conversely, in the 2011 Cote d'Ivoire DHS, bed nets treated with both insecticides were associated with lower odds of malaria infection, suggesting that insecticide resistance may be more localized than widespread (Edi et al., 2012).

As bed net distributions continue to take place, maintaining the efficacy will be of paramount importance. A key component to this effort will be monitoring the insecticide resistance landscape. DHS and MIS surveys, which are conducted regularly within countries, can serve as a key tool in this effort.

3.5 From Human Interventions to Human Movement

While the previous two chapters have addressed challenges to malaria control that largely involve the potential reaction of a mosquito population to human interventions such as agriculture or bed net use, and the effects of such interventions on people themselves, this next chapter focuses on another challenge to malaria control, namely: how does it spread? The process by which a disease spreads throughout a country is largely governed by human movement, and as such the fundamental task in understanding how parasites move is by understanding how people move. Understanding this process, however, is complicated by the lack of data on human movement, and we thus turn to molecular and population genetic tools in an effort to make inferences about it. In particular, as we will see, while data on human movement would allow us to infer parasite movement, the data environment we work in requires us to make inferences about human movement by looking at parasite movement.

CHAPTER 4: EXPLORING SUBPOPULATION STRUCTURE AND GENE FLOW OF MALARIA PARASITES

4.1 Introduction

How does malaria move from place to place? Over short distances, parasite transmission occurs by way of mosquitoes moving through a local population, biting different individuals at different times. Mosquito movement is limited to their local environment, however, suggesting that the spread over larger geographic distances occurs as a result of the movement of individuals (Kaufmann and Briegel, 2004). Understanding such movement is important in the context of malaria control. For example, the movement of infected individuals can reintroduce malaria to settings where it had been eliminated, or lead to the spread of drug-resistant parasites or other parasite population of public health concern (Ferguson et al., 2010; Anderson and Roper, 2005; Roper et al., 2004; Lynch and Roper, 2011).

An important challenge to understanding how malaria spreads, however, is that data on human movement are generally unavailable, and either consist of coarse, census-based estimates of migration between countries, or are constructed from mobile phone records (Tatem and Smith, 2010; Tatem et al., 2014; Ruktanonchai et al., 2016). The former of the two does not allow for inferences about how malaria may spread within a country (an important consideration for national malaria control programs), while the latter will yield a biased view of human movement in settings where mobile phone use is not widespread (such as the DRC), or differs by geography or demographics. Alternatively, population movements can be inferred by measuring genetic markers and estimating associations across geography.

Work to understand parasite population structure in the DRC is limited, but growing. In a survey of 166 parasites drawn from the Demographic and Health Survey (DHS) conducted in the DRC in 2007, for example, 44% of the parasites harbored the haplotype associated with chloroquine failure, although no geographic clustering was apparent (Antonia et al., 2014). A study investigating clustering of sulfadoxine resistance using a sample of 151 parasites from the 2007 DHS found mutations associated with drug resistance were clustered in the northeast of the country, and called for future work to identify

mechanisms by which this parasite population may spread (Taylor et al., 2013). More recently, work to understand the distribution of the parasite population harboring the *pfhrp2* gene deletion found the parasite population was clustered in eastern DRC and in Kinshasa, and was differentiated from parasites that did not harbor the deletion (Parr et al., 2016).

Understanding how these parasites may spread, however, requires understanding underlying processes such as gene flow, which in the case of malaria occurs as a result of human movement. Unfortunately, studies such as those above rely on genetic markers that are under selection, which can yield a distorted view of this process (Anderson et al., 1999; Holderegger et al., 2006). Conversely, one study from the DRC that did use neutral markers hypothesized that the Congo River serves as a corridor through which parasites move, although results did not show clear evidence in support of this hypothesis, with some samples located both near and far from the river appearing to be genetically related, and vice versa (Carrel et al., 2015). One important limitation of this and other studies, however, is that, while they all draw on population-based survey data, they nevertheless rely on a small number of samples obtained from the larger survey, and these are sometimes drawn from a small number of sites. As a result, to date there has not yet been a study that attempts to understand malaria gene flow using data both from across the country, and from the full parasite population available from the surveys. This study attempts to overcome these limitations by analyzing 28 neutral microsatellite markers on malaria parasites infecting 608 children over 5 years of age living in 301 survey clusters sampled as part of the 2013-14 Demographic and Health Survey (DHS) conducted in the DRC.

4.2 Methods

4.2.1 Study Population

Samples were collected from children between the ages of 5 and 15 as part of the 2013-2014 DRC DHS. Heel- or finger-prick blood samples from each child were taken and analyzed by rapid-diagnostic test (RDT), as well as by light microscopy, and then used to prepare dried blood spots (DBS), which were then shipped to the University of North Carolina-Chapel Hill (UNC) (Doctor et al., 2016). Malaria DNA was extracted from DBS, and a real-time PCR assay with a limit of detection of 100 parasites/μL was then used to identify infections with *P. falciparum* malaria using previous published protocols (Doctor et al., 2016; Plowe et al., 1995). Over-5 children were used as this study was also used to develop and test

a new lab protocol (see next section), and over-5 children do not contain survey information other than the survey cluster from which they were sampled, and are thus of greater use as an experimental dataset. These children were identified by subsetting the full DHS dataset to include only those samples with a blood sample (i.e. all children), and then further subsetting those who did not have an age specified (indicating that they were children over 5 years of age).

4.2.2 Microsatellite Analysis

Among those children with PCR-confirmed infections, 28 neutral microsatellite loci located across the *P. falciparum* genome were targeted for amplification using Molecular Inversion Probes (MIPs). MIPs have recently been designed to capture and re-sequence targeted regions of the malaria genome, and allow for cost-effective sequencing of 10s to 100s of candidate markers in one reaction. For each microsatellite locus, we defined each unique nucleotide length detected from the MIPs assay as an allele for use in population genetic analyses (Hathaway et al., *in preparation*).

4.2.3 Population Genetic Analyses

To explore possible genetic differentiation in the malaria parasite population, we proceeded in two ways. First, we sought to ascertain whether or not the population is made up of multiple, partially isolated subpopulations using Principal Components Analysis (PCA) and the Bayesian mixture model approach developed by Pritchard and colleagues (Pritchard et al., 2000). Second, we explore differentiation under the assumption of isolation-by-distance, in which genetic dissimilarity among parasites increases as a function of the geographic distance between them. Inference was based Mantel tests, a global measure of isolation-by-distance (IBD), and by G_{ST}, a measure of genetic differentiation between two pre-defined populations (Mantel, 1967; Nei, 1973). We use the latter because global measures of IBD may mask underlying processes of gene flow or barriers to it, while G_{ST}-based methods allows for exploring where potential gene flow (or barriers to it) may exist.

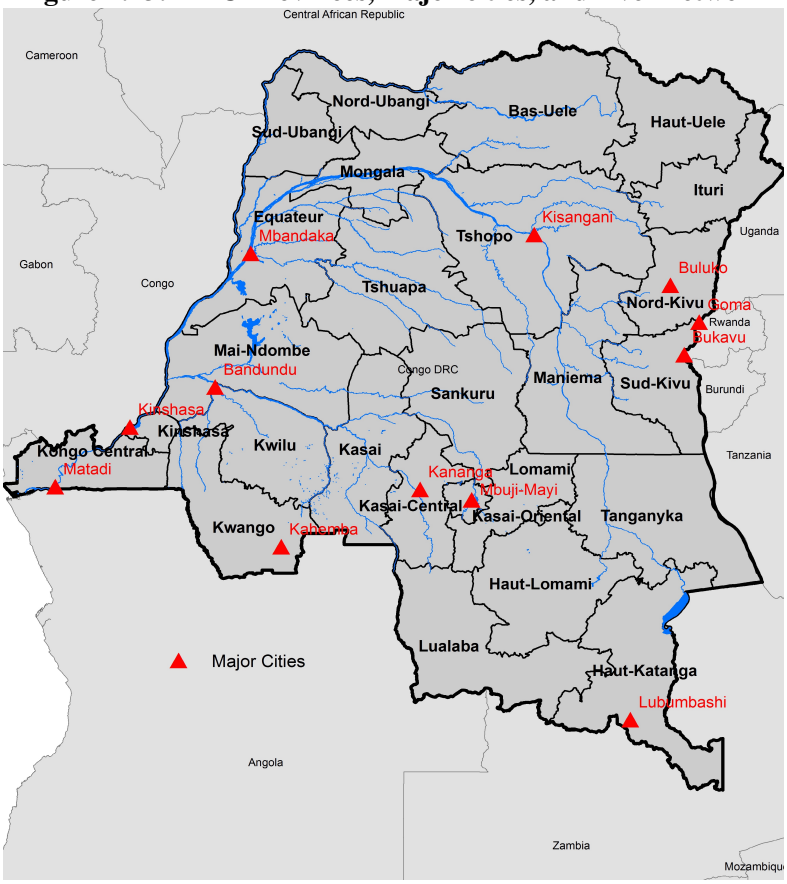
The G_{ST}-based analysis requires that sub-populations be defined a priori, and we choose these sub-populations to reflect hypothesized regions of population movement, as follows. The DRC has a number of urban centers scattered throughout the country, with the largest being Kinshasa (population 10 million), Lubumbashi (1.8 million), and Kisangani (1.6 million), Goma (1 million), and Bukavu (870,000). Economic migration occurs largely between these large urban centers and the rural areas surrounding

them. As such, we compare malaria parasites between: Kinshasa city and surrounding rural Kinshasa, Kongo Central, Kwilu, and Mai-Ndombe provinces; Lubumbashi and Tanganyka, Haut-Lomani, and rural Haut-Katanga provinces; Kisangani and rural Ituri, Bas-U'ele, and Haut-U'ele provinces; Goma and rural North Kivu; Bukavu and rural South Kivu.

While the population assignments above reflect prior beliefs about routes of economic migration and/or circulation regionally within the DRC, we extend the above analysis to consider long-distance population movement, which may occur between major urban centers or along the Congo River, which originates along the DRC-Zambia border and winds its way north across the equator before turning south towards Kinshasa and further out into the Atlantic Ocean. We thus make comparisons between Kinshasa and Tshopo province, which are connected via the Congo River and where Kisangani is located. We further compare parasite populations from Kinshasa and both North and South Kivu (where Goma and Bukavu are located), which may be connected by the river or other network. Finally, we compare two unconnected sites—Lualaba and Haut U'ele and Itrui provinces—which are on opposite sides of the country and not connected via the river network, and may therefore exhibit a pattern of differentiation. Figure 4.13 presents a provincial map together with the locations of DRC's major cities and river network.

Finally, inferences from both Mantel- and G_{ST}-based methods were obtained using randomization-based procedures consisting 10,000 simulations. PCA, Mantel tests, and G_{ST} tests were conducted in R 3.3.1, while Bayesian estimation of the number of sub-populations was done using MavericK 1.0 (Verity and Nichols, 2015).

Figure 4.13: DRC Provinces, major cities, and river network



4.3 Results

We identified 608 children over the age of 5 living in 301 survey clusters who were infected with *P. falciparum* malaria by PCR, out of a total of 1,622 over-5 children sampled in the DHS. Among these 608 children, 1,199 infections from 496 (82%) children living in 262 survey clusters were successfully amplified across multiple loci using MIPs. Samples failing to amplify across all loci had lower parasitemias (i.e. parasite density), on average (difference in mean C_T = 2.4, p $<$ 0.01). As a result, for 39 (13%) survey clusters, no data from MIPs was available. Figure 4.14 shows a map with the locations of the original 608 children, and whether or not MIPs data was obtained from all, some, or none of the kids living in those locations.

Results from PCA and MavericK-based analyses suggest that the parasite population is not meaningfully structured into distinct sub-populations in the DRC, although individuals from two communities (98 and 292) do appear to be different from the population as a whole. Figure 4.15 shows pairwise plots of

the first three principal components from PCA, showing these outliers against a backdrop of no distinct grouping. Figure 4.16 maps these outliers, which occur along the southern border with Angola, and in Tanganyka province near the Tanzanian border. Figure 4.17 plots estimates of the model evidence from the Bayesian analysis, which shows the posterior probability that the overall population is made up of between one and five sub-populations (denoted K). Overwhelmingly, the evidence supports a single population (i.e. K=1 has the highest posterior probability).

Results from Mantel tests showed weak evidence of isolation by distance, with increasing genetic dissimilarity associated with increasing distance. Figure 4.18 plots the observed correlation between genetic and geographic distance against 10,000 permutations of the genetic and geographic distance matrices, together with a plot of genetic versus geographic distance.

Restricting analyses to between major population centers and surrounding areas provides weak evidence of possible gene flow. For example, comparing the capital city of Kinshasa to its rural surrounds showed that the observed G_{ST} (0.016) fell in the lower tail of the distribution of values generated by randomly assigning parasites to the two populations, but that it was still within the range of what would be expected in the absence of strong gene flow. Similar results held for the other comparisons of interest, although it should be noted that the comparison between Goma and surrounding areas was not possible owing to lack of data from Goma. Figure 4.19 shows results from G_{ST}-based analyses for Kinshasa, Kisangani, Lubumbashi, and Bukavu.

Interestingly, sites that were increasingly disconnected geographically yielded similar results to those presented above. For example, comparisons between Kinshasa and Tshopo province, between Kinshasa and North and South Kivu, and between Lualaba and Haut U'ele and Ituri all showed no strong evidence of gene flow or differentiation. Figure 4.20 shows these comparisons.

Figure 4.14: Spatial coverage of MIPs panel

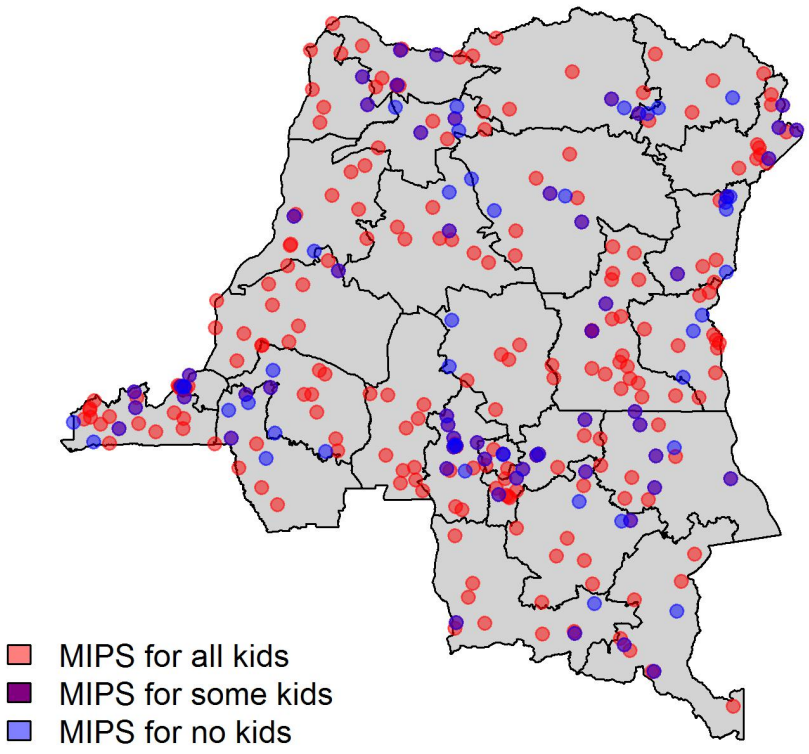
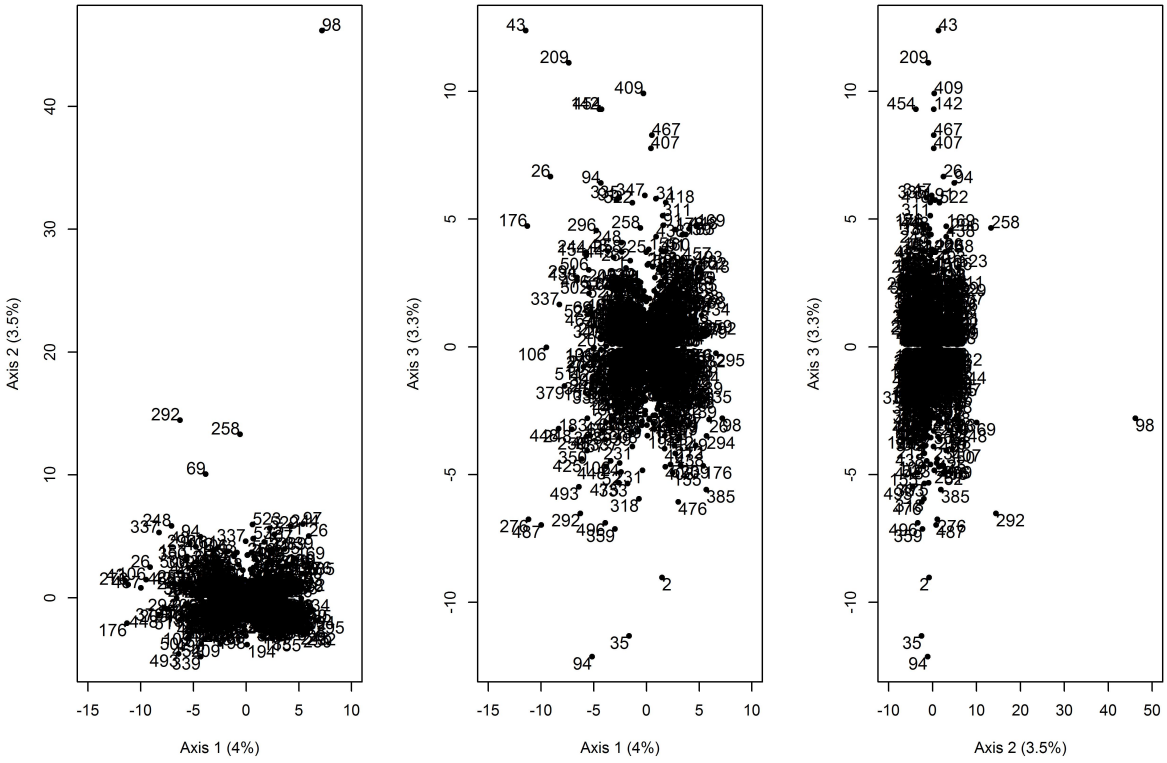


Figure 4.15: Pairwise plots of first 3 PCA axes



Point labels represent the community ID from which study participants originated.

Figure 4.16: PCA Outliers

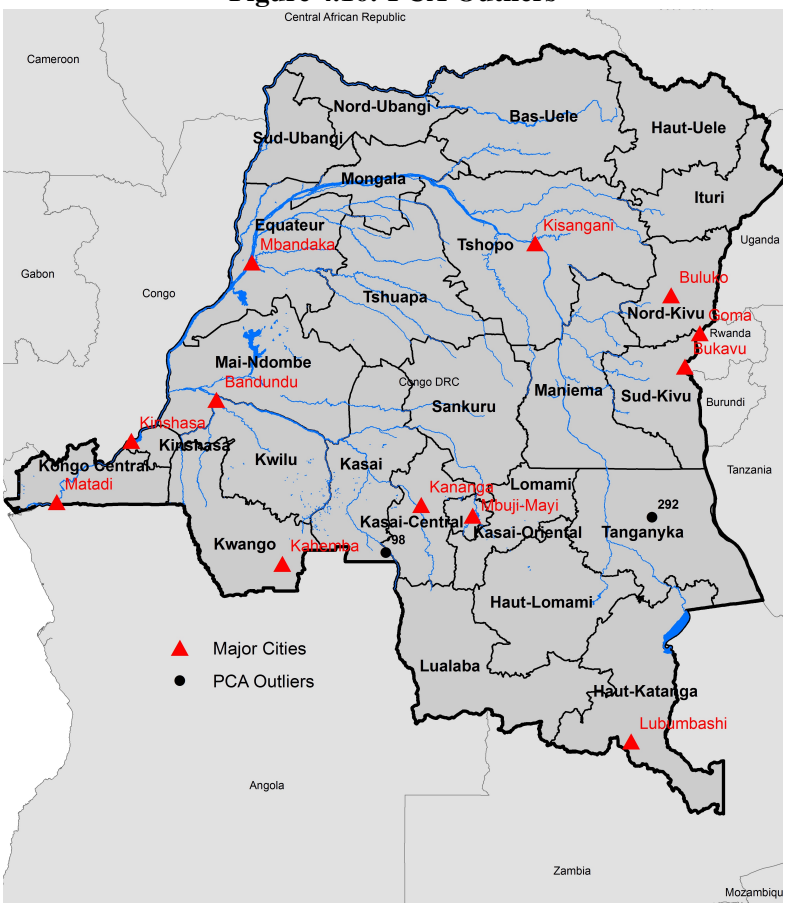
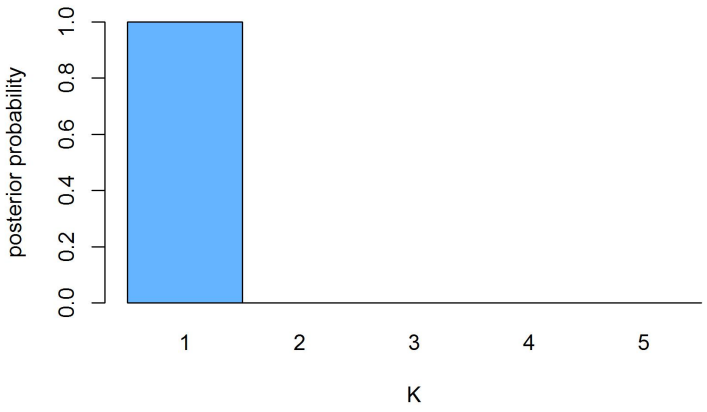


Figure 4.17: Evidence from Bayesian Mixture Model



Plot shows clear evidence in favor of a single population (K=1).

Figure 4.18: Mantel test of isolation by distance

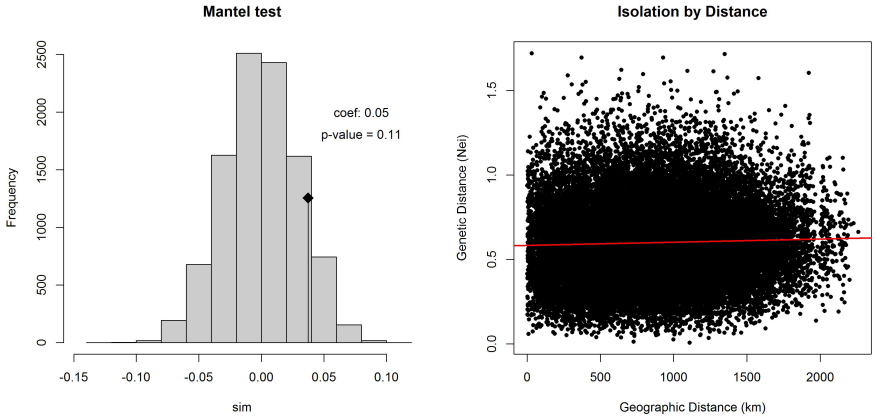
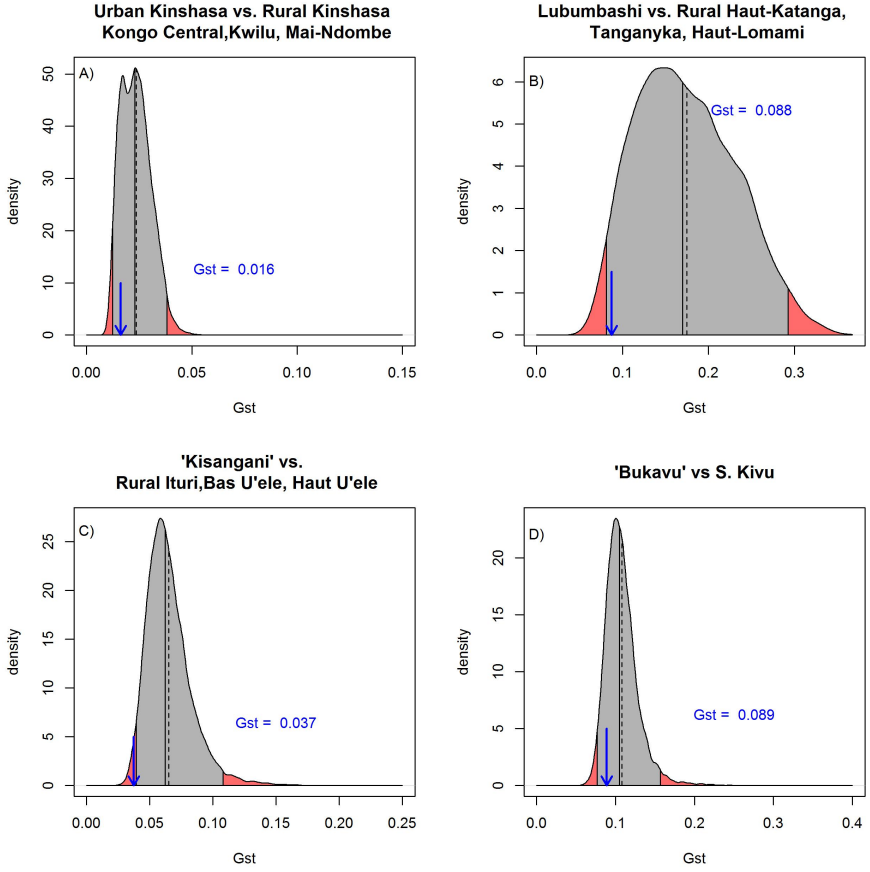
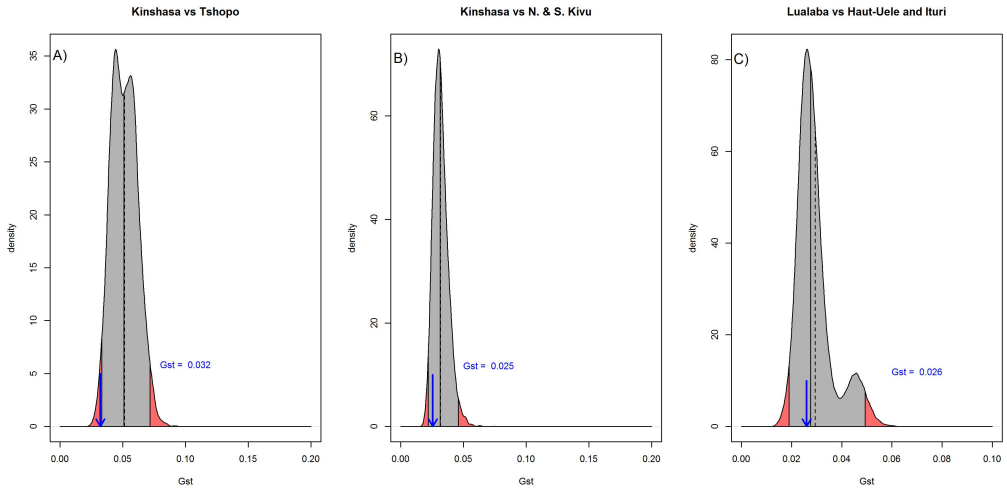


Figure 4.19: G_{ST}-based tests of population structure between urban centers and surrounding regions



Blue arrow = observed G_{ST} statistic.

Figure 4.20: G_{ST}-based tests of population structure between increasingly disconnected regions



Blue arrow = observed G_{ST} statistic.

4.4 Discussion

To our knowledge, this is the first study to attempt to use a population-based sample of the malaria parasite population to understand its sub-population structure and possible mechanisms of gene flow in the DRC. Our results indicate that the malaria population is best characterized as single population that exhibits weak isolation-by-distance, and that no barriers appear present based on results from permutation-based G_{ST} tests. One possible explanation for weak evidence of gene flow would be the country's poor infrastructure. However, the lack of evidence of a clear barrier between populations that are completely disconnected from the transportation network (e.g. Lualaba and Haut-U'ele/Ituri) argues against such a mechanism. Rather, it may be the case that transmission in the DRC is too high to measure gene flow, with high levels of heterozygosity and the accompanying high genetic variability meaning that noise dominates any potential signal in the data. Alternatively, despite the poor infrastructure and low mobility, the parasite population may spread across the country at a rate faster than the mutations occur at the 28 loci under investigation here, thereby hindering the detection of barriers to gene flow.

Importantly, this study has a number of limitations. First, there is evidence of bias favoring amplification of higher-density infections, limiting the generalizability of the sample to the broader parasite population. This bias also resulted in considerable missing data, in which 18% of the original sample could not be analyzed. Second, this study relied on data sampled from over-5 children, and as such this parasite population may not be comparable to the broader parasite population, such as children under 5 or adults, although it is unclear why this may be. Finally, molecular markers may not be adequate instruments to infer population movement, and thereby understand malaria gene flow or barriers to it, owing to the high levels of genetic variability noted above. Nevertheless, the MIPs protocol used here allows for genotyping of malaria parasites at scale, both in terms of number of samples and number of loci. As such, it may still prove useful for other objectives in molecular surveillance, for example of known drug-resistant or other mutations of public health importance.

CHAPTER 5: CONCLUSION

5.1 Summary of Aims 1-3

Malaria transmission is characterized by heterogeneity, which, as noted in the introduction to this dissertation, is frequently described, but not fully understood (Bousema et al., 2010). In this dissertation, I have attempted to focus on key gaps in our knowledge surrounding this heterogeneity, gaps that represent current challenges to malaria control. In Aim 1, I addressed our need to understand the underlying ecology of malaria transmission in the DRC, a prerequisite for malaria elimination (Ferguson et al., 2010). Specifically, I focused on one feature of this ecology—agriculture—and sought to characterize its effect on both the vector population and children under 5 years of age. The results of this work indicate that increases in agriculture in a community increases the probability of malaria infection, and that increased indoor biting among *An. gambiae* mosquitoes may be the mechanism behind the increased risk. Importantly, this work provided further evidence showing the bed nets treated with permethrin are ineffective at reducing malaria risk (Levitz et al., 2017). As a result, if ecological changes such as agricultural expansion lead *An. gambiae*—the most pernicious malaria vector—to bite indoors more frequently, then monitoring the continued success of bed net interventions takes on further salience. Additionally, we observed a vector—*An. paludis*—that is present in high abundance, although its role in transmission is largely unknown.

Understanding the effectiveness of bed nets was the focus of Aim 2 of this dissertation, with particular interest in the potential for insecticide resistance to render bed nets an ineffective tool for reducing transmission. Further, given that bed net interventions are a cornerstone of malaria control, and that there is only one class of insecticides—pyrethroids—approved for use in bed nets, I extended the scope of this dissertation to focus on sub-Saharan Africa, rather than just the DRC. In doing so, I again turned to population-based surveys, and used information about the brand of bed net survey participants slept under to identify the insecticides those nets were treated with, and estimate their effects across 21 countries at various times from 2009 to 2016. Results showed that many countries need to discontinue

use of certain insecticides (such as permethrin in the DRC), while in other countries (i.e. Nigeria), none of the insecticides in use appear to be reducing transmission. Importantly, the evidence in favor (or against) these conclusions varies by country, either owing to the background transmission dynamics, low numbers of children sleeping under nets treated with certain insecticides, or both.

This aim also demonstrated a previously unidentified utility of DHS and MIS surveys, namely their ability to help monitor bed net effectiveness at the population level. This is important because it is impossible to draw a representative sample of the mosquito population, which immediately hinders any efforts to make country-level decisions about bed net interventions based on evidence from entomological surveillance of insecticide resistance. Additionally, current efforts to understand insecticide resistance largely focus on characterizing the prevalence of molecular markers associated with insecticide resistance. As noted, however, use of molecular markers cannot fully describe insecticide resistance, since genotypes (i.e. presence of certain markers) do not necessarily translate to phenotypes (failure of insecticide to kill the mosquito). To be sure, this is not to say that the use of DHS or MIS surveys should be used in place of entomological surveillance efforts. Rather, the two should be considered as complimentary efforts with the same end goal in mind. Indeed, one approach that may hold particular promise is to design entomological surveillance to correspond to the administration of DHS or MIS surveys, such that we can update our knowledge of the insecticide resistance landscape at the same time that we update our knowledge about the effects of different insecticides.

Finally, Aim 3 sought to address another challenge to malaria control: understanding how malaria spreads from place to place. Over long distances, such as between rural villages and urban centers, such spread occurs as a result of human movement, but data on human movement is generally unavailable in a reliable way. Inferences about this process, however, can often be inferred from the parasite population itself, and we thus used molecular and population genetic methods to attempt to understand if the parasite population was structured in any way, and what possible corridors or barriers to movement might give rise to this structure. We considered the Congo River as a potential corridor, and further explored the possiblity that malaria may spread between urban centers and surrounding rural areas as a result of economic migration. We further attempted to identify potential barriers to the spread of malaria by comparing parasite populations that had no apparent connection to the transportation network or other geography, and were on opposite sides of the country. The results from this aim were largely inconclusive. For example, Principal Components Analysis and results from a Bayesian mixture model showed no

signs that the parasite population was structured into distinct sub-populations, although there were two outliers near the border. Additionally, while there was weak evidence in support of isolation by distance across the DRC, analyses using permutation-based G_{ST}'s did not show strong evidence in favor of gene flow where it was expected, nor did we observe strong evidence of barriers to gene flow where it was expected. One reason for this may be that the high levels of malaria transmission in the country lead to high levels of heterozygosity in the parasite population. As a result, such high genetic variability leads to noise that overwhelms any potential signal. This is not to say that the molecular approaches employed have no future in understanding malaria transmission, as one key strength of the molecular methods used is that it allows for parasite genotyping at scale, and as such these methods can be readily employed to monitor molecular markers of public health importance, such as those associated with resistance to different anti-malarial drugs.

5.2 Future Work

This dissertation represents another contribution to a growing body of work in the DRC, work that has been ongoing since the first DHS was conducted in 2007. Before that time, our knowledge of malaria prevalence across the country was based on a map from 1953. When the 2007 DHS was done, the leftover dried blood spots allowed those who came before me to test adults aged 15-59 for malaria and update that map for the first time in over 50 years, and do so using a population-based sample. Working with these data has allowed us to begin to understand, among other things, the distribution of molecular markers of drug resistance in the DRC, the burden of malaria in pregnancy, and the structure of the parasite population (Antonia et al., 2014; Taylor et al., 2013, 2011; Carrel et al., 2015). This dissertation, then, has served to update our knowledge of malaria transmission, both by using the latest round of the DRC DHS, but also by addressing other gaps in our knowledge, primarily in the DRC, but beyond it as well. Nevertheless, important gaps remain. For example, none of this dissertation focused on malaria in an adult population. Such information is not likely to be collected any time soon across the malaria-endemic world, since the malaria rapid diagnostic testing done in DHS and MIS surveys focuses only on the under-5 population. In the DRC, however, adults have recently been tested for malaria by PCR using the leftover dried blood spots used for HIV serology, meaning that the studies presented in this dissertation can be redone, this time focusing on the adult population in the DRC.

Further work is also necessary to understand the links between agriculture and malaria. In particular, we do not yet understand the role of agriculture around urban areas, and its effects may be different from those identified in rural areas owing to the fundamentally different ecology of cities. Additionally, in the work conducted in this dissertation, I did not attempt to identify different types of crops, and instead sought to incorporate the possibility that different crop types have different effects on malaria risk through specifying a model with a spatially-varying coefficient process. This process, however, is likely not sensitive enough to identify the effects of crop variability *within* a location, since the coefficients varied spatially *between* locations. Future work will need to address this limitation, preferably by focusing on collecting data on different crop types.

With regard to Aim 2, future work should focus on trying to address possible misclassification bias that may have occurred in using bed net brand to identify the insecticides used, since some bed net brands are obscure and appear to be country-specific brand names with no clear manufacturer. Additionally, none of the surveys used identified whether or not bed nets were treated with piperonyl butoxide (PBO), a synergistic compound that works to maintain the efficacy of the insecticide. As a result, estimates where we conclude certain insecticides are effective may not be the result of the insecticide alone, but the result of the insecticide in combination with PBO. Ascertaining this was beyond the scope of this dissertation, as it would likely involve extensive conversations (and in-country visits) to various malaria control programs, but given the importance of bed nets to malaria control, future work should address this concern.

With regard to Aim 3, I sought to understand population structure of the malaria parasite population, and possible corridors and barriers to gene flow that may give rise to such structure. Unfortunately, we could draw no firm conclusions, and the best evidence available suggests that the DRC is best characterized as one, large population of parasites. As noted, this may be due to high levels of transmission. As transmission declines, however, it will become increasingly important to understand how the parasite spreads, such that these transmission reductions can be maintained. In that sense, then, continued molecular surveillance efforts must continue, possibly through exploring other markers that may prove useful in measuring gene flow.

Finally, while the substantive focus of this dissertation has been on malaria and challenges to malaria control, population health/medical geography has served as a backdrop against which I conducted this work. In particular, I began this dissertation with a question that was asked 40 years ago, namely:

what can geographers offer that other disciplines cannot or do not? My answer to this question largely centered around the argument that the *fundamental* goal of population health/medical geography should be to characterize and understand heterogeneity, an effort which, to a large degree, requires reorienting our thoughts around how to model data. The example I outlined (and implemented in much greater detail in chapter 2) suggested that effects of exposures on outcomes may vary over space, and that identifying this variability can facilitate public health by allowing us to hypothesize why we observe such variability. Moreover, I sought to connect this objective to the core objective in epidemiology, namely the identification of causal relationships between exposures and health outcomes. In epidemiology, this effort is largely driven by careful construction of directed acyclic graphs (DAGs) to control for potential confounding. In a sense, then, observing relationships that vary spatially can provide insights into where a given DAG fails to capture the data generating process.

Importantly, population health/medical geographers may be able to provide further insights into the utility of a given DAG through spatially modeling the DAG itself. Indeed, recent work in spatial statistics has extended spatially varying coefficient models to include variable selection steps over space, allowing us to understand *where* an exposure (or set of exposures) influences a health outcome, and where it does not (Boehm Vock et al., 2015). As with spatially-varying coefficient models, this effort would serve to help learn about unmeasured confounding and its spatial dimensions. Such an effort fits well within Meade's framework, and further compliments epidemiology's objectives. Moving in this direction, however, will require population health/medical geographers to develop considerable statistical expertise (particularly in Bayesian statistics), to pursue collaborations with statisticians, or both.

Ultimately, and even in the absence of adopting the modeling philosophy outlined here, population health/medical geographers have much to offer in pursuing this future work. Indeed, understanding how humans interact with their environments is a core goal of geography in general, and, when health and disease is of interest, an explicit focus of population health/medical geographers.

APPENDIX I: DERIVATION OF FULL CONDITIONALS FOR AIM 1

6.1 Introduction

The following walks the reader through the construction of the data and probit models used in Chapter 2. I begin by writing down the basic structure for the data and model for a single individual in a single spatial location, and then build up to a model for all individuals across all locations. I then discuss the prior distributions used in the analysis, and finally present derivations of the full conditionals for all model parameters. The motivation for doing so is as follows. First, there is the practical benefit in showing how these models "work." Second, the presentation of statistical methods—including spatially varying coefficient regression—is, not surprisingly, done in statistics journals. These outlets are generally readable only by statisticians, meaning that the vast majority of individuals applying statistical methods cannot easily access material on methods for which there is not a software/"black box" implementation. At present, there is not yet a package that can flexibly fit "multilevel" spatial models (although WinBUGS remains an option here in some settings), or spatially-varying coefficient models specifically. Thus, those wishing to implement these types of models must code them from scratch, which requires knowing what the model actually looks like "under the hood." Thus, it is my hope that walking through the details here will help future students who may have interest in working with multilevel spatial models or spatially-varying coefficient regression models. Finally, it's kinda fun to think through this stuff (i.e. write down a model that I think will increase our understanding of how nature works, and then see if it does).

6.1.1 Motivation for Probit model

Our outcome of interest is each individual's PCR-diagnosed malaria status, a binary indicator taking the value 1 if an individual is infected with malaria, and 0 otherwise. Typically, binary data are handled using logistic regression. However, spatial models for point-referenced data become computationally intensive very quickly as the number of spatial locations increases. This computational burden is further

increased due to the lack of conjugacy between the prior distributions for model parameters and the data likelihood in logistic regression. As such, we adopt a probit specification in which we introduce latent variables that are assumed to follow a normal distribution with unit variance. Such a specification also has a scientific rationale. For example, we can think of these latent variables as a propensity to become infected with malaria, with values above 0 indicating increased propensity to become infected with malaria, and vice versa. To see this connection, observe that we can represent the probability of malaria infection, given covariates, as coming from a linear model:

$$P(Y = 1|\mathbf{X}) = P(\mathbf{X}\beta + \epsilon > 0) \tag{6.1}$$

$$= P(\mathbf{X}\beta > -\epsilon) \tag{6.2}$$

$$= P(\epsilon < \mathbf{X}\beta) \tag{6.3}$$

$$= \Phi(\mathbf{X}\beta) \tag{6.4}$$

where $\Phi(\cdot)$ represents the CDF of a standard normal distribution, and where the move from $(1.1.2)$ to $(1.1.3)$ is possible due to the symmetry of the normal distribution.

The modeling tasks in this aim are more complex than this simple illustration. For example, in our problem we are interested in introducing random effects, up to and including both a spatially varying intercept and spatially varying slope. Further, individuals are not uniquely located in space, but are nested within DHS survey clusters. As such, our intercept and slopes vary over the set of spatial locations (DHS clusters), not individuals, requiring us to map cluster-specific random effects to each survey respondent. We can accomplish this mapping through careful construction of the design matrix for the random effects. Further, this construction will then allow us to construct a separable spatial model, in which slopes and intercept are correlated with each other within spatial locations, but independent across locations, while each given slope/intercept is spatially-structured across locations.

I now turn to showing how the data are constructed, which will then allow me to write down the general form of the models fit in this aim.

6.2 Model Structure

6.2.1 Model for a single individual

I begin by writing down the structure of the model for one individual i within one spatial location s. We have:

$$Y_{is} = \begin{pmatrix} X_{is} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} Z_{is} \end{pmatrix} \begin{pmatrix} \theta_s \end{pmatrix} + \epsilon_{is} \tag{6.5}$$

Here, Y_{is} is the response for individual i (i in $1 \ldots n$) at location s (s in $1 \ldots q$). The design matrix for the constant intercept and slope is given by $\begin{pmatrix} X_{is} \end{pmatrix}$, a row vector of dimension p. The $p \times 1$ vector β contains regression coefficients that are invariant over space and link the covariates to the response. The random effects design matrix is of the same general form as the design matrix for constant intercept and slope, differing only in that our interest here is in a random intercept and, in one model, a spatial random slope, meaning that Z_{is} is either a scalar 1, when only a random intercept is introduced into the model, or vector-valued and of dimension 1×2, when we introduce the agriculture covariate to model a spatially varying effect. This can easily be extended up to the case where all coefficients vary spatially, in which case the design matrix for the spatial random effects in this step would be the same as the constant intercept and slopes. The random effects θ_s that vary across locations are then easily recognized through the index s. For a single individual with a single random effect, θ_s is a scalar, but vector valued as additional random effects are incorporated. Again, this vector would increase in dimension if additional covariates were to be modeled as spatially varying. Finally, ϵ_{is} represents pure white noise error for individual i at location s, and is assumed to follow a standard normal distribution.

6.2.2 Model for individuals within one location

Given the model for one individual in one location, we now construct a model for all individuals within one location. We do this simply by stacking individuals within a location as follows:

Let Y_s be the vector of all individual responses within location s, and X_s and Z_s be the design matrices for these same individuals, such that:

$$\underbrace{\begin{pmatrix} Y_{1,s} \\ Y_{2,s} \\ \vdots \\ Y_{n_s,s} \end{pmatrix}}_{\mathbf{Y_s}} = \underbrace{\begin{pmatrix} X_{1,s} \\ X_{2,s} \\ \vdots \\ X_{n_s,s} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{\mathbf{X_s}\beta} + \underbrace{\begin{pmatrix} Z_{1,s} \\ Z_{2,s} \\ \vdots \\ Z_{n_s,s} \end{pmatrix} \begin{pmatrix} \theta_s \end{pmatrix}}_{\mathbf{Z_s}\theta_\mathbf{s}} + \underbrace{\begin{pmatrix} \epsilon_{1,s} \\ \epsilon_{2,s} \\ \vdots \\ \epsilon_{n_s,s} \end{pmatrix}}_{\epsilon_\mathbf{s}} \tag{6.6}$$

Here, n_s is the n^{th} individual in location s, while $\theta_\mathbf{s}$ is as before, consisting of a scalar or vector of random effects for location s, depending on the model being fit. I now write down a model for the full data.

6.2.3 Model for all individuals in all locations

With a model for each individual i within a spatial location s in hand, we must now extend this for a model for all individuals across all spatial locations. For the response variable, design matrix for the constant intercept and slope, and white noise error, we simply continue stacking the data as before. If we were to do this for the random effects design matrix, however, we would be inducing random effects for each individual, rather than for each spatial location. Thus, we need now to write down a random effects design matrix that will map spatial random effects to each survey respondent within a spatial location. We can do this by constructing a block-diagonal matrix $\mathbf{Z}$ such that each block is the design matrix Z_s from before, meaning that the overall random effects design matrix $\mathbf{Z}$ is $n \times kq$, where k is either 1 (model with only a random intercept) or 2 (model with a random intercept and slope). We can write this model down as follows:

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{pmatrix}}_{\mathbf{Y}} = \underbrace{\begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_q \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{\mathbf{X}\beta} + \underbrace{\begin{pmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & Z_q \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \end{pmatrix}}_{\mathbf{Z}vec(\theta)} + \underbrace{\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_q \end{pmatrix}}_{\epsilon} \tag{6.7}$$

Now, we can see the general form of the latent model can be expressed as follows:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{Z}vec(\theta) + \epsilon \tag{6.8}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of latent normal responses, $\mathbf{X}$ is an $n \times p$ matrix of covariates, β is a $p \times 1$ vector of regression coefficients linking the covariates to the response, $\mathbf{Z}$ is an $n \times kq$ matrix that maps the random effects to each respondent as described above, while $vec(\theta)$ stacks the random effects into a $kq \times 1$ vector as follows:

$$vec(\theta) = \begin{pmatrix} \theta_{\mathbf{1}} \\ \theta_{\mathbf{2}} \\ \vdots \\ \theta_{\mathbf{q}} \end{pmatrix}$$

Writing down the random effects in this way allows us later to write down a separable spatial model, in which we can learn about the correlation between intercepts and slopes *within* a spatial location, the correlation between slopes *across* locations, as well as between intercepts *across* locations. The implicit assumption here, then, is that the correlation between an intercept and slope within a location is not spatially structured. Further, the model with only a random intercept (spatially correlated or otherwise), are a special case of this specification. Finally, ϵ is an $n \times 1$ vector of random error assumed to follow a zero-centered normal distribution with unit variance.

6.3 Prior Distributions

In a Bayesian analysis, our model specification is only complete after we have assigned prior distributions for all model parameters. In this section, I briefly outline the construction of these prior distributions.

6.3.1 Prior distribution for β

Prior distributions for regression coefficients are often specified as a low-precision normal centered around zero. Such a specification is considered as non-informative, meaning we have no strong prior

knowledge about the size of an effect, and allow the model to explore a large parameter space centered around the null value of 0. While this prior is generally a "default" choice, malaria has been studied for over 100 years, with a vast literature from which we can draw prior knowledge and incorporate it into our model where appropriate. Further, we know a lot about regression and effect sizes in general. Consider, for example, the classic example of the effect of smoking on probability of death due to lung cancer. Assume that the probability of death due to lung cancer in non-smokers is 0.01, and is 0.5 in smokers. On the logit scale, this equates to a log odds ratio of 5. This effect size can be thought of as an upper bound on the size of an effect of an exposure on a health outcome, since smoking is arguably the greatest risk factor for death/disease. For these reasons, I choose a standard normal as a default prior for all regression coefficients, after centering and scaling continuous covariates so as to make modeling scale free. A prior such as this is desirable since it restricts the model to sample only reasonable values of the posterior distribution—that is, effect sizes that are plausible.

6.3.2 Prior distribution for $vec(\theta), \phi, \mathbf{H}$

For the spatial random effects, we assume these are generated by a mean-zero Gaussian process. We write down this process for the model with a spatially varying intercept and slope, noting that the model for just a spatial random intercept and an independently varying intercept are special cases:

$$vec(\theta) \sim N_{kq}\big(\mathbf{0}, \mathbf{\Sigma}(\phi) \otimes \mathbf{H}\big) \tag{6.9}$$

where $\mathbf{\Sigma}(\phi)$ is a $q \times q$ matrix of pairwise distances between locations, with the decay in their correlation governed by range parameter ϕ. The spatial variances and the covariance between the intercept and slope at a location is introduced through the $k \times k$ spatial variance-covariance matrix $\mathbf{H}$, which has the variances for the spatial intercept and spatial slope on the diagonals and their covariance on the off-diagonal. For our purposes, k is either 1 or 2, depending on the model being fit. The $\otimes$ denotes the Kronecker product. In the case of a model with just a spatially-varying intercept (i.e. $k = 1$), $\mathbf{H}$ is a scalar, denoted σ^2, and represents the variance of the spatial random intercept. In the case of a model in which the intercept varies independently, $\mathbf{H}$ is again a scalar, σ^2, with $\mathbf{\Sigma}(\phi)$ being replaced with a q-dimensional identity matrix.

The above specification has 2 unknown parameters, which I assign the following prior distributions. For the common spatial range parameter ϕ, I assign the following uniform prior:

$$U\left(\frac{3}{0.1 \times max(distances)}, \frac{3}{0.1}\right)$$

, which yields a prior range of between 100 meters and roughly 225 kilometers. I choose to be more informative about the range parameter owing to identifiability concerns and a preference for learning about the spatial variance over the range (Banerjee et al., 2014). Further, I adopt an exponential correlation function, as this is the most commonly used spatial correlation function used in geostatistical modeling of malaria transmission (see, for example, the work conducted by the Malaria Atlas Project).

For the variance-covariance matrix $\mathbf{H}$, I assign an Inverse Wishart prior distribution, that is:

$$f(\mathbf{H}) \sim InvWish(k+1, \mathbf{B}_0) \tag{6.10}$$

where $\mathbf{B}_0$ is a 2×2 identity matrix. This specification has the appealing feature that the marginal prior distribution of the correlation parameters is $U(-1, 1)$ (Gelman et al., 2014).

In the case where there is only a random intercept (i.e. $k = 1$ in 1.3.1), the variance parameter (spatial or otherwise), is assigned a conjugate $IG(2, 1)$ prior distribution.

6.4 Full Conditional Distributions for all Model Parameters

Given the model formulation developed in section 1.2 and the prior specifications presented in 1.3, I now turn to deriving the full conditionals for all model parameters.

6.4.1 Full Conditional for Latent Variables Y

Let Y_{is} be the latent variable for individual i at spatial location s. These Y_{is} are sampled from their full conditional distribution

$$f(Y_{is}|Y_{is}^*, \mathbf{\Omega}) \propto f(Y_{is}^*|Y_{is}, \mathbf{\Omega}) f(Y_{is}|\mathbf{\Omega}) \tag{6.11}$$

where Y_{is}^* is an indicator of malaria status as defined earlier, and $\mathbf{\Omega}$ is a vector of model parameters. There are two possibilities for the observed indicator Y_{is}^*, either $Y_{is}^* = 0$ (i.e. the respondent does not have malaria), or $Y_{is}^* = 1$, the respondent has malaria. If $Y_{is}^* = 0$, then,

$$f(Y_{is}^*|Y_{is}, \mathbf{\Omega}) = P(Y_{is}^* = 0|Y_{is}, \mathbf{\Omega}) = \begin{cases} 1 & Y_{is} \leq 0, \\ 0 & Y_{is} > 0. \end{cases} \tag{6.12}$$

Thus, $f(Y_{is}|Y_{is}^*, \mathbf{\Omega}) \propto I(Y_i \leq 0) f(Y_{is}|\mathbf{\Omega})$. Similarly, for the case where $Y_{is}^* = 1$, then,

$$f(Y_{is}^*|Y_{is}, \mathbf{\Omega}) = P(Y_{is}^* = 1|Y_{is}, \mathbf{\Omega}) = \begin{cases} 0 & Y_{is} \leq 0, \\ 1 & Y_{is} > 0. \end{cases} \tag{6.13}$$

Thus, $f(Y_{is}|Y_{is}^*, \mathbf{\Omega}) \propto I(Y_{is} > 1) f(Y_{is}|\mathbf{\Omega})$. Given these two quantities, we can write down the full conditional for the latent variable as:

$$f(Y_{is}|Y_{is}^*, \mathbf{\Omega}) \sim \begin{cases} N(X_{is}\beta + Z_{is}\theta_s, 1)[Y_{is} \leq 0] \\ N(X_{is}\beta + Z_{is}\theta_s, 1)[Y_{is} > 0] \end{cases} \tag{6.14}$$

where we can see that the latent variable is distributed according to a truncated normal distribution, truncated from above by 0 when the latent variable is negative (lower propensity for malaria), and truncated from below by 0 when the latent variable is positive (higher propensity for malaria).

6.4.2 Full Conditional for β

The full conditionals for β depend on the distribution of the latent variables, which we assume (as noted earlier) follow a normal distribution as follows:

$$f(\mathbf{Y}|\mathbf{\Omega}) \propto \exp\Bigg\{ -\frac{1}{2}\Bigg[\Big\{\mathbf{Y} - \Big(\mathbf{X}\beta + \mathbf{Z}vec(\theta)\Big)\Big\}^T \times \Big\{\mathbf{Y} - \Big(\mathbf{X}\beta + \mathbf{Z}vec(\theta)\Big)\Big\}\Bigg]\Bigg\} \tag{6.15}$$

where, as before, $\mathbf{\Omega}$ is a vector of model parameters. Note that we omit the inverse of the pure noise variance here since we are specifying the model to have unit variance. Next, note that the full conditional distribution for β depends only on the distribution for the latent variables and the prior, such that:

$$f(\beta|rest) \propto f(\mathbf{Y}|\mathbf{\Omega}) f(\beta|\sigma^2_\beta = 1) \tag{6.16}$$

$$\exp\left\{ = \frac{1}{2}\left[\Big(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}vec(\theta)\Big)^T \Big(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}vec(\theta)\Big)\right]\right\}$$

$$\times \exp\left\{ -\frac{1}{2}(\beta - \mathbf{m})^T \mathbf{I}^{-1}(\beta - \mathbf{m})\right\} \tag{6.17}$$

Let $\gamma = \mathbf{Y} - \mathbf{Z}vec(\theta)$. Then:

$$f(\beta|\cdot) \propto \exp\left\{ -\frac{1}{2}\Big(\gamma - \mathbf{X}\beta\Big)^T \Big(\gamma - \mathbf{X}\beta\Big)\right\}$$

$$\times \exp\left\{ -\frac{1}{2}(\beta - \mathbf{m})^T \mathbf{I}^{-1}(\beta - \mathbf{m})\right\} \tag{6.18}$$

$$\propto \exp\left\{ -\frac{1}{2}\left[\beta^T \Big(\mathbf{X}^T\mathbf{X} + \mathbf{I}^{-1}\Big)\beta - 2\beta^T\mathbf{X}^T\gamma - 2\beta^T\mathbf{m}\right]\right\} \tag{6.19}$$

$$= \exp\left\{\frac{1}{2}\left[\beta^T \Big(\mathbf{X}^T\mathbf{X} + \mathbf{I}^{-1}\Big)\beta - 2\beta^T(\mathbf{X}^T\gamma + \mathbf{m})\right]\right\} \tag{6.20}$$

Observe that expression $(A.4.10)$ has the general form $\exp\big\{ -\frac{1}{2}\big(\beta^T(\mathbf{A})\beta - 2\beta(\mathbf{B})\big)\big\}$, which is the kernel of a multivariate normal distribution with mean $\mathbf{A}^{-1}\mathbf{B}$ and variance $\mathbf{A}^{-1}$. Thus, we can write down the full conditional for β as:

$$f(\beta|\cdot) \propto N_p\big(\mathbb{E}_\beta, \mathbb{V}_\beta\big), \tag{6.21}$$

where $\mathbb{E}_\beta = \big(\mathbf{X}^T\mathbf{X} + \mathbf{I}\big)^{-1}\big(\mathbf{X}^T\gamma + \mathbf{m}\big)$ and $\mathbb{V}_\beta = \big(\mathbf{X}^T\mathbf{X} + \mathbf{I}\big)^{-1}$.

6.4.3 Full Conditional for $vec(\theta)$

As with the full conditional for β, the full conditionals for $vec(\theta)$ depend on the latent variables owing to the conditional nature of the model. As noted, we model the spatial random effects via a zero-centered Gaussian process. We model the variance as separable, such that correlation between intercept and slope within a location is separate from correlations between intercepts and between slopes

across locations. Recall that the prior is specified in 1.3.1 as:

$$f(vec(\theta)|\phi, \mathbf{H}) \sim N_{kq}\Big(0, \mathbf{\Sigma}(\phi) \otimes \mathbf{H}\Big)$$

where $\otimes$ denotes the Kronecker product. To obtain the full conditional distribution, we combine the likelihood and prior:

$$f(vec(\theta)|rest) \propto f(\mathbf{Y}|\mathbf{\Omega}) f(vec(\theta)|\mathbf{H}, \phi) \tag{6.22}$$

$$\propto \exp\Big\{ -\frac{1}{2}\Big(\mathbf{Y} - \Big(\mathbf{X}\beta + \mathbf{Z}vec(\theta)\Big)\Big)^T \tag{6.23}$$

$$\times \Big(\mathbf{Y} - \Big(\mathbf{X}\beta + \mathbf{Z}vec(\theta)\Big)\Big)\Big\} \exp\Big\{ -\frac{1}{2} vec(\theta)^T \Big(\mathbf{\Sigma}(\phi) \otimes \mathbf{H}\Big)^{-1} vec(\theta)\Big\}$$

Let $\mu = \mathbf{Y} - \mathbf{X}\beta$. Then we can combine likelihood and prior:

$$= \exp\Big\{ -\frac{1}{2}\Big[\Big(\mu - \mathbf{Z}vec(\theta)\Big)^T \Big(\mu - \mathbf{Z}vec(\theta)\Big) \tag{6.24}$$

$$+ vec(\theta)^T \Big(\mathbf{\Sigma}(\phi) \otimes \mathbf{H}\Big)^{-1} vec(\theta)\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\Big[\mu^T\mu - \mu^T\mathbf{Z}vec(\theta) - vec(\theta)^T\mathbf{Z}^T\mu + vec(\theta)^T\mathbf{Z}^T\mathbf{Z}vec(\theta) \tag{6.25}$$

$$+ vec(\theta)^T \Big(\mathbf{\Sigma}(\phi) \otimes \mathbf{H}\Big)^{-1} vec(\theta)\Big]\Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\Big[-\mu^T\mathbf{Z}vec(\theta) - vec(\theta)^T\mathbf{Z}^T\mu \tag{6.26}$$

$$+ vec(\theta)^T \Big(\mathbf{Z}^T\mathbf{Z} + (\mathbf{\Sigma}(\phi) \otimes \mathbf{H})^{-1}\Big) vec(\theta)\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\Big[vec(\theta)^T \Big(\mathbf{Z}^T\mathbf{Z} + (\mathbf{\Sigma}(\phi) \otimes \mathbf{H})^{-1}\Big) vec(\theta) - 2vec(\theta)^T\mathbf{Z}^T\mu\Big]\Big\} \tag{6.27}$$

As with the full conditional for β, we see here the kernel of a multivariate normal distribution:

$$f(vec(\theta)|\cdot) \sim N_{kq}(\mathbb{E}_{vec(\theta)}, \mathbb{V}_{vec(\theta)}) \tag{6.28}$$

where

$$\mathbb{E}_{vec(\theta)} = \Big(\mathbf{Z}^T\mathbf{Z} + \Big(\mathbf{\Sigma}(\phi) \otimes \mathbf{H}\Big)^{-1}\Big)^{-1} \mathbf{Z}^T\mu \tag{6.29}$$

and

$$\mathbb{V}_{vec(\theta)} = \left(\mathbf{Z}^T \mathbf{Z} + \left(\mathbf{\Sigma}(\phi) \otimes \mathbf{H} \right)^{-1} \right)^{-1} \tag{6.30}$$

Note that in the case where we only have a spatially varying intercept ($k = 1$), the kronecker operation between $\mathbf{\Sigma}(\phi) \otimes \mathbf{H}$ is replaced with $\sigma^2 \mathbf{\Sigma}(\phi)$. In the case where we have an independently varying intercept, this is further reduced to $\sigma^2 \mathbf{I}_q$.

6.4.4 Full Conditional for H

The full conditional for the spatial variance-covariance matrix $\mathbf{H}$ for the random effects does not depend on the data likelihood, but on the prior distribution of $vec(\theta)$. The prior for the variance-covariance matrix $\mathbf{H}$ is specified as Inverse Wishart:

$$f(\mathbf{H}) \sim InvW(\lambda_0, \mathbf{B}_0) \tag{6.31}$$

where λ_0 is scalar and $\mathbf{B}_0$ is 2×2. To write down the full conditional distribution for $\mathbf{H}$, I rely on the following matrix properties:

1. $vec(A)^T (D \otimes B) vec(C) = tr(A^T B C D^T)$ (By google)
2. $tr(ABCD) = tr(BCDA) + tr(CDAB)$, i.e. the cyclic property of the trace function
3. $det(AB) = det(A) det(B)$
4. $tr(A + B) = tr(A) + tr(B)$
5. $det(A)^n = det(A^n)$
6. $det(A \otimes B) = det(A)^m det(B)^n$, for $A_{n \times n}$ and $B_{m \times m}$

The full conditional for $\mathbf{H}$ is as follows:

$$f(\mathbf{H}|\theta, \phi) \propto f(\theta|\mathbf{\Sigma}(\phi), \mathbf{H}) f(\mathbf{H}) \tag{6.32}$$

$$\propto |\mathbf{\Sigma}(\phi) \otimes \mathbf{H}|^{-1/2} \exp\bigg(-\frac{1}{2} vec(\theta)^T (\mathbf{\Sigma}(\phi)^{-1} \otimes \mathbf{H}^{-1}) vec(\theta) \bigg) \tag{6.33}$$
$$\times |\mathbf{H}|^{-(\lambda_0+p+1)/2} \exp\bigg(-\frac{1}{2} tr(\mathbf{B}_0 \mathbf{H}^{-1}) \bigg)$$

Using Matrix Property I above, we can rewrite the first line in (4) above as a trace function:

$$\propto |\mathbf{\Sigma}(\phi) \otimes \mathbf{H}|^{-1/2} \exp\bigg(-\frac{1}{2} tr(\theta^T \mathbf{H}^{-1} \theta \mathbf{\Sigma}(\phi)^{-1}) \bigg) \tag{6.34}$$
$$\times |\mathbf{H}|^{-(\lambda_0+p+1)/2} \exp\bigg(-\frac{1}{2} tr(\mathbf{B}_0 \mathbf{H}^{-1}) \bigg)$$

where θ is $2 \times q$ matrix of spatial intercepts and slopes. Note that since $\mathbf{\Sigma}(\phi)$ is symmetric, $\mathbf{\Sigma}(\phi)^T = \mathbf{\Sigma}(\phi)$, so I exclude the transpose from property 1. This holds for the inverse of the matrix as well. I now look at the kronecker product coeffient in $(1.4.24)$, where I use property 6 above to assign the proper exponent to the determinant of $\mathbf{H}$, and then drop $\mathbf{\Sigma}(\phi)$, which is absorbed in the proportionality. Doing this, and then combining the terms in the exponent, we have:

$$|\mathbf{H}|^{-q/2} |\mathbf{H}|^{-(\lambda_0+p+1)/2} \exp\bigg(-\frac{1}{2} \big(tr(\theta^T \mathbf{H}^{-1} \theta \mathbf{\Sigma}(\phi)^{-1}) + tr(\mathbf{B}_0 \mathbf{H}^{-1}) \big) \bigg) \tag{6.35}$$

Now, iteratively using the cyclic property (property 2 above) in the first trace function in the exponent, as well as properties 3 and 5 for the product of determinants, we can write down the full conditional as:

$$|\mathbf{H}|^{-(q+\lambda_0+p+1)/2} \exp\bigg(-\frac{1}{2} tr(\mathbf{B}_1 \mathbf{H}^{-1}) + tr(\mathbf{B}_0 \mathbf{H}^{-1}) \bigg) \tag{6.36}$$

where $\mathbf{B}_1 = \theta \mathbf{\Sigma}(\phi)^{-1} \theta^T$. Finally, using property 4 above, we can write the full conditional as:

$$|\mathbf{H}|^{-(q+\lambda_0+p+1)/2} \exp\bigg(-\frac{1}{2} tr((\mathbf{B}_1 + \mathbf{B}_0) \mathbf{H}^{-1}) \bigg) \tag{6.37}$$

which is the kernel of an Inverse Wishart distribution with degrees of freedom equal to $q + \lambda_0$ and scale matrix $\tilde{\mathbf{B}} = \mathbf{B}_1 + \mathbf{B}_0$.

In the case where we are dealing only with a random intercept, for which we specified a $IG(2, 1)$ prior distribution, the full conditional distribution is also inverse gamma. For the spatial random intercept,

we have:

$$f(\sigma^2|\cdot) \sim IG\left(2 + q/2, \frac{vec(\theta)^T \mathbf{\Sigma}(\phi)^{-1} vec(\theta)}{2} + 1\right) \tag{6.38}$$

and for the independently varying intercept, we have:

$$f(\sigma^2|\cdot) \sim IG\left(2 + q/2, \frac{vec(\theta)^T vec(\theta)}{2} + 1\right) \tag{6.39}$$

6.4.5 Full Conditional for ϕ

As with the spatial variance, the range parameter ϕ does not depend on the likelihood function. We can write the full conditional for ϕ as follows:

$$f(\phi|rest) \propto f(vec(\theta)|\mathbf{H}, \phi) f(\phi) \tag{6.40}$$

$$\propto |\mathbf{\Sigma}(\phi) \otimes \mathbf{H}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} vec(\theta)^T \left(\mathbf{\Sigma}(\phi) \otimes \mathbf{H}\right)^{-1} vec(\theta) \right\} \tag{6.41}$$

$$\times \frac{1}{3/.1 - 3/(0.1 \times max(distances))}$$

There is no distributional form here, and a Metropolis step is therefore necessary to update ϕ. This is done by mapping ϕ to the real line and using a Normal proposal density and using Pilot Adaptation to tune the variances of the proposal density.

6.4.6 Posterior Predictive Distribution for Intercept/Slope Processes

Given posterior samples of spatial intercept and slope parameters at observed spatial locations, interest naturally turns to predicting these quantities at unobserved locations—that is, we would like to learn about the underlying spatial surface in the intercept and slope process. This can be done through composition sampling of the posterior predictive distribution.

As a happy aside, here I note that one elegant consequence of working with spatial models implemented using a Gaussian process relates to the oft-pluralized "geographies" that a geographer might study (e.g. "I study the geographies of struggle."). As a geographer myself, I frequently struggle to figure out how many geographies these geographers are talking about, and am always tempted to ask the

simple question: "How many geographies do you study?". Simple questions, however, are often the most difficult to answer, and asking it would, of course, make such a question of me equally valid (provided I ever said 'geographies'), but working with Gaussian processes allows me simply to say "it's uncountable" and move on. Conversely, if I find myself wanting to be more specific, I can also say that it's arbitrary, and that I can draw finite-dimensional realizations of any size I wish from the process, meaning that the number of geographies is entirely up to me. Thus, models incorporating Gaussian processes have the nice property of allowing me to ask the question while at the same time not have to count anything myself. Which is good, because counting is hard.

Turning back to the substantive issue at hand, let θ^* denote the intercept and slope at an unmeasured location. The posterior predictive distribution, then, is:

$$f(vec(\theta^*)|vec(\theta), \mathbf{H}, \phi) \tag{6.42}$$

To derive this quantity, I begin by writing down the joint distribution $f(vec(\theta^*), vec(\theta))$ and then use multivariate normal theory to construct the conditional of interest. The joint distribution can be written as:

$$vec([\theta, \theta^*]) \sim N\left(\begin{pmatrix} \mathbf{0}_{kq} \\ \mathbf{0}_{kq^*} \end{pmatrix}, \quad \begin{pmatrix} \mathbf{\Sigma}(\phi)_{q\times q} \otimes \mathbf{H} & \mathbf{\Sigma}(\phi)^T_{q^*\times q} \otimes \mathbf{H}^T \\ \mathbf{\Sigma}(\phi)_{q^*\times q} \otimes \mathbf{H} & \mathbf{\Sigma}(\phi)_{q^*\times q^*} \otimes \mathbf{H} \end{pmatrix}\right) \tag{6.43}$$

where $\mathbf{0}_{2q}$ is the mean for the spatial random intercepts and slopes at observed locations with length $2q$, and $\mathbf{0}_{2q^*}$ is the $2q^*$-length vector of mean spatial random intercepts and slopes at predicted locations. The variance-covariance matrix consists of four components/submatrices. The (1,1) entry represents the variance-covariance matrix for the intercepts and slopes at observed locations, while the (2,2) entry represents the variance-covariance matrix for the intercepts and slopes at unobserved locations. The off-diagonal entries [(2,1) and (1,2)] represent the variance-covariance matrices for the intercepts and slopes between observed and unobserved locations.

Given the multivariate normal structure written here, we can immediately write down the conditional distribution of $vec(\theta^*)$.

$$f(vec(\theta^*)|vec(\theta), \mathbf{H}, \phi) \sim N(\mathbb{E}_{vec(\theta^*)}, \mathbb{C}_{vec(\theta^*)}) \tag{6.44}$$

where

$$\mathbb{E}_{vec(\theta^*)} = [\mathbf{\Sigma}(\phi)_{q^* \times q} \otimes \mathbf{H}][\mathbf{\Sigma}_{q^* \times q}^T \otimes \mathbf{H}^T]^{-1} vec(\theta) \tag{6.45}$$

$$\tag{6.46}$$

and

$$\mathbb{C}_{vec(\theta^*)} = \left[\mathbf{\Sigma}(\phi)_{q \times q} \otimes \mathbf{H}\right] - \left[\mathbf{\Sigma}(\phi)_{q^* \times q}^T \otimes \mathbf{H}^T\right] \left[\mathbf{\Sigma}(\phi)_{q^* \times q^*} \otimes \mathbf{H}\right]^{-1} \left[\mathbf{\Sigma}(\phi)_{q^* \times q} \otimes \mathbf{H}\right] \tag{6.47}$$

APPENDIX II: SUPPLEMENTARY MATERIALS FOR AIM 1

All models implemented in Chapter 2 were run for 120,000 iterations, with the first 20,000 discarded as burn-in and the Markov chain thinned such that inference about model parameters is based on 10,000 posterior samples. Model convergence was assessed by inspecting traceplots of model parameters, and final inferences are based on the best fitting model, as assessed by Brier score, area under the ROC curve, and DIC. Table 7.6 below shows these fit statistics, showing that the model with just a random intercept and no spatial correlation yielded the best fit.

Table 7.6: Fit statistics for hierarchical probit regression models on agriculture and malaria risk

	Random Intercept	Spatial Random Intercept	Spatial Random Intercept and Slope
Fit Statistic			
Brier Score	0.160	0.161	0.159
ROC curve	0.839	0.836	0.838
DIC	4687	4695	4715

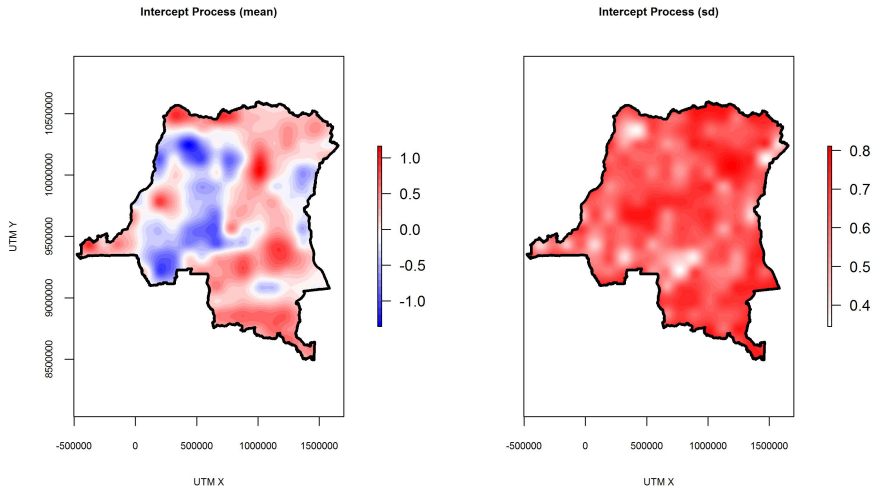
Spatial models were initially compared by randomly withholding a third of the spatial locations and predicting those data out-of-sample. Performance was identical across both models, as can be seen above, and all models were re-fit to the full data, with final inferences presented in the manuscript being based off of the model incorporating a random intercept, as it had the lowest DIC.

While the non-spatial model exhibited the best fit to the data, we show results for the spatial processes from both models here, as these can be suggestive of potential areas of future concern. Supplementary Figure 7.21 below shows the spatial intercept surface, together with corresponding uncertainty.

Considerable variability in the spatial random intercept process persists after accounting for other risk factors, with areas of the DRC exhibiting both strong increased and decreased risk of infection, particularly in northern regions. Notably, however, these estimates are accompanied by considerable imprecision, preventing definitive conclusions about areas of increased or decreased residual risk.

Figure 7.22 below shows the spatial intercept and slope surfaces for the model incorporating both a spatially varying intercept and a spatially varying slope for the effect of agriculture on malaria risk.

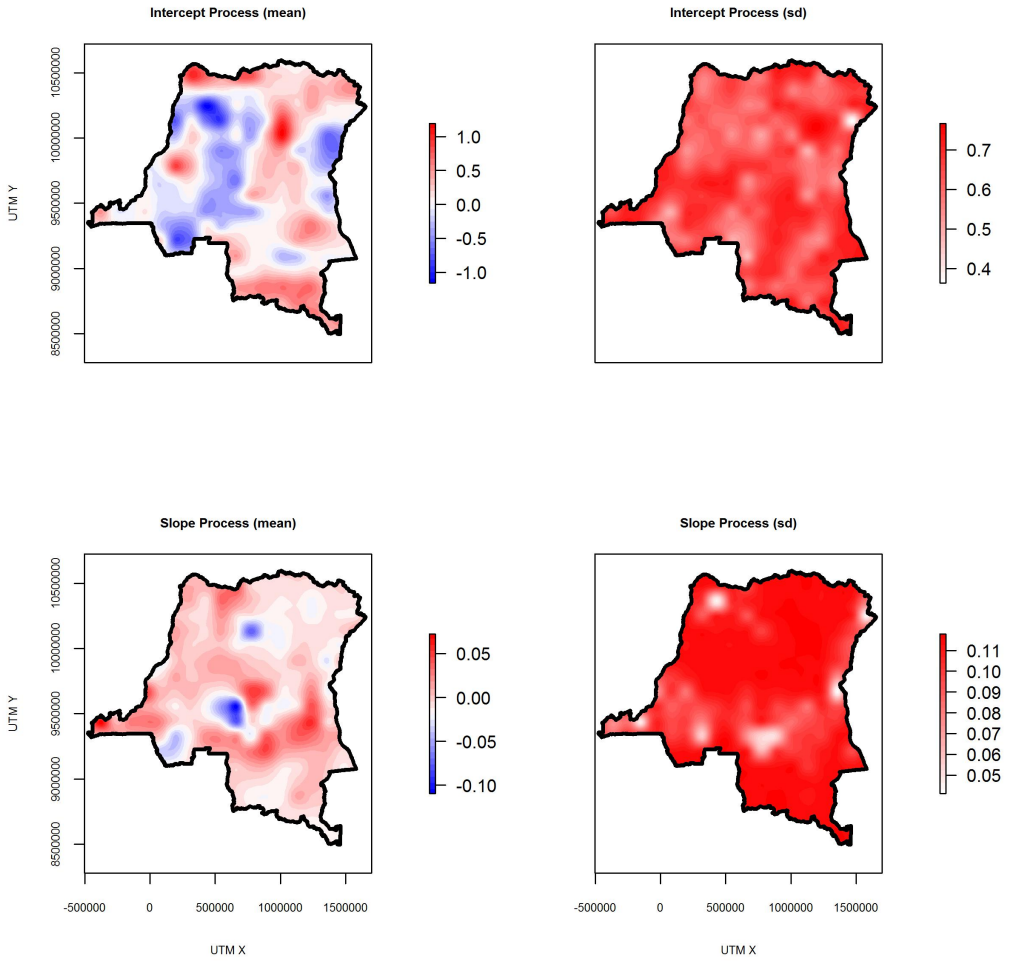
Figure 7.21: Spatial process from the 'spatial random intercept' model



Incorporating the spatial random slope leads to slight attenuation in the intercept process, although the spatial pattern broadly remains. Further, there is slight evidence of possible attenuation of the effect of agriculture in two places in DRC, one in the central-northern region, which is largely forest, and the other in central DRC in what is largely Savannah. This latter area also shows pockets of increased risk. In both cases, however, inferences on the intercept and slope processes are accompanied by considerable imprecision.

Much of this imprecision could potentially be controlled through use of a stronger prior for the variance of the random effects. Indeed, for all models an Inverse Gamma distribution with shape parameter equal to 2 and rate parameter equal to 1 was used. This specification is highly non-informative, as the variance of this parameter is infinite. Much stronger prior information is likely warranted here. This is especially true since this variance parameter governs the range of values that the random effects can take on, and these random effects are adjustments to the intercept and slope processes. As such, we can think of specifying this prior in the same way that we specify the variance of the regression coefficients (β). Future work in this area should consider a specification with a much more informative prior that reflects the range of values that the random effects can take on, and in so doing lead to shrinkage in the random effects, and more precise estimates. Worth noting, moreover, is that control of the variance can be done through both the shape and rate parameters, with the variance governed by a quadratic in

Figure 7.22: Spatial process from the 'spatial random intercept and slope' model



both parameters. More specifically, the variance for an Inverse Gamma distributed random variable is given by:

$$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$$

One can simply plug in values for α and β to obtain the variance of the prior variance desired.

REFERENCES

Afrane, Y. A., Lawson, B. W., Githeko, A. K., and Yan, G. (2005). Effects of microclimatic changes caused by land use and land cover on duration of gonotrophic cycles of anopheles gambiae (diptera: Culicidae) in western kenya highlands. *Journal of medical entomology*, 42(6):974–980.

Afrane, Y. A., Little, T. J., Lawson, B. W., Githeko, A. K., and Yan, G. (2008). Deforestation and vectorial capacity of anopheles gambiae giles mosquitoes in malaria transmission, kenya. *Emerging Infectious Diseases*, 14(10).

Afrane, Y. A., Zhou, G., Lawson, B. W., Githeko, A. K., and Yan, G. (2006). Effects of microclimatic changes caused by deforestation on the survivorship and reproductive fitness of anopheles gambiae in western kenya highlands. *The American journal of tropical medicine and hygiene*, 74(5):772–778.

Ageep, T. B., Cox, J., M'oawia, M. H., Knols, B. G., Benedict, M. Q., Malcolm, C. A., Babiker, A., and El Sayed, B. B. (2009). Spatial and temporal distribution of the malaria mosquito anopheles arabiensis in northern sudan: influence of environmental factors and implications for vector control. *Malaria Journal*, 8(1):123.

AIRS (2014). Drc enhanced entomological monitoring final report october 1, 2013-october 31, 2014. Technical report, USAID.

Ali, M., Emch, M., Von Seidlein, L., Yunus, M., Sack, D. A., Rao, M., Holmgren, J., and Clemens, J. D. (2005). Herd immunity conferred by killed oral cholera vaccines in bangladesh: a reanalysis. *The Lancet*, 366(9479):44–49.

Anderson, R. M., May, R. M., and Anderson, B. (1992). *Infectious diseases of humans: dynamics and control*, volume 28. Wiley Online Library.

Anderson, T., SU, X.-Z., Bockarie, M., Lagog, M., and Day, K. (1999). Twelve microsatellite markers for characterization of plasmodium falciparum from finger-prick blood samples. *Parasitology*, 119(02):113–125.

Anderson, T. J. and Roper, C. (2005). The origins and spread of antimalarial drug resistance: lessons for policy makers. *Acta tropica*, 94(3):269–280.

Antonia, A. L., Taylor, S. M., Janko, M., Emch, M., Tshefu, A. K., and Meshnick, S. R. (2014). A cross-sectional survey of plasmodium falciparum pfcrt mutant haplotypes in the democratic republic of congo. *The American journal of tropical medicine and hygiene*, 90(6):1094–1097.

Arcaya, M., Brewster, M., Zigler, C. M., and Subramanian, S. (2012). Area variations in health: A spatial multilevel modeling approach. *Health & place*, 18(4):824–831.

Asidi, A., N'Guessan, R., Akogbeto, M., Curtis, C., and Rowland, M. (2012). Loss of household protection from use of insecticide-treated nets against pyrethroid-resistant mosquitoes, benin. *Emerging infectious diseases*, 18(7):1101–6.

Auchincloss, A. H., Gebreab, S. Y., Mair, C., and Diez Roux, A. V. (2012). A review of spatial methods in epidemiology, 2000–2010. *Annual review of public health*, 33:107–122.

Awolola, T., Oduola, O., Strode, C., Koekemoer, L., Brooke, B., and Ranson, H. (2009). Evidence of multiple pyrethroid resistance mechanisms in the malaria vector anopheles gambiae sensu stricto from nigeria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 103(11):1139–1145.

Awolola, T., Oyewole, I., Amajoh, C., Idowu, E., Ajayi, M., Oduola, A., Manafa, O., Ibrahim, K., Koekemoer, L., and Coetzee, M. (2005). Distribution of the molecular forms of anopheles gambiae and pyrethroid knock down resistance gene in nigeria. *Acta tropica*, 95(3):204–209.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press.

Bhatt, S., Weiss, D., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K., Moyes, C., Henry, A., Eckhoff, P., et al. (2015). The effect of malaria control on plasmodium falciparum in africa between 2000 and 2015. *Nature*, 526(7572):207–211.

Bingham, G., Strode, C., Tran, L., Khoa, P. T., and Jamet, H. P. (2011). Can piperonyl butoxide enhance the efficacy of pyrethroids against pyrethroid-resistant aedes aegypti? *Tropical Medicine & International Health*, 16(4):492–500.

Boehm Vock, L. F., Reich, B. J., Fuentes, M., and Dominici, F. (2015). Spatial variable selection methods for investigating acute health effects of fine particulate matter components. *Biometrics*, 71(1):167–177.

Bongaarts, J. and Casterline, J. (2013). Fertility transition: is sub-saharan africa different? *Population and Development review*, 38(s1):153–168.

Boserup, E. (2005). *The conditions of agricultural growth: The economics of agrarian change under population pressure*. Transaction Publishers.

Bousema, T., Drakeley, C., Gesase, S., Hashim, R., Magesa, S., Mosha, F., Otieno, S., Carneiro, I., Cox, J., Msuya, E., et al. (2010). Identification of hot spots of malaria transmission for targeted malaria control. *Journal of Infectious Diseases*, 201(11):1764–1774.

Carrel, M., Janko, M., Mwandagalirwa, M. K., Morgan, C., Fwamba, F., Muwonga, J., Tshefu, A. K., Meshnick, S., and Emch, M. (2016). Changing spatial patterns and increasing rurality of hiv prevalence in the democratic republic of the congo between 2007 and 2013. *Health & place*, 39:79–85.

Carrel, M., Patel, J., Taylor, S. M., Janko, M., Mwandagalirwa, M. K., Tshefu, A. K., Escalante, A. A., McCollum, A., Alam, M. T., Udhayakumar, V., et al. (2015). The geography of malaria genetics in the democratic republic of congo: A complex and fragmented landscape. *Social Science & Medicine*, 133:233–241.

CDC (2015). Malaria disease. `https://www.cdc.gov/malaria/about/disease.html`. Accessed: 2017-04-29.

Chaix, B., Merlo, J., and Chauvin, P. (2005a). Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilisation in france. *Journal of Epidemiology and Community Health*, 59(6):517–526.

Chaix, B., Merlo, J., Subramanian, S., Lynch, J., and Chauvin, P. (2005b). Comparison of a spatial perspective with the multilevel analytical approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in malmö, sweden, 2001. *American journal of epidemiology*, 162(2):171–182.

Corbel, V., Akogbeto, M., Damien, G. B., Djenontin, A., Chandre, F., Rogier, C., Moiroux, N., Chabi, J., Padonou, G. G., and Henry, M.-C. (2012). Combination of malaria vector control interventions in pyrethroid resistance area in benin: a cluster randomised controlled trial. *The Lancet infectious diseases*, 12(8):617–626.

Czeher, C., Labbo, R., Arzika, I., and Duchemin, J.-B. (2008). Evidence of increasing leu-phe knockdown resistance mutation in anopheles gambiae from niger following a nationwide long-lasting insecticide-treated nets implementation. *Malaria Journal*, 7(1):189.

Darriet, F., Robert, V., Vien, N. T., Carnevale, P., Organization, W. H., et al. (1984). Evaluation of the efficacy of permethrin impregnated intact and perforated mosquito nets against vectors of malaria.

Davies, T., Field, L., Usherwood, P., and Williamson, M. (2007). Ddt, pyrethrins, pyrethroids and insect sodium channels. *IUBMB life*, 59(3):151–162.

de Araújo, V. E. M., Pinheiro, L. C., de Mattos Almeida, M. C., de Menezes, F. C., Morais, M. H. F., Reis, I. A., Assunção, R. M., and Carneiro, M. (2013). Relative risk of visceral leishmaniasis in brazil: a spatial analysis in urban area. *PLoS Negl Trop Dis*, 7(11):e2540.

DESA, U. (2015). World population prospects: The 2015 revision, key findings and advance tables. *Working PaperNo*.

DHS (2017). Dhs program overview.

Djogbénou, L., Pasteur, N., Akogbeto, M., Weill, M., and Chandre, F. (2011). Insecticide resistance in the anopheles gambiae complex in benin: a nationwide survey. *Medical and veterinary entomology*, 25(3):256–267.

Djouaka, R. F., Bakare, A. A., Coulibaly, O. N., Akogbeto, M. C., Ranson, H., Hemingway, J., and Strode, C. (2008). Expression of the cytochrome p450s, cyp6p3 and cyp6m2 are significantly elevated in multiple pyrethroid resistant populations of anopheles gambiae ss. from southern benin and nigeria. *BMC genomics*, 9(1):538.

Doctor, S. M., Liu, Y., Whitesell, A., Thwai, K. L., Taylor, S. M., Janko, M., Emch, M., Kashamuka, M., Muwonga, J., Tshefu, A., et al. (2016). Malaria surveillance in the democratic republic of the congo: comparison of microscopy, pcr, and rapid diagnostic test. *Diagnostic microbiology and infectious disease*, 85(1):16–18.

Edi, C. V., Koudou, B. G., Jones, C. M., Weetman, D., and Ranson, H. (2012). Multiple-insecticide resistance in anopheles gambiae mosquitoes, southern cote divoire. *Emerging infectious diseases*, 18(9):1508.

FAO (2017). Democratic republic of the congo.

Ferguson, H. M., Dornhaus, A., Beeche, A., Borgemeister, C., Gottlieb, M., Mulla, M. S., Gimnig, J. E., Fish, D., and Killeen, G. F. (2010). Ecology: a prerequisite for malaria elimination and eradication. *PLoS Med*, 7(8):e1000303.

Fontenille, D. Lochouarn, L. (1999). The of the malaria vectorial system in africa. *Parassitologia*, 41:267–271.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gimnig, J. E., Ombok, M., Kamau, L., and Hawley, W. A. (2001). Characteristics of larval anopheline (diptera: Culicidae) habitats in western kenya. *Journal of Medical Entomology*, 38(2):282–288.

Graham, K., Kayedi, M., Maxwell, C., Kaur, H., Rehman, H., Malima, R., Curtis, C., Lines, J., and Rowland, M. (2005). Multi-country field trials comparing wash-resistance of permanet and conventional insecticide-treated nets against anopheline and culicine mosquitoes. *Medical and veterinary entomology*, 19(1):72–83.

Hajat, A., Diez-Roux, A. V., Adar, S. D., Auchincloss, A. H., Lovasi, G. S., O'Neill, M. S., Sheppard, L., and Kaufman, J. D. (2013). Air pollution and individual and neighborhood socioeconomic status: evidence from the multi-ethnic study of atherosclerosis (mesa). *Environmental Health Perspectives (Online)*, 121(11-12):1325.

Hancioglu, A. and Arnold, F. (2013). Measuring coverage in mnch: tracking progress in health for women and children using dhs and mics household surveys. *PLoS medicine*, 10(5):e1001391.

Hargreaves, K., Koekemoer, L., Brooke, B., Hunt, R., Mthembu, J., and Coetzee, M. (2000). Anopheles funestus resistant to pyrethroid insecticides in south africa. *Medical and veterinary entomology*, 14(2):181–189.

Hay, S. I., Guerra, C. A., Gething, P. W., Patil, A. P., Tatem, A. J., Noor, A. M., Kabaria, C. W., Manh, B. H., Elyazar, I. R., Brooker, S., et al. (2009). A world malaria map: Plasmodium falciparum endemicity in 2007. *PLoS Med*, 6(3):e1000048.

Hay, S. I. and Snow, R. W. (2006). The malaria atlas project: developing global maps of malaria risk. *PLoS Med*, 3(12):e473.

Henry, M.-C., Assi, S.-b., Rogier, C., Dossou-Yovo, J., Chandre, F., Guillet, P., and Carnevale, P. (2005). Protective efficacy of lambda-cyhalothrin treated nets in anopheles gambiae pyrethroid resistance areas of cote divoire. *The American journal of tropical medicine and hygiene*, 73(5):859–864.

Hernández-Avila, J. E., Rodríguez, M. H., Betanzos-Reyes, A. F., Danis-Lozano, R., Méndez-Galván, J. F., Velázquez-Monroy, O. J., and Tapia-Conyer, R. (2006). Determinant factors for malaria transmission on the coast of oaxaca state, the main residual transmission focus in mexico. *salud pública de méxico*, 48(5):405–417.

Holderegger, R., Kamm, U., and Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, 21(6):797–807.

Hunter, J. M. et al. (1974). *The geography of health and disease: papers of the first Carolina Geographical Symposium (1974)*. University of North Carolina at Chapel Hill, Dept. of Geography.

Ijumba, J. and Lindsay, S. (2001). Impact of irrigation on malaria in africa: paddies paradox. *Medical and veterinary entomology*, 15(1):1–11.

Ijumba, J., Shenton, F., Clarke, S., Mosha, F., and Lindsay, S. (2002). Irrigated crop production is associated with less malaria than traditional agricultural practices in tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96(5):476–480.

Jayne, T. S., Chamberlin, J., and Headey, D. D. (2014). Land pressures, the evolution of farming systems, and development strategies in africa: A synthesis. *Food Policy*, 48:1–17.

Karch, S. and Mouchet, J. (1992). Anopheles paludis, vecteur important du paludisme au zaïre. *Bulletin de la Société de pathologie exotique*, 85(5):388–389.

Kaufmann, C. and Briegel, H. (2004). Flight performance of the malaria vectors anopheles gambiae and anopheles atroparvus. *Journal of vector ecology*, 29:140–153.

Kearns, R. A. (1993). Place and health: towards a reformed medical geography. *The Professional Geographer*, 45(2):139–147.

Kearns, R. A. and Joseph, A. E. (1993). Space in its place: developing the link in medical geography. *Social Science & Medicine*, 37(6):711–717.

Kelly-Hope, L., Ranson, H., and Hemingway, J. (2008). Lessons from the past: managing insecticide resistance in malaria control and eradication programmes. *The Lancet infectious diseases*, 8(6):387–389.

Krefis, A. C., Schwarz, N. G., Krüger, A., Fobil, J., Nkrumah, B., Acquah, S., Loag, W., Sarpong, N., Adu-Sarkodie, Y., Ranft, U., et al. (2011). Modeling the relationship between precipitation and malaria incidence in children from a holoendemic area in ghana. *The American journal of tropical medicine and hygiene*, 84(2):285–291.

Kristan, M., Fleischmann, H., Della Torre, A., Stich, A., and Curtis, C. (2003). Pyrethroid resistance/-susceptibility and differential urban/rural distribution of anopheles arabiensis and an. gambiae ss malaria vectors in nigeria and ghana. *Medical and veterinary entomology*, 17(3):326–332.

Lawson, A. B. (2013). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. CRC press.

Lengeler, C. et al. (2004). Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database Syst Rev*, 2(2).

Levitz, L., Janko, M., Mwandagalierwa, K., Thwai, K., Likwela, J., Tshefu, A., Emch, M., and Meshnick, S. (2017). Effect of individual- and community-level bed net usage on malaria prevalence among under-fives in the democratic republic of congo. *submitted*.

Lindblade, K., Gimnig, J., Kamau, L., Hawley, W., Odhiambo, F., Olang, G., Ter Kuile, F., Vulule, J., and Slutsker, L. (2006). Impact of sustained use of insecticide-treated bednets on malaria vector species distribution and culicine mosquitoes. *Journal of Medical Entomology*, 43(2):428–432.

Lyimo, E., Takken, W., and Koella, J. (1992). Effect of rearing temperature and larval density on larval survival, age at pupation and adult size of anopheles gambiae. *Entomologia experimentalis et applicata*, 63(3):265–271.

Lynch, C. and Roper, C. (2011). The transit phase of migration: circulation of malaria and its multidrug-resistant forms in africa. *PLoS Med*, 8(5):e1001040.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1):209–220.

Mathias, D. K., Ochomo, E., Atieli, F., Ombok, M., Bayoh, M. N., Olang, G., Muhia, D., Kamau, L., Vulule, J. M., Hamel, M. J., et al. (2011). Spatial and temporal variation in the kdr allele l1014s in anopheles gambiae ss and phenotypic variability in susceptibility to insecticides in western kenya. *Malaria journal*, 10(1):10.

May, J. M. (1950). Medical geography: its methods and objectives. *Geographical Review*, 40(1):9–41.

May, J. M. et al. (1959). *The Ecology of Human Disease.* MD Publications Inc.

Meade, M. and Emch, M. (2010). Medical geography. 3rd.

Meade, M. S. (1977). Medical geography as human ecology: the dimension of population movement. *Geographical Review*, pages 379–393.

Messina, J. P., Taylor, S. M., Meshnick, S. R., Linke, A. M., Tshefu, A. K., Atua, B., Mwandagalirwa, K., and Emch, M. (2011). Population, behavioural and environmental drivers of malaria prevalence in the democratic republic of congo. *Malaria journal*, 10(1):1.

Moores, G. and Bingham, G. (2005). Use of'temporal synergism'to overcome insecticide resistance. *Outlooks on Pest Management*, 16(1):7–9.

Morland, K., Wing, S., and Roux, A. D. (2002). The contextual effect of the local food environment on residents diets: the atherosclerosis risk in communities study. *American journal of public health*, 92(11):1761–1768.

Munga, S., Minakawa, N., Zhou, G., Mushinzimana, E., Barrack, O.-O. J., Githeko, A. K., and Yan, G. (2006). Association between land cover and habitat productivity of malaria vectors in western kenyan highlands. *The American journal of tropical medicine and hygiene*, 74(1):69–75.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, 70(12):3321–3323.

N'Guessan, R., Corbel, V., Akogbéto, M., and Rowland, M. (2007). Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, benin. *Emerging infectious diseases*, 13(2):199–206.

Paaijmans, K., Takken, W., Githeko, A., and Jacobs, A. (2008). The effect of water turbidity on the near-surface water temperature of larval habitats of the malaria mosquito anopheles gambiae. *International journal of biometeorology*, 52(8):747–753.

Paaijmans, K. P., Blanford, S., Bell, A. S., Blanford, J. I., Read, A. F., and Thomas, M. B. (2010). Influence of climate on malaria transmission depends on daily temperature variation. *Proceedings of the National Academy of Sciences*, 107(34):15135–15139.

Paaijmans, K. P., Blanford, S., Chan, B. H., and Thomas, M. B. (2011). Warmer temperatures reduce the vectorial capacity of malaria mosquitoes. *Biology letters*, page rsbl20111075.

Padonou, G. G., Sezonlin, M., Ossé, R., Aizoun, N., Oké-Agbo, F., Oussou, O., Gbédjissi, G., and Akogbéto, M. (2012). Impact of three years of large scale indoor residual spraying (irs) and insecticide treated nets (itns) interventions on insecticide resistance in anopheles gambiae sl. in benin. *Parasites & vectors*, 5(1):72.

Parr, J. B., Verity, R., Doctor, S. M., Janko, M., Carey-Ewend, K., Turman, B. J., Keeler, C., Slater, H. C., Whitesell, A. N., Mwandagalirwa, K., et al. (2016). Pfhrp2-deleted plasmodium falciparum parasites in the democratic republic of congo: A national cross-sectional survey. *Journal of Infectious Diseases*, page jiw538.

Patel, J. C., Taylor, S. M., Juliao, P. C., Parobek, C. M., Janko, M., Gonzalez, L. D., Ortiz, L., Padilla, N., Tshefu, A. K., Emch, M., et al. (2014). Genetic evidence of importation of drug-resistant plasmodium falciparum to guatemala from the democratic republic of the congo. *Emerging infectious diseases*, 20(6):932.

Pates, H. and Curtis, C. (2005). Mosquito behavior and vector control. *Annu. Rev. Entomol.*, 50:53–70.

Patz, J. A., Daszak, P., Tabor, G. M., Aguirre, A. A., Pearl, M., Epstein, J., Wolfe, N. D., Kilpatrick, A. M., Foufopoulos, J., Molyneux, D., et al. (2004). Unhealthy landscapes: policy recommendations on land use change and infectious disease emergence. *Environmental health perspectives*, pages 1092–1098.

Patz, J. A., Graczyk, T. K., Geller, N., and Vittor, A. Y. (2000). Effects of environmental change on emerging parasitic diseases. *International journal for parasitology*, 30(12):1395–1405.

Perez-Heydrich, C., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M., and Emch, M. E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics*, 70(3):731–741.

Pinto, J., Lynd, A., Vicente, J. L., Santolamazza, F., Randle, N. P., Gentile, G., Moreno, M., Simard, F., Charlwood, J. D., do Rosário, V. E., et al. (2007). Multiple origins of knockdown resistance mutations in the afrotropical mosquito vector anopheles gambiae. *PLoS One*, 2(11):e1243.

Plowe, C. V., Djimde, A., Bouare, M., Doumbo, O., and Wellems, T. E. (1995). Pyrimethamine and proguanil resistance-conferring mutations in plasmodium falciparum dihydrofolate reductase: polymerase chain reaction methods for surveillance in africa. *American Journal of Tropical Medicine and Hygiene*, 52(6):565–568.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.

Prothero, R. M. (1977). Disease and mobility: a neglected factor in epidemiology. *International journal of epidemiology*, 6(3):259–267.

Protopopoff, N., Van Bortel, W., Marcotty, T., Van Herp, M., Maes, P., Baza, D., D'Alessandro, U., and Coosemans, M. (2007). Spatial targeted vector control in the highlands of burundi and its impact on malaria transmission. *Malaria Journal*, 6(1):158.

Protopopoff, N., Verhaeghen, K., Van Bortel, W., Roelants, P., Marcotty, T., Baza, D., DAlessandro, U., and Coosemans, M. (2008). A significant increase in kdr in anopheles gambiae is associated with an intensive vector control intervention in burundi highlands. *Tropical medicine & international health*, 13(12):1479–1487.

Ranson, H., Abdallah, H., Badolo, A., Guelbeogo, W. M., Kerah-Hinzoumbé, C., Yangalbé-Kalnoné, E., Sagnon, N., Simard, F., and Coetzee, M. (2009). Insecticide resistance in anopheles gambiae: data from the first year of a multi-country study highlight the extent of the problem. *Malaria Journal*, 8(1):299.

Ranson, H., NGuessan, R., Lines, J., Moiroux, N., Nkuni, Z., and Corbel, V. (2011). Pyrethroid resistance in african anopheline mosquitoes: what are the implications for malaria control? *Trends in parasitology*, 27(2):91–98.

Roper, C., Pearce, R., Nair, S., Sharp, B., Nosten, F., and Anderson, T. (2004). Intercontinental spread of pyrimethamine-resistant malaria. *Science*, 305(5687):1124–1124.

Ruktanonchai, N. W., DeLeenheer, P., Tatem, A. J., Alegana, V. A., Caughlin, T. T., zu Erbach-Schoenberg, E., Lourenço, C., Ruktanonchai, C. W., and Smith, D. L. (2016). Identifying malaria transmission foci for elimination using human mobility data. *PLoS Comput Biol*, 12(4):e1004846.

Russell, T. L., Govella, N. J., Azizi, S., Drakeley, C. J., Kachur, S. P., and Killeen, G. F. (2011). Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural tanzania. *Malaria journal*, 10(1):1.

Singh, B., Bobogare, A., Cox-Singh, J., Snounou, G., Abdullah, M. S., and Rahman, H. A. (1999). A genus-and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies. *The American journal of tropical medicine and hygiene*, 60(4):687–692.

Sinka, M. E., Bangs, M. J., Manguin, S., Coetzee, M., Mbogo, C. M., Hemingway, J., Patil, A. P., Temperley, W. H., Gething, P. W., Kabaria, C. W., et al. (2010). The dominant anopheles vectors of human malaria in africa, europe and the middle east: occurrence data, distribution maps and bionomic précis. *Parasites & vectors*, 3(1):1.

Stoddard, S. T., Morrison, A. C., Vazquez-Prokopec, G. M., Soldan, V. P., Kochel, T. J., Kitron, U., Elder, J. P., and Scott, T. W. (2009). The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl Trop Dis*, 3(7):e481.

Stresman, G. H. (2010). Beyond temperature and precipitation: ecological risk factors that modify malaria transmission. *Acta Tropica*, 116(3):167–172.

Stump, A. D., Atieli, F. K., Vulule, J. M., and Besansky, N. J. (2004). Dynamics of the pyrethroid knockdown resistance allele in western kenyan populations of anopheles gambiae in response to insecticide-treated bed net trials. *The American journal of tropical medicine and hygiene*, 70(6):591–596.

Tatem, A. J., Huang, Z., Narib, C., Kumar, U., Kandula, D., Pindolia, D. K., Smith, D. L., Cohen, J. M., Graupe, B., Uusiku, P., et al. (2014). Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria journal*, 13(1):52.

Tatem, A. J. and Smith, D. L. (2010). International population movements and regional plasmodium falciparum malaria elimination strategies. *Proceedings of the National Academy of Sciences*, 107(27):12222–12227.

Taylor, S. M., Antonia, A. L., Parobek, C. M., Juliano, J. J., Janko, M., Emch, M., Alam, M. T., Udhayakumar, V., Tshefu, A. K., and Meshnick, S. R. (2013). Plasmodium falciparum sulfadoxine resistance is geographically and genetically clustered within the dr congo. *Scientific reports*, 3:1165.

Taylor, S. M., Van Eijk, A. M., Hand, C. C., Mwandagalirwa, K., Messina, J. P., Tshefu, A. K., Atua, B., Emch, M., Muwonga, J., Meshnick, S. R., et al. (2011). Quantification of the burden and consequences of pregnancy-associated malaria in the democratic republic of the congo. *Journal of Infectious Diseases*, 204(11):1762–1771.

Trape, J.-F., Tall, A., Diagne, N., Ndiath, O., Ly, A. B., Faye, J., Dieye-Ba, F., Roucher, C., Bouganali, C., Badiane, A., et al. (2011). Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. *The Lancet infectious diseases*, 11(12):925–932.

Verity, R. and Nichols, R. (2015). Estimating k in genetic mixture models. *bioRxiv*, page 022988.

Vittor, A. Y., Pan, W., Gilman, R. H., Tielsch, J., Glass, G., Shields, T., Sánchez-Lozano, W., Pinedo, V. V., Salas-Cobos, E., Flores, S., et al. (2009). Linking deforestation to malaria in the amazon: characterization of the breeding habitat of the principal malaria vector, anopheles darlingi. *The American journal of tropical medicine and hygiene*, 81(1):5–12.

Vulule, J., Beach, R., Atieli, F., McAllister, J., Brogdon, W., Roberts, J., Mwangi, R., and Hawley, W. (1999). Elevated oxidase and esterase levels associated with permethrin tolerance in anopheles gambiae from kenyan villages using permethrin-impregnated nets. *Medical and veterinary entomology*, 13(3):239–244.

WHO (2005). World malaria report 2005. Technical report, World Health Organization and UNICEF.

WHO et al. (2007). Insecticide-treated mosquito nets: a position statement. 2007. *Geneva: WHO-Global Malaria Programme Google Scholar*.

WHO et al. (2011). Guidelines for monitoring the durability of long-lasting insecticidal mosquito nets under operational conditions.

WHO et al. (2016). World malaria report 2015: summary.

Wongsrichanalai, C., Pickard, A. L., Wernsdorfer, W. H., and Meshnick, S. R. (2002). Epidemiology of drug-resistant malaria. *The Lancet infectious diseases*, 2(4):209–218.

Yadouleton, A. W., Padonou, G., Asidi, A., Moiroux, N., Bio-Banganna, S., Corbel, V., N'guessan, R., Gbenou, D., Yacoubou, I., Gazard, K., et al. (2010). Insecticide resistance status in anopheles gambiae in southern benin. *Malaria journal*, 9(1):83.

Yang, G.-J., Brook, B. W., Whelan, P. I., Cleland, S., and Bradshaw, C. J. (2008). Endogenous and exogenous factors controlling temporal abundance patterns of tropical mosquitoes. *Ecological Applications*, 18(8):2028–2040.

Zhang, Y., Bi, P., and Hiller, J. E. (2008). Climate change and the transmission of vector-borne diseases: a review. *Asia-Pacific Journal of Public Health*, 20(1):64–76.