

Estimating the patterns and consequences of malaria transmission dynamics on fine spatial scales

INAUGURALDISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

Von

Josephine Malinga

Basel, 2021

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

<https://edoc.unibas.ch/>

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von
Prof. Jurg Utzinger, Dr. Amanda Ross, Dr. Lucy Okell.

Basel den 19. Februar 2019

Prof. Dr. Martin Spiess

Dekan

To my parents,
William and Ruth Mallingah.

And my all-time favourites,
Alice and Derrick Malinga.

Summary

Plasmodium falciparum is the leading cause of malaria infection and a major cause of morbidity and mortality across the globe, particularly in the African region. The burden of malaria is unevenly distributed, with some countries, districts or even households within villages harboring a disproportionately higher burden. There is an intricate relationship between the mosquito vector, humans and the parasites they carry, and how they interact with the environment. Small movements on a fine-scale lead to the patterns observed in the community. Quantifying transmission dynamics on a fine-scale, how malaria infections spread locally and the processes leading to the observed spatial and temporal distribution patterns is important for many aspects of malaria epidemiology, in particular, the design of targeted interventions against malaria, the design of studies to evaluate the effectiveness of vector control in the field, and the parameterization of mathematical models to predict the likely impact of interventions for settings where data is not available.

Mathematical and statistical models have been developed to quantify fine scale malaria transmission dynamics and investigate the effects of interventions. Since data on the spread of vectors and parasites is challenging to collect, it is not available from many endemic settings for analytic methods to provide estimates, or to validate model predictions. Due to variability between settings, findings from one setting cannot be easily generalized. There is thus a need to involve methods that can extract information from imperfect but available datasets, to make the most of the existing data sources from settings with a variety of characteristics.

The overall aim of this thesis was to use statistical and mathematical modelling approaches to characterize fine scale malaria transmission dynamics and their consequences on the measurement of heterogeneity on a local scale for targeted interventions.

Chapter 2 used an established comprehensive simulator of malaria epidemiology developed at the Swiss Tropical and Public Health Institute (Swiss TPH) to predict the proportion of malaria infections that are in mosquitoes and humans and how this varies by setting specific characteristics. A substantial proportion of infections was predicted to be in mosquitoes, to vary with setting specific characteristics, and in response to interventions. The predictions also highlighted the role of the dynamics of infections in humans and mosquitoes following the introduction or scale-up of interventions.

In Chapter 3, a statistical model which takes into account movement between houses in a village to estimate how far and where mosquitoes fly to in the presence of spatial repellents was developed. This was a secondary use of data on mosquito densities. The method

evaluation using simulation showed that the model could be used as a potential tool to gain information on mosquito movement, estimating the distance between the houses the mosquitoes were repelled from and the houses they move to, the proportion of mosquitoes repelled, and the proportion of repelled mosquitoes moving to another house as opposed to somewhere outside. However, the trial data needs to contain sufficient information to be able to disentangle the effects of the underlying processes and provide accurate estimates for all the parameters. We found that additional data on the total number of mosquitoes and sufficient numbers of mosquitoes repelled were required in the case of the motivating trial. Findings from the simulations could inform the design of studies and help quantify criteria for trial settings.

In Chapter 4, a simulation method was developed and applied to data on parasite genotypes from Kilifi County, Kenya. A previous study found an interaction between time and geographical distance on the genetic difference between pairs of parasite genotypes: genetic differences were lower for pairs of parasite genotypes which were evaluated within a shorter time interval and found within a shorter geographic distance apart. A stochastic individual-based model of malaria infections, people and homesteads was developed and fitted to the genetic differences in order to investigate hypotheses and parameter values consistent with the observed interaction.

The observed interaction could be reproduced by the individual-based model. Although hypothesis about immunity to previously seen genotypes, and or a limit on the number of current infections per individual could not be ruled out, they were not necessary to account for the observed interaction. The mean geographical distance between parent and offspring infections was estimated to be 0.40km (95%CI 0.24 – 1.20), in the base model. This was the first modeling study that we know of which has attempted to estimate parameter values and test hypotheses from malaria genotyping data with a low coverage of infections in a setting with moderate transmission. The findings glean some insights on how simulation can be used in quantifying factors driving transmission, and in estimating unknown parameters when analytic methods are limited.

The work in Chapter 5 uses the simulation model developed in Chapter 4 to investigate how the method chosen, local seasonality and movement of infections influence the detection of areas of higher transmission on fine spatial scales for targeted interventions. Our findings show that the identification of hotspots was less accurate when there was a gentle decay in risk from the hotspot boundary, the hotspot was irregularly shaped, there was seasonality in the area or when the mean distance between parent and offspring infections was longer. The findings highlight the importance of setting characteristics, the choice of outcome, and method of detection on the accuracy of identifying areas of higher transmission for targeted

interventions. The underlying fine scale transmission dynamics should be taken into account when performing and interpreting analyses of heterogeneity for targeted interventions.

Taken as a whole, this thesis provides information on the characteristics of transmission dynamics on a fine scale. It highlights that a substantial proportion of malaria infections are in mosquitoes, and places emphasis on the role that vectors, and humans play in the spread of infections and the implications of fine scale movement for the measurement of heterogeneity for targeted interventions. The estimates have implications for the design and evaluation of malaria control and elimination interventions.

Acknowledgements

First and foremost, I would like to thank my supervisor PD Dr. Amanda Ross. It is impossible to appreciate her enough for taking me on as her student. Her support, constant supervision, and mentorship for the last three years have been invaluable. I'm greatly indebted to her for my growth, in both my professional career and personally. It was an extreme honour.

I would like to express my gratitude to Professor Marcel Tanner and Professor Jurg Utzinger for a warm welcome to the institute, and for being part of my PhD supervisory committee.

Many thanks to everyone in the Biostatistics unit, for their support and encouraging insights. Special thanks to Christian Schindler for giving me the opportunity to assist in tutoring students during the Biostatistics courses. It was a wonderful experience.

I am deeply grateful to Christine Mensch and all the colleagues in the training department for their immense help. I would also like to thank Nora, Dagmar, and Laura for all their continued support. With special thanks to Christian Heuss for the cover photo.

I thank all the students and staff who supported me fully, assisting where necessary, during my tenure as a PhD Student representative. During my PhD, I also had the privilege of being a mentee in the Antelope Program for female researchers at the University of Basel. The life skills, career advice and coaching provided are irreplaceable. I huge appreciation to Dr. Lucy Tusting who accepted my request to be my external expert during this program. It was my pleasure.

I would like to thank a few people; Professor Phillip Bejon, Dr Irene Omedo and Dr Polycarp Mogeni from the KEMRI/Wellcome Trust Programme in Nairobi for your help with the datasets, modeling and continued correspondence, Dr Emelda Okiro and Dr Dejan Zurovac, for your constant belief in me and continued mentorship, Dr Sarah Moore and Dr Marta Maia from the Ifakara Health Institute for providing us with datasets and interesting discussions.

I will always treasure the great moments that I shared with the friends I was so fortunate to make during my PhD. The world is a better place with you in it. Sammy Khaghayi, Oliver Baerenbold, Henry Ntuku, Tessa Oraro, Angela Lazarova, Marta Palmeirim, Astrid Knoblauch, Nerina Vischer, Francis Mhimbira, Wyvine Bapolisi, Giovanni Francesco, Miriam Karinja, Sokhna Thiam, Lea Multerer, Christine Burli, Manuella Runge, Yeromin Mlacha, Nancy Matowo, Hala Allabadi, Shala Mhlanga, Anne-Christine Heedeger, Anna Fesser, Harry Mapesi, Tamsin Lee, Manuela Runge and Corine Karema.

To my parents William and Ruth, my sister Alice, my brother Derrick and my dear friends, Patricia, Racheal, Antoinette, Sandra, Chemutai, Salome, Sylvia, Ngatia, Becky, Betty, Deola and Luke. Thank you for being there since my day one! I appreciate you all!

I recognize that this PhD would not have been possible without funding from Gottfried und Julia Bangerter-Rhyner Stiftung, the Novartis Foundation for Medical Biological Research, the Amt für Ausbildungsbeiträge of the Kanton Basel-Stadt, the Rudolf Geigy Stiftung and the Bill and Melinda Gates Foundation.

Contents

Summary	iv
Acknowledgements	vii
Contents	ix
1. Introduction	1
1.1. Transmission and Life Cycle	1
1.2. Heterogeneity in fine-scale malaria transmission.....	2
1.3. The need to understand malaria transmission on a local scale	3
1.4. Investigating fine scale malaria dynamics using mathematical and statistical models 5	
1.5. Objectives of the Thesis	6
1.6. References.....	7
2. What proportion of <i>Plasmodium falciparum</i> and <i>Plasmodium vivax</i> malaria infections are in mosquitoes?.....	12
2.1. Abstract	13
2.2. Background.....	14
2.3. Methods.....	15
2.4. Results.....	19
2.5. Discussion.....	28
2.6. Conclusion.....	29
2.7. References.....	30
3. Can trials of spatial repellents be used to estimate mosquito movement within a village?	36
3.1. Abstract	37
3.3. Methods.....	39
3.4. Results.....	46
3.5. Discussion.....	52
3.6. Conclusion.....	54
3.7. References.....	55

4. Investigating the drivers of the spatio-temporal patterns of genetic differences between <i>Plasmodium falciparum</i> malaria infections in Kilifi County, Kenya	59
4.1. Abstract	60
4.2. Introduction	61
4.3. Methods	63
4.5. Results	72
4.6. Discussion	78
4.7. Conclusion	80
4.8. References	81
4.9. Supplementary information	86
5. Detecting malaria hotspots: the role of movement, seasonality, shape and statistical method	88
5.1. Summary	89
5.2. Introduction	90
5.3. Methods	91
5.4. Results	95
5.5. Discussion	100
5.6. Conclusion	101
5.7. References	102
6. Discussion	105
6.1. Summary of Main Findings	106
6.2. The use of mathematical and statistical models	107
6.3. The importance of setting characteristics	108
6.4. Potential impact of recent advances in the detection of infections	108
6.5. Implications of fine scale movement	109
6.6. Outlook and future work	111
6.8. References	112

List of Figures

Figure 1.1 Life Cycle of the <i>Plasmodium</i> malaria Parasites.....	2
Figure 2.1 Predicted <i>P.falciparum</i> infections in mosquitoes and humans by transmission intensity for different seasonal patterns	20
Figure 2.2 Seasonal transmission and the predicted proportion of <i>P.falciparum</i> infections in mosquitoes	20
Figure 2.3 Predicted <i>P.falciparum</i> infections in mosquitoes and humans by transmission intensity for different human blood indices	21
Figure 2.4 Predicted <i>P.falciparum</i> infections in mosquitoes and humans by transmission intensity for different model variants.....	22
Figure 2.5 Predicted <i>P.vivax</i> infections in mosquitoes and humans by transmission intensity for different seasonal patterns.....	23
Figure 2.6 Validation of the predictions for the number of infections in humans and mosquitoes with observed data.....	25
Figure 2.7 The predicted impact of a mass campaign of long-lasting insecticidal nets on the proportion of infections in mosquitoes. <i>P.falciparum</i> (top row) and <i>P.vivax</i> (bottom row)...	26
Figure 2.8 The predicted impact of mass drug administration on the proportion of infections in mosquitoes. <i>P.falciparum</i> (top row) and <i>P.vivax</i> (bottom row)	27
Figure 3.1 Examples of the distributions of geographical distances given by the normal kernel.....	42
Figure 3.2 The ability of the model to return known parameter values.	47
Figure 3.3 The ability of the model to return the known values for β , the proportion of mosquitoes repelled from houses using spatial repellents.	48
Figure 3.4 The ability of the model to return known values for φ , the proportion of mosquitoes repelled that go to households as opposed to elsewhere.....	49

Figure 3.5 Estimated and Known values for λ , the parameter for distance between households moved by the mosquito. (mean = $\lambda \sqrt{2/\pi}$).....	50
Figure 3.6 The ability of the model to return known values for λ , the parameter for distance between households moved by mosquitoes	50
Figure 4.1 Examples of the half-normal distribution probability density function for positive values of the distance between parent and offspring infections. The dotted lines mark the mean of the distribution; Red line = 0.40km, Blue line = 1.20km.	66
Figure 4.2 Ability of the method to recover known parameter values from simulated data .	72
Figure 4.3 Patterns of the log likelihood for different values of recombination for the base model	73
Figure 4.4 Patterns of the log likelihood by σ , the parameter for the distance for the different model variants.....	75
Figure 4.5 Patterns of the log likelihood by σ , the parameter for the distance for data simulated from mixture distributions.....	75
Figure 4.6 Predicted effect of time and distance interaction on the number of SNPs different between pairs of infections.	76
Figure 4.7 Plot of residuals by geographic distance.....	77
Figure 4.8 Patterns of the log likelihood by σ , the parameter for distance for different values of the input parameters.	78
Figure 5.1 Examples of hotspots with varied features.....	93
Figure 5.2 An example of simulated areas of higher underlying transmission and detected hotspots of prevalence	96
Figure 5.3 Simulated of detected hotspots with varying distance between parent and offspring infection after a period of 1 year	97
Figure 5.4 Example of simulated hotspots with by season	97

List of Tables

Table 2.1 Scenarios simulated*	17
Table 2.2 Vector parameters	18
Table 2.3 Data sources for model validation.....	24
Table 3.1 Quantities in the models.....	44
Table 3.2 Simulated scenarios of trial characteristics to evaluate the method.....	45
Table 3.3 Trial Characteristics for the three villages	46
Table 3.4 Parameter estimates using the trial data	51
Table 4.1 Quantities in the simulation model	66
Table 4.2 Inputs to the model	68
Table 4.3 Simulated scenarios.....	70
Table 4.4 Estimated mean distance between parent and offspring infections for each model variants.....	74
Table 5.1 Simulated scenarios.....	91
Table 5.2 Methods included in the evaluation.....	94
Table 5.3 Assessment of measures to detect hotspots.....	99

Abbreviations

ACT	artemisinin combination therapies
EIP	extrinsic incubation period
EIR	entomological inoculation rate
GIS	geographic information system
HBI	human blood index
HDSS	health demographic surveillance system
IBD	identity by descent
IBS	identity by state
IRS	indoor residual spraying
ITN	insecticide treated bed nets
LLIN	long lasting insecticidal net
MAF	minor allele frequency
MDA	mass drug administration
MMRR	mosquito mark-release-recapture
MOI	multiplicity of infection
PCR	polymerase chain reaction
RBC	red blood cells
RCD	reactive case detection
RDT	rapid diagnostic test
SNP	single nucleotide polymorphism
SR	sporozoite rate
WHO	world health organization

Chapter 1

1. Introduction

Malaria in humans is caused by a single celled protozoan parasite of the *Plasmodium* genus. *Plasmodium falciparum* is the most virulent of the four main species that cause malaria, is predominantly found in the African region, and together with *Plasmodium vivax* account for most of the malaria cases and deaths reported worldwide (1,2). According to the World Health Organization (WHO), approximately 3 billion people are still at risk of malaria infection globally (2).

Between 2000 and 2017, significant declines in the burden of malaria have been observed from many endemic settings (2,3). This has mainly been attributed to the substantial investments and efforts towards malaria control and case management, including the scale up of effective interventions such as insecticide treated bed nets (ITNs), indoor residual spraying (IRS) and artemisinin combination therapies (ACT's) (2–6). However, malaria still remains one of the greatest global health challenges (7,8). In many sites which have seen decreases, malaria transmission continues and in some sites, the burden of malaria morbidity, mortality and incidence rates has either remained stable or increased over time (2,8,9).

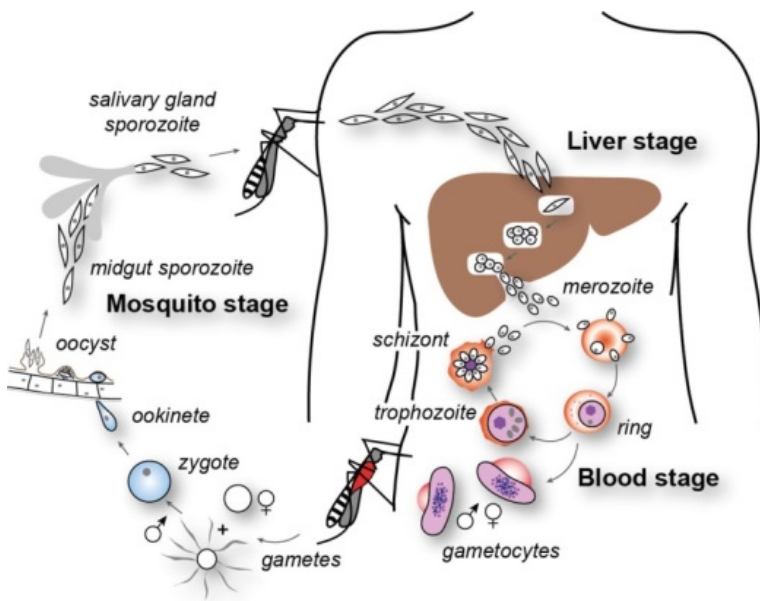
Understanding the complex relationship between malaria vectors, hosts and parasites, and factors affecting their interaction is imperative, if the recent successes in reducing malaria transmission are to be sustained (3,10,11).

1.1. Transmission and Life Cycle

Plasmodium falciparum malaria is transmitted through the bite of a female *Anopheles* mosquito. Transmission may occur when a mosquito bites an infected host during a blood meal and after 10 – 15 days the mosquito may transfer parasites to another host (12). After a blood meal, sporozoites from the mosquito's salivary glands are inoculated into the bloodstream from where they migrate to the liver of the human host. The hepatocytes (liver cells) become infected and following the development of parasites, rupture to release merozoites which then invade the red blood cells (RBCs). The sequence of reproduction continues as infected RBCs rupture, releasing thousands of newly formed merozoites into the blood stream. The clinical symptoms associated with malaria including fever and chills arise during this stage. Only a small proportion of the merozoites differentiate into the sexual forms

of the parasite, the male and female gametocytes, which are then taken up by the mosquito during a subsequent blood meal. In the mosquito gut, the gametocytes develop into male and female gametes and then fuse to form a zygote. The zygote then develops into an oocyst which then bursts to release sporozoites which migrate to the mosquito's salivary glands, completing the transmission cycle (Figure 1). During the sexual development stages in the mosquito, gametes from genetically distinct parasite clones could recombine, giving rise to an offspring that is different from both parents.

Figure 1.1 Life Cycle of the *Plasmodium* malaria Parasites



Adapted from Cowman *et al* (13)

For *Plasmodium vivax* the parasite can stay dormant in the form of hypozoites in the liver cells for days, months or even years after first inoculation causing relapses of parasitemia and incidences of disease (14).

1.2. Heterogeneity in fine-scale malaria transmission

Spatial and temporal variations in the transmission and consequences of malaria infections have been observed at varying scales. Marked heterogeneities exist in both malariological outcomes and the factors that drive transmission at regional, national and local levels. Even at micro-epidemiological scales such as within villages and between individuals within households, the risk of malaria infection is heterogeneous (15–19). At such scales, some individuals or a cluster of households within a study site tend to carry a disproportionately higher burden, with individuals in the surrounding areas having fewer or no instances of infection (16,18–22). This is partly due to fine scale variations in the abundance and behaviour

of vector populations (23), which is driven by environmental variables such as distance to the nearest mosquito breeding site or vegetation (24–26), and human behavioural, economic and genetic factors such as access to treatment, knowledge of malaria signs and symptoms, socio-economic status, structural features of housing, and the coverage of interventions (3,27,28).

A focus of malaria transmission is a defined geographical area situated within a region which is formerly or currently malarious, and is characterized by epidemiological conditions necessary to support the transmission of malaria (22). Malaria hotspots or pockets of transmission are smaller areas within a focus of transmission where the level of transmission is significantly higher than the average in the neighbouring areas (20,22).

1.3. The need to understand malaria transmission on a local scale

Ultimately, the spatial and temporal patterns observed for different metrics at the community level are made up of tiny movements of individual mosquitoes, human hosts and the parasites inside them. Vector dispersal is the most frequent pathway through which infections spread (29–31), and mosquitoes have been shown to fly for distances ranging from a few hundred meters to more than a few kilometres (32,33). Human movement could also contribute substantially to the transfer of parasites, beyond the range of vector dispersal (34). These fine-scale transmission dynamics have implications for many areas in malaria epidemiology.

As transmission declines, patchy spatial patterns, driven mainly by heterogeneous exposure to mosquito bites become more evident and might require a shift in efforts towards focused control (35,36). Blanket interventions have proved efficacious in recent decades (3), but the changing epidemiology of malaria following declines in transmission over time has resulted into persistence heterogeneous patterns which could be targeted with interventions (25,37,38).

At fine resolutions, this distance is mediated mostly by mosquito dispersal in search of blood meals and ovipositing sites. For instance, if mosquitoes fly for only short distances then transmission is only within a hotspot and the effects of targeted interventions can only be observed within this area. If mosquitoes disperse for distances further than the hotspot boundary, then targeted interventions could potentially have community wide effects. Understanding how malaria infections spread locally and the processes leading to the observed spatial and temporal distribution patterns is important for the design of interventions aiming to reduce and interrupt transmission by targeting foci where there is fine scale heterogeneity.

Estimates of fine-scale transmission dynamics are also important for parameterizing mathematical models for predicting the impact of interventions. Since Ross and MacDonald's models in the early 1900's (39,40), many models which investigate malaria transmission dynamics have assumed an enclosed system of humans, vectors and parasites in which movement of mosquitoes over geographical space has not been considered. This uncertainty raises a need to develop spatially explicit frameworks to model the movement of vectors, and as a consequence, the malaria parasites they carry. In recent decades, fine scale spatial heterogeneities have been incorporated in both statistical and mathematical models (41–44), for instance, including the interaction of mosquito dispersal with environmental heterogeneity and vector control interventions like spatial repellents (44,45)(44). Models which account for some aspects of mosquito movement have been developed (33,44,45). Some of these studies have quantified the connectivity between adjacent regions, either movement between patches or zones, particularly those of varying risk, providing information which could potentially guide the choice of interventions (44,45).

Ignoring vector dispersal in the design and analysis of trials of the effects of intervention strategies might underestimate community wide effects or fail to detect differences between study arms due to contamination (5). A further implication of the distances between parent and offspring infections is the measurement of hotspots for targeting interventions. The methods to detect heterogeneity may perform less well when there are longer distances between parent and offspring infections. Results from hotspot targeted interventions are varied (46,47) which may be in part be due to the characteristics and measurement of the hotspots themselves.

Fine-scale movement of infections may also impact the estimation of the spread of drug resistance (3,48). Mixing of mosquito sub-populations is heterogeneous and this influences the rate of gene flow (49).

There still remains uncertainty on how best to quantify the spread of malaria parasites across local scales, and how relevant this is to the measurement of heterogeneity in fine-scale transmission and in evaluating the effect of interventions. However, there is limited information on patterns of transmission dynamics on a fine scale.

If we are interested in reducing or interrupting malaria transmission, then we need information on how infections are spreading. Specifically, there is a need to know i) where the malaria infections are, (ii) how far infections spread and (iii) the consequences of these processes on the measurement of heterogeneity for targeted interventions.

1.4. Investigating fine scale malaria dynamics using mathematical and statistical models

Epidemiological and statistical methods have been applied in the estimation and quantification of malaria at fine scale, often using routine surveillance or survey health data to obtain realistic estimates (50–54). Mathematical modelling on the other hand provides a framework through which we can explore the anticipated effects of different aspects of malaria epidemiology or interventions against malaria, on a range of outcome measures [47–50]. Mathematical models include both observable and unobservable transmission dynamics, from which biologically plausible mechanisms can be developed. Models which have both a biological structure and are fitted to data contain both mathematical and statistical elements, with the advantage of realistic parameter estimates of meaningful quantities.

Both mathematical and statistical approaches have previously been used to characterize aspects of transmission using data such as the duration of infections (55,56), quantifying heterogeneity, and estimating transmission levels using measures such as the prevalence or number of infectious bites per person per year, reproductive number (R_0), and the effects of control measures.

More recently, efforts have been made to estimate the spread of infections. Using genomic data, cross-sectional studies have found fine-scale spatial clustering of related malaria parasites, showing very little parasite movement (57–61), others have found a high degree of parasite mixing (62), which is an indication of infection flow. Genetic markers provide an avenue through which parasites can be identified and tagged for further monitoring and analysis, for instance, to improve surveillance systems (e.g. detecting and tracking drug resistance), distinguish between local and imported transmission, and can also be used as an independent measure to validate changes in the transmission intensity and the impact of control programs [53,54]. Network models have been developed to investigate transmission pathways in Swaziland [55] and Zanzibar [56] or using identity by descent (IBD) or identity by state (IBS) models to estimate relatedness between parasite genotypes [57–59].

One limitation is with data sources. Data that are sufficient for these methods in terms of the proportion of malaria infections sampled and low transmission are available from only a few sites. Similarly, vector dispersal studies are challenging to carry out and have limitations. These are laborious and costly, and methods for collecting and sampling mosquitoes such as mark and recapture studies (MMRR) and genetic studies are not standardized. It is hence not easy to generalize across settings, since the epidemiology of malaria varies with site specific characteristics. Therefore, methods that can extract information from imperfect but available

datasets are needed. To make the most of the available data sources, there is a need to involve models with both mathematical and statistical aspects.

This thesis uses both statistical and mathematical modelling approaches to develop models and fit them to data, to estimate quantities which would have otherwise not been possible without adding biological structure. I use available data from observational studies with genotyped samples, field-trials of spatial repellents for mosquitoes and surveys. I validate the models using simulation and the predictions from the models using field data.

1.5. Objectives of the Thesis

The overall aim of this thesis was to use statistical and mathematical modelling approaches to characterize fine scale malaria transmission dynamics and their consequences on the measurement of fine scale heterogeneity.

The specific objectives were:

1. Predict the proportion of malaria infections that are in mosquitoes and humans (Chapter 2)
2. Develop and validate a statistical model to estimate movement of vectors within three villages from data collected for a trial of spatial repellents in Kilombero Valley, Tanzania (Chapter 3)
3. Develop an individual-based stochastic simulation model of malaria infections to investigate characteristics of transmission, which drive the observed spatial and temporal patterns of genetic differences between *Plasmodium falciparum* malaria infections in Kilifi County, Kenya (Chapter 4)
4. Assess and evaluate the effect of movement of infections on the performance of different common statistical measures of heterogeneity in malaria transmission (Chapter 5)

1.6. References

1. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis*. 2004 Jun;4(6):327–36.
2. WHO. World Malaria Report 2018. WHO, 2018.
3. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015 Oct 8;526(7572):207–11.
4. Fegan GW, Noor AM, Akhwale WS, Cousens S, Snow RW. Effect of expanded insecticide-treated bednet coverage on child survival in rural Kenya: a longitudinal study. *Lancet*. 2007 Sep 22;370(9592):1035–9.
5. Lines J, Kleinschmidt I. Combining malaria vector control interventions: some trial design issues. *Pathog Glob Health*. 2013 Jan;107(1):1–4.
6. Takken W. Do insecticide-treated bednets have an effect on malaria vectors? *Trop Med Int Health TM IH*. 2002 Dec;7(12):1022–30.
7. Murray CJ, Rosenfeld LC, Lim SS, Andrews KG, Foreman KJ, Haring D, et al. Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet*. 2012 Feb;379(9814):413–31.
8. WHO. WHO | World Malaria Report 2015. WHO, 2016
9. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of *Plasmodium falciparum* malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. *Lancet*. 2014 May 17;383(9930):1739–47.
10. Arrow KJ, Panosian C, Gelband H. *The Parasite, the Mosquito, and the Disease*. National Academies Press (US); 2004.
11. MacDonald G, Section WHOM, Conference on Malaria in Africa (1955: Lagos N. Epidemiological basis of malaria control [Internet]. Geneva: Geneva: World Health Organization; 1955
12. Beier JC. Malaria parasite development in mosquitoes. *Annu Rev Entomol*. 1998;43:519–43.
13. Cowman AF, Berry D, Baum J. The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *J Cell Biol*. 2012 Sep 17;198(6):961–71.

14. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, et al. Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infect Dis*. 2009 Sep 1;9(9):555–66.
15. Bejon P, Williams TN, Liljander A, Noor AM, Wambua J, Ogada E, et al. Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. *PLoS Med*. 2010 Jul 6 ;7(7).
16. Kreuels B, Kobbe R, Adjei S, Kreuzberg C, von Reden C, Bäter K, et al. Spatial Variation of Malaria Incidence in Young Children from a Geographically Homogeneous Area with High Endemicity. *J Infect Dis*. 2008 Jan 1;197(1):85–93.
17. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, et al. Identification of hotspots of malaria transmission for targeted malaria control. *J Infect Dis*. 2010;201:1764–74.
18. Gaudart J, Poudiougou B, Dicko A, Ranque S, Toure O, Sagara I, et al. Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC Public Health*. 2006;6:286.
19. Ghebreyesus TA, Haile M, Witten KH, Getachew A, Yohannes AM, Yohannes M, et al. Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey. *BMJ*. 1999 Sep 11;319(7211):663–6.
20. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, et al. Identification of Hot Spots of Malaria Transmission for Targeted Malaria Control. *J Infect Dis*. 2010 Jun 1;201(11):1764–74.
21. Greenwood BM. The microepidemiology of malaria and its importance to malaria control. *Trans R Soc Trop Med Hyg*. 1989;83 Suppl:25–9.
22. WHO. Guidelines on the Elimination of Residual Foci of Malaria Transmission. World Health Organization; 2007. 49 p.
23. Machault V, Gadiaga L, Vignolles C, Jarjaval F, Bouzid S, Sokhna C, et al. Highly focused anopheline breeding sites and malaria transmission in Dakar. *Malar J*. 2009;8:138.
24. Greenwood, BM. The microepidemiology of malaria and its importance to malaria control. *Trans Roy Soc Trop Med Hyg*. 1989;83 Suppl:25–9.
25. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, et al. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med*. 2012;9:e1001165.
26. Gebreyesus, TA, Haile, M, Witten, KH, et al. Incidence of malaria among children living in dams in northern Ethiopia: community based incidence survey. *BMJ*. 1999;319:663–6.

27. Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW. Urbanization, malaria transmission and disease burden in Africa. *Nat Rev Microbiol.* 2005;3(1):81–90.
28. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of Malaria in Africa. *PLOS Med.* 2005 Nov 8;2(12):e340.
29. Midega, JT, Mbogo, CM, Mwambi, H, Wilson, MD, Ojwang, G, Mwangangi, JM, et al. Estimating dispersal and survival of *Anopheles gambiae* and *Anopheles funestus* along the Kenyan coast by using mark-release-recapture methods. *J Med Entomol.* 2007;44:923–9.
30. McCall, PJ, Mosha, FW, Njunwa, KJ, Sherlock, K. Evidence for memorized site-fidelity in *Anopheles arabiensis*. *Trans Roy Soc Trop Med Hyg.* 2001;95:587–90.
31. Constantini, C, Li, S-G, Della Torre, A, Sagnon, N, Coluzzi, M, Taylor, CE. Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan savanna village. *Med Vet Entomol.* 1996;10:203–19.
32. Thomson MC, Connor SJ, Quiñones ML, Jawara M, Todd J, Greenwood BM. Movement of *Anopheles gambiae s.l.* malaria vectors between villages in The Gambia. *Med Vet Entomol.* 1995 Oct;9(4):413–9.
33. Thomas CJ, Cross DE, Bøgh C. Landscape Movements of *Anopheles gambiae* Malaria Vector Mosquitoes in Rural Gambia. *PLoS ONE.* 2013 Jul 18;8(7).
34. Buckee CO, Wesolowski A, Eagle NN, Hansen E, Snow RW. Mobile phones and malaria: Modeling human and parasite travel. *Travel Med Infect Dis.* 2013 Jan;11(1):15–22.
35. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. *Bull World Health Organ.* 2000;78(12):1401–11.
36. Bousema T, Baidjoe A. Heterogeneity in malaria transmission: underlying factors and implications for disease control. In: Takken W, Koenraadt CJM, editors. *Ecology of parasite-vector interactions.* Wageningen Academic Publishers; 2013. p. 197–220. (Ecology and control of vector-borne diseases).
37. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proc Natl Acad Sci U S A.* 1997 Jan 7;94(1):338–42.
38. Carter R, Mendis KN, Roberts D. Spatial targeting of interventions against malaria. *Bull World Health Organ.* 2000;78(12):1401–11.
39. Ross R. *The prevention of malaria.* New York, 1910.

40. MacDonald G. *The Epidemiology & Control of Malaria*. 1st Ed. Oxford University Press, 1957.
41. Smith DL, Dushoff J, McKenzie FE. The Risk of a Mosquito-Borne Infection in a Heterogeneous Environment. *PLOS Biol*. 2004 Oct 26;2(11):e368.
42. Bomblies A, Duchemin J-B, Eltahir EAB. A mechanistic approach for accurate simulation of village scale malaria transmission. *Malar J*. 2009 Oct 2;8:223.
43. Bomblies A, Duchemin J-B, Eltahir EAB. Hydrology of malaria: Model development and application to a Sahelian village. *Water Resour Res*. 2008;44(12).
44. Lutambi AM, Penny MA, Smith T, Chitnis N. Mathematical modelling of mosquito dispersal in a heterogeneous environment. *Math Biosci*. 2013 Feb;241(2):198–216.
45. Lutambi AM, Chitnis N, Briët OJT, Smith TA, Penny MA. Clustering of Vector Control Interventions Has Important Consequences for Their Effectiveness: A Modelling Study. *PLOS ONE*. 2014 May 13;9(5):e97065.
46. Bousema T, Stresman G, Baidjoe AY, Bradley J, Knight P, Stone W, et al. The Impact of Hotspot-Targeted Interventions on Malaria Transmission in Rachuonyo South District in the Western Kenyan Highlands: A Cluster-Randomized Controlled Trial. *PLoS Med*. 2016 Apr 12;13(4). A
47. Stresman GH, Mwesigwa J, Achan J, Giorgi E, Worwui A, Jawara M, et al. Do hotspots fuel malaria transmission: a village-scale spatio-temporal analysis of a 2-year cohort study in The Gambia. *BMC Med*. 2018 Sep 14;16(1):160.
48. Tanner M, Greenwood B, Whitty CJM, Ansah EK, Price RN, Dondorp AM, et al. Malaria eradication and elimination: views on how to translate a vision into reality. *BMC Med*. 2015 Jul 25;13(1):167.
49. Hedrick PW. Population genetics of malaria resistance in humans. *Heredity*. 2011 Oct;107(4):283–304.
50. Baidjoe AY, Stevenson J, Knight P, Stone W, Stresman G, Osoti V, et al. Factors associated with high heterogeneity of malaria at fine spatial scale in the Western Kenyan highlands. *Malar J*. 2016 Jun 4;15.
51. Auchincloss AH, Gebreab SY, Mair C, Roux AVD. A Review of Spatial Methods in Epidemiology, 2000–2010. *Annu Rev Public Health*. 2012 Apr;33:107–22.
52. Sturrock HJW, Novotny JM, Kunene S, Dlamini S, Zulu Z, Cohen JM, et al. Reactive Case Detection for Malaria Elimination: Real-Life Experience from an Ongoing Program in Swaziland. *PLOS ONE*. 2013 May 20;8(5):e63830.

53. Alegana VA, Atkinson PM, Lourenço C, Ruktanonchai NW, Bosco C, Erbach-Schoenberg E zu, et al. Advances in mapping malaria for elimination: fine resolution modelling of *Plasmodium falciparum* incidence. *Sci Rep*. 2016 Jul 13;6:29628.
54. Sturrock HJ, Cohen JM, Keil P, Tatem AJ, Le Menach A, Ntshalintshali NE, et al. Fine-scale malaria risk mapping from routine aggregated case data. *Malar J*. 2014 Nov 3;13(1):421.
55. Sama W, Killeen G, Smith T. Estimating the Duration of *Plasmodium falciparum* Infection from Trials of Indoor Residual Spraying. *Am J Trop Med Hyg*. 2004 Jun 1;70(6):625–34.
56. Bretscher MT, Maire N, Chitnis N, Felger I, Owusu-Agyei S, Smith T. The distribution of *Plasmodium falciparum* infection durations. *Epidemics*. 2011 Jun;3(2):109–18.
57. Forsyth, KP, Anders, RF, Cattani, JA, Alpers, MA. Small area variation in prevalence of an S-antigen serotype of *Plasmodium falciparum* in villages of Madang, Papua New Guinea. *Am J Trop Med Hyg*. 1989;4:344–50.
58. Bogueau, H, Renaud, F, Bouchiba, H, Durand, P, Assi, S-B, Henry, M-C, et al. Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas.
59. Konaté, L, Zwetyenga, J, Rogier, C, Bischoff, E, Fontenille, D, Tall, A, et al. Variation of *Plasmodium falciparum* msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. *Trans Roy Soc Trop Med Hyg*. 1999;93 Suppl 1:21–8.
60. Bejon, P, Turner, L, Lavstsen, T, Cham, G, Olotu, A, Drakeley, CJ, et al. Serological evidence of spatial clusters of *Plasmodium falciparum* parasites. *PLoS ONE*. 2011;6:e21711.
61. Kemere, J, Hsiang, MS, Ali, AS, Msellem, MI, Makame, MH, Baltzell, KA, et al. Evidence for local malaria transmission in the wet season and imported malaria in the dry season in Zanzibar. *Am Soc Trop Med Hyg Annu Meet*. 2011;
62. Kyes, S, Harding, R, Black, G, Craig, A, Peshu, N, Newbold, C, et al. Limited spatial clustering of individual *Plasmodium falciparum* alleles in field isolates from coastal Kenya. *Am J Trop Med Hyg*. 57:521–8.

Chapter 2

2. What proportion of *Plasmodium falciparum* and *Plasmodium vivax* malaria infections are in mosquitoes?

Josephine Malinga^{1, 2}, Jo Lines³, Amanda Ross^{1, 2*}

¹ Swiss Tropical and Public Health Institute, 4002, Basel, Switzerland, ² University of Basel, 4001, Basel, Switzerland, ³ London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom

This Chapter has been submitted for publication at *Malaria Journal*.

2.1. Abstract

Background

Designing effective malaria intervention strategies calls for knowledge of where the infections are. Little is known about the proportions of infections that are in humans and mosquitoes, or how these vary with setting-specific characteristics, both before and after the introduction of interventions.

Methods

We use comprehensive simulation models of malaria epidemiology which have been extensively fitted to data and validate where possible using observed data.

Results

The predicted proportion of infections that is in mosquitoes was substantial and varied by transmission intensity. For *Plasmodium falciparum*, the proportion was 60% in areas of intense transmission declining to a floor of approximately 20% for low transmission. The proportion varied by *Plasmodium* species and the tendency of the vector species to feed on animals as well as humans. The proportion of infections in mosquitoes changed in different ways over time following the introduction of interventions for vectors and humans, and included a sharp spike following mass drug administration.

Conclusion

A substantial proportion of infections are in mosquitoes. Quantifying weak spots in individual interventions enables optimal combination strategies to be developed, and knowledge of the proportion helps place decisions on resources in context.

Keywords: *P.falciparum*, *P.vivax*, malaria, mosquito, mass drug administration, long-lasting insecticide-treated nets

2.2. Background

Knowledge of how malaria parasites and infections are distributed between humans and mosquitoes, both before and after the introduction or scale-up of interventions, would inform the design of effective strategies to interrupt malaria transmission.

Although the numbers of parasites in people and mosquitoes would be an informative measure, they are difficult to quantify. Low densities of parasites are challenging to measure [1,2] and data are not available on the numbers of parasites sequestered away from peripheral blood [3,4], in the liver cells or in mosquitoes leaving a limited evidence base for calculations. A commonly reported measure is the prevalence, and both the prevalence of blood-stage infection in humans and the sporozoite rate (SR) in mosquitoes are straightforward to quantify. However, this does not provide information on the overall distribution between people and mosquitoes since the relative numbers of people and vectors are not taken into account.

A measure for which there is some data as a basis for quantification is the number of infections. We define this as the number of successfully established individual infections through separate inoculations which have not yet been cleared. It includes inoculations leading to liver-stage or blood-stage parasites in humans, and to the presence of oocysts in mosquitoes. The number of infections is more sensitive than prevalence to transmission intensity. With the numbers of infections in mosquitoes and humans, we can calculate the proportion of infections that are in mosquitoes and thus provide a measure of the relative load between humans and mosquitoes. For many policy questions, this measure has similar implications to that of the total parasite load. For example, showing where the infections or parasites are following mass drug administration (MDA) both provide information on weak spots in the intervention. Similarly, both many parasites and many infections in mosquitoes will not transmit further due to the life span and the time needed to become infectious following an infected bite.

Despite the fundamental nature and practical relevance, information on the proportion of infections that are in mosquitoes and how this varies by setting specific characteristics is limited. It is challenging to estimate the component measures such as the number of infections in humans and the number of infected host-seeking mosquitoes per adult in real field settings. The observed number of infections, or multiplicity of infection (MOI), is quantified using genotyping techniques which suffer from low detectability [5], tend to be restricted to younger age-groups rather than represent the whole community and does not detect low density infections or infections that are only in the liver. However, in some settings,

data on the mean MOI in the whole community using high-resolution genotyping together with the number of infected host-seeking mosquitoes per adult are available. In addition, mathematical models provide a framework to synthesize data on different aspects of malaria epidemiology and leverage it to provide predictions for settings where data on the numbers of infections in humans and mosquitoes do not exist.

We use simulation models to predict the proportion of *Plasmodium falciparum* and *Plasmodium vivax* infections that are in mosquitoes. We use OpenMalaria [6], an established comprehensive simulator of *P.falciparum* malaria epidemiology and control which has been extensively fitted to and validated using field datasets, and a recent simulation model of *P.vivax* [7], which has also been parameterized and validated using field data. We predict the proportions by transmission intensity, seasonality, mosquito species, and after the introduction of interventions aimed at control in the vector and the human, long-lasting insecticidal nets (LLIN) and mass drug administration (MDA).

2.3. Methods

Model for *P.falciparum* in humans

OpenMalaria comprises an ensemble of discrete-time individual-based models of malaria in humans [8,9] linked to a deterministic model of malaria in mosquitoes [10,11]. These models have been extensively fitted and validated using different field data sets [6]. The simulated population of humans are updated at each five-day time step via model components representing dynamics of new infections to humans [12], blood-stage parasite densities [13], acquired immunity [13], morbidity and mortality [14–16], and infectivity to mosquitoes [17,18]. The parameter values for the model components of the individual-based model were estimated by fitting to data from a total of 61 malaria field studies of different aspects of malaria epidemiology [6]. An ensemble of simulation model variants has previously been described in detail [9]. The present work uses a subset of six of the model variants previously used for large-scale predictions of the impact of the pre-erythrocytic vaccine, RTS,S/AS01 [9,19].

Model for *P.vivax* in humans

The model for *P.vivax* in humans is a simple individual-based model updated at five-day time-steps for dynamics of new infections, primary infection, relapse, clearance of blood-stage infection, acquired immunity, morbidity, and infectivity to mosquitoes [7]. The parameter values were estimated by statistical analyses of longitudinal cohort data [7,20].

Model component for mosquitoes

The individual-based models for malaria in humans are linked to a compartmental model of vector dynamics (10,11). The vector model describes the life stages of the female mosquito. After emergence from a breeding site, mosquitoes mate and the females search for blood meals which are necessary for egg development. They enter a sequence of feeding cycles with discrete phases for host-seeking, encountering a host, searching for a resting place, resting and ovipositing, if they survive. The mosquitoes are infected according to their feeding behaviour and human infectiousness. The models can accommodate multiple mosquito species with varying periodical emergence rates, and non-human hosts (10,11). Mosquitoes can be either infected or uninfected, but they cannot have multiple infections. A mosquito taking up multiple infections in a single blood meal is counted for the purposes of this paper as one infection. Mosquitoes have been reported to feed on multiple hosts within a feeding cycle (21–23) and in subsequent cycles. However, the low probability of a mosquito being infected per blood meal and the short mosquito lifespan limit potential bias.

Parameter values for each part of the feeding cycle, survival, and the human blood index (HBI) are input separately for each mosquito species. The seasonal emergence rate is calculated as that necessary to produce the input seasonal inoculation entomological rate from the specific vector parameters combined with the vector and human models. The model output explicitly includes the number of infected host-seeking mosquitoes. We adjust this to include resting mosquitoes using the predicted mean durations of the host-seeking and resting phases.

Definition of infections in humans and mosquitoes

We define an infection as having occurred through a separate inoculation and not having been cleared at the time of measurement. For humans, the infection can be liver-stage, blood-stage or both. A *P.vivax* infection in a human is counted once per inoculation regardless of the number of relapses arising from the inoculation.

Different diagnostic methods have different limits of detection: for this study, the numbers of infections in humans and mosquitoes are estimated assuming the use of a perfect diagnostic, successfully detecting all individual infections. To predict the mean number of infections in mosquitoes, we do not model gametocytes or sporozoites explicitly but consider infected mosquitoes as having oocysts, sporozoites or both. They are infected but may or may not be infectious.

Predicted measures

We predict the mean number of infections in mosquitoes in a time-step for one human, a , the mean number of infections in a time-step per human, b , and the proportion of infections

which are in mosquitoes as $a/(a + b)$. If there was a single infection in a mosquito then, at that time-point, the proportion would be one and if the single infection was in a human, the proportion would be zero. To compare the proportions by transmission intensity easily, in seasonal settings we average the proportion over one year.

Scenarios simulated

We predict the proportion of infections in mosquitoes and the impact of transmission intensity, seasonality, mosquito species and interventions, LLIN and MDA (Table 2.1).

Table 2.1 Scenarios simulated*

	<i>Plasmodium falciparum</i>	<i>Plasmodium vivax</i>
Annual EIR	0.5,1,2,5, 10 ,20,50,100,200	0.5,1,2,5, 10 ,20,50,100,200
Seasonality	Constant , Garki,Nigeria (24) Kilifi Town, Kenya (25,26)	Constant , Maprik, Papua New Guinea (27), Hobe, Ethiopia (28)
Mosquito species	An.gambiae , <i>An.arabiensis</i>	An.farauti (29) , <i>An.arabiensis</i> , <i>An.gambiae</i>
Case-management	Low coverage	Low coverage
LLIN	None , 70% coverage	None , 70% coverage
MDA	None , blood-stage drug 70% coverage	None , MDA (blood-stage drug) 70% coverage, MDA (liver-stage) 55%, MDA (both liver and blood stages) 55%,
Model variants	Base model , remaining 5 model variants in ensemble**	Base model , model variant reflecting uncertainty in duration of blood-stage infection

*The reference scenario is shown in bold.

**The *P.falciparum* model variants are based on varying assumptions about heterogeneity in transmission, immunity decay and heterogeneity in co-morbidity.

The parameter values for the mosquito species and the long-lasting insecticidal nets follow those of Stuckey and colleagues (30) which were derived from Briet *et al* (31) and Chitnis *et al* (32) and updated for Yukich *et al* (33) (Table 2.2). We used a health system with a low coverage (less than 5%) of uncomplicated case management with artemisinin-based combination therapies (ACTs) (30). For the mass drug administration, we used a simple model component with a blood-stage drug which cleared blood-stage infections for 30 days. The *P.vivax* liver-stage drug was assumed to clear all hypnozoites in one five-day time-step. The coverage of the liver-stage drug was assumed to be lower to allow for individuals not being treated due to mutations in CYP2D6 and GP6D. The human age distribution follows that of a

typical rural African population based on Ifakara, Tanzania (2002) (13,34). All scenarios are simulated with a population of 10,000 individuals.

Table 2.2 Vector parameters

	<i>An.gambiae ss</i>	<i>An.arabiensis</i>	<i>An.farauti</i>
Duration of resting period in days	2	2	2
Extrinsic incubation period	12	12	10
Proportion of mosquitoes' host-seeking on same day as ovipositing	0.313	0.313	0.728
Maximum proportion of day spent host-seeking	0.33	0.33	0.33
Probability that mosquito survives the feeding cycle	0.623	0.623	0.577
Probability that mosquito successfully bites chosen host	0.95	0.95	0.95
Probability that mosquito successfully escapes host and finds a resting place after biting	0.95	0.95	0.95
Probability of mosquito successfully resting after finding a resting place	0.99	0.99	0.99
Probability of mosquito successfully laying eggs given that it has rested	0.88	0.88	0.88
Human Blood Index, the proportion of mosquitoes which fed on human blood during the last feed	0.85	0.5	0.72

Validation

Validation of saturation in the number of infections in humans with transmission intensity

To validate the relationship between the number of infections in humans and transmission intensity as far as possible, we collected data on the observed MOI, prevalence and EIR for study locations where all three were available. In order to represent the community MOI we excluded studies where the sample included only febrile patients, but include studies which comprise of asymptomatic individuals only or surveys with asymptomatic and symptomatic individuals sampled randomly. We also excluded studies with a limited age range. We used literature searches by Karl et al (35), Ehle *et al* (submitted), Arnot (36) and Owusu-Agyei *et al* (37) as the basis of the search for *P.falciparum*. For *P.vivax*, there are few data available with

both multiplicity of infection and transmission intensity, and the number of liver-stage infections is not known. To gain the mean MOI in all individuals, rather than those with infections only, we adjusted the mean MOI reported for individuals with at least one infection using the prevalence by polymerase chain reaction (PCR). If the prevalence by PCR was not reported, then we estimate it from the rapid diagnostic test (RDT) or microscopy prevalence using the algorithm by Okell *et al* (38). We plot the observed mean MOI against age for settings with different transmission intensities.

Validation of the relationship between the proportion of mosquitoes that are infectious and transmission intensity

For some settings, both sporozoite rates and EIR are available for the same study locations (39). We use this data to validate the model component of the relationship between the proportion of mosquitoes that are infectious and transmission intensity. Infectious mosquitoes are a fraction of infected mosquitoes. We plot the observed proportion of mosquitoes with sporozoites against transmission intensity for the different study sites: we acknowledge that there is variation between sites in the approaches for measuring the components of the sporozoite rate and EIR and in site characteristics in this ecological analysis (39). We compare the plotted relationship to predictions of the proportion of host-seeking mosquitoes that are infectious for one reference scenario with no seasonality with different EIR.

2.4. Results

The relationship between the proportion of infections that are in mosquitoes and transmission intensity

Both the predicted number of *P.falciparum* infections that are in mosquitoes (Figure 2.1a) and in humans (Figure 2.1b) increase at higher transmission intensities. In humans, the mean number of infections saturates principally due to the relationship between the transmission intensity and the force of infection [12], but also to a lesser extent due to acquired immunity, morbidity, mortality, and treatment. In mosquitoes, there is little saturation since even at high transmission intensities the proportion of mosquitoes infected is predicted to never become high. As a consequence, the proportion of infections that are in mosquitoes increases with increasing transmission intensity (Figure 2.1c). Taking the mean over one year, the pattern is not substantially affected by the seasonality of the setting (Figure 2.1). At very low transmission intensities (not shown), the proportion of infections in mosquitoes maintains the floor. Simulations were run with transmission intensities down to 0.001 infectious bites per person per year, but interruption of transmission prohibits lower values.

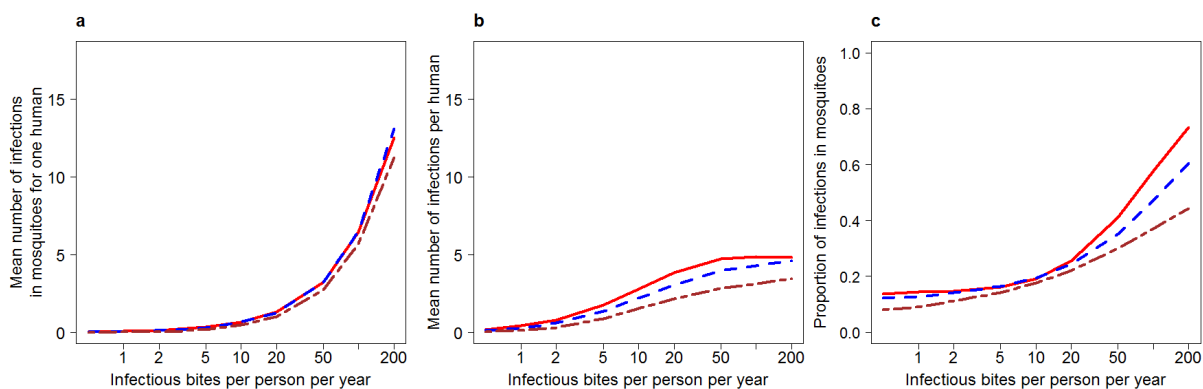


Figure 2.1 Predicted *P.falciparum* infections in mosquitoes and humans by transmission intensity for different seasonal patterns

a) mean number of infections in mosquitoes for one human; b) mean number of infections per human; c) proportion of infections that are in mosquitoes. The values are the mean over one year and use the OpenMalaria base model with *An.gambiae* mosquitoes. Red solid line: constant transmission; blue dashed line: Kilifi seasonality; brown two-dash line: Garki seasonality.

There is a seasonal pattern in the predicted proportion of infections that are in mosquitoes which peaks slightly earlier than the transmission intensity (Figure 2.2).

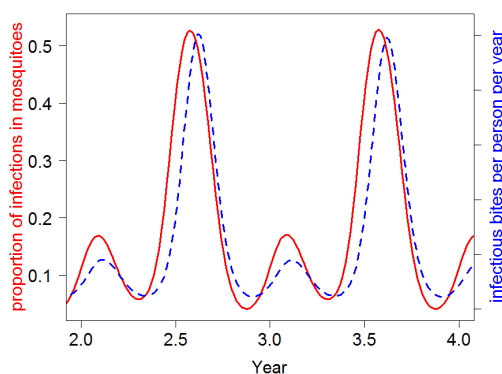


Figure 2.2 Seasonal transmission and the predicted proportion of *P.falciparum* infections in mosquitoes

Blue dashed line: transmission intensity, red solid line: proportion of infections in mosquitoes. Simulations were run using the base model with the reference scenario with a seasonal pattern for Kilifi.

The species of the mosquito affects the predicted proportion of infections that are in mosquitoes (not shown). We found that this was driven chiefly by the human blood index (HBI). *An.arabiensis* has been parameterized with a markedly lower HBI than *An.gambiae* and

so a higher number of infected host-seeking mosquitoes would be required to produce the same transmission intensity. For all levels of transmission intensity, the proportion of *P.falciparum* infections that are in mosquitoes decreases as the HBI increases (Figure 2.3).

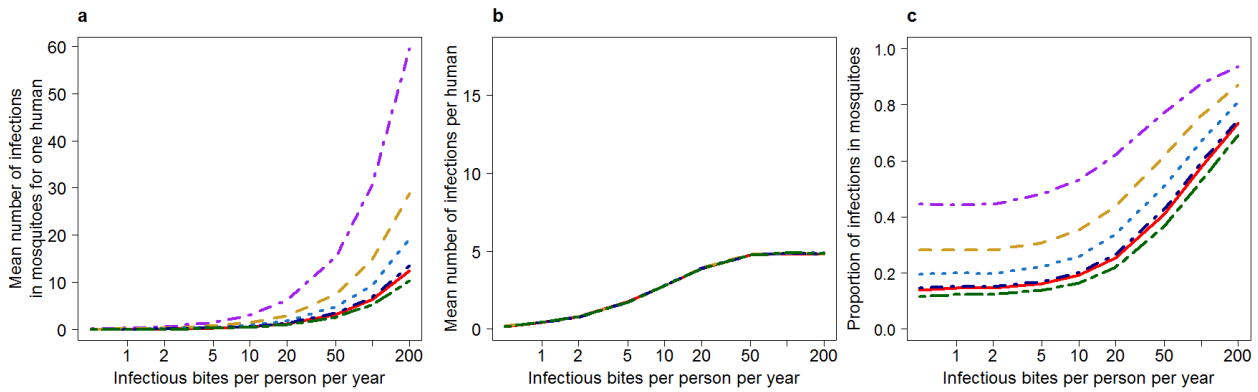


Figure 2.3 Predicted *P.falciparum* infections in mosquitoes and humans by transmission intensity for different human blood indices

a) mean number of infections in mosquitoes for one human; b) mean number of infections per human; c) proportion of infections that are in mosquitoes. The reference scenario with no seasonality was used and values are the mean for one year. Solid red line: base model with HBI=0.85; purple dotdash lines: HBI=0.2; yellow dashed lines: HBI=0.4; light blue dotted line: HBI=0.6; dark green twodash line; HBI=1.0

In addition to the *OpenMalaria* base model, we used a further five model variants with varying assumptions about heterogeneity in transmission, immunity decay, and heterogeneity in co-morbidity [9,19]. These assumptions predominantly affect humans and lead to variation in the predicted mean number of infections per human (Figure 2.4b). The highest numbers of infections per person are predicted for the model variant with heterogeneity in transmission since heterogeneity rather than saturation of successful inocula accounts for the relationship between transmission intensity and force of infection. The patterns of the predicted proportion of infections in mosquitoes are nevertheless reasonably similar for all of the model variants (Figure 2.4c).

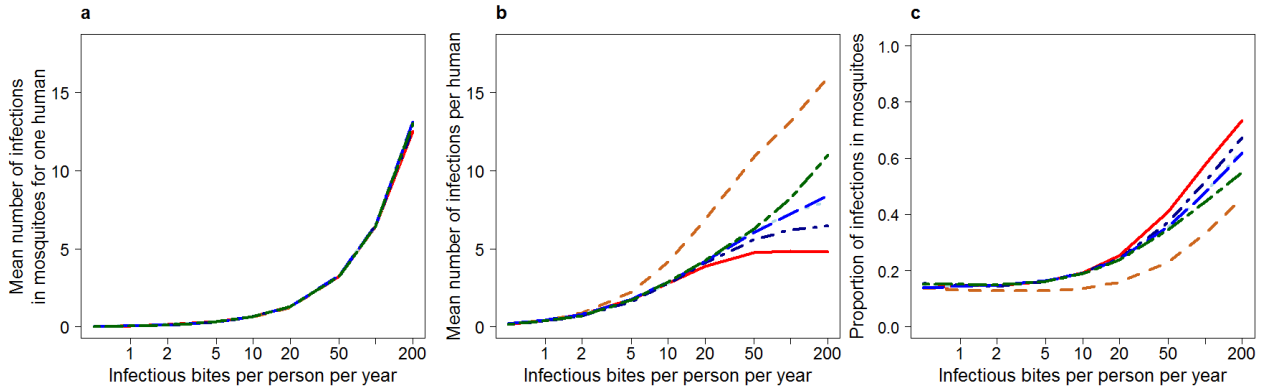


Figure 2.4 Predicted *P.falciparum* infections in mosquitoes and humans by transmission intensity for different model variants

a) mean number of infections in mosquitoes for one human; b) mean number of infections per human; c) proportion of infections that are in mosquitoes. The reference scenario with no seasonality was used and values are the mean for one year. Solid red line: base model variant; yellow dashed line: heterogeneity in transmission (OpenMalaria model variant number R0068); blue lines: decay in immunity in effective cumulative exposure (R0131 light blue dotted line), proxies (R0132 dark blue dotdash line) or both (R0133 mid-blue long dash line); green two-dash line: heterogeneity in co-morbidity (R670).

The pattern for *P.vivax* is similar to that of *P.falciparum* with a floor in the predicted proportion of infections in mosquitoes at low transmission intensities (Figure 2.5). The numbers of infections are greater due to mosquito species, liver-stage infections and because the model component for the relationship between the force of infection and transmission intensity in *P.vivax* is based on the model variant for *P.falciparum* which uses heterogeneity in transmission [7]. The *P.vivax* model variant reflecting uncertainty in the duration of blood-stage infections predicted similar patterns (not shown).

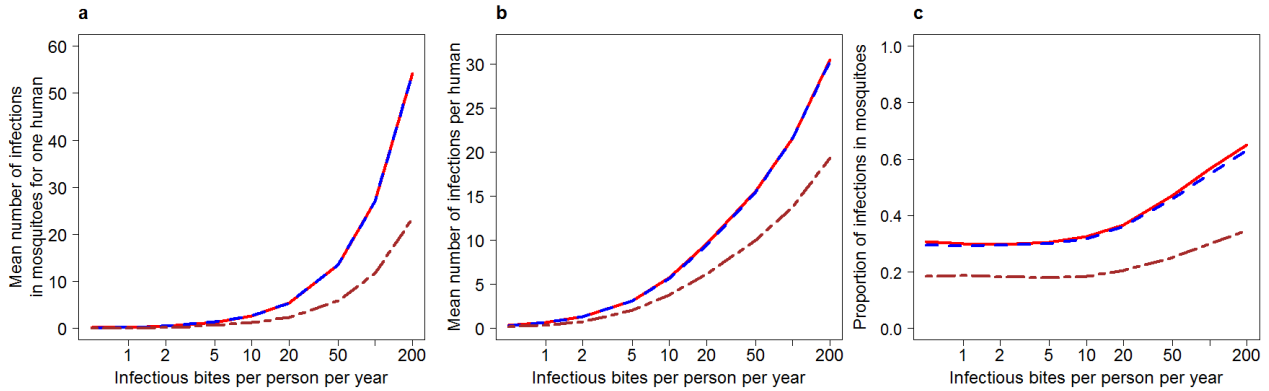


Figure 2.5 Predicted *P. vivax* infections in mosquitoes and humans by transmission intensity for different seasonal patterns

a) mean number of infections in mosquitoes for one human; b) mean number of infections per human c) proportion of infections that are in mosquitoes. Infections in humans were counted once whether they were liver-stage, blood-stage or both. Red solid line: base model with constant seasonality with *An. farauti*; Blue dashed line: PNG Seasonality with *An. farauti*; Brown dotdash line: Ethiopia seasonality with *An. arabiensis*

Validation

We use available data to validate (i) the relationship between the number of *P. falciparum* infections in humans and transmission intensity and (ii) the relationship between the proportion of mosquitoes that are infectious and transmission intensity (Table 2.3).

Table 2.3 Data sources for model validation

Study	Study site	Age-range	MOI period	Time	Study design	Marker	EIR	EIR year	EIR ref
Engelbrecht <i>et al</i> (40)	Kaduna State, Nigeria	5-70y	Cross-sectional survey		104 asymptomatic individuals in cross-sectional survey at end of the rainy season	<i>msp2</i>	High	-	(40)
Smith <i>et al</i> (41)	Kilombero, Tanzania	0-82y	1989-1996		1677 individuals from multiple surveys	<i>msp2</i>	400	1994	(41)
Owusu <i>et al</i> (42)	Navrongo, Ghana	All ages	June-July 2000		308 individuals in a cross-sectional survey	<i>msp2</i>	300	2000	(42)
Konaté <i>et al</i> (43)	Dielmo, Senegal	1-84y	October 1994		144 villagers, survey at the end of transmission season	<i>msp2</i>	120	1994	(43)
Bereczky <i>et al</i> (44)	Rufiji, Tanzania	1-84y	March & April 1999		873 individuals in a cross-sectional survey	<i>msp2</i>	High	-	
Färnert <i>et al</i> (45)	Chonyi, Kenya	0-11y	October 2000		264 children, cool dry season	<i>msp2</i>	22-53	1997-98	(45,46)
Mayor <i>et al</i> (47,48)	Manhica, Mozambique	4mo-83y	1997-1999		826 samples in 5 cross-sectional surveys	<i>msp2</i>	15	1998	(47)
Cortes <i>et al</i> (49)	Wosera, Papua New Guinea	All ages	June & November 2000		628 samples from cross-sectional surveys	<i>msp2</i>	12		(50)
Färnert <i>et al</i> (45)	Ngerenya, Kenya	0-11y	October 2000		273 children, cool dry season	<i>msp2</i>	10	1992-93	(26,45)
Roper <i>et al</i> (51)	Daraweesh, Sudan	9-36y	1993-1996		Cohort of 106 residents	<i>msp1</i>	1	-	(52)

The mean observed multiplicity of infection per individual varies substantially by transmission intensity and age (Figure 2.6a and Table 2.2). The decrease in older ages may reflect acquired immunity lowering the density of individual genotypes to below the limit of detection. This makes it difficult to compare MOI by transmission intensity; however we use the saturation in peak MOI with transmission intensity to validate the saturation in the predicted relationship between the number of infection in humans and the EIR.

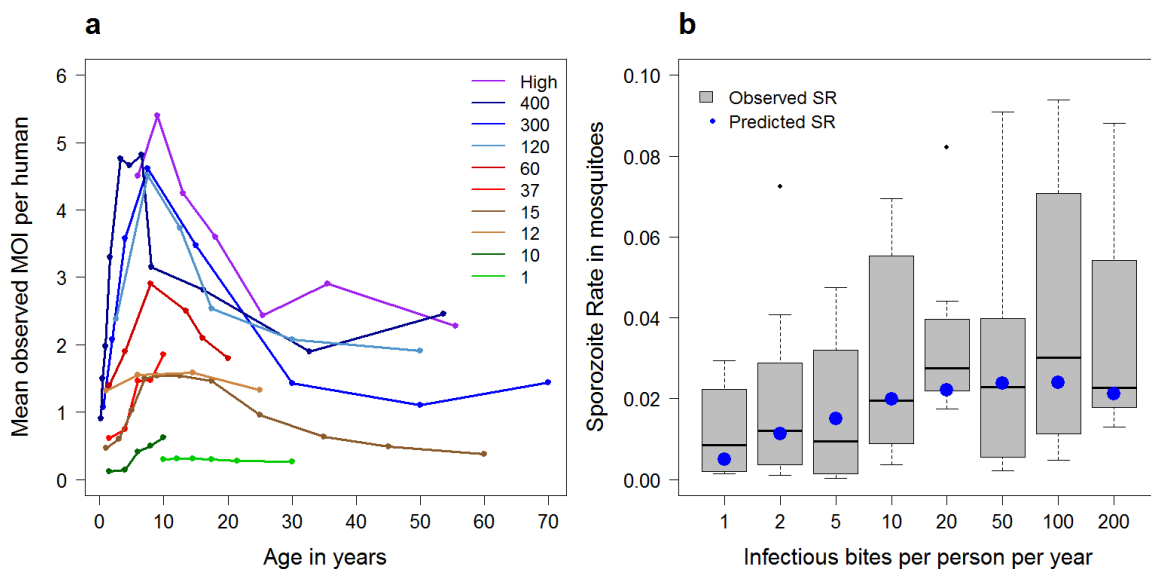


Figure 2.6 Validation of the predictions for the number of infections in humans and mosquitoes with observed data

a) Observed mean number of *P.falciparum* detected infections by age and transmission intensity. The colour bands correspond to the study specific EIR. The points represent the mean number of detected infections in all individuals (including those with no infections). b) Observed (boxplots) and predicted (blue dots) sporozoite rates in mosquitoes with transmission intensity. The observed data was obtained from different endemic settings with differing seasonality profiles and study periods. A reference scenario for the predictions was used with no seasonality. The slight decrease for the EIR of 200 is likely to be due to a reduction in infectiousness to mosquitoes due to acquired immunity in people living in high transmission settings.

Both the observed and predicted sporozoite rates in mosquitoes were low and slightly increased with increasing transmission intensity (Figure 2.6b). There is a slight decrease for both observed and predicted SR for an EIR of 200. This may be due to lower infectiousness of humans with high levels of acquired immunity in high transmission settings.

Predicted impact of long-lasting insecticidal nets

In both *P.falciparum* and *P.vivax*, an LLIN campaign with 70% coverage leads to a substantial decrease in the numbers of infections in both mosquitoes (Figure 2.7a,d) and in humans (Figure 2.7b,e). The decrease is sharper in mosquitoes due to the longevity of infections in humans, and this leads to an initial decrease in the proportion of infections in mosquitoes (Figure 2.7c,f). Over time, the numbers of infections in mosquitoes and humans increase. This is partly due to reduced effectiveness of the net through wear and tear and insecticide decay, and partly due to reduced exposure in humans leading to reduced acquired immunity and thus higher infectiousness to mosquitoes.

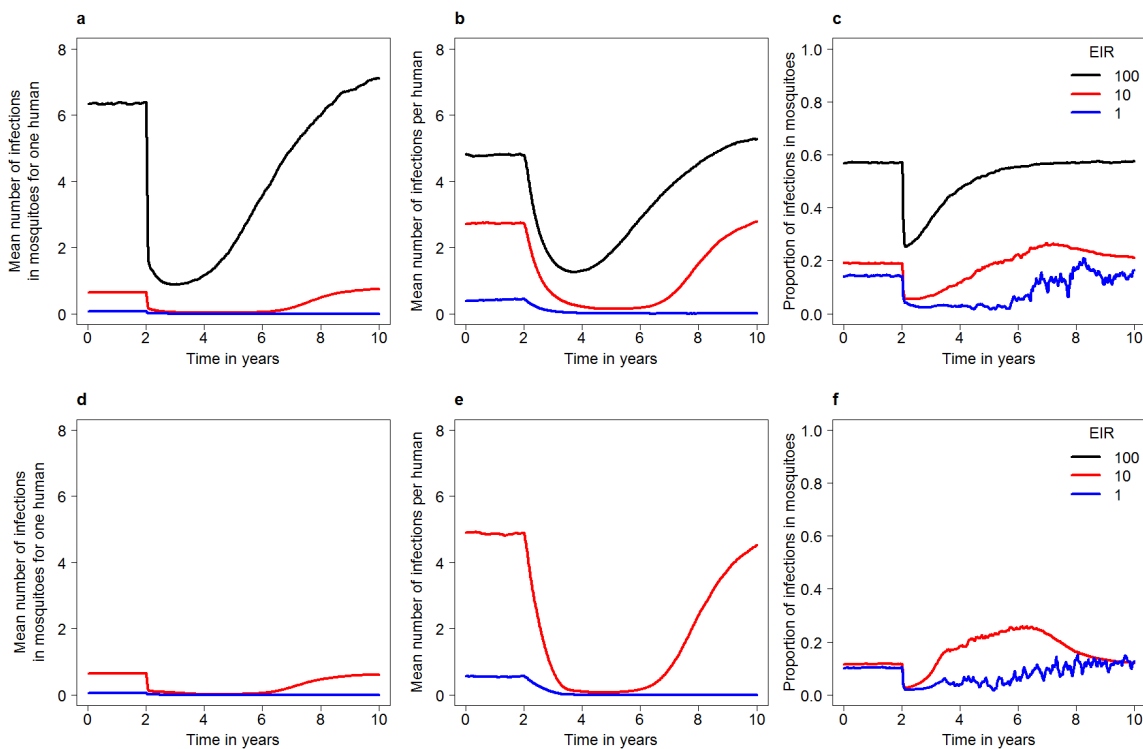


Figure 2.7 The predicted impact of a mass campaign of long-lasting insecticidal nets on the proportion of infections in mosquitoes. *P.falciparum* (top row) and *P.vivax* (bottom row) a & d) mean number of infections in mosquitoes for one human; b & e) mean number of infections per human; c & f) proportion of infections that are in mosquitoes. The reference scenarios with constant seasonality were used and values were averaged over one year. Solid

black line: EIR 100; red line EIR 10; blue line EIR 1. The LLIN campaign had 70% coverage and took place at year 2.

Predicted effect of mass drug administration

For *P.falciparum*, MDA with a blood-stage drug leads to a sharp drop in the number of infections in humans and mosquitoes (Figure 2.8a,b). The drop is immediate in humans and after 10 to 15 days in mosquitoes, leading to a dramatic spike in the proportion of infections that are in mosquitoes just after administration (Figure 2.8c).

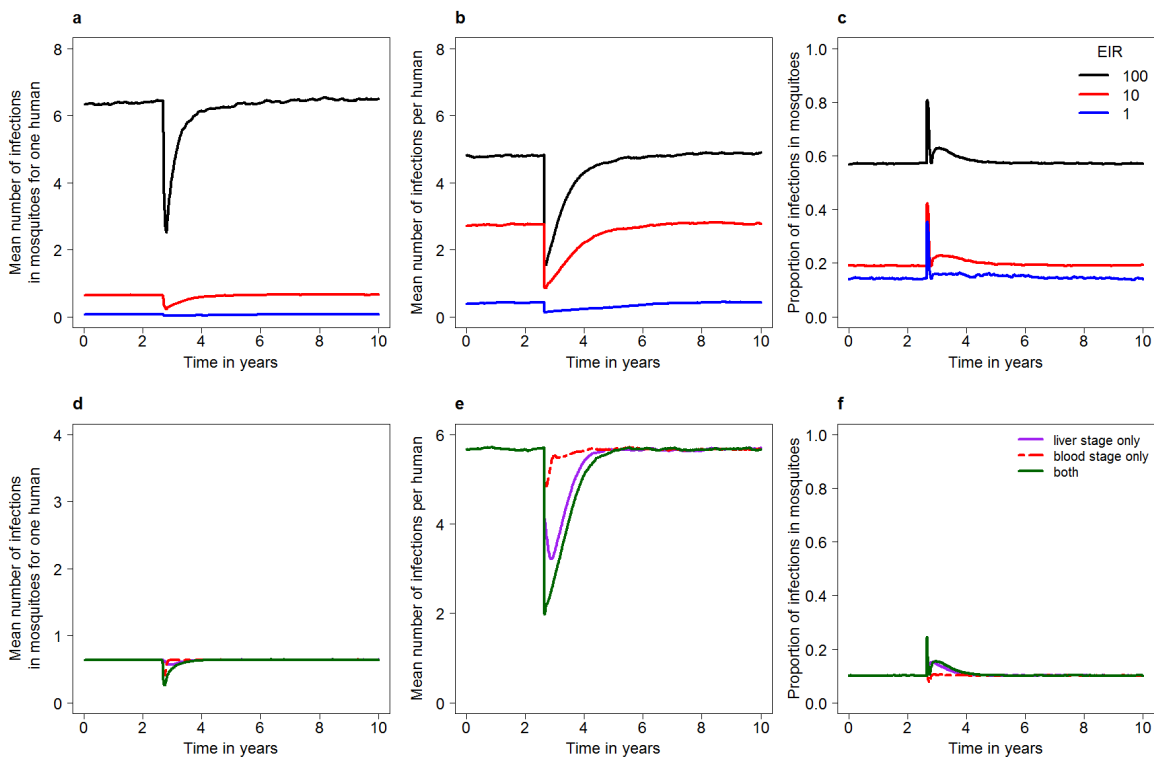


Figure 2.8 The predicted impact of mass drug administration on the proportion of infections in mosquitoes. *P.falciparum* (top row) and *P.vivax* (bottom row)

a & d) mean number of infections in mosquitoes for one human; b & e) mean number of infections per human; c & f) proportion of infections which are in mosquitoes. Top row (*P.falciparum*): Solid black line: EIR 100; red line EIR 10; blue line EIR 1. The MDA campaign had 70% coverage and took place at year 2. Bottom row (*P.vivax*): Constant transmission, EIR=10. Solid purple line: liver stage only; Two-dash red line: blood stage only; dark green line: both blood and liver stage. The MDA campaigns had coverage of 70% for blood-stage and 55% for liver-stage drugs and took place at year 2.

In a seasonal setting, the timing of MDA affects the height of the spike (not shown). The absolute increase in the spike is greatest if MDA is given around the time of peak transmission

and the relative increase (as a ratio) is greatest when the proportion of infections in mosquitoes is low, in the dry season.

For *P.vivax*, MDA with a liver-stage drug or both blood-and liver-stage drugs produces substantial decreases in the numbers of infections in mosquitoes and humans (Figure 2.8d, e). The spike in the proportion of infections that are in mosquitoes is smaller than that for *P.falciparum* since a larger proportion of infections are in humans (Figure 2.8f). MDA with a blood-stage drug alone is predicted to have a limited impact since the liver-stage parasites are unaffected. Unusually for MDA, with the blood-stage drug only there is a decrease in the proportion of infections in mosquitoes just after administration, as a consequence of clearance of the blood-stage parasites.

2.5. Discussion

The predictions indicate that the proportion of infections in mosquitoes is substantial, for *P.falciparum* this ranges between 20-60%, decreases with decreasing transmission intensity and reaches a floor for low transmission intensities. This pattern is driven by saturation in the numbers of infections in humans at higher transmission intensities, and this saturation has been validated as far as possible using observed data. The predicted proportion of infections additionally depends on setting specific characteristics such as mosquito species and seasonality, and, in the case of *P.vivax*, relapse patterns.

Transmission-blocking immunity is not included in our predictions and could alter the relationships. In humans, this would lead to more infections in humans required to infect the same number of mosquitoes, and vice versa in mosquitoes. We expect that the pattern of the relationship between the proportion of infections in mosquitoes by transmission intensity would remain similar.

The predictions highlight the role of the dynamics of infections in humans and mosquitoes following the introduction or scale-up of an intervention. The interventions simulated LLIN and MDA, target the vector and the parasites in the human respectively, and were predicted to have different effects on the proportion of infections in mosquitoes. LLIN initially decreases the proportion, while MDA shows a striking spike for *P.falciparum* and for *P.vivax* liver-stage drugs. Using predictions to identify weak spots in individual interventions and their timing may aid the optimization of combinations of interventions for malaria control and elimination.

Previous predictions of the impact of MDA on *P.vivax* also show that blood-stage drugs alone are not effective at interrupting transmission and that a liver-stage intervention with high coverage is needed [33–35]. The predicted spike in the proportion of infections in mosquitoes

following MDA suggests that, in low transmission settings where MDA may have a chance of interrupting transmission for *P.falciparum* (56), vector control or ivermectin preceding a round of MDA could potentially be beneficial in slowing the rate of re-infection. However, in practice, further challenges facing MDA are the difficulties in achieving a high coverage in humans (57), and in controlling the threat of resurgence from imported infections.

There is an on-going debate about whether infections with low parasite densities should be detected and treated, given limited resources for malaria control, and uncertainty over their relevance to transmission and morbidity (58,59). New, highly sensitive diagnostic tests have been developed for use in humans (60). Our prediction that a substantial proportion of infections are to be found in mosquitoes helps inform the rational allocation of resources by informing the utility of ultra-diagnostic testing to find and clear the last malaria infections in humans and highlighting the shifting dynamics of infections in humans and mosquitoes in response to interventions. The use of existing vector control tools and effective anti-malarials has been shown to lead to substantial declines in the burden of malaria (61), and quantifying where infections are would help in identifying the weak spots in single interventions to allow the effective design and allocation of optimal combination strategies, especially in areas where resources are constrained.

2.6. Conclusion

The proportion of infections in mosquitoes is predicted to be substantial, and to vary with transmission intensity, human blood index, and in response to interventions. These predictions could inform the design of effective strategies with integrated interventions and diagnostic tools to reduce and interrupt malaria transmission and prevent resurgence.

2.7. References

1. The malERA Consultative Group on Diagnoses and Diagnostics. A Research Agenda for Malaria Eradication: Diagnoses and Diagnostics. PLOS Med. 2011 Jan 25;8(1):e1000396.
2. Elimination T malERA RCP on T for M. malERA: An updated research agenda for diagnostics, drugs, vaccines, and vector control in malaria elimination and eradication. PLOS Med. 2017 Nov 30;14(11):e1002455.
3. Ochola LB, Marsh K, Lowe B, Gal S, Pluschke G, Smith T. Estimation of the sequestered parasite load in severe malaria patients using both host and parasite markers. Parasitology. 2005 Oct;131(Pt 4):449–58.
4. Dondorp AM, Desakorn V, Pongtavornpinyo W, Sahassananda D, Silamut K, Chotivanich K, et al. Estimation of the Total Parasite Biomass in Acute Falciparum Malaria from Plasma PfHRP2. PLOS Med. 2005 Aug 23;2(8):e204.
5. Bretscher MT, Valsangiacomo F, Owusu-Agyei S, Penny MA, Felger I, Smith T. Detectability of *Plasmodium falciparum* clones. Malar J. 2010;9:234.
6. Smith, T, Maire, N, Ross, A, Penny, M, Chitnis, N, Schapira, A, et al. Towards a comprehensive simulation model of malaria epidemiology and control. Parasitology. 2008;135(13):1507–16.
7. Ross, A, Koepfli, C, Timinao, L, Hardy, D, Thuring, T, Lin, E, et al. The relationship between *Plasmodium vivax* acute illness and primary infection and relapse in a cohort of children in Papua New Guinea Amanda Ross, Cristian Koepfli, Lincoln Timinao, Diggory Hardy, Tobias Thuring, Lin E, Kiniboro B, Nicolas Senn, Peter Siba , Ingrid Felger, Ivo Mueller, Marcel Tanner. Prep.
8. Smith, T, Killeen, GF, Maire, N, Ross, A, Molineaux, L, Tediosi, F, et al. Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of *Plasmodium falciparum* malaria: Overview. Am J Trop Med Hyg. 2006;75(Suppl 2):1–10.
9. Smith, T, Ross, A, Maire, N, Chitnis, N, Studer, A, Hardy, D. Ensemble modeling of the likely public health impact of a pre-erythrocytic malaria vaccine. PLoS Med. 2012;9(1):e1001157.
10. Chitnis, N, Smith, T, Steketee, R. A mathematical model for the dynamics of malaria in mosquitoes feeding on a heterogeneous host population. J Biol Dyn. 2008;2:259–85.

11. Chitnis, N, Hardy, D, Smith, T. A periodically-forced mathematical model for the seasonal dynamics of malaria in mosquitoes. *Bull Math Biol.* 2012;74:1098–1124.
12. Smith, T, Maire, N, Dietz, K, Killeen, GF, Vounatsou, P, Molineaux, L, et al. Relationship between the entomologic inoculation rate and the force of infection for *Plasmodium falciparum* malaria. *Am J Trop Med Hyg.* 2006;75(Suppl 2):11–8.
13. Maire, N, Smith, T, Ross, A, Owusu-Agyei, S, Dietz, K, Molineaux, L. A model for natural immunity to asexual blood stages of *Plasmodium falciparum* malaria endemic areas. *Am J Trop Med Hyg.* 2006;75(Suppl 2):19–31.
14. Smith, T, Ross, A, Maire, N, Rogier, C, Trape, JF, Molineaux, L. An epidemiologic model of the incidence of acute illness in *Plasmodium falciparum* malaria. *Am J Trop Med Hyg.* 2006;75(Suppl 2):56–62.
15. Ross, A, Smith, T. The effect of malaria transmission intensity on neonatal mortality in endemic areas. *Am J Trop Med Hyg.* 2006;75(Suppl 2):74–81.
16. Ross, A, Maire, N, Molineaux, L, Smith, T. An epidemiologic model of severe morbidity and mortality caused by *Plasmodium falciparum*. *Am J Trop Med Hyg.* 2006;75(Suppl 2):63–73.
17. Ross, A, Killeen, G, Smith, T. Relationships between host infectivity to mosquitoes and asexual parasite density in *Plasmodium falciparum*. *Am J Trop Med Hyg.* 2006;75(Suppl 2):32–7.
18. Killeen, GF, Ross, A, Smith, T. Infectiousness of malaria-endemic human populations to vectors. *Am J Trop Med Hyg.* 2006;75(Suppl 2):38–45.
19. Penny, MA, Verity, R, Bever, CA, Sauboin, C, Galaktionova, K, Flasche, S, et al. Public health impact and cost-effectiveness of the RTS,S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models. *Lancet.* 2016;387(10016):367–75.
20. Ross A, Koepfli C, Schoepflin S, Timinao L, Siba P, Smith T, et al. The Incidence and Differential Seasonal Patterns of *Plasmodium vivax* Primary Infections and Relapses in a Cohort of Children in Papua New Guinea. *PLoS Negl Trop Dis.* 2016 May 4;10(5):e0004582.
21. Norris, LC, Fornadel, CM, Hung, WC, Pineda, FJ, Norris, DE. Frequency of multiple blood meals taken in a single gonotrophic cycle by *Anopheles arabiensis* mosquitoes in Macha, Zambia. *Am J Trop Med Hyg.* 2010;83:33–7.

22. Scott, TW, Githeko, AK, Fleischer, A, Harrington, LC, Yan, G. DNA profiling of human blood in Anophelines from lowland and highland sites in western Kenya. *Am J Trop Med Hyg.* 2006;75:231–7.
23. Soremekun, S, Maxwell, C, Zuwakuu, M, Chen, C, Michael, E, Curtis, C. Measuring the efficacy of insecticide treated bed nets: the use of DNA fingerprinting to increase the accuracy of personal protection estimates in Tanzania. *Trop Med Int Health.* 2004;9:663–72.
24. Molineux, L, Grammiccia, G. The Garki Project. Geneva: World Health Organization; 1980.
25. Mbogo, CN, Snow, R, Kabiru, EW, Ouma, J, Githure, JI, Marsh, K, et al. Low-level *Plasmodium falciparum* transmission and the incidence of severe malaria infections on the Kenyan coast. *Am J Trop Med Hyg.* 1993;49:245–53.
26. Mbogo, CN, Snow, R, Khamala, CP, Kabiru, EW, Ouma, J, Githure, JI, et al. Relationships between *Plasmodium falciparum* transmission by vector populations and the incidence of severe disease at nine sites on the Kenyan coast. *Am J Trop Med Hyg.* 1995;52:201–6.
27. Ross, A, Koepfli, C, Schoepflin, S, Timinao, L, Siba, P, Smith, T, et al. The incidence and differential seasonal patterns of *Plasmodium vivax* primary infections and relapses in a cohort of children in Papua New Guinea. *PLoS Negl Trop Dis.* 2016;10(5):e0004582.
28. Animut A, Balkew M, Gebre-Michael T, Lindtjørn B. Blood meal sources and entomological inoculation rates of anophelines along a highland altitudinal transect in south-central Ethiopia. *Malar J.* 2013 Feb 23;12:76.
29. Sinka, ME, Bangs, MJ, Manguin, S, Rubio-Palis, Y, Chareonviriyaphap, T, Coetzee, M, et al. A global map of dominant malaria vectors. *Parasit Vectors.* 2012;5:69.
30. Stuckey, EM, Stevenson, JC, Cooke, MK, Owaga, C, Mambe, E, Onando, G, et al. Simulation of malaria epidemiology and control in the highlands of Kenya. *Malar J.* 2012;11:357.
31. Briet, OJ, Hardy, D, Smith, TA. Importance of factors determining the effective lifetime of a mass long-lasting insecticidal net distribution: a sensitivity analysis. *Malar J.* 2012;11:20.
32. Chitnis, N, Smith, T, Schapira, A. Parameter values for transmission model. Unpubl Work Swiss TPH. 2010;
33. Yukich JO, Chitnis N. Modelling the implications of stopping vector control for malaria control and elimination. *Malar J.* 2017 Oct 13;16:411.

34. INDEPTH Network. Population, health and survival at INDEPTH sites. Ottawa, Ontario, Canada: International Development Research Centre; 2002.
35. Karl S, White MT, Milne GJ, Gurarie D, Hay SI, Barry AE, et al. Spatial Effects on the Multiplicity of *Plasmodium falciparum* Infections. PLOS ONE. 2016 Oct 6;11(10):e0164054.
36. Arnot D. Unstable malaria in Sudan: the influence of the dry season: Clone multiplicity of *Plasmodium falciparum* infections in individuals exposed to variable levels of disease transmission. Trans R Soc Trop Med Hyg. 1998 Nov 1;92(6):580–5.
37. Owusu-Agyei S, Smith T, Beck H-P, Amenga-Etego L, Felger I. Molecular epidemiology of *Plasmodium falciparum* infections among asymptomatic inhabitants of a holoendemic malarious area in northern Ghana. Trop Med Int Health TM IH. 2002 May;7(5):421–8.
38. Okell LC, Bousema T, Griffin JT, Ouédraogo AL, Ghani AC, Drakeley CJ. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. Nat Commun. 2012 Dec 4;3:1237.
39. Hay SI, Rogers DJ, Toomer JF, Snow RW. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, internet access and review. Trans R Soc Trop Med Hyg. 2000;94(2):113–27.
40. Engelbrecht, F, Tögel, E, Beck, HP, Enwezor, F, Oettli, A, Felger, I. Analysis of *Plasmodium falciparum* infections in a village community in Northern Nigeria: determination of msp2 genotypes and parasite -specific IgG responses. Acta Trop. 2000;74:63–71.
41. Smith, T, Beck, HP, Kitua, A, Mwankuyse, S, Felger, I, Fraser-Hurt, N, et al. Age dependence of the multiplicity of *Plasmodium falciparum* infections and of other malariological indices in an area of high endemicity. Trans Roy Soc Trop Med Hyg. 1999;93(Suppl 1):15–20.
42. Owusu-Agyei, S, Smith, T, Beck, HP, Amenga-Etego, L, Felger, I. Molecular epidemiology of *Plasmodium falciparum* infections among asymptomatic inhabitants of a holoendemic malarious area in northern Ghana. Trop Med Int Health. 2002;7:421–8.
43. Konaté, L, Zwetyenga, J, Rogier, C, Bischoff, E, Fontenille, D, Tall, A, et al. Variation of *Plasmodium falciparum* msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. Trans Roy Soc Trop Med Hyg. 93 (Suppl 1):21–8.

44. Berezcky S, Liljander A, Rooth I, Faraja L, Granath F, Montgomery SM, et al. Multiclonal asymptomatic *Plasmodium falciparum* infections predict a reduced risk of malaria disease in a Tanzanian population. *Microbes Infect.* 2007 Jan 1;9(1):103–10.
45. Färnert, A, Williams, TN, Mwangi, TW, Ehlin, A, Fegan, G, Macharia, A, et al. Transmission-dependent tolerance to multiclonal *Plasmodium falciparum* infection. *J Infect Dis.* 2009;200:1166–75.
46. Mbogo, CM, Mwangangi, JM, Nzovu, J, Gu, W, Yan, G, Gunter, JT, et al. Spatial and temporal heterogeneity of *Anopheles* mosquitoes and *Plasmodium falciparum* transmission along the Kenyan coast. *Am J Trop Med Hyg.* 2003;68:734–42.
47. Mayor, A, Saute, F, Aponte, JJ, Almeda, J, Gomez-Olivé, FX, Dgedge, M, et al. *Plasmodium falciparum* multiple infections in Mozambique, its relation to other malariological indices and to prospective risk of malaria morbidity. *Trop Med Int Health.* 2003;8:3–11.
48. Nhacolo, AQ, Nhalungo, DA, Saco, CN, Aponte, JJ, Thompson, R, Alonso, P. Levels and trends of demographic indices in southern rural Mozambique: evidence from demographic surveillance in Manhica district. *BMC Public Health.* 2006;6:291.
49. Cortés; A, Mellombo, M, Benet, A, Lorry, K, Rare, L, Reeder, JC. *Plasmodium falciparum*: distribution of msp2 genotypes among symptomatic and asymptomatic individuals from the Wosera region of Papua New Guinea. *Exp Parasitol.* 2004;106:22–9.
50. Smith T, Hii J, Genton B, Müller I, Booth M, Gibson N, et al. Associations of peak shifts in age-prevalence for human malarias with bednet coverage. *Trans R Soc Trop Med Hyg.* 2001 Jan 1;95(1):1–6.
51. Roper, C, Richardson, W, Elhassan, IM, Giha, H, Hviid, L, Satti, GMH, et al. Seasonal changes in the *Plasmodium falciparum* population in individuals and their relationship to clinical malaria: a longitudinal study in a Sudanese village. *Parasitology.* 1998;116:501–10.
52. Arnot, D. Unstable malaria in Sudan: the influence of the dry season. *Trans Roy Soc Trop Med Hyg.* 1998;92:580–5.
53. Robinson L, Wampfler R, Betuela I, Karl S, White M, Li Wai Suen, CSI, et al. Strategies for understanding and reducing the *Plasmodium vivax* and *Plasmodium ovale* hypnozoite reservoir in Papua New Guinean children: a randomised placebo-controlled trial and mathematical model. *PLoS Med.* 2015;12(10):e1001891.

54. Ishikawa, H, Ishii, A, Nagai, N, Ohmae, H, Harada, M, Sugura, S, et al. A mathematical model for the transmission of *Plasmodium vivax* malaria. *Parasitol Int.* 2003;52:81–93.
55. Roy, M, Bouma, MJ, Ionides, EL, Dhiman, RC, Pascual, M. The potential elimination of *Plasmodium vivax* malaria by relapse treatment: insights from a transmission model and surveillance data from NW India. *PLoS Negl Trop Dis.* 2013;7:e1979.
56. Pemberton-Ross P, Chitnis N, Pothin E, Smith TA. A stochastic model for the probability of malaria extinction by mass drug administration. *Malar J.* 2017 Sep 18;16:376.
57. World Health Organization. Guidelines for the Treatment of Malaria. 3rd ed. World Health Organization; 2015.
58. Slater, HC, Ross, A, Ouédraogo, AL, White, LJ, Nguon, C, Walker, PG, et al. Assessing the impact of next-generation rapid diagnostic tests on *Plasmodium falciparum* malaria elimination strategies. *Nature.* 2015;528(7580):S94-101.
59. Hofmann NE, Gruenberg M, Nate E, Ura A, Rodriguez-Rodriguez D, Salib M, et al. Assessment of ultra-sensitive malaria diagnosis versus standard molecular diagnostics for malaria elimination: an in-depth molecular community cross-sectional study. *Lancet Infect Dis.* 2018 Oct 1;18(10):1108–16.
60. Hofmann N, Mwingira F, Shekalaghe S, Robinson LJ, Mueller I, Felger I. Ultra-Sensitive Detection of *Plasmodium falciparum* by Amplification of Multi-Copy Subtelomeric Targets. *PLOS Med.* 2015 Mar 3;12(3):e1001788.
61. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature.* 2015 Oct 8;526(7572):207–11.

Chapter 3

3. Can trials of spatial repellents be used to estimate mosquito movement within a village?

Josephine Malinga,^{1,2} Marta Maia,^{1,2,3,4} Sarah Moore,^{1,2,3} Amanda Ross^{1,2}

¹ *Swiss Tropical and Public Health Institute, Basel, Switzerland*

² *University of Basel, Basel, Switzerland*

³ *Ifakara Health Institute, Tanzania*

⁴ *KEMRI/Wellcome Trust Research Programme, Kilifi, Kenya*

This Chapter has been published:

Malinga, J., Maia, M., Moore, S. *et al.* Can trials of spatial repellents be used to estimate mosquito movement? *Parasites Vectors* **12**, 421 (2019).

3.1. Abstract

Background

Knowledge of mosquito movement would aid the design of effective intervention strategies against malaria. However, data on mosquito movement through mark-recapture or genetics studies are challenging to collect and so are not available for many sites. An additional source of information may come from secondary analyses of data from trials of repellents where household mosquito densities are collected. Using the study design of published trials, we developed a statistical model which can be used to estimate the movement between houses for mosquitoes displaced by a spatial repellent. The method uses information on the different distributions of mosquitoes between houses when no households are using spatial repellents compared to when there is incomplete coverage. The parameters to be estimated are the proportion of mosquitoes repelled, the proportion of those repelled that go to another house and the mean distance moved between houses. Estimation is by maximum likelihood.

Results

We evaluated the method using simulation and found that data on the seasonal pattern of mosquito densities was required, which could be additionally collected during a trial. The method was able to provide accurate estimates from simulated data, except when the setting has few mosquitoes overall, few repelled, or the coverage with spatial repellent is low. The trial that motivated our analysis was found to have too few mosquitoes caught and repelled for our method to provide accurate results.

Conclusion

We propose that the method could be used as a secondary analysis of trial data to gain estimates of mosquito movement in the presence of repellents for trials with sufficient numbers of mosquitoes caught and repelled and with coverage levels which allow sufficient numbers of houses with and without repellent. Estimates from this method may supplement those from mark-release-recapture studies, and be used in designing effective malaria intervention strategies, parameterizing mathematical models and in designing trials of vector control interventions.

3.2. Introduction

There has been an increase in interest in the movement of vectors and people, and how each contributes to the spread of malaria infections (1–4). The flow of infections within and between households in an area has implications for interventions such as targeting areas or transmission foci and reactive case detection (5,6). Mosquito movement is the main mode of spread of malaria parasites within a community. Hence, information on how mosquitoes move can help inform the design of intervention strategies and aid in the parameterization of mathematical models to predict their likely impact (7). It can also inform the design of cluster-randomised controlled trials (cRCTs) to estimate the effect of new tools (8).

There is limited information on the movement of mosquitoes between households. Vector dispersal has been estimated at different spatial and temporal scales using mosquito mark-release-recapture (MMRR) and genetic markers (9–14). Previous MMRR studies have shown that approximately 80% or more of mosquitoes recaptured are within three kilometres of release points up to two weeks after release (7,15–17), including those emerging from breeding sites (18–20). Long-range movement between villages, or farther, is only occasionally observed (17,21,22). Distances moved by the mosquito vary by vector species, distribution of host habitats, wind direction and the use of vector control interventions (15,17,18) among other factors.

Both MMRR and genetic methods have limitations. MMRR is dependent on the recapture success, which is affected by factors ranging from geographical landscapes and climate, vector population structure and behaviour, collection effort as a function of distance from release points (10) and how systematic the sampling is, in addition to ethical implications regarding the release of potential disease vectors back into the environment. Population genetic studies using microsatellites and other molecular variants to define fine scale genetic patterns of vectors might be subject to resolution effects, masking patterns (12), and are very costly, limiting the number of mosquitoes that can be analysed. These studies are challenging to carry out and further sources of data would be valuable.

A potential source of data on mosquito movement which has not been fully harnessed is trials of repellents. To estimate the effect of topical and spatial repellents, mosquito densities in households with and without repellents (23–32) have been compared. Some studies have estimated the extent to which mosquitoes are diverted to houses without repellents when there is less than full coverage in a study area. They report the possibility of diversion to non-users (31,33), no change in mosquito densities collected (33,34), while some experimental trials outline the impact of the repellents on the mosquito olfactory cues and delayed feeding

(35,36). These studies have not estimated the distance between households that the mosquitoes were diverted. We sought to determine if data from the trials with diversion could be used to estimate fine-scale movement of mosquitoes in the presence of spatial repellents as a secondary analysis and whether modifications to the trial design would be necessary to achieve this.

Mosquito movement is likely to be altered by the presence of repellents. Spatial repellents such as transfluthrin induce orthokinesis, where the mosquito moves in a random fashion until it moves into a lower concentration when it resumes natural flight (36,37). Therefore, estimates of mosquito movement in the presence of spatial repellents complement those from other data sources.

We develop a statistical model for estimating the movement of mosquitoes that are repelled. We validate the model using simulation to determine the characteristics of a study under which the model could provide accurate estimates of the parameter values. We apply the model to observed data from a trial in Tanzania where the main objective was to investigate whether mosquitoes are diverted from users to non-users of spatial repellents in an area of residual transmission and incomplete spatial repellent coverage (33).Methods

3.3. Methods

Trial design

We use a trial of spatial repellents from Tanzania described previously in Maia *et al* (33). Briefly, the study was conducted in three villages, each with 30 households. The distance between any two villages was greater than two kilometres while households within villages were on average within 0.1km to 0.3km of each other. The study took place over 24 weeks between December 2012 and June 2013. Three coverage scenarios with mosquito coils containing 0.03% transfluthrin were rotated every two weeks among the villages: 1) 100% coverage, 2) no coverage, 3) incomplete coverage with 80% of the households using coils. Coils were distributed and used on each day of the week. Blank coils were used as a placebo. Mosquito densities were collected outdoors under the kitchen thatch roof as well as indoors for three consecutive days each week using Prokopok aspirating devices (38). There were a total of 72 collection days from each household. The presence of a spatial repellent in a household was defined as a combination of two features, availability of a coil with transfluthrin, and observed compliance to coil use. Compliance was assessed by inspecting the ashes produced the previous night.

The original study compared the numbers of mosquitoes collected in households in four groups: households using repellent in weeks with complete coverage in a village, households using repellent in a village when there was incomplete coverage, households not using repellent in a village when there was incomplete coverage, and households in a village when there was zero coverage. For our analysis, we use *Anopheles arabiensis* mosquitoes since they were the most repelled by the active coils.

Statistical Model

a) Model Strategy

We develop a statistical model with the aim of estimating the geographic distances between households that the mosquitoes diverted by the repellent move from and to. Movement of individual mosquitoes cannot be determined, but we can estimate the population parameters such as the distance between houses that the diverted mosquitoes move from and to.

We define the baseline distribution of the proportion of mosquitoes in each house as the distribution of mosquitoes when there is 0% coverage. The proportions may vary between houses and must sum to 1. The total number of mosquitoes per day can vary throughout the study period but we assume that the proportions in each house remain the same in the absence of repellent use. In the case of unfed mosquitoes emerging from breeding sites, this assumption is unlikely to be true. Seasonal patterns in rainfall may vary emergence rates from breeding sites, and newly emerged mosquitoes may cluster in houses closest to a breeding site. Therefore, we restrict the analysis to blood-fed mosquitoes only. Mosquitoes in general take a few days for their first blood-meal (39,40) allowing time for dispersal away from the breeding site. We assume that the distributions of mosquito densities which differ from the baseline distribution when a proportion of households use repellent reflect the movement of the diverted mosquitoes (Figure 1).

The model derives the expected proportion of mosquitoes in each house based on the baseline distribution of mosquitoes between houses when there is zero coverage and the excess outgoing and incoming mosquitoes for each house when some households use spatial repellents. The parameters that govern the outgoing and incoming mosquitoes to be estimated are, the proportion of mosquitoes diverted when repellent is used (β), the parameter for distance between households moved by the diverted mosquitoes (λ), and the proportion of those diverted that go to another house as opposed to elsewhere such as vegetation (φ).

Model A is the base model. Model B is a small extension of model A in the case where data on the seasonal pattern of mosquito densities in the absence of spatial repellents is available.

Model A

Let N_t be the total number of mosquitoes collected from all households on day t . We assume that in the 0% coverage scenario, the proportion of mosquitoes in each house h in a village is given by; $C_1, C_2 \dots C_h$

The proportion of mosquitoes diverted by the repellent is represented by β . We use a and b to denote the house that a mosquito is potentially diverted from and to. Of the total number of mosquitoes in houses on day t , N_t , the proportion diverted from house a , O_{at} , is given by the proportion in the house in the absence of intervention, C_a , multiplied by the proportion diverted, β , so that

$$O_{at} = C_a \beta s_{at} \quad (1)$$

where s_{at} is equal to 1 in a house with repellent use on that day and 0 if the repellent was not used. $O_{at} = \text{zero}$ if no spatial repellent was used.

Diverted mosquitoes may move to another house with probability φ or to somewhere outside the houses with probability $(1 - \varphi)$. Conditional on moving from house a to another house, the probability that a mosquito moves to house b , $Pr(M_{abt} | M_{a.t})$, depends on a function f of the distance in kilometres, d_{ab} , and the repellent status in house b on day t . The probability is scaled so that the probabilities of moving to each house in the village, conditional on moving to a house, sum to one.

$$Pr(M_{abt} | M_{a.t}) = \frac{f(d_{ab}) s_{bt}}{\sum_b f(d_{ab}) s_{bt}} \quad (2)$$

We set the function f , which describes the chance of the mosquito moving to a house depending on distance, to a normal kernel (Figure 3.1), which is similar to diffusion and represents the distance travelled by a random walk in a fixed time period, so that,

$$f(d_{ab}) = \exp\left(-\frac{1}{2} \lambda^{-2} d_{ab}^2\right) \quad (3)$$

where λ , the parameter for distance, is to be estimated. Other distributions may also be used.

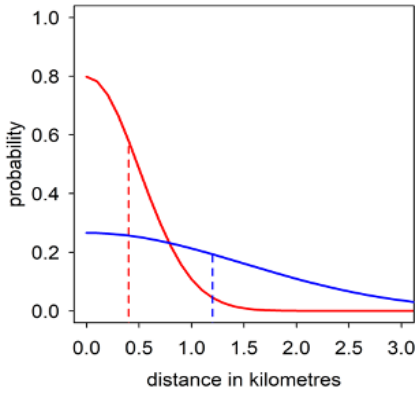


Figure 3.1 Examples of the distributions of geographical distances given by the normal kernel. The dotted lines represent the mean distance; Red line: 0.40km and Blue line: 1.20km. The mean is estimated from the parameter for distance, λ , where $\text{mean} = \lambda \sqrt{2/\pi}$.

The proportion of all mosquitoes on day t who are diverted to house b from house a , I_{abt} , is given by multiplying the probability for being diverted from house a given repellent use, O_{at} , the probability of being diverted to a house rather than elsewhere, φ , and the conditional probability of moving to house b given that the mosquito has moved to another house from house a so that,

$$I_{abt} = Pr(M_{abt} | M_{a.t}) \varphi O_{at} \quad (4)$$

The proportion moving to house b from all other houses, I_{bt} , is then summed over all houses,

$$I_{bt} = \sum_a I_{abt} \quad (5)$$

The proportion of mosquitoes in house h of all mosquitoes on day t , P_{ht} , is then given by the baseline proportion (that would occur if there is zero coverage), C_h , minus the proportion of diverted mosquitoes, O_{ht} , and adding the proportion of incoming mosquitoes, I_{ht} ,

$$P_{ht} = C_h - O_{ht} + I_{ht} \quad (6)$$

The observed densities, Y_{ht} , are based on mosquitoes in houses only (as opposed to those diverted elsewhere) and include houses with missing data. To correspond to the observed densities, we set the predicted proportions to zero in houses with missing mosquito densities. This makes no assumption about their actual values. We then re-scale the proportions to sum to one.

$$Q_{ht} = \frac{P_{ht} w_{ht}}{\sum_h P_{ht} w_{ht}} \quad (7)$$

where w_{ht} is an indicator set to 0 if the house has a missing mosquito density on that day and 1 if the data is present.

The observed densities follow a multinomial distribution around the predicted probabilities.

$$Y_{ht} \sim Mn(Q_{1t}, \dots, Q_{Ht}, N_t) \quad (8)$$

Model B

If data on the total number of mosquitoes including those diverted elsewhere is available, then there is potentially more information with which to disentangle the effects of repellency, movement and diversion elsewhere. Mosquitoes diverted elsewhere are not sampled in the houses, but this information can be gained by having data on the seasonal pattern of mosquito densities in the absence of spatial repellents, either from another control village or from a rotation of coverage levels which allows the seasonal pattern to be estimated.

The model can be modified to incorporate information on the proportion of mosquitoes that are diverted elsewhere. In this case, the observed data is fitted to a multinomial distribution with the expected proportion in each house, Q'_{ht} and an additional category for the expected proportion of mosquitoes that were diverted elsewhere, Q'_{et} . These probabilities are scaled to sum to one. The total number of mosquitoes is N'_t .

$$Y_{ht} \sim Mn(Q'_{1t}, \dots, Q'_{Ht}, Q'_{et}, N'_t) \quad (9)$$

Q'_{et} is calculated by subtracting N_t (the total number of mosquitoes sampled from households) from N'_t (total number of mosquitoes including those which are diverted elsewhere).

Quantities in the models

Table 3.1 Quantities in the models

Quantity	Description
Included in model A and model B	
N_t	number of mosquitoes caught in all houses on day t
C_h	baseline proportion of mosquitoes in house h of mosquitoes in all houses, when there is zero coverage
P_{ht}	proportion of mosquitoes in house h on day t of those in all houses
O_{at}	proportion of mosquitoes diverted from house a on day t of those in all houses
I_{bt}	proportion of mosquitoes diverted to house b on day t of those in all houses
$Pr(M_{a,t})$	probability of mosquito being diverted from house a on day t
$Pr(M_{abt})$	probability that diverted mosquitoes move from house a to house b
Q_{ht}	predicted proportion of mosquitoes in house h on day t of those in all houses
β	proportion of mosquitoes diverted of those in houses using repellent
φ	proportion of mosquitoes moving to another house of those diverted
λ	parameter for distance between households for diverted mosquitoes (mean = $\lambda \sqrt{2/\pi}$)
s_{at}	presence of spatial repellent in house a on day t
d_{ab}	distance between house a and house b
Additionally, included in model B	
N'_t	number of mosquitoes caught in houses and those diverted elsewhere
Q'_{et}	predicted proportion of mosquitoes diverted elsewhere on day t of mosquitoes in houses or diverted elsewhere
Q'_{ht}	predicted proportion of mosquitoes in house h on day t of those in all houses or diverted elsewhere

Implementation

The statistical model was written in C++. The simulations were run on sciCORE (<http://scicore.unibas.ch/>) scientific computing core facility at the University of Basel. We used Nelder-Mead optimization (41,42) to maximize the multinomial log likelihood in order to estimate the parameters of interest; the proportion of mosquitoes repelled from houses with spatial repellents (β), the parameter for distance between households moved by the diverted mosquitoes (λ), and the proportion of mosquitoes repelled that go to a household as opposed to elsewhere (φ).

Model Validation

We evaluated the ability of the models to recover known values using simulated data. We assess the method under different conditions to establish at what level of coverage, the proportion of mosquitoes repelled from households using repellents, the proportion of

mosquitoes repelled going to households as opposed to elsewhere, and mean total number of mosquitoes collected per day, the model is able to reproduce accurate parameter values.

We base the scenarios of trial characteristics on the design of the trial of spatial repellents from Tanzania. We specify a reference scenario in which the model could work well and vary each of the input parameters in turn to determine the values at which the model no longer works well (Table 3.2). We simulate trial datasets of observed numbers of fed mosquitoes for each household per day using our underlying model assumptions and random variation. Since there is stochasticity in the simulations, we simulate 100 datasets for each scenario, and estimate the parameter values for each dataset.

For simplicity, the total number of mosquitoes collected per day remains constant and there is no seasonality.

Table 3.2 Simulated scenarios of trial characteristics to evaluate the method

Quantity	Value
β , proportion diverted from houses using repellent	0.10, 0.30, 0.50, 0.80
φ , proportion of those diverted that go to another house	0.20, 0.50, 0.80
λ , parameter for distance of movement for diverted mosquitoes (km) (mean = $\lambda \sqrt{2/\pi}$)	0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.80, 1.00
N_t , number of mosquitoes on day t in houses (Model A)	10, 100, 1000
N'_t , number of mosquitoes on day t including those diverted elsewhere (Model B)	10, 100, 1000
number of experimental days	72
number of days with zero coverage	18
number of households with spatial repellent out of 30 per day	6, 15, 24, 28

***The reference scenario is indicated by bold font.**

Ethics Statement

Ethical approval for the trial was given by the Ifakara Institutional Review Board (IHI-IRB), from the Tanzanian National Institute for Medical Research (NIMR) and from the Interventions Research Ethics Committee of the London School of Hygiene and Tropical Medicine (33). Permission to publish the data was obtained from the Tanzania National Institute of Medical Research.

3.4. Results

Trial Data

The trial characteristics for the three villages are summarized in Table 3.3.

Table 3.3 Trial Characteristics for the three villages

	Uwata	Matete	Igima
Number of mosquitoes collected per day			
median (90% central range)	6 (2 - 20)	2 (0 - 9)	0 (0 - 3)
Distance between all pairs of households (km)			
median (90% central range)	0.31 (0.14 - 0.50)	0.21 (0.07 - 0.30)	0.14 (0.09 - 0.21)
Compliance to repellent use in each treatment arm			
Complete coverage	90%	89%	93%
Incomplete coverage ¹	90%	90%	93%

¹The denominator is the total number of households allocated the treatment. There were 30 households in each study village.

Model validation

We use simulation to assess how well the method was able to recover known parameter values. The simulations are based on the design of the trial and using the house coordinates for one village, Uwata.

Model A worked well for some parameters, but not others. The estimates are reasonable for β , the proportion of mosquitoes repelled from households with repellents, across the range of values for distance (Figure 3.2a), the proportion of diverted mosquitoes that go to another house (Figure 3.2b), for different coverage levels (Figure 3.2c), and numbers of mosquitoes collected per day (Figure 3.2d). However, for λ , the parameter for distance of diversion between houses, model A estimates were accurate only for scenarios where 80% of mosquitoes were repelled (Figure 3.2e), 80% of mosquitoes repelled went to a house (Figure 3.2f) and with a coverage of around 50% (Figure 3.2g). The estimates for φ , the proportion of mosquitoes repelled that go to households as opposed to elsewhere were poor for all scenarios (Figure 3.2 (bottom row)).

For Model A, there was little information in the simulated datasets to disentangle the effects of the parameter for distance moved by mosquitoes that were repelled, the proportion repelled and the proportion of mosquitoes repelled that go to households as opposed to elsewhere.

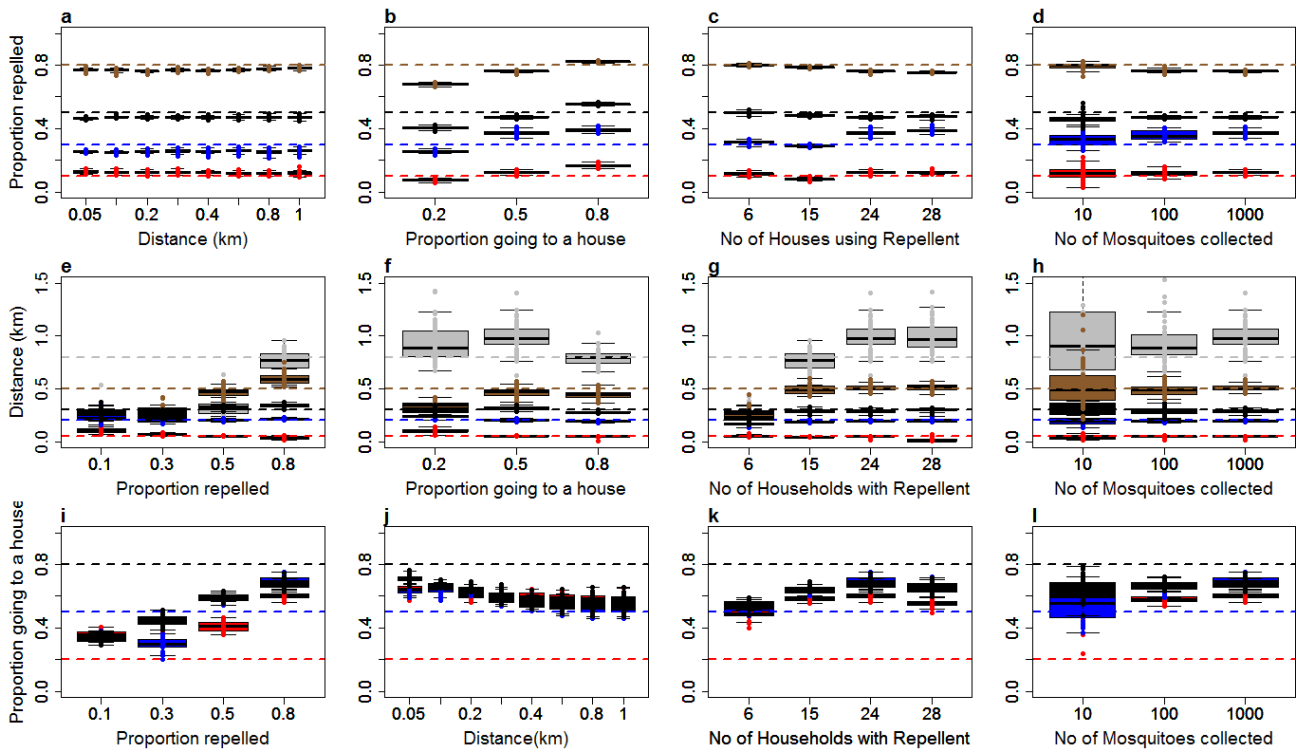


Figure 3.2 The ability of the model to return known parameter values.

Estimated values for β , the proportion of mosquitoes repelled from houses using spatial repellents (top row) a) Estimates by parameter for distance between households moved, b) by proportion of mosquitoes repelled that go to households as opposed to elsewhere, c) by the number of households using spatial repellents (out of the total of 30), d) by the total number of mosquitoes collected from all households per day, **for φ , the proportion of mosquitoes repelled that go to households as opposed to elsewhere (middle row)** e) By the proportion of mosquitoes repelled, f) by the parameter for distance between households moved by diverted mosquitoes, g) by the number of households using spatial repellents on any given day, h) by the total number of mosquitoes collected from all households per day, **for λ , the parameter for distance between households moved by mosquitoes (bottom row)** i) by the proportion of mosquitoes repelled from houses using a spatial repellent, j) by the proportion of mosquitoes repelled that go to households as opposed to elsewhere k) by the number of households using spatial repellents out of a total of 30, l) by the total number of mosquitoes collected per day from all households. The horizontal lines represent the different known values to be returned coded by colour: Red (0.1), Blue (0.3), Black (0.5), and

Grey (0.8). The boxplots represent the estimated values from 100 simulated datasets for each scenario. The reference scenario is based on the trial design and trial house coordinates and is given in Table 3.2. We alter one characteristic at a time.

For Model B, we extend the model to include data on mosquitoes that were diverted elsewhere. For the method evaluation, this can be simulated easily but for a trial, data on the seasonal pattern in the absence of repellents would be required.

For Model B, we extend the model to include data on mosquitoes that were diverted elsewhere. For the method evaluation, this can be simulated easily but for a trial, data on the seasonal pattern in the absence of repellents would be required.

The model returned the correct values for β , the proportion of mosquitoes repelled from households with spatial repellents, for all levels assessed for the parameter for distance between households (Figure 3.3a), the proportion of mosquitoes repelled that go to households as opposed to elsewhere (Figure 3.4b), the coverage of households using spatial repellents (Figure 3.3c) and the total number of mosquitoes collected per day (Figure 3.3d).

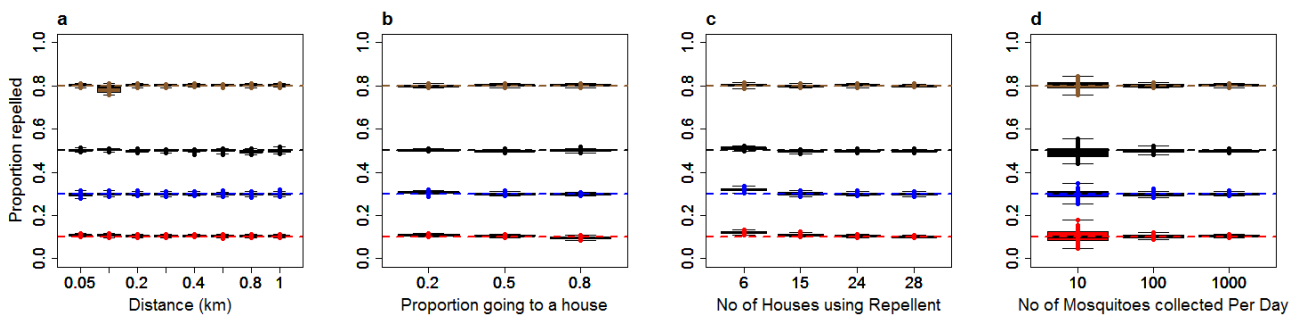


Figure 3.3 The ability of the model to return the known values for β , the proportion of mosquitoes repelled from houses using spatial repellents.

a) By parameter for distance between households moved, b) by proportion of mosquitoes repelled that go to households as opposed to elsewhere, c) by the number of households using spatial repellents (out of the total of 30), d) by the total number of mosquitoes collected from all households per day. The horizontal lines represent the different known values to be returned coded by colour: Red (0.1), Blue (0.3), Black (0.5), and Grey (0.8). The boxplots represent the estimated values from 100 simulated datasets for each scenario. The reference scenario is based on the trial design and trial house coordinates and is given in Table 3.2. We alter one characteristic at a time.

Estimates for φ , the proportion of mosquitoes repelled that go to households as opposed to elsewhere, were also reproduced precisely over the range of parameter for distance (Figure 3.4b), coverage (Figure 3.4c) and total number of mosquitoes collected per day (Figure 3.4d). But, estimates were less precise if the proportion of mosquitoes repelled was low (Figure 3.4a). There is too little information provided by the relatively small number of mosquitoes repelled to produce accurate estimates.

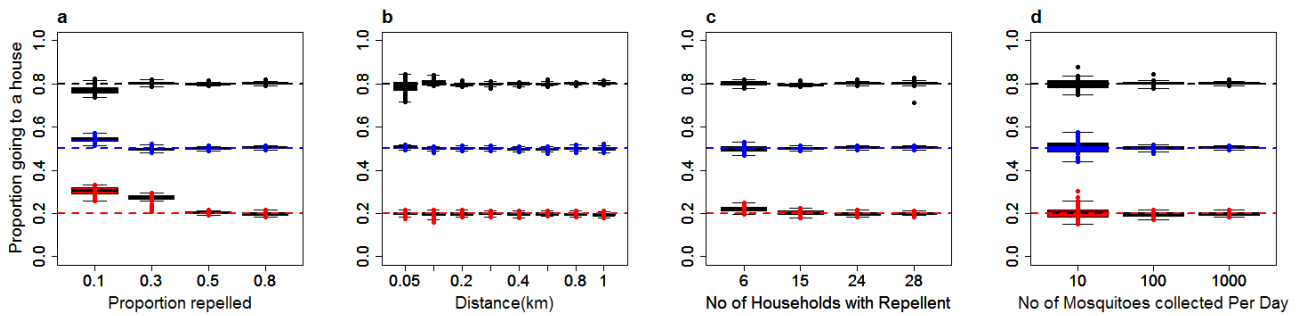


Figure 3.4 The ability of the model to return known values for φ , the proportion of mosquitoes repelled that go to households as opposed to elsewhere

a) By the proportion of mosquitoes repelled, b) By the parameter for distance between households moved by diverted mosquitoes, c) by the number of households using spatial repellents on any given day, d) by the total number of mosquitoes collected from all households per day. The horizontal lines represent the different known values to be returned coded by colour: Red (0.2), Blue (0.5), and Black (0.8). The boxplots represent the estimated values from 100 simulated datasets for each scenario. The reference scenario is based on the trial design and trial house coordinates and is given in Table 3.2. We alter one characteristic at a time.

Parameter for distance between households moved by the mosquito: Estimates for λ , the parameter for distance between households that the mosquitoes were diverted, were accurate when the known values were shorter, but not when the parameter for distance was greater than 0.8km (Figure 3.5a). This is likely to be due to the configuration of the trial village where more than 90% of distances between pairs of houses were less than 800 metres apart (Figure 3.5b).

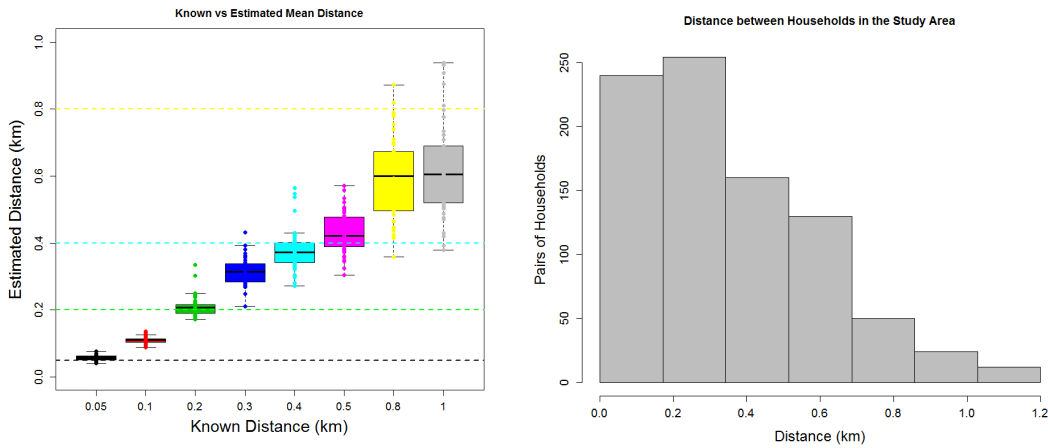


Figure 3.5 Estimated and Known values for λ , the parameter for distance between households moved by the mosquito. (mean = $\lambda \sqrt{2/\pi}$)

a. Estimated (box plots) and actual (dotted lines) simulated parameter for distances between households moved by mosquitoes diverted by the spatial repellent. b. Distribution of the distances between all pairs of households in the study area. The boxplots in (a) represent the estimated values from 100 simulated datasets for each scenario, and the colours represent the different parameters for distances. The reference scenario is based on the trial design and trial house coordinates and is given in Table 3.2.

For the parameter for distance of 800m or less, the model estimates were accurate for scenarios where more than 30% of mosquitoes were repelled (Figure 3.6a), 50% or more mosquitoes repelled went to a house (Figure 3.6b), and with sufficient coverage (Figure 3.6c) and a higher number of mosquitoes collected per day (Figure 3.6d). One hundred mosquitoes per day (3 per house) provided precise estimates, but for 10 (0.3 per house) the precision was less. If there is low coverage, or few mosquitoes are repelled, then there is little information in the dataset to estimate the parameter for distance.

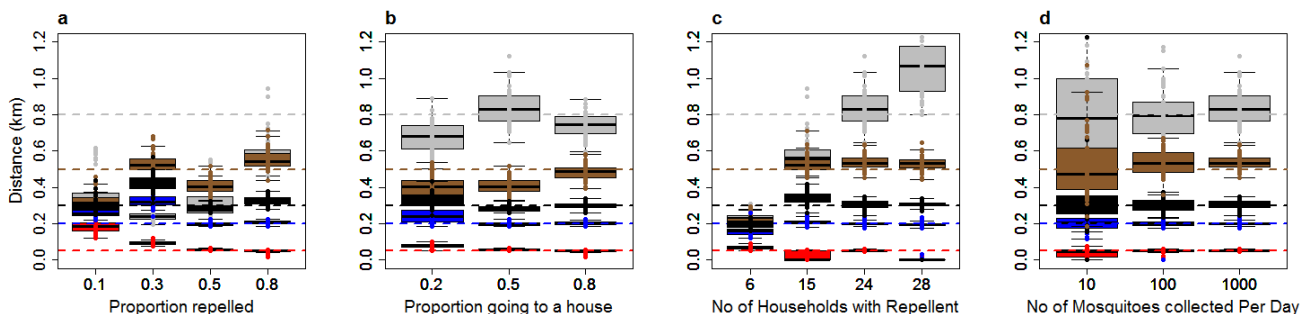


Figure 3.6 The ability of the model to return known values for λ , the parameter for distance between households moved by mosquitoes

a) By the proportion of mosquitoes repelled from houses using a spatial repellent, b) by the proportion of mosquitoes repelled that go to households as opposed to elsewhere c) by the number of households using spatial repellents out of a total of 30, d) by the total number of mosquitoes collected per day from all households. The horizontal lines represent the different known values to be returned coded by colour. Red (0.05km), Blue (0.2km), Black (0.3km), Brown (0.5km), Grey (0.8km). The boxplots represent the estimated values from 100 simulated datasets for each scenario. The reference scenario is based on the trial design and trial house coordinates and is given in Table 2. We alter one characteristic at a time.

Application to data from Kilombero Valley, Tanzania

We applied the method to the observed trial data on collected mosquito densities from Kilombero, Tanzania (Table 3.4). The trial data have characteristics which, from the method evaluation, indicate that the model would not provide accurate estimates. The mosquito densities were low and there was a very low proportion of mosquitoes repelled. There was no evidence of an impact on mosquito abundance of the spatial repellent (33). It has been suspected that the transfluthrin concentration in the coils might have been too low to repel mosquitoes in this particular study.

We applied the model to two of the villages, Uwata and Matete. Due to the extremely low numbers of mosquitoes collected in Igima, it was excluded from further analysis. The estimates are consistent with the study findings that a very low proportion of mosquitoes were diverted by the spatial repellents.

Table 3.4 Parameter estimates using the trial data

Parameter Estimate ¹	Uwata	Matete
	Est. (95%CI)	Est. (95%CI)
proportion of mosquitoes repelled	0.04 (0.03 - 0.04)	0.04 (0.03 - 0.05)
parameter for distance moved between households (mean = $\lambda \sqrt{2/\pi}$)	0.12 (0.09 - 0.35)	0.04 (0.00 - 0.14)
proportion of mosquitoes moving to households of those repelled	0.88 (0.54 - 1.00)	0.87 (0.04 - 1.00)
Log likelihood	-1716.51	-686.02

¹Blood-fed *Anopheles arabiensis* mosquitoes were used in this analysis

3.5. Discussion

We have developed a statistical model to estimate the proportion of mosquitoes repelled from households using spatial repellents, of these, the proportion that are diverted to another house and how far apart the houses are. The evaluation of the method suggests that although Model A, the model without information on seasonality, works well for some parameters, it does not provide accurate estimates for others. The method only works well when there is information on the total number of mosquitoes in the study area, including those diverted elsewhere as opposed to only those diverted to households (Model B). Taken together, these results suggest that trials of repellents could potentially be used to estimate mosquito movement, as long as the trial design is modified so that information on the total number of mosquitoes or the seasonality pattern is available.

Findings from this study may help quantify the criteria for trial settings seeking to estimate mosquito movement by providing insights on what type of data needs to be collected. Our results show that trial data needs to contain sufficient information for the different variables. We found that estimates were not precise if there was a low coverage with repellents (less than 50%) and a low proportion of mosquitoes were repelled from households using repellent (less or equal to 30%). Estimates for longer distances moved between households by the repelled mosquitoes (greater than 800m) were also imprecise: this is expected since the houses in the village were closely arranged. Estimates were reasonably precise if 100 mosquitoes per day (3 per house) were caught, but less precise if this was reduced to 10 mosquitoes (0.3 per house). Simulation could be further exploited to refine the trial design, by investigating factors such as the number of mosquito collections and number of houses when specific trials are being planned. The code is available from the authors on request.

Estimates of mosquito movement in the presence of interventions can inform the design of trials of interventions where effectiveness is affected by movement generally. Mosquito movement has been shown to affect the effectiveness of interventions; the effectiveness can be attenuated through contamination from different study arms, or community wide effects conferred to the surrounding areas (8). In previous trials of bed nets in some African settings, the failure to observe any significant differences between the intervention and control study villages could partially be attributed to the movement of mosquitoes between villages which might have led to the underestimation of the intervention effect (8,22). In Tanzania, mosquitoes were diverted to non-users in trials of topical and spatial repellents in sites with incomplete coverage (31,33), highlighting the need for feasible allotment strategies if complete coverage is hard to achieve. These estimates of mosquito movement can also be used to parameterize mathematical models for assessing the anticipated impacts of

intervention strategies where data is not available. It is not clear how much mosquito movement varies depending on whether spatial repellents are present or absent, but as more studies are carried out, further estimates will become available and potentially allow comparison in similar settings.

The need to estimate mosquito movement from as many sources of data as possible stems from the low number of datasets designed specifically to measure mosquito movement. This is compounded by the need to have estimates from different settings and in the presence of different interventions due to the lack of generalizability. The distances travelled are highly dependent on the setting, due to factors like the vector species and environmental features such as vegetation, breeding sites, wind direction and the spatial distribution of households (13,18).

There are some limitations with our modelling strategy. We did not take into account the number of consecutive evenings that the spatial repellent had been used within each two-week period of intervention or placebo but rather assumed that the effect was constant over time. This may not be correct and could be validated by estimating any trend in mosquito densities among the houses over the fortnight in trials with sufficiently large numbers of mosquitoes. The model could be extended to take further time detail into account, for example by using the estimates of the previous day for the distribution of mosquitoes between houses as the baseline proportions of the current day. Validation could be carried out using further datasets and approaches such as individual-based simulation modelling of mosquito movement (Denz et al, *in prep*). Our model could also be extended to test hypotheses about mosquito movement, such as whether mosquitoes prefer to move to the first house they encounter without repellents or any other. Incorporating data on the number of hosts as a measure of attractiveness or on excess mosquito mortality could refine the estimates.

The low proportion of mosquitoes repelled from households using spatial repellent estimated by the method was consistent with results published previously (33), where there were no significant differences in the number of mosquitoes collected in households with and those without spatial repellents. The reason for the lack of repellency is likely to be the concentration of transfluthrin which would have been too low for substantial action in natural settings with free air movements (37). The deterrence and repellency effects of transfluthrin are dose-dependent with substantial protective effects seen at higher concentrations than those used in the current study e.g. 0.03% Transfluthrin coils used indoors (35). Using blank coils as a placebo may also reduce differences between houses with spatial repellents and those without due to the effects of smoke.

3.6. Conclusion

We have developed a statistical model as a potential tool to gain information on mosquito movement from trials of repellents. If the design of trials of repellents is modified to provide information on the total number of mosquitoes using the seasonal pattern, then the method is able to reproduce known values from simulated datasets well. Further work to validate the method in field settings is needed. Estimates of mosquito movement can inform the design of both intervention strategies and trials of interventions where effectiveness is affected by movement generally, and in particular estimates of movement in the presence of spatial repellents may inform decisions on implementation and allocation.

3.7. References

1. Churcher TS, Cohen JM, Novotny J, Ntshalintshali N, Kunene S, Cauchemez S. Measuring the path toward malaria elimination. *Science*. 2014 Jun 13;344(6189):1230–2.
2. Reiner Jr RC, Menach AL, Kunene S, Ntshalintshali N, Hsiang MS, Perkins TA, et al. Mapping residual transmission for malaria elimination. *eLife*. 2015 Dec 29;4:e09520.
3. Ruktanonchai NW, DeLeenheer P, Tatem AJ, Alegana VA, Caughlin TT, Erbach-Schoenberg E zu, et al. Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. *PLOS Comput Biol*. 2016 Apr 4;12(4):e1004846.
4. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the Impact of Human Mobility on Malaria. *Science*. 2012 Oct 12;338(6104):267–70.
5. Lines J, Kleinschmidt I. Combining malaria vector control interventions: some trial design issues. *Pathog Glob Health*. 2013 Jan;107(1):1–4.
6. Thomas CJ, Cross DE, Bøgh C. Landscape Movements of *Anopheles gambiae* Malaria Vector Mosquitoes in Rural Gambia. *PLoS ONE*. 2013 Jul 18;8(7).
7. Mandal S, Sarkar RR, Sinha S. Mathematical models of malaria - a review. *Malar J*. 2011 Jul 21;10:202.
8. Killeen GF, Knols BG, Gu W. Taking malaria transmission out of the bottle: implications of mosquito dispersal for vector-control interventions. *Lancet Infect Dis*. 2003 May 1;3(5):297–303.
9. Cianci D, Van Den Broek J, Caputo B, Marini F, Della Torre A, Heesterbeek H, et al. Estimating Mosquito Population Size from Mark—Release—Recapture Data. *J Med Entomol*. 2013 May 1;50(3):533–42.
10. Guerra CA, Reiner RC, Perkins TA, Lindsay SW, Midega JT, Brady OJ, et al. A global assembly of adult female mosquito mark-release-recapture data to inform the control of mosquito-borne pathogens. *Parasit Vectors*. 2014;7:276.
11. Villela DAM, Codeço CT, Figueiredo F, Garcia GA, Maciel-de-Freitas R, Struchiner CJ. A Bayesian Hierarchical Model for Estimation of Abundance and Spatial Density of *Aedes aegypti*. *PLOS ONE*. 2015 Apr 23;10(4):e0123794.
12. Rašić G, Filipović I, Weeks AR, Hoffmann AA. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics*. 2014 Apr 11;15:275.

13. Gillies MT. Studies on the dispersion and survival of *Anopheles gambiae* Giles in East Africa, by means of marking and release experiments. Bull Entomol Res. 1961 Apr;52(1):99–127.
14. McCall PJ, Moshia FW, Njunwa KJ, Sherlock K. Evidence for memorized site-fidelity in *Anopheles arabiensis*. Trans R Soc Trop Med Hyg. 2001 Dec;95(6):587–90.
15. Ejercito A, Urbino CM. Flight range of gravid and newly emerged anopheles. Bull World Health Organ. 1951;3(4):663–71.
16. Costantini C, Li S-G, Torre AD, Sagnon N, Coluzzi M, Taylor CE. Density, survival and dispersal of *Anopheles gambiae* complex mosquitoes in a West African Sudan savanna village. Med Vet Entomol. 1996 Jul 1;10(3):203–19.
17. Takken W, Charlwood JD, Billingsley PF, Gort G. Dispersal and survival of *Anopheles funestus* and *A. gambiae s.l.* (Diptera: Culicidae) during the rainy season in southeast Tanzania. Bull Entomol Res. 1998 Oct;88(5):561–6.
18. Midega JT, Mbogo CM, Mwnambi H, Wilson MD, Ojwang G, Mwangangi JM, et al. Estimating dispersal and survival of *Anopheles gambiae* and *Anopheles funestus* along the Kenyan coast by using mark-release-recapture methods. J Med Entomol. 2007 Nov;44(6):923–9.
19. Manga L, Fondjo E, Carnevale P, Robert V. Importance of low dispersion of *Anopheles gambiae* (Diptera: Culicidae) on malaria transmission in hilly towns in south Cameroon. J Med Entomol. 1993 Sep;30(5):936–8.
20. Prince JAAL, Griffiths THD. Flight of Mosquitoes: Studies on the Distance of Flight of *Anopheles Quadrimaculatus*. Public Health Rep 1896-1970. 1917;32(18):656–9.
21. Baber I, Keita M, Sogoba N, Konate M, Diallo M, Doumbia S, et al. Population Size and Migration of *Anopheles gambiae* in the Bancoumana Region of Mali and Their Significance for Efficient Vector Control. PLoS ONE. 2010 Apr 21;5(4).
22. Thomson MC, Connor SJ, Quiñones ML, Jawara M, Todd J, Greenwood BM. Movement of *Anopheles gambiae s.l.* malaria vectors between villages in The Gambia. Med Vet Entomol. 1995 Oct;9(4):413–9.
23. Chen-Hussey V, Carneiro I, Keomanila H, Gray R, Bannavong S, Phanalasy S, et al. Can Topical Insect Repellents Reduce Malaria? A Cluster-Randomised Controlled Trial of the Insect Repellent N,N -diethyl- m -toluamide (DEET) in Lao PDR. PLOS ONE. 2013 Aug 14;8(8):e70664.

24. Deressa W, Yihdego YY, Kebede Z, Batisso E, Tekalegne A, Dagne GA. Effect of combining mosquito repellent and insecticide treated net on malaria prevalence in Southern Ethiopia: a cluster-randomised trial. *Parasit Vectors*. 2014 Mar 28;7:132.
25. Moore SJ, Davies CR, Hill N, Cameron MM. Are mosquitoes diverted from repellent-using individuals to non-users? Results of a field study in Bolivia. *Trop Med Int Health TM IH*. 2007 Apr;12(4):532–9.
26. Hill N, Zhou HN, Wang P, Guo X, Carneiro I, Moore SJ. A household randomized, controlled trial of the efficacy of 0.03% transfluthrin coils alone and in combination with long-lasting insecticidal nets on the incidence of *Plasmodium falciparum* and *Plasmodium vivax* malaria in Western Yunnan Province, China. *Malar J*. 2014 May 31;13:208.
27. Kawada H, Temu EA, Minjas JN, Matsumoto O, Iwasaki T, Takagi M. Field Evaluation of Spatial Repellency of Metofluthrin-Impregnated Plastic Strips Against *Anopheles gambiae* Complex in Bagamoyo, Coastal Tanzania. *J Am Mosq Control Assoc*. 2008 Sep 1;24(3):404–9.
28. Mng'ong'o FC, Sambali JJ, Sabas E, Rubanga J, Magoma J, Ntamatungiro AJ, et al. Repellent Plants Provide Affordable Natural Screening to Prevent Mosquito House Entry in Tropical Rural Settings—Results from a Pilot Efficacy Study. *PLoS ONE* [Internet]. 2011 Oct 12 [cited 2016 Jun 10];6(10).
29. Moore SJ, Darling ST, Sihuincha M, Padilla N, Devine GJ. A low-cost repellent for malaria vectors in the Americas: results of two field trials in Guatemala and Peru. *Malar J*. 2007 Aug 1;6:101.
30. Sangoro O, Turner E, Simfukwe E, Miller JE, Moore SJ. A cluster-randomized controlled trial to assess the effectiveness of using 15% DEET topical repellent with long-lasting insecticidal nets (LLINs) compared to a placebo lotion on malaria transmission. *Malar J*. 2014;13:324.
31. Maia MF, Onyango SP, Thele M, Simfukwe ET, Turner EL, Moore SJ. Do topical repellents divert mosquitoes within a community? Health equity implications of topical repellents as a mosquito bite prevention tool. *PloS One*. 2013;8(12):e84875.
32. Wilson AL, Chen-Hussey V, Logan JG, Lindsay SW. Are topical insect repellents effective against malaria in endemic populations? A systematic review and meta-analysis. *Malar J*. 2014;13:446.
33. Maia MF, Kreppel K, Mbeyela E, Roman D, Mayagaya V, Lobo NF, et al. A crossover study to evaluate the diversion of malaria vectors in a community with incomplete

- coverage of spatial repellents in the Kilombero Valley, Tanzania. *Parasit Vectors*. 2016;9:451.
34. Syafruddin D, Bangs MJ, Sidik D, Elyazar I, Asih PB, Chan K, et al. Impact of a Spatial Repellent on Malaria Incidence in Two Villages in Sumba, Indonesia. *Am J Trop Med Hyg*. 2014 Dec 3;91(6):1079–87.
 35. Ogoma SB, Lorenz LM, Ngonyani H, Sangusangu R, Kitumbukile M, Kilalangongono M, et al. An experimental hut study to quantify the effect of DDT and airborne pyrethroids on entomological parameters of malaria transmission. *Malar J*. 2014;13:131.
 36. Ogoma SB, Ngonyani H, Simfukwe ET, Mseka A, Moore J, Maia MF, et al. The mode of action of spatial repellents and their impact on vectorial capacity of *Anopheles gambiae sensu stricto*. *PLoS One*. 2014;9(12):e110433.
 37. Ogoma SB, Moore SJ, Maia MF. A systematic review of mosquito coils and passive emanators: defining recommendations for spatial repellency testing methodologies. *Parasit Vectors*. 2012;5:287.
 38. Maia MF, Robinson A, John A, Mgando J, Simfukwe E, Moore SJ. Comparison of the CDC Backpack aspirator and the Prokopack aspirator for sampling indoor- and outdoor-resting mosquitoes in southern Tanzania. *Parasit Vectors*. 2011;4:124.
 39. Charlwood JD, Pinto J, Sousa CA, Ferreira C, Petrarca V, Rosario V do E. 'A mate or a meal' – Pre-gravid behaviour of female *Anopheles gambiae* from the islands of São Tomé and Príncipe, West Africa. *Malar J*. 2003 May 7;2:9.
 40. Gillies MT, Wilkes TJ. A study of the age-composition of populations of *Anopheles gambiae* Giles and *A. funestus* Giles in North-Eastern Tanzania. *Bull Entomol Res*. 1965 Dec;56(2):237–62.
 41. Nelder JA, Mead R. A Simplex Method for Function Minimization. *Comput J*. 1965 Jan 1;7(4):308–13.
 42. O'Neill R. Algorithm AS 47: Function Minimization Using a Simplex Procedure. *J R Stat Soc Ser C Appl Stat*. 1971;20(3):338–45.

Chapter 4

4. Investigating the drivers of the spatio-temporal patterns of genetic differences between *Plasmodium falciparum* malaria infections in Kilifi County, Kenya

Josephine Malinga^{1, 2}, Polycarp Mogeni³, Irene Omedo³, Kirk Rockett⁴, Christina Hubbart⁴, Anne Jeffreys⁴, Thomas Williams^{3, 5}, Dominic Kwiatkowski^{4, 6}, Philip Bejon^{3, 7}, Amanda Ross^{1, 2*}

¹ *Swiss Tropical and Public Health Institute, Basel, Switzerland*

² *University of Basel, Basel, Switzerland*

³ *Kenya Medical Research Institute-Wellcome Trust Research Programme, Kilifi, Kenya*

⁴ *Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK*

⁵ *Department of Medicine, South Kensington Campus, Imperial College London, London, UK*

⁶ *Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK*

⁷ *Centre for Tropical Medicine & Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK*

This Chapter has been published:

Malinga, J., Mogeni, P., Omedo, I. *et al.* Investigating the drivers of the spatio-temporal patterns of genetic differences between *Plasmodium falciparum* malaria infections in Kilifi County, Kenya. *Sci Rep* **9**, 19018 (2019).

4.1. Abstract

Knowledge of how malaria infections spread locally is important both for the design of targeted interventions aiming to interrupt malaria transmission, and the design of trials to assess the interventions. A previous analysis of 1602 genotyped *Plasmodium falciparum* parasites in Kilifi, Kenya collected over 12 years, found an interaction between time and geographic distance: the mean number of single nucleotide polymorphism (SNP) differences was lower for pairs of infections which were both a shorter time interval and shorter geographic distance apart. We determine whether the empiric pattern could be reproduced by a simple model, and what mean geographic distances between parent and offspring infections and hypotheses about genotype-specific immunity or a limit on the number of infections would be consistent with the data. We developed an individual-based stochastic simulation model of households, people and infections. We parameterized the model for the total number of infections, and population and household density observed in Kilifi. The acquisition of new infections, mutation, recombination, geographic location and clearance were included. We fit the model to the observed numbers of SNP differences between pairs of parasite genotypes. The patterns observed in the empiric data could be reproduced. Although we cannot rule out genotype-specific immunity or a limit on the number of infections per individual, they are not necessary to account for the observed patterns. The mean geographic distance between parent and offspring malaria infections for the base model was 0.40km (95%CI 0.24 – 1.20), for a distribution with 58% of distances shorter than the mean. Very short mean distances did not fit well, but mixtures of distributions were also consistent with the data. For a pathogen which undergoes meiosis in a setting with moderate transmission and a low coverage of infections, analytic methods are limited but an individual-based model can be used with genotyping data to estimate parameter values and investigate hypotheses about underlying processes.

4.2. Introduction

Ultimately, the spatial and temporal patterns of malaria infections at the community level are made up of movements of individual mosquitoes, human hosts and the parasites inside them. Understanding how malaria infections spread locally and the processes leading to the observed spatial and temporal distribution patterns is important for the design of interventions aiming to reduce and interrupt transmission by targeting foci, to estimate the spread of drug resistance, or to prevent resurgence ^{1,2}.

Investigating genetic patterns can give insights into how infections spread ³⁻⁷. The availability of high-throughput genotyping methods has enabled the genetic characterization of the malaria parasite, identifying polymorphisms on the genome that can be used to infer the degree of genetic relatedness between infections ^{6,8}. Using single nucleotide polymorphism (SNP) genotyping, some studies show clustering of parasites into distinct sub-populations between Thai and African ⁹⁻¹¹, West and East African ^{10,12}, and Asia and African isolates ^{10,12} while others report persistent clustering of genetically identical parasites across years in Senegal ⁹. On a sub-national scale, a study from school surveys in western Kenya, an area of high malaria transmission, found no spatial structure to genetic relatedness over geographical distance in a well mixing parasite population ⁴. Studies have also used genetic information to estimate the risk of imported infections ^{7,13}, to estimate transmission networks in Swaziland ¹⁴, to provide confirmatory signals of the decline of transmission in Senegal ^{15,16}, and to describe the relatedness of parasite pairs between proximal clinics at the Thai-Myanmar border ¹⁷ and focal transmission in Zambia ¹⁸. However, there are a few studies focusing on parasite genotype patterns in the community on a fine scale, such as within a district.

A recent study investigated *Plasmodium falciparum* genotypes in two districts in Kenya, and in one district in The Gambia ³. All possible pairs of observed parasite genotypes within each site were formed and the number of SNPs different calculated. In the sites with longitudinal data, the effect of the time interval on the pairwise SNP differences was modified by geographic distance. For short distances, the number of SNPs different increased with time, but saturated as distance between parasites increased ³.

While these studies provide information on the distribution of malaria parasite genotypes over space and time with qualitative inferences on parasite mixing, the underlying parameters and processes leading to these patterns have not been well established and the geographic distances between parent and offspring infections remain unclear.

In many malaria endemic areas, vector movement is likely to be the most frequent mechanism for the spread of infections. Studies have shown that mosquitoes tend to fly relatively short

distances from their breeding sites in search of blood meals ^{19,20}, depending on the distance to human settlements, the availability of easily accessible hosts, and whether any barriers to movement exist ^{20,21}. In Kilifi, mosquitoes were recaptured in households within 0.7km of larval habitats up to 14 days after release ¹⁹, while in other settings flight distance from release points including breeding sites, was reported to range from less than 100m to 3km ²²⁻²⁵, or further ²⁶. However, there is limited information on the distances between households moved by the mosquito from the initial bite when the mosquito becomes infected to the subsequent bites when further humans are infected. After a blood meal, the mosquito looks for a resting place to digest the blood and develop eggs and after two to three days, searches for a breeding site to oviposit before seeking another host. The extrinsic incubation period (EIP) takes approximately 10 to 12 days for the mosquito to become infectious, corresponding to roughly two or three feeding cycles. Human movement may also inform malaria distribution patterns ^{27,28}, especially in populations with low transmission and high rates of imported infections ^{28,29}. Hence, simply knowing the dispersal distance from breeding sites does not lead to a calculation from which we can predict the mixing of parasite genotypes.

The interaction between time and geographic distance on genetic differences found by Omedo *et al* raises questions about the processes behind it. The explanation may be stochastic drift or may require additional processes such as genotype-specific immunity. The effective repertoire of antigenic variation may be reduced by immunity raised from the previous infections with the same or similar genotype. Deliberate infections of the same strain in naïve adults as malaria therapy led to shorter patent infections with lower parasite densities than infections with a different strain ^{30,31}. It is not also known if there is a limit on the number of infections that one person can have at a time. There is saturation in the mean multiplicity of infection (MOI) detected with transmission intensity ^{32,33} but this may be due to the limits of detection. This would decrease the chance of large numbers of similar infections circulating in the same household. Recombination may occur in the mosquito during meiosis if gametocytes from multiple infections are drawn up in a blood meal. Recombination is likely to be dependent on the intensity of transmission and the relative gametocyte densities of the infections ^{34,35} and may play a role in the distribution of parasite genotypes across time and space, mainly through reshuffling of alleles that are already present.

Our aim was to assess which processes and parameter values are consistent with the observed spatial and temporal patterns of parasite genetic differences observed in Kilifi County, Kenya ³. We develop a stochastic simulation model and fit it to the observed number of SNPs different between pairs of parasite genotypes. We estimate the distance between parent and offspring malaria infections, and investigate the role of recombination, imported infections,

and hypotheses about functions of immunity. We evaluated the ability of this method to recover known parameter values using simulated data.

4.3. Methods

i. Data Sources

Study site. The study was conducted within the area covered by the Kilifi Health and Demographic Surveillance System (KHDSS) on the Kenyan Coast. The area covers 891km² with a population of approximately 260,000 people³⁶. Most of the residents live off small scale farming and fishing. Malaria transmission in Kilifi is seasonal³⁷, with long rains between April and June and short rains between October and December. Malaria prevalence reported from both health facilities and community surveys declined over the period of the study^{38,39}, attributed mainly to increased coverage of control interventions, improved case management and environmental changes^{38,40}.

Sample collection and genotyping. The details of sample collection and genotyping process have been described elsewhere³. Briefly, 1602 *P.falciparum* infected blood samples were collected over 14 years between 1998 and 2011 in Kilifi. Of these, 1259 (79%) samples were from hospital cases in children between three months and 13 years old, 195 were from community surveys with participants between 3 weeks and 85 years old and 148 were from short term laboratory cultured samples³. The residence location of the participant providing the sample was recorded. The samples were genotyped using the Sequenom MassARRAY iPLEX platform⁴¹. Up to 256 SNPs were typed in each sample, across the 14 parasite chromosomes³. Sample success rates and genotyping pass rates were determined for each sample and for each SNP. SNPs and samples with high failure rates (>30%) were excluded from the final analyses³.

Data preparation. For the present study, the number of SNPs included, n_s , is 53. We selected bi-allelic SNPs with a minor allele frequency (MAF) of greater than 5%. We denote the presence of a major allele at each SNP position by 1 and the minor allele by 0. Each infection is then characterized by a string of zeros and ones to represent the observed alleles at each of the selected loci. The base-10 number is generated from this base-2 number to give a unique identifier for each infection genotype. We assume that the genotypes represent single infections.

Calculating pairwise time, distance and SNP differences. Following Omedo *et al*³, the observed parasite genotypes were paired and the number of SNPs different between them

used as the measure of genetic difference. Pairwise differences in time in days and distance in kilometres were also computed.

ii. Modelling Strategy

We aimed to determine which processes and parameter values were consistent with the observed patterns of genetic differences over time and distance in Kilifi, where the observed numbers of SNPs different were fewer for parasite pairs with both a shorter time interval and a shorter geographic distance apart³.

We developed a simple individual-based stochastic simulation model describing how malaria infections are transmitted over time and space. The acquisition and clearance of individual blood-stage infections in people living within homesteads was simulated. The site-specific inputs such as house density, number of people per house, the total number of infections and prevalence over time were derived from published estimates from Kilifi^{36,38}. We extended the base model to a set of model variants including the effects of imported infections and hypotheses about whether there needs to be genotype-specific acquired immunity or a limit on the number of current infections per individual for model predictions to be consistent with the patterns in the data.

We objectively measured how well the model variants and parameter values fit the data using a likelihood function which compares the numbers of SNPs different in pairs of blood samples to those from pairs of simulated infections drawn from the same time points and closest locations. We estimated the mean distance between parent and offspring infections and the probability of recombination in multiply infected individuals.

iii. Base Model

We refer to the simplest simulation model as the “base model”. The simulation model was seeded with a number of initial current infections. Households and people within households were randomly selected for each initial infection. The initial infection genotypes were generated by randomly selecting alleles based on the MAF in the observed data at each locus for the selected SNPs.

At each five-day time-step, each current infection could give rise to new infections. The number of new infections for an infection i of age a at time t , n_{iat} , was stochastic and was drawn from a Poisson distribution with mean μ_{at} , so that $n_{iat} \sim \text{Poisson}(\mu_{at})$. The mean number of new infections, μ_{at} , was the product of the relative infectiousness to mosquitoes of an infection of age a , I_a , and a constant, m_t , so that $\mu_{at} = I_a m_t$ (Table 1). Values of m_t were fixed to generate the number of infections which was set to the estimated prevalence in Kilifi

multiplied by the number of individuals and the MOI corresponding to the prevalence and estimated by a systematic review (39, Malinga *et al submitted*). I_a followed a decay function which was derived from the estimated period of higher infectiousness of an infection ⁴².

Each new infection was assigned a genotype. The genotype was the same as that of the parent infection except for two processes, mutation and recombination. Mutations may occur in the new infection, where each allele mutated with probability p_m . In individuals with more than one current infection, recombination occurred with probability p_r . If a new infection was the product of recombination, then the infection with which it recombines was drawn at random from the remaining infections in the same individual. The two infections may have the same genotype, in which case selfing occurs. Recombination hotspots for different chromosomes have been reported ^{43,44}, but we assumed that the probability of a break is equal along the length of the chromosome and each chromosome could break in only one place ⁴³. The new recombinant infection had one genotype. We recognize that meiosis results in four different genotypes but including all four would substantially increase complexity for little gain in accuracy: infections are expected to lead to zero or one subsequent infection due to declining transmission, the chance of selecting a pair of simulated infections arising from the same recombinant infection is small and the number of infections is tailored to that of Kilifi thus either co-infections or separate infections would allow a similar number of recombinations. The effect of the model not capturing the full complexity of the recombinations in the observed data would likely be conservative by increasing the imprecision of the estimates. We assumed a constant probability of clearance of each infection in a five-day time-step, p_c .

Each new infection was also assigned a homestead. The homestead was selected randomly, with probability according to a function of the distance from the location of the current infection. We assumed that the probability follows a normal kernel, simulating diffusion. The cumulative distance of Brownian motion at a specified time is given by the kernel for positive distances of a normal distribution with mean of zero and a standard deviation denoted by σ . Therefore, the mean distance between parent and offspring infections is given by $\sigma \sqrt{2/\pi}$, and 58% of distances are shorter than this value (Figure 4.1). The probabilities are not scaled so that an isolated household has a lower chance of transmitting to other households than does a house which is close to other houses. An individual within each homestead was selected at random to host the new infection.

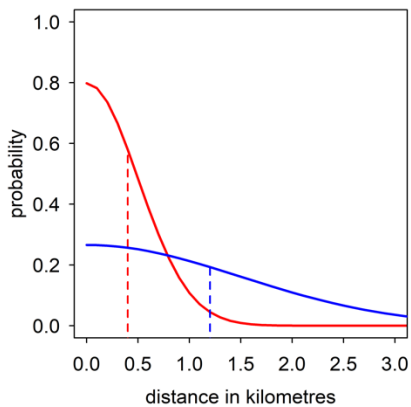


Figure 4.1 Examples of the half-normal distribution probability density function for positive values of the distance between parent and offspring infections. The dotted lines mark the mean of the distribution; Red line = 0.40km, Blue line = 1.20km.

Table 4.1 Quantities in the simulation model

Quantity	Description	Units of measurement
n_s	number of SNPs	positive integer
a	age of an infection	five-day time-step
t	time from start of simulation	five-day time-step
n_{iat}	number of new infections for infection i of age a at time-step t	Integer, greater or equal to zero
μ_{at}	mean number of new infections for an infection of age a at time-step t	Numeric, greater or equal to zero
m_t	mean number of new infections for an infection at time-step t	Numeric, greater or equal to zero
I_a	relative infectiousness to mosquitoes of an infection of age a	Numeric, $0 \leq I_a \leq 1$
p_m	probability of mutation for one allele per infection cycle	Probability
p_c	probability of clearance per five-day time-step	Probability
p_r	probability of recombination conditional on multiple infections in a host	Probability
σ	parameter for distance between parent and offspring infection (mean = $\sigma \sqrt{2/\pi}$)	Kilometres

iv. Model variants

We included alternative model variants to investigate the effect of imported infections and whether hypotheses about specific immunity are supported by the observed data. We simulated (i) the base model, (ii) the base model with imported infections, whose alleles were chosen with probability equal to the individual allele frequencies, (iii) the base model with a limit on the number of current infections: new infections cannot establish in a person if they already have the maximum permitted number of infections, (iv) the base model with acquired immunity to recently seen genotypes: a new infection cannot establish in a person if they have already seen a number of genotypes with a limited number of SNPs different in a recent time period, (v) the base model with heterogeneity in transmission at the household level, and (vi) the base model with each combination of pairs of (ii) – (v).

We did not simulate acquired immunity in individuals in general, but only specific acquired immunity to recently seen genotypes since it may affect the spatial and temporal distribution of parasite genotypes.

v. Model Inputs

The area simulated. Population and household density estimates were derived from the KHDSS surveillance reports (Table 4.2). We required 20 homesteads per kilometer squared, with eight residents per homestead. Using the estimated parasite prevalence in Kilifi ³⁸, and an estimate for MOI of 1.5 corresponding to the prevalence ³², we would require approximately 200 malaria infections per square kilometer at the beginning of the study period declining to 25 by the end. For the whole study area of 891km², however, this would have high computational demands. Instead, we simulated an area of 9km by 9km. This is the largest size that could be run within 24 hours for the base model and comfortably covers the range of zero to 3km where the interaction between time and distance was observed ³. We included 1800 homesteads each with eight people. The number of simulated infections present in the 9km by 9km square at the same time is around 13,000 at the beginning of the study period declining to 2,000 at the end of the study.

At the edge of the grid, infections have a restriction on the direction in which they can move. To avoid artificially pushing too many infections back towards the centre of the grid, we specified that current infections in a 5% border at the edge of the 9km by 9km grid give rise to half as many new infections per time-step. We used the central 6km by 6km area for fitting, allowing a 3km border to minimize edge effects on the pattern of genetic differences.

Table 4.2 Inputs to the model

Parameter	Value	reference
<i>Kilifi Characteristics</i>		
population density	260,000 people	36
area size (in km ²)	891	36
median number of people per homestead (IQR)	8 (6 - 11)	
median homestead density per km ² (IQR)	20 (7 - 58)	(a)
estimated parasite prevalence	60% in 1998; 10% in 2011	38
number of infections per km ²	200 in 1998, 25 in 2011	(b)
mean number of new infections per five-day time-step (μ_t)	Varies	(c)
<i>General fixed inputs</i>		
probability of clearance per five-day time-step, p_c	0.025	(d) ^{45,46}
probability of mutation per SNP, p_m	2.0×10^{-7}	(e)
relative infectiousness of an infection by age of infection, I_a	Varies	(f) ⁴²

- (a) The co-ordinates for households in the simulated 9km by 9km grid were randomly drawn from two independent uniform distributions.
- (b) The expected number of infections was derived from microscopy and RDT prevalence data and multiplied by the MOI corresponding to the prevalence in a systematic review (39, Malinga *et al submitted*).
- (c) The values of μ_t are fixed according one of the phases of the simulation to produce the correct numbers of infections estimated for Kilifi in (b).
- (d) Assuming an exponential decay function, a probability of clearance of 0.025 per five-day time-step corresponds to a mean duration of infection of 200 days ^{46,46-48}.
- (e) The probability of a mutation per base pair per generation has been estimated to be 1.7×10^{-9} ⁴⁹. We multiplied this by the estimated number of generations in the liver stage, for the gametocytes and in the asexual blood-stage, each of 48 hours, before the mature gametocytes are taken up in a blood meal. The probability of at least one mutation per SNP per infection cycle is then approximately 2.0×10^{-7} .
- (f) Assumes an exponential decay function estimated from a period of higher infectiousness of an infection (probability of infection > 5%) ⁴².

Number of infections and model run time. The was initialized with 13,000 current infections, distributed randomly across homesteads and to individuals within the homesteads.

The model running period was divided into three phases. To achieve the required numbers of infections over time, values of m_t , the multiplier for the number of new infections for a current infection in time-step t , were set to be constant within but to vary between the different phases of the simulation. Value of m_t also needed to vary slightly according to the

input parameter values, to reach the required numbers of infections in the study area. The initial phase was made up of 1000 five-day time-steps with m_t set to lead to an increasing number of current infections. This phase was not necessary if the number of initial infections was equal to the number required at the start of the study period. The second phase had 1000 time steps with m_t set to keep the number of infections constant as a warm-up period to allow the structure in the genetic differences to develop, and the study period phase covering the 850 five-day time steps for the observation period with m_t set to reproduce the decline in the number of infections observed in Kilifi.

Sensitivity analysis. We conducted sensitivity analyses to investigate the influence of input parameter values where the real value was unclear. These include varying the amount of clustering of the distribution of households within the study area, the number of initial current infections, the warm-up period and the size of the simulation area.

There is also uncertainty surrounding the genotyping of infections from samples collected from individuals with multiple *P.falciparum* infections. It may be that a dominant infection is captured during genotyping, or that the process selects different alleles from different infections. For the base model, we assume that only one of the infection genotypes is sampled from individuals with multiple or recombinant infections. However, we investigate the effect of detecting a mix of alleles for different SNPs from different infections in the same person.

vi. Fitting the models to the data

The model was fitted to the data by comparing the number of SNPs different between simulated infections to the number of SNPs different in the observed infections. The simulated infections were matched to the observed infections on time-step and location.

The study area was divided into 6km by 6km grid squares. The locations of each observed infection in each square were transformed from GPS co-ordinates into the co-ordinates of the 6km by 6km area, and the date that the sample was collected converted into the time-step of the simulation. We nest the 6km by 6km square within the 9km by 9km simulated square. The 3km buffer serves to provide the correct spatial structure for infections within the nested area.

For each observed infection, a number of simulated infections were selected at exactly the same time-step. Infected individuals closest to the location of the observed infection were identified, and a maximum of one infection per individual was sampled. To balance the effects of stochasticity, the requirement for selected infections to lie close to the observed location and computational demands, five simulated infections were selected for each observed infection, and each simulation was repeated with ten random seeds.

Within each 6km by 6km square k , we formed all possible pairs of observed infections, and for each pair j , calculated the observed number of SNPs different, d_{jk} . We then calculated the number of SNPs different in the simulated data, q_{jrk} within each 6km by 6km square k , for each pair j , for each of the 10 replicates of the simulated infections r . We gained the mean proportion of SNP differences for the replicates \bar{q}_{jk}/n_s . The Binomial log likelihood was used to quantify the amount of support from the data, calculated by summing over all the pairs j within each square k and over all the squares.

$$ll = \sum_k \sum_j d_{jk} \log\left(\frac{\bar{q}_{jk}}{n_s}\right) + (n_s - d_{jk}) \log\left\{1 - \left(\frac{\bar{q}_{jk}}{n_s}\right)\right\}$$

The observed pairs are not independent given that one observed parasite genotype contributes to multiple pairs of data. Therefore, 95% confidence intervals were obtained by using bootstrap resampling of the observed number of SNP differences.

vii. Method evaluation

We evaluated how precisely the method could estimate parameter values. We simulated data with known values for the σ , the parameter for distance between parent and offspring infections, and probability of recombination, p_r , and then applied the method to the simulated data to see how well the known values can be recovered. We used the reference scenario with the base model (Table 4.3), but with a constant rather than declining number of infections over the study period.

viii. Scenarios simulated for the Kilifi setting

We simulate all combinations of the different parameter values as a grid search (Table 4.3) and calculate the log likelihoods.

Table 4.3 Simulated scenarios

Parameters to be estimated by a grid search	Value
parameter for the distance between parent and offspring infections (in kilometers)**	0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 1.00, 1.20, 1.50, 2.00, 2.50, 3.00
probability of recombination resulting from multiply infected individuals	0.01, 0.50, 1.00
Parameters included in the model variants	Value*
imported infections per 1000 people per year (model variant (ii))	0, 5, 10

total number of current infections per person (model variant (iii))	no limit , 20
maximum number of recently seen similar genotypes for new infection to establish (model variant (iv))	no limit , 10
number of SNPs different for defining 'similar' genotypes (model variant (iv))	0 , 10
number of time steps for counting recently seen similar genotypes (model variant (iv))	0 , 40
heterogeneity between houses in transmission (model variant (v))	No heterogeneity , log normal distribution with mean μ_{at} and standard deviation of 0.01

*The base model scenario is indicated by bold font

**The mean of a half normal distribution is given by $\sigma \sqrt{2/\pi}$

The individual based stochastic simulation model was created in R statistical software (version 3.02). The simulations were run on sciCORE (<http://scicore.unibas.ch/>) scientific computing core facility at the University of Basel.

4.5. Results

i. Method evaluation

The method was able to recover known values from simulated data for σ , the parameter for distance between parent and offspring infections, and the probability of recombination (Figure 4.2). The model could reproduce reasonable estimates for σ , but the estimates were less precise if the σ was 2.0km or greater. For longer distances, the size of the area simulated may limit the accuracy.

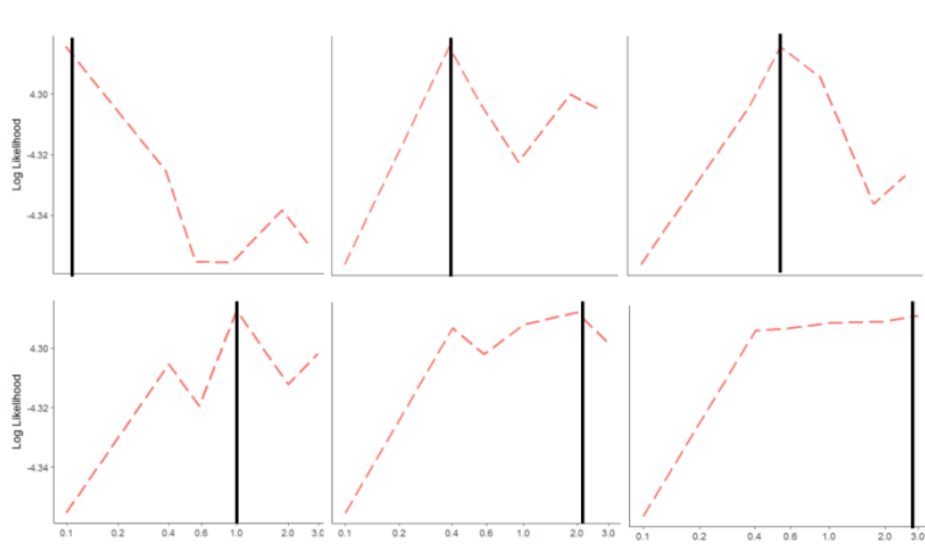


Figure 4.2 Ability of the method to recover known parameter values from simulated data

Black solid line: indicates the true parameter value for σ . Red dashed line: the log likelihood. The log likelihood is a measure of the support from the data for a parameter value. The estimated parameter value is the one which coincides with the maximum log likelihood. In this figure, the estimated parameter values are correct since they align with the black lines. The method and simulated data used the base model with the probability of recombination in multiply infected individuals set to 0.5 (details of the base model are given in the Methods section).

ii. Application to data from Kilifi County

Reproducing the number of infections. We were able to reproduce the estimated number of infections in the study area and their decline over the study period, by altering the input mean number of new infections per current infection (Supplementary Information). The required value of this input varied by scenario, where longer mean distances and adding

immunity functions increased the number of new infections required per current infection per time-step.

Model variants and parameter values consistent with the observed data. We fit each model variant to the observed data on the numbers of SNPs different for pairs of parasites. The aim was to determine which model variants and parameter estimates were consistent with the observed interaction between the time and geographical distance.

Different values for the probability of recombination in multiply infected individuals resulted in similar values for the best-fitting mean distance at the maximum log likelihood (Figure 4.3). The probability of recombination had an impact on the log likelihood only for very short mean distances of parasite movement, where it increased variation in the genotypes and increased the log likelihood. Hence, we adopted one value (0.5) for all further analyses shown.

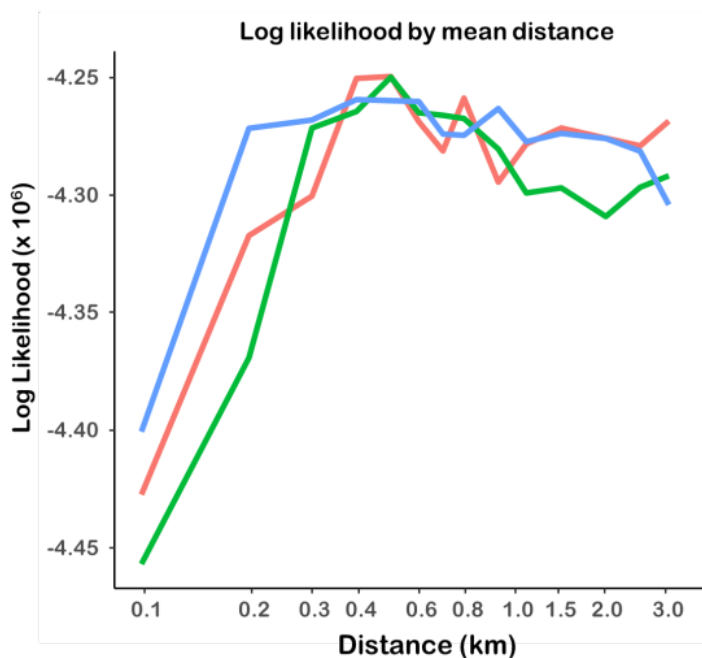


Figure 4.3 Patterns of the log likelihood for different values of recombination for the base model

The x-axis shows the value of σ , the parameter for the distance between parent and offspring infections. The log-likelihood is a measure of support from the data for the parameter values. Red solid line: the base model with probability of recombination in multiply infected individuals set to 0.5; Green line: 0.0; Blue line: 1.0.

The estimates of the mean distance between parent and offspring infections vary slightly by model variant, and the confidence intervals are wide (Table 4.4).

Table 4.4 Estimated mean distance between parent and offspring infections for each model variants

Model Variant**	mean distance in km* (95%CI)
Base model	0.40 (0.24 – 1.20)
Model variant (ii)	0.32 (0.24 – 1.20)
Model variant (iii)	0.80 (0.40 – 1.99)
Model variant (iv)	0.24 (0.24 – 1.20)
Model variant (v)	0.32 (0.24 – 1.20)

*The mean of a half-normal distribution is $\sigma \sqrt{2/\pi}$. In all cases, the probability of a short distance is highest and 58% of parent-offspring distances are shorter than the mean value (Figure 4.1)

**Model variant (ii): the base model with imported infections; Model variant (iii): the base model with a limit on the number of current infections per person; Model variant (iv): the base model with immunity to recently seen similar genotypes; Model variant (v): the base model with heterogeneity in transmission at the household level.

The maximum log likelihood values, a measure of the goodness-of-fit, were similar for the model variants. The differences in the log likelihoods between model variants were greatest for the short mean distances, where the likelihoods tended to increase when the model variants were likely to increase the variation between the genotypes (Figure 4.4). While it was not possible to rule out any model variant, the base model alone was sufficient to reproduce the observed patterns. This suggests that stochastic drift is sufficient to account for the observed interaction.

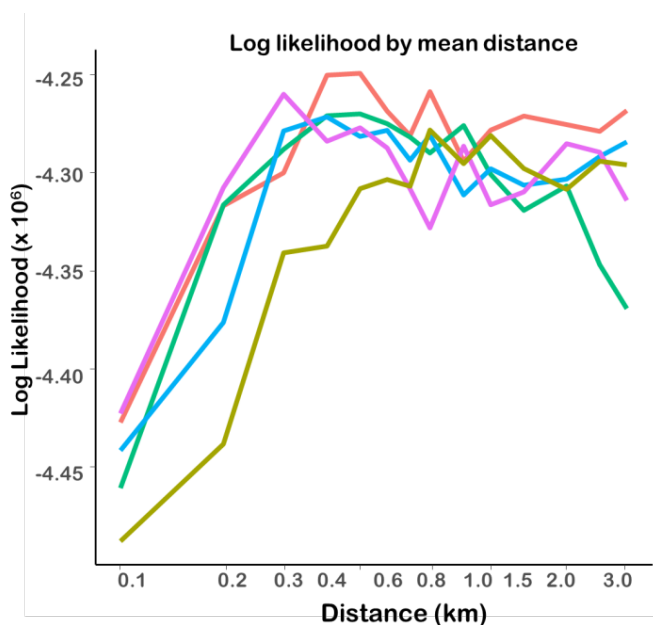


Figure 4.4 Patterns of the log likelihood by σ , the parameter for the distance for the different model variants.

Red line: the base model; Blue line: the base model with imported infections; Green line: the base model with heterogeneity in transmission at the household level; Purple line: the base model with immunity to recently seen genotypes, Brown line: the base model with a limit on the number of current infections per person.

We explored the possibility that the distribution of distances comprises a mixture of short and longer mean distances between parent and offspring infections. This may arise due to a mix of human and vector movement, a mix of local and imported infections, because the normal kernel implicitly assumes the same length of time between the mosquito bites infecting the mosquito and the human and this is likely to vary, or due to the influence of wind or interventions on mosquito dispersal¹⁹. We simulated data using the base model, but assuming a mixture distribution for the mean distance with half of the distances from a distribution with a short and half from a distribution with a longer mean distance. Applying the method to the simulated data produced a log likelihood profile which was also consistent with the Kilifi data indicating that it is a potential explanation (Figure 4.5), although the exact values for the mixture proportions and the two means are unknown and would be difficult to estimate from this data.

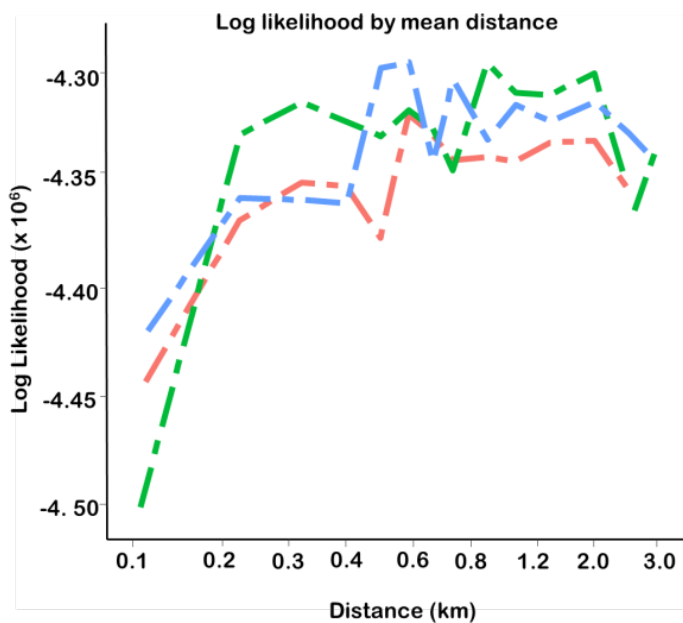


Figure 4.5 Patterns of the log likelihood by σ , the parameter for the distance for data simulated from mixture distributions.

Data was simulated assuming a 50:50 mixture distribution for short and longer mean distances of movement. Red two-dash line: 0.05 & 3.0km; Green two-dash line: 0.1 & 2.5km; Blue two-dash line: 0.4 & 4.0km.

Reproducing time-distance patterns in observed data. For each model variant, the observed interaction shown by Omedo and colleagues³ could be reproduced for at least some parameter values. The number of SNPs different between pairs of malaria infections increased with geographical distance for short time periods, but this trend was less apparent for longer time periods. The predictions for the base model are shown (Figure 4.6), and these patterns were similar for all model variants including the variants of model combinations (ii) – (iv) with imported infections, functions of acquired immunity and heterogeneous transmission at the household level.

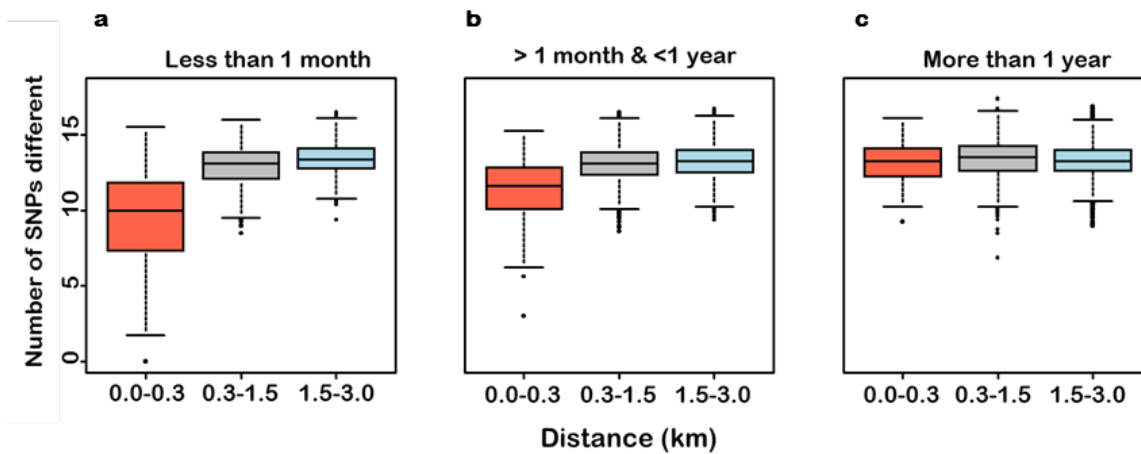


Figure 4.6 Predicted effect of time and distance interaction on the number of SNPs different between pairs of infections.

These predictions are from the base model with the best-fitting value for mean geographical distance (0.40km)

The residual plot showed that the model fitted well with no systematic patterns in the residuals (Figure 4.7).

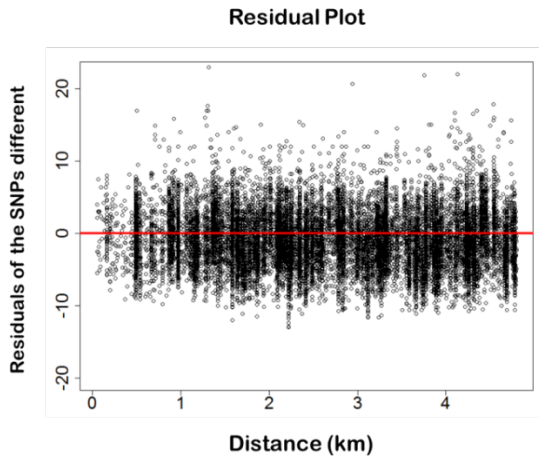


Figure 4.7 Plot of residuals by geographic distance.

The base model was used with the best fitting value for mean geographic distance (0.40km).

Sensitivity analysis of the impact of uncertain input values. We explored the impact of varying the inputs in the base model on the parameter estimates: the number of initial infections, the warm-up period of the simulation, and the degree of clustering of the homesteads. The log likelihood values, measuring how well the model fits, were slightly lower overall for the clustered homesteads and with a shorter warm-up period. A smaller number of initial infections lead to slightly lower log likelihood values for short mean distances, due to the decrease in variation in the genotypes. However, none of these factors substantially influenced the estimated distance between parent and offspring infections (Figure 4.8). The effect of assuming that the observed genotypes are a mix of alleles from multiple infections within individuals was to increase uncertainty in the parameter estimates (not shown).

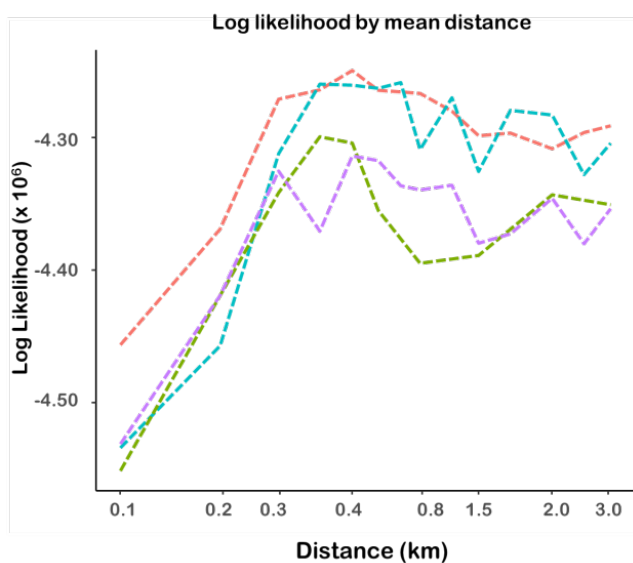


Figure 4.8 Patterns of the log likelihood by σ , the parameter for distance for different values of the input parameters.

Red dashed line: reference scenario for sensitivity analysis (base model with constant transmission, 13,000 initial infections with only the warm-up and model run-in period, uniform distribution of households), Blue dashed line: reference with a lower number of initial infections (1000 initial infections with an additional run-in period), Green dashed line: reference scenario with half the warm-up period (500 time-steps), Purple dashed line: reference with 50% clustering of households in the study area.

4.6. Discussion

We have developed an individual-based stochastic simulation model of malaria infections to determine processes and parameter values which are consistent with the observed patterns of genetic differences in Kilifi County, Kenya. The method was able to recover known parameter values reasonably well from simulated data and could reproduce the number of infections in the study area during the study period.

The observed spatial and temporal patterns of genetic differences could be reproduced for some parameter values for all of the model variants. The log likelihood for the best-fitting parameter values was similar for most of the variants. While we cannot rule out effects of imported infections, genotype-specific acquired immunity or a limit on the number of current infections per person, they were not necessary to account for the observed patterns in the data. This suggests that stochastic drift is a sufficient explanation.

Studies have found that the genotypes of people from the same house are more similar than those from further apart in a village⁵⁰⁻⁵² and that index and secondary cases within a radius of 140m in a study of reactive case detection (RCD) study in Zambia were more genetically similar than cases from the radii of other index cases¹⁸. One might conclude, based on these and the data of Omedo *et al*, that short distances between parent and offspring infections are more likely. The range of possible values for the mean distance between parent and offspring infections in our study ranged between 0.24km (0.24km – 1.20km) and 0.80km (0.40km – 1.99km) for the different model variants. These results are consistent with these previous findings, since with the distribution used, 58% of the distances are less than the mean.

Our results also indicated that mixtures of short and longer mean distances were consistent with the observed data, which could potentially represent vector and human movement. It would be challenging to determine the exact mixture of distributions given that the patterns

of the log likelihoods are similar for different combinations of values. The mixture of distributions we used to investigate whether this would be consistent with the data was composed of two normal kernels. If one kernel should represent human movement, then another distribution may be more appropriate since it is likely to be incorrect if people only tend to stay elsewhere overnight for distances too far to get home or if human movement tends to follow major roads rather than random movements in all directions.

Findings from this study have implications for the assessment and planning of malaria interventions which are affected by parasite movement such as vector control, for setting radii for targeted interventions, for designing trials to evaluate interventions and for informing estimates of the spread of drug resistance^{53,54}. While targeting interventions to areas of relatively higher risk on fine scales has been considered⁵⁴⁻⁵⁶, our findings highlight the need to consider the movement of infections between untargeted and targeted zones. A modelling study investigating the contribution of RCD towards malaria elimination in Zambia emphasised the need to limit importation of infections from connected high transmission areas, since they could reduce the overall benefits⁵⁷.

There are some limitations to our method. Some of the parameter inputs in the study area were simplified, we did not consider differences according to age in the population, nor did we assume any population turn-over. Our rationale was that this simplifies the model and that although general acquired immunity affects the parasite densities, it does not affect the spatial distribution of genotypes as long as the age-groups are geographically mixed in the population. In one model variant, we account for genotype-specific immunity, but this is limited to recent exposure.

During the study period, the area experienced moderate declining to low transmission intensity and the proportion of infections sampled out of the total number in the area was small: we estimate using the reported prevalence^{38,39} and the relationship between prevalence and MOI³² that this proportion is less than 1%. In addition, the number of SNPs sampled was not extensive. Therefore measures of identity by descent (IBD)⁵⁸⁻⁶⁰ or network models of transmission^{61,62} could not be used.

The sampling may have affected the patterns seen, but this is not known. The samples came from hospital admissions in children aged three months to 13 years and community surveys in individuals aged three weeks to 85 years. We assume that the genotypes in those sampled are no different from those in individuals not sampled. The community surveys were carried out in specific locations but were not otherwise restricted. The hospital sampling may have affected the locations and socio-economic status (SES) of the individuals included. Individuals with higher SES may be over-represented because they may be more likely to present for

treatment or under-represented because they may be less likely to have co-morbidities which can increase the risk of severe illness. The distances travelled by the mosquito could potentially be greater in areas of higher SES due to greater protection by LLIN or housing.

This is the first study that we know of which has attempted to estimate parameter values or test hypotheses from malaria genotyping data with a low coverage of infections in a setting with moderate transmission. Few methods are available and although our method is a blunt tool, we have gleaned some insights. Our findings raise questions about whether there is a mixture of more than one distribution for the distances between parent and offspring infections, and a potential influence of immunity on the spatial and temporal distribution patterns of genotypes. It may be possible to gain more definitive results by incorporating additional data sources to isolate mosquito dispersal or human mobility patterns. To fully utilise the information from genotyping tools, purpose-designed datasets have been recommended⁵, ideally with a greater number of SNPs and greater coverage of infections.

4.7. Conclusion

The observed interaction between time and space in the patterns of genetic differences between pairs of *Plasmodium falciparum* infections in Kilifi can be reproduced by an individual-based simulation model. The model did not need any assumptions about genotype-specific acquired immunity or a limit on the number of current infections per person in order to reproduce the patterns, suggesting that stochastic drift is sufficient to account for the interaction.

The estimate of the mean distance between parent and offspring infections was 0.40km (95%CI 0.24 – 1.20) for the base model. The pattern of results was also consistent with a mixture of distributions with short and longer mean distances.

Estimates of the spread of infections have implications for the design and evaluation of malaria control and elimination interventions. Simulation models fitted to genotyping data can be used as an analytic tool to glean insights in settings with moderate transmission and a low coverage of infections where methods are limited.

Availability of data

The code and data analysed during this study are available at:

<https://github.com/josemalinga/Patterns-of-genetic-differences-in-Kilifi>

4.8. References

1. Tanner, M. *et al.* Malaria eradication and elimination: views on how to translate a vision into reality. *BMC Med.* **13**, 167 (2015).
2. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).
3. Omedo, I. *et al.* Micro-epidemiological structuring of *Plasmodium falciparum* parasite populations in regions with varying transmission intensities in Africa. *Wellcome Open Res.* **2**, 10 (2017).
4. Omedo, I. *et al.* Geographic-genetic analysis of *Plasmodium falciparum* parasite populations from surveys of primary school children in Western Kenya. *Wellcome Open Res.* **2**, 29 (2017).
5. Wesolowski, A. *et al.* Mapping malaria by combining parasite genomic and epidemiologic data. *bioRxiv* 288506 (2018) doi:10.1101/288506.
6. Tessema, S. K. *et al.* Applying next-generation sequencing to track falciparum malaria in sub-Saharan Africa. *Malar. J.* **18**, 268 (2019).
7. Tessema, S. K. *et al.* Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. *eLife* **8**, e43510 (2019).
8. Wesolowski, A. *et al.* Mapping malaria by combining parasite genomic and epidemiologic data. *bioRxiv* 288506 (2018) doi:10.1101/288506.
9. Daniels, R. *et al.* A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar. J.* **7**, 223 (2008).
10. Neafsey, D. E. *et al.* Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.* **9**, R171 (2008).
11. Nkhoma, S. C. *et al.* Close kinship within multiple-genotype malaria parasite infections. *Proc R Soc B* **279**, 2589–2598 (2012).
12. Campino, S. *et al.* Population Genetic Analysis of *Plasmodium falciparum* Parasites Using a Customized Illumina GoldenGate Genotyping Assay. *PLOS ONE* **6**, e20251 (2011).
13. Chang, H.-H. *et al.* Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. *eLife* **8**, e43481 (2019).
14. Reiner, R. C., Jr *et al.* Mapping residual transmission for malaria elimination. *eLife* **4**, e09520 (2015).

15. Daniels, R. F. *et al.* Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc. Natl. Acad. Sci.* **112**, 7067–7072 (2015).
16. Daniels, R. *et al.* Genetic Surveillance Detects Both Clonal and Epidemic Transmission of Malaria following Enhanced Intervention in Senegal. *PLOS ONE* **8**, e60780 (2013).
17. Taylor, A. R. *et al.* Quantifying connectivity between local *Plasmodium falciparum* malaria parasite populations using identity by descent. *PLOS Genet.* **13**, e1007065 (2017).
18. Pringle, J. C. *et al.* Genetic evidence of focal *Plasmodium falciparum* transmission in a pre-elimination setting in Southern Province, Zambia. *J. Infect. Dis.* (2018).
19. Midega, J. T. *et al.* Estimating dispersal and survival of *Anopheles gambiae* and *Anopheles funestus* along the Kenyan coast by using mark-release-recapture methods. *J. Med. Entomol.* **44**, 923–929 (2007).
20. Prince, J. A. A. L. & Griffitts, T. H. D. Flight of Mosquitoes: Studies on the Distance of Flight of *Anopheles Quadrimaculatus*. *Public Health Rep. 1896-1970* **32**, 656–659 (1917).
21. Thomas, C. J., Cross, D. E. & Bøgh, C. Landscape Movements of *Anopheles gambiae* Malaria Vector Mosquitoes in Rural Gambia. *PLoS ONE* **8**, (2013).
22. Thomson, M. C. *et al.* Movement of *Anopheles gambiae* s.l. malaria vectors between villages in The Gambia. *Med. Vet. Entomol.* **9**, 413–419 (1995).
23. Gillies, M. T. Studies on the dispersion and survival of *Anopheles gambiae* Giles in East Africa, by means of marking and release experiments. *Bull. Entomol. Res.* **52**, 99–127 (1961).
24. Takken, W., Charlwood, J. D., Billingsley, P. F. & Gort, G. Dispersal and survival of *Anopheles funestus* A. *gambiae* s.l. (Diptera: Culicidae) during the rainy season in southeast Tanzania. *Bull. Entomol. Res.* **88**, 561–566 (1998).
25. Baber, I. *et al.* Population Size and Migration of *Anopheles gambiae* in the Bancoumana Region of Mali and Their Significance for Efficient Vector Control. *PLoS ONE* **5**, (2010).
26. Huestis, D. L. *et al.* Windborne long-distance migration of malaria mosquitoes in the Sahel. *Nature* **574**, 404–408 (2019).
27. Tatem, A. J. & Smith, D. L. International population movements and regional *Plasmodium falciparum* malaria elimination strategies. *Proc. Natl. Acad. Sci.* **107**, 12222–12227 (2010).
28. Pindolia, D. K. *et al.* Human movement data for malaria control and elimination strategic planning. *Malar. J.* **11**, 205 (2012).

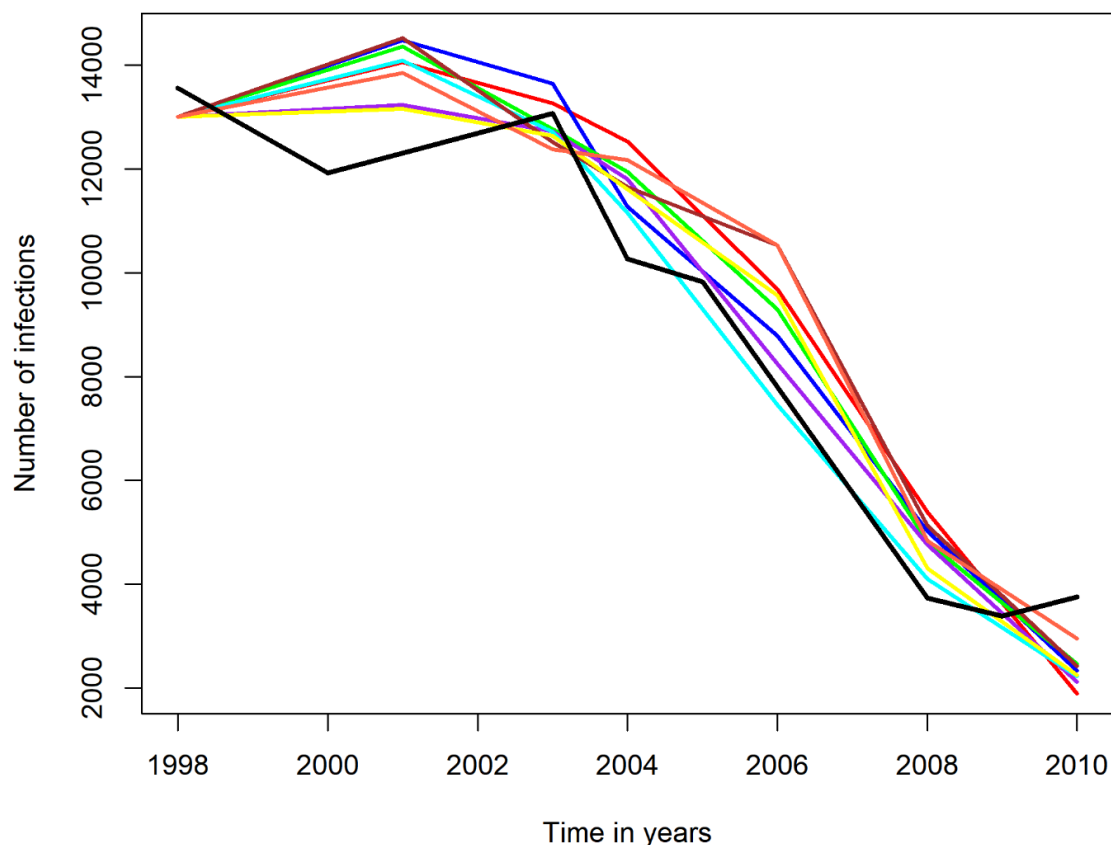
29. Searle, K. M. *et al.* Distinct parasite populations infect individuals identified through passive and active case detection in a region of declining malaria transmission in southern Zambia. *Malar. J.* **16**, 154 (2017).
30. Boyd, M. F., Stratman-Thomas, W. K. & Kitchen, S. F. On Acquired Immunity to *Plasmodium falciparum*. *Am. J. Trop. Med.* **16**, (1936).
31. Jeffery, G. M. Epidemiological significance of repeated infections with homologous and heterologous strains and species of *Plasmodium*. *Bull. World Health Organ.* **35**, 873–882 (1966).
32. Karl, S. *et al.* Spatial Effects on the Multiplicity of *Plasmodium falciparum* Infections. *PLOS ONE* **11**, e0164054 (2016).
33. Malinga, J., Ross, A. & Lines, J. What proportion of *Plasmodium falciparum* and *Plasmodium vivax* malaria infections are in mosquitoes? *In press*.
34. Jeffery, G. M. & Eyles, D. E. Infectivity to mosquitoes of *Plasmodium falciparum* as related to gametocyte density and duration of infection. *Am. J. Trop. Med. Hyg.* **4**, 781–789 (1955).
35. Conway, D. J. *et al.* High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* **96**, 4506–4511 (1999).
36. Scott, J. A. G. *et al.* Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *Int. J. Epidemiol.* **41**, 650–657 (2012).
37. Snow, R. W. *et al.* Relationships between *Plasmodium falciparum* Transmission by Vector Populations and the Incidence of Severe Disease at Nine Sites on the Kenyan Coast. *Am. J. Trop. Med. Hyg.* **52**, 201–206 (1995).
38. Snow, R. W. *et al.* Changing Malaria Prevalence on the Kenyan Coast since 1974: Climate, Drugs and Vector Control. *PLoS ONE* **10**, (2015).
39. Mogeni, P. *et al.* Age, Spatial, and Temporal Variations in Hospital Admissions with Malaria in Kilifi County, Kenya: A 25-Year Longitudinal Observational Study. *PLoS Med.* **13**, (2016).
40. O'Meara, W. P. *et al.* Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya., Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet Lond. Engl. Lancet* **372**, **372**, 1555, 1555–1562 (2008).
41. Gabriel, S., Ziaugra, L. & Tabbaa, D. SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. in *Current Protocols in Human Genetics* (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142905.hg0212s60.

42. Johnston, G. L., Smith, D. L. & Fidock, D. A. Malaria's Missing Number: Calculating the Human Component of R_0 by a Within-Host Mechanistic Model of *Plasmodium falciparum* Infection and Transmission. *PLOS Comput. Biol.* **9**, e1003025 (2013).
43. Jiang, H. *et al.* High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. *Genome Biol.* **12**, R33 (2011).
44. Mu, J. *et al.* Recombination Hotspots and Population Structure in *Plasmodium falciparum*. *PLoS Biol.* **3**, (2005).
45. Sama, W., Killeen, G. & Smith, T. Estimating the Duration of *Plasmodium falciparum* Infection from Trials of Indoor Residual Spraying. *Am. J. Trop. Med. Hyg.* **70**, 625–634 (2004).
46. Bretscher, M. T. *et al.* The distribution of *Plasmodium falciparum* infection durations. *Epidemics* **3**, 109–118 (2011).
47. Felger, I. *et al.* The Dynamics of Natural *Plasmodium falciparum* Infections. *PLOS ONE* **7**, e45542 (2012).
48. Sama, W., Dietz, K. & Smith, T. Distribution of survival times of deliberate *Plasmodium falciparum* infections in tertiary syphilis patients. *Trans. R. Soc. Trop. Med. Hyg.* **100**, 811–816 (2006).
49. Bopp, S. E. R. *et al.* Mitotic Evolution of *Plasmodium falciparum* Shows a Stable Core Genome but Recombination in Antigen Families. *PLOS Genet.* **9**, e1003293 (2013).
50. Tessema, S. K. *et al.* Phylogeography of var gene repertoires reveals fine-scale geospatial clustering of *Plasmodium falciparum* populations in a highly endemic area. *Mol. Ecol.* **24**, 484–497 (2015).
51. Bejon, P. *et al.* Serological Evidence of Discrete Spatial Clusters of *Plasmodium falciparum* Parasites. *PLoS ONE* **6**, (2011).
52. Amambua-Ngwa, A. *et al.* Long-distance transmission patterns modelled from SNP barcodes of *Plasmodium falciparum* infections in The Gambia. *Sci. Rep.* **9**, 1–9 (2019).
53. Hawley, W. A. *et al.* Community-Wide Effects Of Permethrin-Treated Bed Nets On Child Mortality And Malaria Morbidity In Western Kenya. *Am. J. Trop. Med. Hyg.* **68**, 121–127 (2003).
54. Bousema, T. *et al.* Hitting Hotspots: Spatial Targeting of Malaria for Control and Elimination. *PLoS Med.* **9**, (2012).

55. Woolhouse, M. E. J. *et al.* Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 338–342 (1997).
56. Bousema, T. *et al.* Identification of Hot Spots of Malaria Transmission for Targeted Malaria Control. *J. Infect. Dis.* **201**, 1764–1774 (2010).
57. Gerardin, J. *et al.* Effectiveness of reactive case detection for malaria elimination in three archetypical transmission settings: a modelling study. *Malar. J.* **16**, 248 (2017).
58. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. hmmlBD: software to infer pairwise identity by descent between haploid genotypes. *bioRxiv* 188078 (2017).
59. Henden, L., Lee, S., Mueller, I., Barry, A. & Bahlo, M. Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLOS Genet.* **14**, e1007279 (2018).
60. Li, B. *et al.* Leveraging Identity-by-Descent for Accurate Genotype Inference in Family Sequencing Data. *PLoS Genet.* **11**, (2015).
61. Reiner Jr, R. C. *et al.* Mapping residual transmission for malaria elimination. *eLife* **4**, e09520 (2015).
62. Welch, D., Bansal, S. & Hunter, D. R. Statistical inference to advance network models in epidemiology. *Epidemics* **3**, 38–45 (2011).

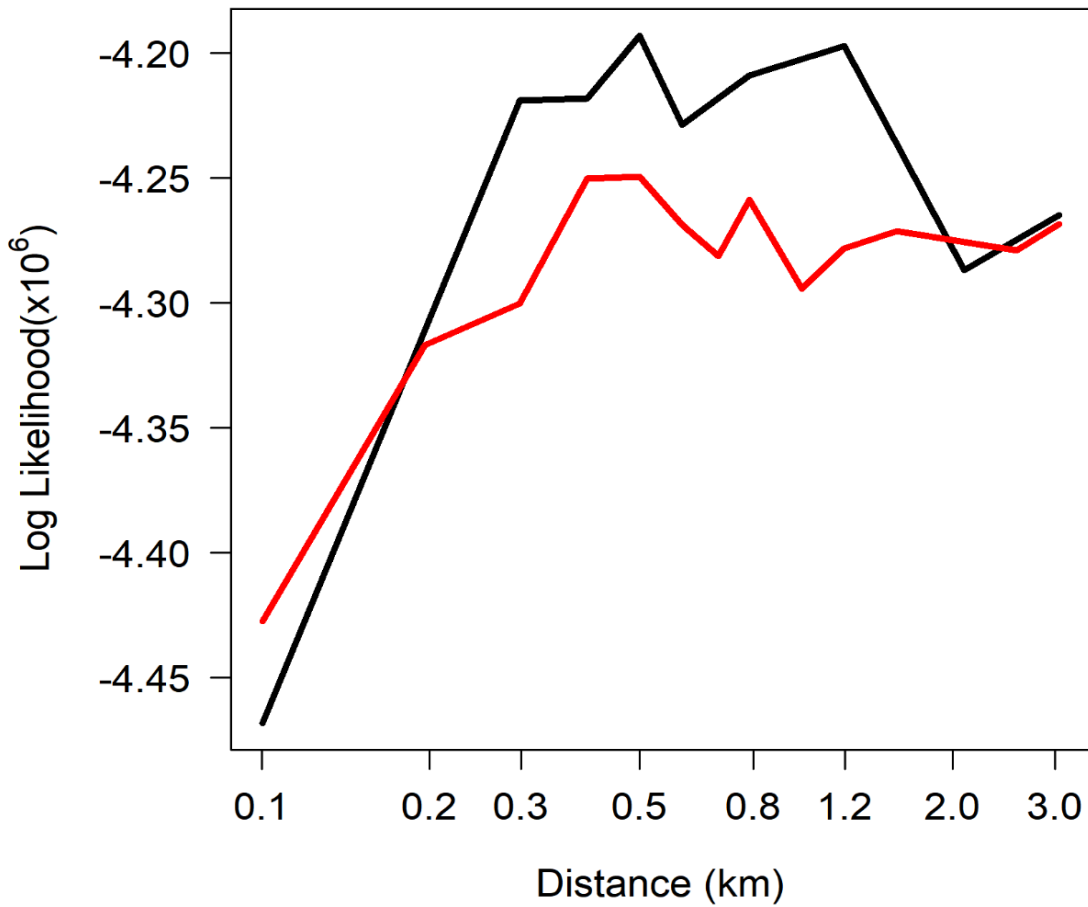
4.9. Supplementary information

Supplementary Figure 1: Estimated and simulated total number of malaria infections in the study area over time



Black solid line: the estimated total number of infections in Kilifi over time. The estimated total number of infections is a product of the estimated prevalence in Snow *et al*, population and house density in the Kilifi HDSS, and the multiplicity of infections estimated from Karl *et al*. Red line: the base model with σ , the parameter for distance set to 0.1km; Blue line: 0.3km; Green line: 0.4km; Purple line: 0.5km; Cyan line: 0.8km; Brown line: 1.2km; Yellow line: 2.0km Tomato line: 3.0km. (The mean distance is given by $\sigma\sqrt{2/\pi}$, giving 0.24, 0.32, 0.40, 0.64, 0.96, 1.60, 2.39km).

Supplementary Figure 2: Log likelihood by distance for different assumptions about recombination in multiply infected individuals



The x-axis shows the value of σ , the parameter for distance between parent and offspring infections (where the mean distance is given by $\sigma\sqrt{2/\pi}$). The log-likelihood is a measure of support from the data for the parameter values. Red solid line: the base model with probability of recombination in multiply infected individuals set to 0.5; Black line: the base model with probability of recombination in multiply infected individuals set to 0.5 and allowing four sibling genotypes per recombinant infection.

Chapter 5

5. Detecting malaria hotspots: the role of movement, seasonality, shape and statistical method

Josephine Malinga,^{1,2} Yeromin Mlacha,^{1,2,3} Penelope Vounatsou,^{1,2} Amanda Ross^{1,2}

¹ Swiss Tropical and Public Health Institute, Basel, Switzerland

² University of Basel, Basel, Switzerland

³ Ifakara Health Institute, Tanzania

5.1. Summary

As malaria transmission declines in many settings, there has been increasing interest in variation in the risk of infection at local scales such as within and between villages. Methods to detect hotspots frequently rely on statistical significance and do not consider underlying processes which may affect both the ability to detect and the choice of intervention strategy, such as the geographical distance between parent and offspring infections, season, the relationship between the malariological outcome and underlying transmission intensity, or the shape and gradient of the hotspot. We assess the influence of these features on the performance of different methods to identify pockets of higher transmission intensity.

We used an individual-based simulation model of malaria transmission dynamics. We simulated an area of 100km² with 2000 households and set pockets of higher transmission with different characteristics and transmission dynamics. We took cross-sectional surveys of the simulated area for a commonly used outcome for quantifying heterogeneity, the prevalence of malaria infection. We applied the detection methods, both Kulldorf's circular and elliptic, and Tango's flexibly shaped spatial scan statistic and a Bayesian geospatial model of the prevalence and calculated the sensitivity and specificity for households lying inside the simulated pockets of higher transmission.

The size of the detected hotspot was larger when there was a greater distance between parent and offspring infections, and in the high transmission season. Overall, the sensitivity was lower when there was a gentle decay in risk from the hotspot boundary, the hotspot was irregularly shaped or with a greater mean distance between parent and offspring infections. The specificity was reasonably high for all characteristics and methods. The Kulldorff's scan statistic with elliptical or circular windows and the Bayesian model both had greater sensitivity than Tango's flexible scan. Only the Bayesian model gave estimates of how much higher the prevalence was in the pocket of higher transmission.

The method of detection, characteristics of the areas of higher transmission and the underlying transmission dynamics affected the detection of areas of higher transmission intensity. These factors should be considered when interpreting analyses of heterogeneity.

5.2. Introduction

Malaria transmission is unevenly distributed with variations in the incidence of infection and clinical illness on all spatial scales. Major factors driving the heterogeneity are the abundance and behaviour of vector populations¹, environmental characteristics²⁻⁴, human behavioural, economic and genetic factors⁵, and the coverage of interventions⁶. Interest in heterogeneity in transmission and infection on local scales, such as between villages and between households within villages, has grown⁷⁻¹¹. There have been reductions in malaria transmission over the last decade¹² leading to increasing interest in targeting interventions to pockets of higher transmission in the search for cost-effective strategies^{3,6,13}. Targeted indoor residual spraying (IRS)¹⁴, focal mass drug administration (fMDA), test and treat (fTAT), and packages of interventions¹⁵ have been trialled. There has also been interest in the impact of the pockets of higher transmission intensity on the surrounding areas¹⁵⁻¹⁷.

This raises the question of how best to identify these pockets of higher transmission. Although many studies have described heterogeneity, there is uncertainty in how it should be defined and measured at fine scales^{6,15,18,19}. Measures of global clustering such as Moran's I and Tango's excess events test (EET) provide information on whether there is significant spatial autocorrelation or clustering in a given study area, but do not identify where the clusters are^{20,21}. The targeting ratio, which is a measure of the number of individuals that would need to be sampled randomly to the number who actually receive an intervention to cover the same proportion of the infections²² has been proposed for parameterizing mathematical models of the impact of reactive case detection (RCD), but similarly does not identify where the clusters are. Measures of local clustering such as Kulldorf's spatial scan statistic (SaTScan) and Tango's flexibly shaped spatial scan statistic (Tango FSS) aim to uncover the size and location of any possible cluster²³⁻²⁵, often referred to as hotspots. Geostatistical models have also been developed to provide estimates at each location, accounting for spatial autocorrelation.

These methods have shortcomings. The detection of hotspots uses the difference in risk or prevalence within and outside the hotspot and often relies on hypothesis tests^{20,23}. These produce only an indication that there is an area of higher transmission, yet decisions on implementation usually require estimates of how much higher the transmission is. Statistical significance depends in part on the amount of data available¹⁹, and may be affected by how steep the gradient in transmission around the edges of the hotspot is or heterogeneity within the hotspot^{9,26}. The shape used by the detection method may also affect the results²⁴⁻²⁹. Several hotspot detection methods apply a circular scan window, while areas of higher transmission may be irregularly shaped, for example running alongside a river.

The epidemiology of malaria and underlying dynamic processes are frequently not well considered. Analyses tend to use cross-sectional surveys as snapshots in time and do not consider that the pattern may alter by the time that interventions can practically be implemented. For instance, seasonality may lead to the contraction or increase in the size of the hotspot. They also do not consider the mean geographical distance between parent and offspring infections, carried by the vector or human, which affects how far and fast infections spread. This distance may play a role in the apparent size, gradient at the edge, and persistence over time of the hotspot and whether the prevalence in the surrounding areas is increased. Knowledge of the underlying processes may influence the choice of intervention strategy.

The choice of malariological outcome affects the hotspots detected³⁰⁻³². If the aim is to interrupt transmission, the most useful outcome for targeted interventions would be the underlying transmission intensity. Analyses of heterogeneity commonly use more easily measurable outcomes such as prevalence or clinical episodes. However, where the infections are may not necessarily be where transmission occurs. The relationship between the spatial distributions of transmission intensity and prevalence may vary seasonally, and with vector density and dispersal and human movement^{3,9,18}. The limit of detection of the diagnostic tests also affects the detection of hotspots³⁰.

In this study, we use an individual-based spatial stochastic simulation model of malaria infection dynamics to assess the extent to which different statistical methods can accurately detect pockets of higher transmission and also investigate the influence of movement, shape, gradient and seasonality. We aim to inform the interpretation of analyses for the detection of hotspots.

5.3. Methods

Strategy

We use a previously developed individual-based spatial simulation model of malaria transmission dynamics³³. We simulate hotspots with different characteristics by altering the spatial patterns of the underlying transmission intensity. We systematically vary the shape, gradient and number of hotspots, the seasonality, and the mean geographic distance between parent and offspring infections. We then take cross-sectional surveys of the simulated individuals and whether they have infections or not. We apply the different statistical methods to these simulated surveys in order to evaluate their ability to identify the households known to lie in the areas of higher transmission intensity.

Simulation model

A simple spatial individual-based simulation model of *Plasmodium falciparum* malaria transmission dynamics has previously been developed ³³. Infections, individuals and households are simulated. The original model included genotypes of infections and was fitted to data from Kilifi County, Kenya to estimate the mean geographic distance between parent and offspring infections. In the current version, genotype information is unnecessary and has been removed. The simulation is updated at every five-day time-step. Each current infection may lead to a number of new infections at each time-step, with the number drawn from a Poisson distribution. The location of a new infection is selected with probability determined by the distance of the household from the parent infection. Infections clear with a constant probability, equivalent to a mean duration of 200 days ^{34,35}.

Scenarios

We simulated a 10km by 10km square with 2000 households each with 8 people following the population density of Kilifi ³³. The households were randomly distributed across the simulated square. The total population in each square is 16,000 people. The number of initial infections was set to correspond to a prevalence of 5%, 10%, or 20%, and the initial infections are distributed at random. We set areas of higher transmission to contain approximately 200 households (Table 5.1).

The mean number of new infections per current infection per time-step can be modified to generate the desired number of simulated infections to match a given overall prevalence. The prevalence was assumed to remain constant over the simulated study period, with the exception of scenarios with seasonal patterns.

Table 5.1 Simulated scenarios

<i>Model Inputs</i>	
Area simulated	10km by 10 km
Number of households	2000
Number of people per household	8
Initial prevalence	5%, 10% , 20%
Number of seeds per scenario	10
Run-in period	500 (5-day time steps)
Geographic distance between parent and offspring infections	0.1km, 0.3km , 2.0km
Seasonal pattern	constant, Kilifi (35), Garki district Nigeria (36)

Spatial patterns of transmission intensity

Number of households in areas of higher transmission per scenario	200
Increase in transmission intensity in areas of higher transmission	50% increase
Number of clusters	single , multiple
Shape of clusters	circular , irregular
Gradient at edge of cluster	step function , gentle decay

*reference scenario highlighted in bold

We assessed how well the statistical methods detected areas of higher transmission in areas with different features: 1) single and multiple circular areas of higher transmission, 2) areas of higher transmission with steep and gentle gradients at the edge, and 3) areas of transmission with varying shapes (circular, irregular). We created variations in transmission intensity in the simulated area by modifying the mean number of new infections per current infection for households within hotspots.

For single areas of higher transmission intensity, we randomly selected a single household and then took the closest 199 households based on their co-ordinates. We replicate the process for the other features: for multiple pockets of higher transmission intensity, we randomly select more than one cluster of households in the study area, for irregular ones, we superimpose a river-shaped feature for simplicity, and pick the 200 households within and around it, and for a gentle gradient, we first create a single core area with 100 households and then create an exponential decay surface of risk as one moves away from the boundary of the core for the remaining 100 households (Figure 5.1).

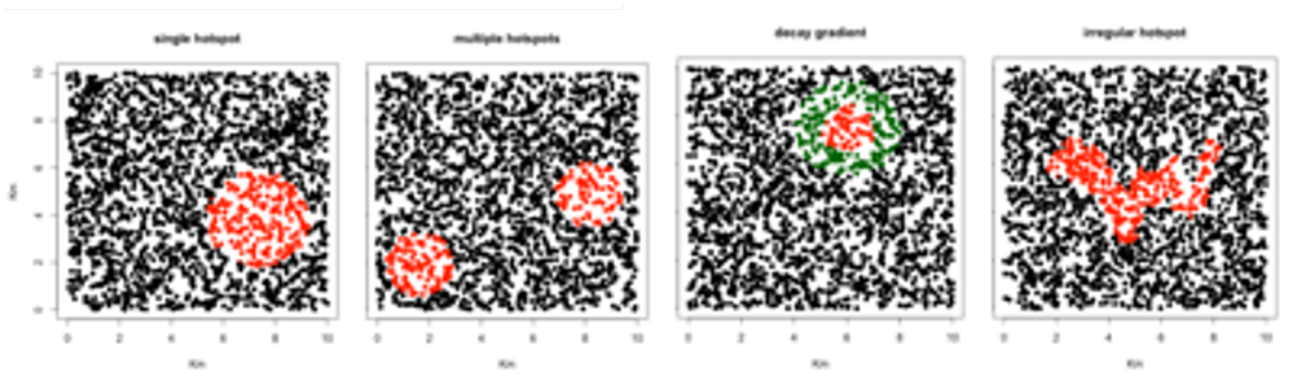


Figure 5.1 Examples of hotspots with varied features

a. Single hotspot, b. multiple hotspots, c. single hotspot with a decay gradient at the edges, d. irregularly shaped hotspot. Black dots: households outside the area of higher transmission. Red dots: household within the area of higher transmission. Green dots: households within the gradient zone. The prevalence of malaria infections is 10% s, and the transmission level

(defined by the mean number of new infections per infection per timestep) is 50% higher within the hotspot or core of the hotspot.

Methods to be assessed

We focus on measures of local clustering. There are a number of methods reported in literature that have been applied in the detection of malaria hotspots. For this study, we select commonly used measures: the Kulldorf's spatial scan statistics (SaTscan) circular and elliptic versions, the flexible shaped scan statistic (Tango FSS) and a simple Bayesian model (Table 5.2).

Table 5.2 Methods included in the evaluation

Method	Description	Reference	property
SatScanO	Kulldorf's spatial scan statistic ¹	²³	circular window
SatScanE	Kulldorf's spatial scan statistic ¹	²⁴	circular and elliptic windows
Tango FSS	Flexible spatial scan statistic ²	²⁵	irregular window
Geospatial model	Simple Bayesian model ³	³⁷	estimates prevalence at each household

¹ The Kulldorf's spatial scan statistic is estimated using the SaTscan software for both the circular and elliptic versions ^{23,24}. The method takes each household as a centroid (central point within a cluster) and then constructs multiple circles or both circles and ellipses of varying sizes (with radius and angles specified by the user) around it, as it moves across the geographical area. The observed number of cases for each location and size of the window are counted and compared with the average across the population using a log likelihood statistic. A cluster is identified when the number of cases within the window (circle or ellipse) is significantly higher than the average across the study area.

² The Tango flexibly spatial scan statistic is estimated using the FleXScan software developed by Tango and Tanashi to detect irregularly shaped clusters ²⁵. The statistic uses an irregularly shaped scan window which is created by connecting a geographical unit (household) with its adjacent neighbours and calculating the prevalence within the scan window. The number of nearest neighbours is specified by the user. A cluster is identified when the number of cases within the window (irregular) is significantly higher than the average across the study area.

³ A simple Bayesian geo-statistical model provides estimates of the prevalence for each household ³⁷. The model has a logistic link function and includes a random effect for location and a matern covariance function, where the covariance only depends on the distance between any two points.

The cluster size used by the methods for detecting hotspots was arbitrarily set to 200 as a size potentially appropriate for a targeted intervention. In this case, this is approximately the same size as the underlying pockets of higher transmission themselves. If a specific targeted intervention is to be triggered, then the size should depend on the smallest area that would be practical and cost-effective to implement. This may depend on setting characteristics, such as the range of vector dispersal and the density of households. We avoid administrative boundaries. These are arbitrary in the context of malaria transmission and frequently, areas of varying transmission intensities are found within administration units and areas with similar transmission intensities are divided by boundaries^{16,38}.

Evaluating the methods performance in detecting areas of higher transmission

We took cross-sectional surveys of the prevalence of infected simulated individuals at three months, one year and three years. We chose three time-points since the patterns of heterogeneity are not stable over time when there is constant seasonality. Only the results at one year are presented in this paper. The higher transmission areas tend to have an increasing prevalence over time and the lower transmission areas see the prevalence reduce. This effect is not necessarily observed when there is seasonality.

We calculated the sensitivity (the proportion of households that are in areas of higher transmission that are identified as being so) and the specificity (the proportion of households outside areas of higher transmission that are correctly identified), and the agreement between the estimated and known transmission intensities per household.

The spatial scan methods allow the identification of households inside and outside the detected hotspots. The geospatial model estimates the prevalence, from which we can estimate areas of higher transmission. To compare the evaluation measures with the methods for identifying hotspots, we choose an arbitrary cut-off of a 50% increase in the predicted prevalence from the baseline value to define households inside hotspots (Table 1). For the hotspot with a gradient in risk at the edge, we arbitrarily set the underlying

5.4. Results

The simulated areas of higher transmission and the detected hotspots are shown for one seed and for one method as an example (Figure 5.2)

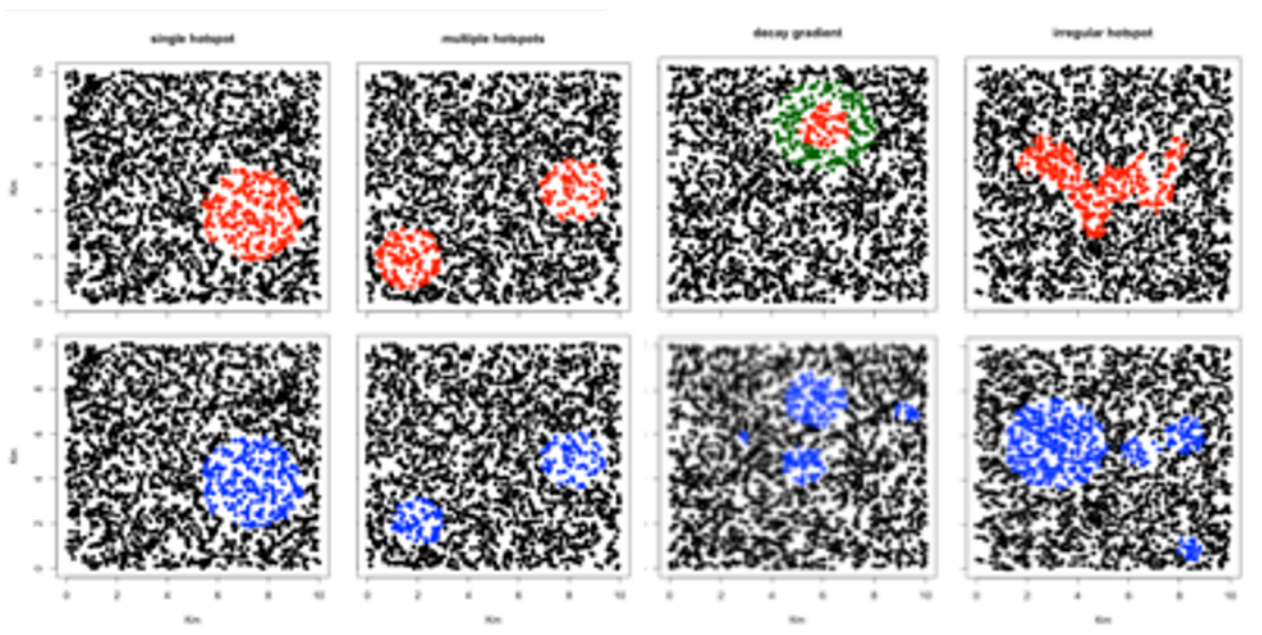


Figure 5.2 An example of simulated areas of higher underlying transmission and detected hotspots of prevalence

a. Single hotspot, b. multiple hotspots, c. single hotspot with a decay gradient at the edges, d. irregularly shaped hotspot, Red dots: True hotspots in underlying transmission intensity, Brown dots: households within the gradient zone (top row), Blue dots: Detected hotspots in prevalence (bottom row). The simulated data was taken as a cross-sectional survey after a period of 1 year with constant transmission (no seasonality) and an initial prevalence of 10%. Hotspots were detected using the SaTScan statistic with a circular scan window and run for one seed.

In general, the shape and number of the underlying areas of higher transmission affected the accuracy of the detected hotspots. A gradient at the edge of the area of higher transmission, a longer distance between parent and offspring infections, presence of multiple clusters or irregular shaped clusters tended to decrease the accuracy. Some small hotspots were detected outside the area of higher transmission intensity due to stochasticity in the predicted prevalence.

When the mean distance between parent and offspring infections was longer, the size of the detected hotspot for prevalence tended to be larger than the underlying area of higher transmission intensity (Figure 5.3). Over time with constant transmission, the prevalence hotspot completely disappeared when the mean distances were longer.

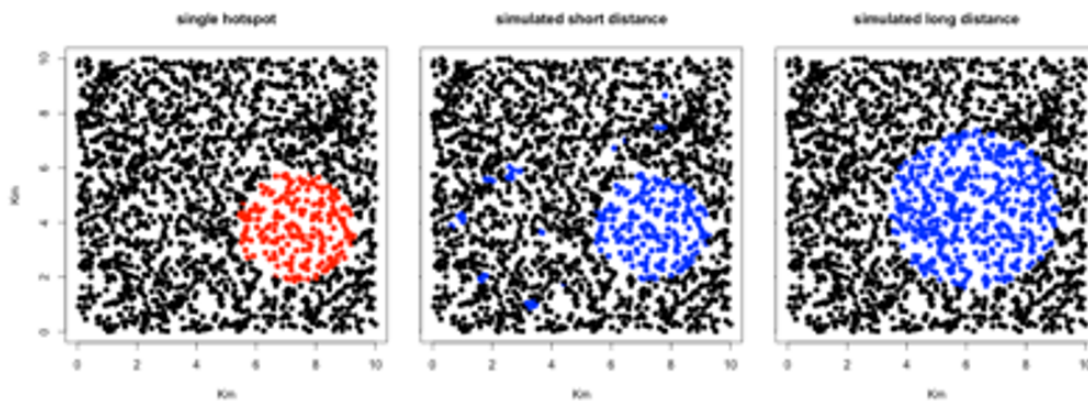


Figure 5.3 Simulated of detected hotspots with varying distance between parent and offspring infection after a period of 1 year

a. Single hotspot, b. detected hotspot using simulated data assuming the mean distance between parent and offspring infections is 100m, c. assuming the mean distance between parent and offspring infections is 2.0km. Red dots: households within the simulated area of higher transmission. Blue dots: Households within the hotspots detected using SaTScan with a circular window.

The size of the detected hotspot for prevalence also varied by season, tending to be larger during the wet compared to the dry season (Figure 5.4).

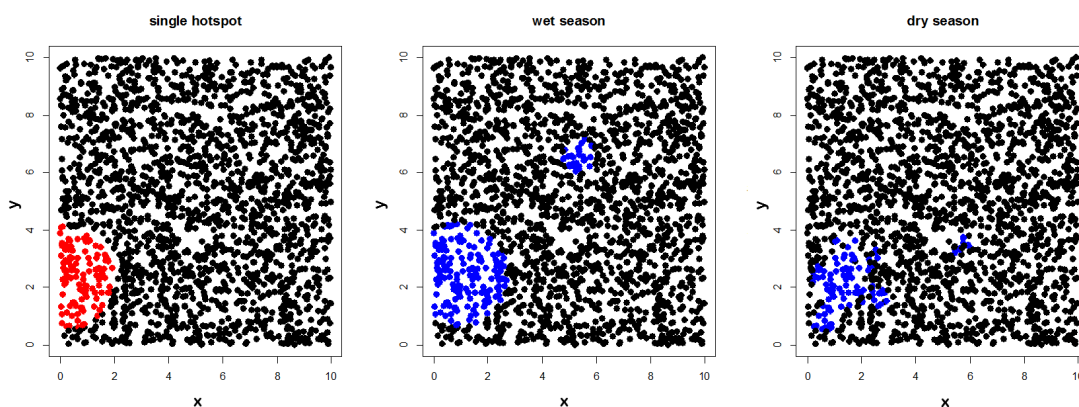


Figure 5.4 Example of simulated hotspots with by season

a. Single known hotspot, b. dry season, c. wet season. The seasonal profile was obtained from Kilifi, Kenya (40). Red dots: household within the area of higher transmission. Blue dots: Detected hotspot using SaTScan with a circular window after 1 year.

The results for all seeds and all methods are summarized in Table 3. The results for sensitivity varied by method and by the characteristics of the area of higher transmission (Table 3).

Characteristics which lead to a lower sensitivity in general were those that were influenced by a change in risk from the center of the hotspot for instance the decay gradient scenario or the mean distance of movement. All methods showed a poor sensitivity for a long mean distance between parent and offspring infections.

Both the SaTScan circular and elliptic methods had a high sensitivity for detecting households within areas of higher transmission except when there was a decay gradient at the edge of the hotspot or long mean distance between parent and offspring infection after the first year.

The Tango FSS method had low sensitivity. For the Bayesian model, where we had an arbitrary cut-off for the predicted prevalence per household, had a high sensitivity particularly for irregularly shaped hotspots, except when there was a decay gradient at the edge of the hotspot or long mean distance between parent and offspring infection.

All of the methods had a reasonably high specificity for all the different underlying features.

In practice, areas of higher transmission are likely to be irregular, have multiple hotspots, may have a gradient at the boundary and decision-makers are likely to require estimates of how much higher the prevalence is inside the hotspot. A method which performs reasonably well on all of these criteria taken together is the geospatial model.

Table 5.3 Assessment of measures to detect hotspots

	Single Circular hotspot %(95%CI)	Multiple hotspots %(95%CI)	Hotspot with decay %(95%CI)	Irregular-shaped hotspot %(95%CI)	Short mean distance %(95%CI)	Long mean distance %(95%CI)	With Kilifi seasonality %(95%CI)
Sensitivity							
SaTScan O ¹	85(72,98)	88(75, 100)	34 (28 , 39)	78 (66 , 89)	55 (26 , 83)	47 (11 , 83)	89 (76 , 100)
SaTScan E ²	85 (72 , 98)	81 (69 , 93)	34 (28 , 39)	81 (69 , 93)	53 (23 , 84)	46 (1 , 91)	89 (76 , 100)
Tango FSS	54 (45 , 63)	57 (48 , 66)	22 (12 , 32)	72 (55 , 90)	42 (33 , 51)	6 (3 , 8)	65 (44 , 86)
Specificity							
SaTScan O	90 (76 , 100)	92 (79 , 100)	92 (79 , 100)	85 (72 , 97)	87 (74 , 100)	76 (60 , 91)	90 (77 , 100)
SaTScan E	90 (77 , 100)	92 (79 , 100)	92 (79 , 100)	87 (75 , 100)	91 (78 , 100)	82 (70 , 94)	87 (75 , 100)
Tango FSS	92 (79 , 100)	92 (80 , 100)	46 (1 , 91)	89 (76 , 100)	90 (77 , 100)	92 (80,100)	73 (47 , 99)
PPV							
SaTScan O	67 (34 , 100)	78 (66 , 91)	86 (72 , 100)	51 (39 , 63)	44 (8 , 80)	27 (4 , 50)	52 (44 , 59)
SaTScan E	68 (35 , 100)	74 (58 , 90)	87 (73 , 100)	59 (51 , 67)	56 (25 , 88)	29 (5 , 53)	39 (34 , 45)
Tango FSS	76 (53 , 100)	86 (74 , 99)	85 (70 , 100)	66 (54 , 79)	46 (24 , 68)	7 (2 , 12)	70 (59 , 81)

¹refers to the circular version of the Kulldorf spatial scan statistic

²refers to the elliptic version of the Kulldorf spatial scan statistic

5.5. Discussion

This study adds to the knowledge of how the features of heterogeneity in malaria transmission and the choice of method affects the accuracy of detection of pockets of higher transmission or prevalence.

We found that the characteristics of the areas of higher transmission affected the performance of the detection methods: identification was less precise for all methods when there was decay in risk from the edge of the underlying area of higher transmission, the area was irregularly shaped, there was seasonality or when the mean distance between parent and offspring infections was longer than 2km.

In practice, most areas will contain multiple, irregularly shaped hotspots with seasonality and movement and will require knowledge of the how much higher the transmission is within the hotspots, sizes and locations of hotspots, and the spread of infections. For such areas, Kulldorf's circular and elliptic methods had reasonable sensitivity but the geospatial model was the only method to estimate the prevalence.

Areas of higher transmission which tend to have long distances between parent and offspring infections, and no seasonality or interventions to mitigate the spread, appeared to be less pronounced over time in our simulations. Movement of infections can create a gradient of a change in risk as one moves away from the boundary. It is possible that an area of higher transmission with a gentle decay at the edge could appear similar to one with a steep gradient and longer distances of movement when looking at prevalence, so appearances alone may not reveal the underlying processes. There is a need for information on how far infections spread in an area, as well as information on the apparent heterogeneity. Future research may allow a separate measure for movement to be used in conjunction with the measure of heterogeneity to inform the likely impact of interventions.

Malaria transmission has been reported to persist within hotspots even during the dry seasons^{3,6,9}. The simulated hotspot sizes contracted and expanded with seasons. This may have implications for the area for targeted interventions at different times of the year. For instance, there may be a rationale to adapt the radius to be considered in the case of reactive case detection (RCD) according to season.

The prevalence of asymptomatic infections, clinical episodes and serological markers have been used as the outcomes of measure for detecting hotspots at different scales for different purposes^{16,30,39}. If the aim is to detect areas of higher transmission intensity, then the choice of outcome may affect the results. We found that the relationship between the detected

prevalence hotspot and the area of higher underlying transmission varied by shape and gradient of the hotspot, and the season.

5.6. Conclusion

The choice of outcome, method of detection and features affected the accuracy of hotspot detection. The identification of hotspots was less accurate when there was decay in gradient with risk from the hotspot boundary, the hotspot was irregularly shaped, there was seasonality or when the mean distance between parent and offspring infections was longer. The SatScan method with elliptical or circular windows performed better than Tango FSS in our tests. The choice of method and underlying transmission dynamics should be taken into account when performing and interpreting analyses of heterogeneity for targeted interventions.

5.7. References

1. Machault, V. *et al.* Highly focused anopheline breeding sites and malaria transmission in Dakar. *Malar. J.* 8, 138 (2009).
2. Greenwood, BM. The microepidemiology of malaria and its importance to malaria control. *Trans Roy Soc Trop Med Hyg* 83 Suppl, 25–29 (1989).
3. Bousema, T *et al.* Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med* 9, e1001165 (2012).
4. Gebreyesus, TA, Haile, M, Witten, KH & *et al.* Incidence of malaria among children living in dams in northern Ethiopia: community based incidence survey. *BMJ* 319, 663–6 (1999).
5. Sachs, J. & Malaney, P. The economic and social burden of malaria. *Nature* 415, 680–685 (2002).
6. Carter, R., Mendis, K. N. & Roberts, D. Spatial targeting of interventions against malaria. *Bull. World Health Organ.* 78, 1401–1411 (2000).
7. Bejon, P. *et al.* Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. *PLoS Med.* 7, (2010).
8. Kreuels, B. *et al.* Spatial Variation of Malaria Incidence in Young Children from a Geographically Homogeneous Area with High Endemicity. *J. Infect. Dis.* 197, 85–93 (2008).
9. Bousema, T *et al.* Identification of hotspots of malaria transmission for targeted malaria control. *J Infect Dis* 201, 1764–1774 (2010).
10. Gaudart, J. *et al.* Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC Public Health* 6, 286 (2006).
11. Ghebreyesus, T. A. *et al.* Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey. *BMJ* 319, 663–666 (1999).
12. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 526, 207–211 (2015).
13. Sturrock, H. J. W. *et al.* Mapping Malaria Risk in Low Transmission Settings: Challenges and Opportunities. *Trends Parasitol.* 32, 635–645 (2016).
14. Pinchoff, J. *et al.* Targeting indoor residual spraying for malaria using epidemiological data: a case study of the Zambia experience. *Malar. J.* 15, 11 (2016).

15. Bousema, T. *et al.* The Impact of Hotspot-Targeted Interventions on Malaria Transmission in Rachuonyo South District in the Western Kenyan Highlands: A Cluster-Randomized Controlled Trial. *PLoS Med.* 13, (2016).
16. Stresman, G. H. *et al.* Do hotspots fuel malaria transmission: a village-scale spatio-temporal analysis of a 2-year cohort study in The Gambia. *BMC Med.* 16, 160 (2018).
17. Stresman, G., Bousema, T. & Cook, J. Malaria Hotspots: Is There Epidemiological Evidence for Fine-Scale Spatial Targeting of Interventions? *Trends Parasitol.* 35, 822–834 (2019).
18. Bousema, T. & Baidjoe, A. Heterogeneity in malaria transmission: underlying factors and implications for disease control. in *Ecology of parasite-vector interactions* (eds. Takken, W. & Koenraadt, C. J. M.) 197–220 (Wageningen Academic Publishers, 2013). doi:10.3920/978-90-8686-744-8_11.
19. Stresman, G. H. *et al.* Impact of metric and sample size on determining malaria hotspot boundaries. *Sci. Rep.* 7, 45849 (2017).
20. Tango, T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat. Med.* 14, 2323–2334 (1995).
21. Zhou, X. & Lin, H. Moran's I. in *Encyclopedia of GIS* 725–725 (Springer US, 2008). doi:10.1007/978-0-387-35973-1_817.
22. Chitnis, N. *et al.* The theory of reactive intervention in the elimination and control of malaria. *In prep.*
23. Kulldorff, M. A spatial scan statistic. *Commun. Stat. - Theory Methods* 26, 1481–1496 (1997).
24. Kulldorff, M., Huang, L., Pickle, L. & Duczmal, L. An elliptic spatial scan statistic. *Stat. Med.* 25, 3929–3943 (2006).
25. Tango, T. & Takahashi, K. A flexibly shaped spatial scan statistic for detecting clusters. *Int. J. Health Geogr.* 4, 11 (2005).
26. Stresman, G. H. *et al.* Impact of metric and sample size on determining malaria hotspot boundaries. *Sci. Rep.* 7, (2017).
27. Duczmal, L., Kulldorff, M. & Huang, L. Evaluation of Spatial Scan Statistics for Irregularly Shaped Clusters. *J. Comput. Graph. Stat.* 15, 428–442 (2006).
28. Kulldorff, M., Heffernan, R., Hartman, J., Assunção, R. & Mostashari, F. A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Med.* 2, (2005).

29. Huang, L., Pickle, L. W. & Das, B. Evaluating spatial methods for investigating global clustering and cluster detection of cancer cases. *Stat. Med.* 27, 5111–5142 (2008).
30. Mogeni, P. *et al.* Detecting Malaria Hotspots: A Comparison of Rapid Diagnostic Test, Microscopy, and Polymerase Chain Reaction. *J. Infect. Dis.* 216, 1091–1098 (2017).
31. Kangoye, D. T. *et al.* Malaria hotspots defined by clinical malaria, asymptomatic carriage, PCR and vector numbers in a low transmission area on the Kenyan Coast. *Malar. J.* 15, (2016).
32. Bejon, P. *et al.* Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. *PLoS Med.* 7, (2010).
33. Malinga, J. *et al.* Investigating the drivers of the spatio-temporal patterns of genetic differences between Plasmodium falciparum malaria infections in Kilifi County, Kenya. *Sci. Rep.* 9, 19018 (2019).
34. Bretscher, M. T. *et al.* The distribution of Plasmodium falciparum infection durations. *Epidemics* 3, 109–118 (2011).
35. Felger, I. *et al.* The Dynamics of Natural Plasmodium falciparum Infections. *PLOS ONE* 7, e45542 (2012).
36. Mbogo, C. N., Baya, N. M., Ofulla, A. V., Githure, J. I. & Snow, R. W. The impact of permethrin-impregnated bednets on malaria vectors of the Kenyan coast. *Med. Vet. Entomol.* 10, 251–259 (1996).
37. Giorgi, E. & Diggle, P. J. PreVMap: An R Package for Prevalence Mapping. *J. Stat. Softw.* 78, 1–29 (2017).
38. Srivastava, A., Nagpal, B., Joshi, P., Paliwal, J. & Dash, A. Identification of malaria hot spots for focused intervention in tribal state of India: a GIS based approach. *Int. J. Health Geogr.* 8, 30 (2009).
39. Bejon, P. *et al.* A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. *eLife* 3, (2014).

Chapter 6

6. Discussion

Ultimately, the spatial and temporal patterns observed for different metrics at the community level are made up of tiny movements of individual mosquitoes, human hosts and the parasites inside them. Information at such fine scales, are relevant for the bigger picture, in particular, the design of intervention strategies aiming to reduce and interrupt malaria transmission, the design of studies to evaluate their effectiveness in the field, and the parameterization of mathematical models to predict their likely impact for settings where data is not available.

This thesis aimed at describing the development and use of both statistical and mathematical modeling approaches to characterize fine scale malaria transmission dynamics, and their consequences on the measurement of heterogeneity on a local scale. Taken as a whole, this thesis highlights the role that vectors, and humans play in the spread of infections and the implications of fine scale movement for the measurement of heterogeneity and targeting interventions.

A detailed discussion of the findings was given separately in each chapter. In this discussion, the main contributions of the thesis are summarized and set in the context of malaria epidemiology. The strengths and limitations of the approaches used and possible implications for malaria control and are also outlined.

6.1. Summary of Main Findings

In Chapter 2, *OpenMalaria* was used to predict the proportion of malaria infections that are in mosquitoes and how this varies by characteristics of the setting. The main findings show that a substantial proportion of infections are in mosquitoes, that this varies with transmission intensity, and highlights weak spots in individual malaria interventions and their timing.

Chapter 3 describes the development of a method to estimate mosquito movement as a secondary analysis of data from trials of spatial repellents. The method was evaluated by simulation and found to work well but only if there was sufficient data, and if data on the seasonal pattern of mosquito densities in the village in the absence of repellents could additionally be obtained. Findings from the simulations could inform the design of studies and help quantify criteria for trial settings.

The work in Chapter 4 aims to characterize and estimate the processes underlying the empiric patterns of genetic differences between *P.falciparum* parasites by time and geographic distance. Methods are limited in the case where transmission intensity is moderate, and the coverage of sampled infections is low. The chapter describes the development of an individual-based stochastic simulation model of households, people and malaria infections and the process of fitting it to genotyping data; with the aim of estimating the mean distance between parent and offspring infections and testing hypotheses about the underlying processes. The model could reproduce the observed patterns in the study area. The findings suggest that random drift (as opposed to selection pressure) is sufficient to explain the empiric patterns, but other hypotheses cannot be ruled out. The mean distance between parent and offspring infections was estimated to be 0.5km (95%CI 0.3 – 1.5). The findings glean some insights on how simulation can be used in quantifying factors driving transmission, and in estimating unknown parameters when analytic methods are limited.

Chapter 5 extends the simulation model developed in Chapter 4, to investigate and predict the consequences of fine-scale movement dynamics on the measurement of heterogeneity in prevalence on a local scale. The mean distance between parent and offspring infections was found to be an important determinant on the shape, size and stability of hotspots. This study also found that underlying features of fine scale malaria transmission (e.g. seasonality) could have a substantial influence on the ability of common methods to identify areas of higher transmission. This has important implications especially when decisions about where to implement interventions are to be made.

6.2. The use of mathematical and statistical models

Mathematical models provide a representation of the underlying mechanisms of a system. This can allow us to quantify unobservable processes. They do not necessarily involve data. Statistical models describe the distribution of data as a function of unknown parameters, without necessarily understanding the mechanisms that lead to the observed data. This thesis contains models which have both mathematical and statistical aspects, both representing the underlying biological mechanisms and being fitted to data.

Mathematical models have long been used to provide predictions on aspects of malaria transmission that cannot be accurately measured [1,2], allowing the quantification of different dynamics, including predicting the effect of various interventions [3–6]. In Chapter 2, the proportion of malaria infections that are in mosquitoes and how this varies by characteristics of the setting were predicted using *OpenMalaria*. The findings highlight weak spots in individual interventions and their timing. Current tools against malaria are imperfect and so interventions are frequently combined to gain more leverage. Quantifying the shifting proportions of infections in mosquitoes before and after the introduction of different interventions would inform the design of optimal combination strategies for malaria control and elimination.

In Chapter 3, a model which incorporates data on household locations within villages and numbers of mosquitoes collected over time was used to estimate mosquito movement. The model included both mathematical and statistical elements, having both structure for mosquito movement and being fitted to data. In Chapter 4, an individual-based simulation model is developed to estimate parameters of infection movement, and again was fitted to data on using pairs of parasite genotypes.

In Chapter 5, the model developed in Chapter 4 is used to characterize fine scale malaria transmission dynamics in order to evaluate statistical methods for measuring heterogeneity on a local scale. The underlying processes leading to the relative difference in transmission (for instance within and outside pockets of transmission) is hard to quantify in real field settings. Simulation models with biological realism provide an avenue through which existing statistical methods can be evaluated. In Chapter 5, the simulation and statistical models are used separately but in the same study.

Table 6.1: Table of Methods

Chapter	Objective	Model description
Chapter 2	Prediction of the distribution of infections between humans and mosquitoes	Comprehensive simulation model of malaria epidemiology, fitted to data
Chapter 3	Estimate vector movement	Model with mathematical structure, fitted to data
Chapter 4	Estimate infection movement	Simulation model with processes represented, fitted to data
Chapter 5	Evaluation of methods for detecting hotspots	Statistical methods for detecting hotspots, evaluated using a simulation model of malaria transmission

6.3. The importance of setting characteristics

The findings in this thesis highlight the importance of having sufficient knowledge on setting-specific characteristics before decisions on interventions or the design of studies are made. In each chapter, changes in the characteristics of model inputs or assumptions lead to different results; mosquito species (Chapters 2 and 3), human blood index (Chapter 2), transmission intensity (Chapter 2 and 4), season (Chapters 2, 3 and 5), house clustering (Chapter 3 and 4), movement of humans and mosquitoes (Chapter 5).

To be able to generalize findings across settings, there is need for additional datasets or purpose-designed studies to collect this data in the field, since the epidemiology of malaria varies with site specific characteristics.

6.4. Potential impact of recent advances in the detection of infections

Recent advances have seen decreases in the limits of detection in parasite densities [7]. This raises the question of whether this would affect the findings in this thesis. OpenMalaria was fitted to data which did not include sub-patent infections. The numbers of infections in both humans and mosquitoes may be underestimated, but it is not known if this would affect the predicted proportion of infections that are in mosquitoes. If infections have been missed in

the genotyping data in Chapter 4 then, since the analysis was designed to accommodate a low coverage of infections, as long as the assumption that detectability is not related to the genotype, the main findings would hold. In Chapter 5, the simulations assume perfect detection. This is unlikely, even in the case of highly sensitive tests, and so the tendency would be that the detection of hotspots would have slightly less power and the prevalence would be underestimated by geostatistical models.

6.5. Implications of fine scale movement

Based on the findings of this thesis, information on fine scale movement of mosquitoes and humans, and the parasites they carry is important, for malaria epidemiology, to inform the design and analysis of studies, and in evaluating the effect of interventions.

Understanding how malaria infections spread locally and the processes leading to the observed spatial and temporal distribution patterns is important for the design of interventions aiming to reduce and interrupt transmission by targeting foci where there is fine scale heterogeneity. In Chapter 5, the distance between parent and offspring malaria infections was shown to have substantial effect on the abilities model to detect areas of higher transmission. The findings show that areas of transmission are harder to detect when the mean distance between parent and offspring infections was long as opposed to the small, isolated clusters observed when this distance was shorter.

Vector dispersal has been shown to negate the effects of interventions especially if mosquitoes are moving between households in different study arms or to confer a community wide effect when interventions are targeted to those at higher risk. Disentangling these effects in exploratory analysis can be challenging and more tools are needed. It is vital to understand the role in mosquito or infection movement in the spread of infection to estimate an efficient spatial scope to be covered with interventions.

Fine scale variations in malaria transmission are driven predominantly by mosquitoes. As transmission declines, heterogeneous patterns become more evident and the range across which mosquitoes fly can be used to estimate the relatedness of infections. This could be valuable in estimating the spread of drug resistance.

Information of fine scale movement can also be used in the parameterization of mathematical models and predict their likely impact.

Some issues

Data Availability:

Ideally, instead of describing what data can or cannot be used for, models should be extensible to answer different research questions since type and quality of data varies substantially from one site to another.

Some of the methods in this thesis were developed in response to lack of sufficient data to provide plausible estimates. In Chapter 3 and 4, we use simulation to validate the model predictions but fit the model to the available data. While these data are imperfect, the findings highlight the potential value of using all the information available (including previous trials) to estimate parameters that cannot be observed in the field. For instance, the movement of individual mosquitoes is hard to estimate - and the statistical model in Chapter 3 can be used to estimate dispersal parameters. The mean distance between parent and offspring infections is also challenging to estimate, but a simulation model that quantifies the underlying processes can be applied.

The methods developed can be used to inform purpose-designed studies, in terms of how much data one would need to collect. Findings from Chapter 3 underline the type of data that would be needed to provide accurate estimates of mosquito movement in a village. Simulation models could be used to determine what sampling strategy is needed for a study on parasite movement using genotyping.

Validation:

The model used in Chapter 2 has been extensively fitted to field data and this has previously been described [8]. We used the available field data to validate some of the model predictions (Chapter 2 and 3). Using trial data to validate model predictions can be challenging, as some of the trial parameters might be unknown. In Chapter 3 to 5, validations of the methods were based on reproducing known initial values, and this lends confidence to the results. There is no consensus on how to evaluate uncertainty and goodness of fit for models which do not conform to standard statistical distributions, and it still remains unclear on the best ways to quantify stochasticity for model predictions [9].

6.6. Outlook and future work

The methods developed, and the findings from this thesis have the potential for application to inform the design of interventions.

Future work may tackle the need for many estimates from many sites to allow a body of evidence tailored to setting characteristics to grow. There is limited information on vector movement, and few studies estimating the geographical distance between parent and offspring infections. Since setting characteristics are important, it is difficult to generalize from a few settings. However, if methods can be developed to allow secondary analysis of existing data sources as well as purpose-designed studies carried out, then there could be a body of information on how the distances relate to setting characteristics and how vector distances relate to infection distances. This information would be valuable for allowing generalizations, stratifying interventions and intervention designs by setting characteristics.

Following the changing malaria epidemiology in many endemic settings, future studies should take into account the underlying site-specific transmission dynamics when performing and interpreting analyses of heterogeneity for targeted interventions. In studies aiming to identify areas of higher risk or burden, there should be greater awareness that the characteristics of the area or the methods themselves might influence the accuracy of detection.

6.8. References

1. Chitnis N, Smith T, Steketee R. A mathematical model for the dynamics of malaria in mosquitoes feeding on a heterogeneous host population. *J Biol Dyn.* 2008 Jul 1;2(3):259–85.
2. Lutambi AM, Penny MA, Smith T, Chitnis N. Mathematical modelling of mosquito dispersal in a heterogeneous environment. *Math Biosci.* 2013 Feb;241(2):198–216.
3. Mandal S, Sarkar RR, Sinha S. Mathematical models of malaria - a review. *Malar J.* 2011 Jul 21;10:202.
4. Bretscher MT, Griffin JT, Ghani AC, Okell LC. Modelling the benefits of long-acting or transmission-blocking drugs for reducing *Plasmodium falciparum* transmission by case management or by mass treatment. *Malar J.* 2017 Aug 16;16:341.
5. Okell LC, Drakeley CJ, Bousema T, Whitty CJM, Ghani AC. Modelling the Impact of Artemisinin Combination Therapy and Long-Acting Treatments on Malaria Transmission Intensity. *PLOS Med.* 2008 Nov 25;5(11):e226.
6. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, et al. Reducing *Plasmodium falciparum* Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies. *PLOS Med.* 2010 Aug 10;7(8):e1000324.
7. The malERA Refresh Consultative Panel on Tools for Malaria Elimination. malERA: An updated research agenda for diagnostics, drugs, vaccines, and vector control in malaria elimination and eradication. *PLOS Med.* 2017 Nov 30;14(11):e1002455.
8. Smith, T, Maire, N, Ross, A, Penny, M, Chitnis, N, Schapira, A, et al. Towards a comprehensive simulation model of malaria epidemiology and control. *Parasitology.* 2008;135(13):1507–16.
9. Knutti R. The end of model democracy? *Clim Change.* 2010 Oct 1;102(3):395–404.

JOSEPHINE KALUNDE MALINGA

Bio Data:

Gender: Female
Date of Birth: 23rd August 1989
Nationality: Kenyan

Education:

2016 to 2019: **PhD Candidate (Epidemiology and Biostatistics)**
Swiss Tropical and Public Health Institute and University of Basel,
Switzerland
Supervisor: PD Dr. Amanda Ross, Prof. Juerg Utzinger

2013 to 2015: **M.Sc. Biometry (Biostatistics)**
University of Nairobi, Kenya
Supervisor: Dr Nelson Owuor

2008 to 2012: **B.Sc. Applied Statistics with Information Technology**
Maseno University, Kisumu, Kenya

Short Courses:

2016/2017: Introduction to High Performance Computing (University of Basel)
Writing to be Published (University of Basel)

May 2014: Good Clinical Practice (GCP) (KEMRI/ Walter Reed Project)

Work Experience:

PhD Candidate: Swiss Tropical and Public Health Institute

2016 to 2019:

My PhD project involves the integration of statistics, mathematical modeling, development and use of simulation models, bioinformatics, and geospatial systems, to investigate the heterogeneous nature of malaria transmission, vector behavior and dispersal, and the distribution of parasite populations over time and space.

- Review of different epidemiological models for application to current research
- Evaluation of malaria interventions by modelling of *Plasmodium* dynamics in vectors & humans
- Using genomics data to model *Plasmodium falciparum* parasite mixing over time and space
- Application of geospatial and geostatistical approaches to evaluate methods for measuring heterogeneity in infectious diseases over time and space
- Descriptive and Exploratory data analysis using R-Gui, C++, STATA
- Drafting of manuscripts for publication in peer reviewed journals
- Assist in tutoring during lectures for junior biostatistics student

Research - Statistics and Epidemiology: KEMRI/Wellcome Trust Research Programme, Kenya

Jan 2013 to Feb 2016:

- Developing research questions, formulating and conducting literature reviews
- Assist in study design, sample size calculation, randomization process and statistical analysis plans
- Assist in proposal development, protocol writing and ethics application processes
- Sampling and mapping of survey health facilities using statistical software and spatial systems
- Development of data collection and visualization tools and platforms (questionnaires, applications)
- Monitoring of daily field data collection procedures and management of data storage databases
- Collection, collating and cleaning of data and associated meta-data.
- DHIS2, IDSR, MIS data extraction, analysis and reporting
- Quantitative and qualitative statistical data analysis and interpretation using STATA/R/SPSS
- Documentation and archiving of data and interpretation of results
- Training and Supervision of field staff and data entry clerks
- Report writing and drafting of manuscripts for publication in peer reviewed journals

Research Consulting/Volunteer: 2013 to date

Sample size calculation, randomization, exploratory and descriptive data analysis using STATA/R

- HIV CASCADE analysis and intervention studies
- Maternal and Child Health studies
- Water, Sanitation and Hygiene (WASH) studies
- Early Childhood Development and nutrition studies

Research Output List

Malinga J, Lines J, Ross A. (2020). What proportion of Plasmodium falciparum and Plasmodium vivax malaria infections are in mosquitoes? In press.

Malinga, J., Maia, M., Moore, S. et al. Can trials of spatial repellents be used to estimate mosquito movement? Parasites Vectors 12, 421 (2019).

Malinga, J., Mogeni, P., Omedo, I. et al. Investigating the drivers of the spatio-temporal patterns of genetic differences between Plasmodium falciparum malaria infections in Kilifi County, Kenya. Sci Rep 9, 19018 (2019).

Githinji S, Oyando R, **Malinga J**, Waqo E, Soti D, Rono J, Snow RW, Buff AM, Noor AM (2017). *Completeness of malaria indicator data reporting via the District Health Information Software 2 in Kenya, 2011 - 2015*. Malaria Journal 16: 344.

Talisuna AO, Oburu A, Githinji S, **Malinga J**, Amboko B, Bejon, P, Jones C, Snow RW, Zurovac D (2017). *Efficacy of text-message reminders on paediatric malaria treatment adherence and their post-treatment return to health facilities in Kenya: A randomized controlled trial*. Malaria Journal 16: 46.

Githinji S, Noor AM, **Malinga J**, Macharia PM, Kiptui R, Omar A, Njagi K, Waqo E, Snow RW (2016). *A national health facility survey of malaria infection among febrile patients in Kenya, 2014*. Malaria Journal 15:591.

Zurovac D, Guintran JO, Wesley D, Naket E, **Malinga J**, Taleo, G (2015). *Health systems readiness and management of febrile outpatients under low malaria transmission in Vanuatu*. Malaria Journal

14:489.

Githinji S, Jones C, **Malinga J**, Snow RW, Talisuna AO, Zurovac D (2015). *Development of a text messaging intervention to improve treatment adherence and post-treatment review of children with uncomplicated malaria in Western Kenya*. *Malaria Journal* 14:320.

Talisuna AO, Zurovac D, Githinji S, Oburu A, **Malinga J**, et al. (2015). *Efficacy of Mobile Phone Short Message Service (SMS) Reminders on Malaria Treatment Adherence and Day 3 Post-Treatment Reviews (SMS-RES-MAL) in Kenya: A Study Protocol*. *Journal of clinical Trials* 5:217.

Reports:

Githinji S, **Malinga J**, Waqo E, Soti D, Rono J, et al. *Exploring routinely reported malaria data in Kenya; a data quality assessment*. Nairobi: National Malaria Control Programme, Ministry of Health; **2016**.

Machini B, Nyandigisi A, Kigen S, Memusi D, Kimbui R, **Malinga J**, et al. *Monitoring outpatient malaria case management under the 2010 diagnostic and treatment policy in Kenya-Progress 2010–2015*. National Malaria Control Programme, Ministry of Health; **2015**.

Githinji S, **Malinga J**, Kiptui R, Ahmed O, Kiambu N, Ejersa W, Noor AM. *Prevalence of confirmed malaria infections among febrile patients in public health facilities in Kenya*. A report for the National Malaria Control Programme, Ministry of Health, Kenya. **May 2015**.

Emelda A Okiro, Abdisalan M Noor, **Josephine Malinga**, Bernard Mitto, Clara W Mundia, Don Mathanga, Themba Mzilahowa & Robert W Snow (2013). *An epidemiological profile of malaria and its control in Malawi*. A report prepared for the Ministry of Health Malawi, the Roll Back Malaria Partnership and the Department for International Development, UK. March, **2014**.

Institutional responsibilities:

- Teaching Support – MSc and PhD biostatistics students (University of Basel and Swiss TPH)
- PhD Student Representative (Swiss TPH)
- Staff Representative (KEMRI/WT)

Awards and Fellowships:

- antelope Program for female doctoral and Post-doctoral students for career development: University of Basel. *Amount: ~ CHF 8,000*
- 3 year PhD Scholarship from: Gottfried und Julia Bangerter-Rhyner Stiftung, Novartis Foundation for Medical Biological Research, Amt für Ausbildungsbeiträge, Kanton Basel-Stadt and Rudolf Geigy Stiftung. *Amount: ~ CHF 160,000*.
- 2 year Masters Scholarship from: Higher Education Loans Board of Kenya. *Amount: ~ USD 2000*

Skills:

- Language: Swahili (native), English (fluent), German (beginner)
- Programming Skills (R-Gui, STATA, C++)
- Proficient use of statistical analysis software (STATA, C++ and R-Gui)
- Proficient use of geospatial tools such as ArcGIS, QGIS and R