Quantifying the Impacts of Variation in Entomological and Epidemiological Determinants of Malaria Transmission

Charles Frederick Whittaker

Thesis Submitted for the
Degree of Doctor of Philosophy

Department of Infectious Disease Epidemiology
Imperial College London
May 2022

Abstract

Malaria epidemiology is characterised by extensive heterogeneity that manifests across a range of spatial and temporal scales. This heterogeneity is driven by a diversity of factors spanning the human host, the parasite, the mosquito vector and the environment. Together, variation in these factors lead to marked differences in the epidemiology of malaria across different settings; in where malaria is concentrated, how malaria is transmitted and who is most at-risk. These differences have material consequences for the impact of control interventions aimed at combatting the disease, underscoring the crucial need to better understand and quantify the factors underlying heterogeneity in malaria epidemiology and transmission dynamics. In this thesis, I use a combination of statistical and mathematical modelling to further our understanding of how variation in the epidemiological and entomological determinants of malaria transmission drives heterogeneity in dynamics across settings and explore the implications of this variation for control efforts.

Accurate ascertainment of malaria infections represents a crucial component of malaria surveillance and control. Previous work has revealed the often-substantial prevalence of infections with parasite densities lower than the threshold of detection by microscopy (so called “submicroscopic” infections). The drivers of these infections remain uncertain, despite their established relevance to onwards transmission. In Chapter 2, I carry out a systematic literature review and meta-analysis exploring the prevalence of submicroscopic malaria infections and how this varies between settings. My results highlight extensive variation between settings, with much of this driven by a combination of both historical and current levels of transmission. Crucially, these results highlight significant variation in the prevalence of submicroscopic infections even across settings characterised by similar current levels of transmission, with implications for the utility of control efforts specifically targeting this infected sub-group depending on the context.

Within communities, the distribution of malaria infections is frequently characterised by extensive spatial heterogeneity, which can make identification and treatment of infections challenging. In Chapter 3, using a regression-based approach, I characterise the fine-scale spatial clustering of malaria infections at the household level across a diverse range of sub-Saharan African settings through systematic analysis of 57 Demographic and Health Surveys spanning 23 countries. My results highlight that malaria infections cluster within households, and that the extent of this clustering becomes significantly more pronounced as transmission declines – a factor which will affect the comparative impact of household-targeting or whole-community based control strategies and result in their appropriateness depending closely on the levels of transmission characterising a setting.

In addition to this spatial heterogeneity, malaria transmission dynamics are also frequently characterised by extensive temporal heterogeneity, a phenomenon underpinned by the (often annual) temporal fluctuations in the size of the mosquito populations responsible for transmission. Many questions remain surrounding the drivers of these dynamics however, questions that are rarely answerable from individual entomological studies (focussed on only a single location or species). In Chapter 4 I carry out a systematic literature review to collate anopheline mosquito time-series data from across India and develop a statistical framework capable of characterising the dominant temporal patterns in this dataset. The results demonstrate extensive diversity in the timing and extent of seasonality across mosquito species, but also show that this diversity can be clustered into a small number of "dynamical archetypes", each shaped and driven by a largely unique set of environmental factors including rainfall, temperature, proximity to water bodies and patterns of land use.

In Chapter 5, I apply this framework to time-series data from across South Asia and the Middle East for the highly efficient vector *Anopheles stephensi*, to better understand the factors shaping its seasonal dynamics and the likely impact of its recent establishment in the Horn of Africa. My results reveal significant differences in the extent of seasonality across *Anopheles stephensi* populations, with dynamics frequently differing between rural and urban settings, suggesting structural differences in how these environments shape patterns of vector abundance and potentially warranting different vector control strategies depending on predominant patterns of land-use. Integrating these seasonal profiles into a mathematical model of malaria transmission highlights the crucial need for an understanding of the timing of seasonal peaks in vector density if control interventions like IRS are to be most effectively deployed.

Overall, the results presented here highlight some of the drivers influencing spatial and temporal heterogeneity in malaria epidemiology, quantifies how they contribute to the diverse malaria dynamics observed across different settings, and explores the implication of this variation for effective control of the disease.

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Statement of Originality

I declare that the work presented here is my own work, and that all work carried out by others is appropriately referenced or described below.

In Chapter 2, Dr Lucy Okell shared the results of a previous systematic review of submicroscopic malaria infections (Okell et al., 2012), which formed the starting point for my work. The work presented in Chapter 2 has been published as the following paper:

- **Charles Whittaker**, Hannah Slater, Rebecca Nash, Teun Bousema, Chris Drakeley, Azra C. Ghani, and Lucy C. Okell. "Global patterns of submicroscopic Plasmodium falciparum malaria infection: insights from a systematic review and meta-analysis of population surveys." *The Lancet Microbe* 2, no. 8 (2021): e366-e374 (https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(21)00055-0/fulltext). The data and code used to carry out these analyses can be found at https://github.com/cwhittaker1000/submicroscopic_malaria.

The work presented in Chapter 3 was conducted as part of a collaboration with Dr Gillian Stresman and Dr Jackie Cook, both at the London School of Hygiene and Tropical Medicine (LSHTM). This thesis chapter presents only my work – specifically, analyses carried out by myself, with supervision from Dr Hannah Slater and the LSHTM team. The complete work (including analyses undertaken by the LSHTM team) is published and available via the reference below:

- Gillian Stresman, **Charles Whittaker**, Hannah C. Slater, Teun Bousema, and Jackie Cook. "Quantifying Plasmodium falciparum infections clustering within households to inform household-based intervention strategies for malaria control programs: an observational study and meta-analysis from 41 malaria-endemic countries." *PLoS Medicine* 17, no. 10 (2020): e1003370. https://doi.org/10.1371/journal.pmed.1003370

In Chapter 4, Dr Marianne Sinka and Dr Samuel Pironon shared raster maps of vector presence/absence for the Indian mosquito species considered. The work presented in Chapter 4 has been published as the following paper:

- **Charles Whittaker**, Peter Winskill, Marianne Sinka, Samuel Pironon, Claire Massey, Daniel Weiss, Michele Nguyen, Peter Gething, Ashwani Kumar, Azra Ghani and Samir Bhatt. "A novel statistical framework for exploring the population dynamics and seasonality of mosquito populations". *Proceedings of the Royal Society B (Biological Sciences). 289.1972 (2022):* https://doi.org/10.1098/rspb.2022.0089. The data and code used to carry out these analyses can be found at https://github.com/cwhittaker1000/anopheleseasonality.

In Chapter 5, Dr Ellie Sherrard Smith shared efficacy estimates for different indoor residual spray compounds from (Sherrard-Smith et al., 2018). Dr Arran Hamlet shared estimates of *Anopheles stephensi* bionomic parameters from (Hamlet et al., 2022).

Table of Contents

Chapter 1 Introduction 11
Malaria Global Epidemiology and Burden 11
Parasite Natural History and Lifecycle 14
Diagnosis and Treatment of Malaria 19
Interventions and the Control of Malaria 20
Thesis Aims 36
Chapter 2 Global Patterns of Submicroscopic *Plasmodium falciparum* Malaria Infection: Insights from a Systematic Review and Meta-Analysis of Population Surveys 38
Introduction **38**
Methods **40**
Search strategy and selection criteria 40
Data extraction 40
ANOVA and Tukey's honest significant difference 40
Log-linear regression model formulation 42
Historical and current regional transmission intensity stratification 44
Estimation of contributions to onwards transmission 45
Results 46
Discussion 60
Chapter 3 Quantifying *Plasmodium falciparum* Infection Clustering Within Households to Inform Household-Based Intervention Strategies for Malaria Control Programs 67
Introduction 67
Methods 69
DHS Data: Overview of Data Collation, Survey Selection Criteria and Index Household Definition 69
DHS Data: Detailed Information On Data Collation and Processing Workflow 70
Analysis and Bayesian Regression Modelling of DHS Data 79
Results 82
Discussion 89
Chapter 4 The Ecological Structure of Mosquito Population Seasonal Dynamics 92
Introduction 92
Methods 94
Systematic Review of Indian Entomological Literature 94
Data Extraction, Collation and Initial Processing 95
Time-Series Fitting and Interpolation 99
Time-Series Characterisation and Clustering by Features 102
Statistical Modelling and Prediction of Seasonal Modality 104
Results 106

Discussion 115
Chapter 5 Seasonal Dynamics of an Emerging African Malaria Vector *Anopheles stephensi* and the Implications for Malaria Control 125
Introduction 125
Methods 127
Systematic Review of *Anopheles stephensi* Literature 127
Systematic Review Data Extraction, Collation and Initial Processing 128
Study Geolocation and Environmental Covariate Extraction 130
Time-Series Fitting and Interpolation 130
Statistical Modelling and Prediction of Cluster Membership 131
Transmission Modelling of *Anopheles stephensi*-Driven Malaria Dynamics and Control 133
Results 148
Discussion 165
Chapter 6 Discussion 170
Thesis Aims and Objectives 170
Summary of Findings 170
Limitations and Future Directions 172
Conclusions 175
References 177

List of Figures

Figure 1.1 Global Map Indicating the Dominant Vector Species Responsible for Malaria Transmission. 12
Figure 1.2 The Incidence of Malaria Globally. 13
Figure 1.3 Estimated Country Share of Malaria In 2020. 15
Figure 1.4 Lifecycle of *Plasmodium falciparum* and *Plasmodium vivax* across both the mosquito and human host. 17
Figure 1.5 Trends In Indigenous Malaria Cases in E-2020 Countries, 2010-2018. 21
Figure 2.1: Systematic review overview, workflow and selection of eligible studies. 41
Figure 2.2: Prevalence of infection by PCR *vs* microscopy in 267 prevalence survey pairs and model fits. 47
Figure 2.3: Prevalence of infection by PCR versus microscopy and model fits for previously collated data and data newly identified as part of this review. 49
Figure 2.4: Comparison of Empirically Observed Microscopy Prevalence and Microscopy Prevalence Predicted by Bayesian Regression Modelling. 50
Figure 2.5: Comparison of different model structures and their capacity to fit the collated data. 51
Figure 2.6: Comparing the prevalence ratio across different sampling seasons. 52
Figure 2.7: Comparing the prevalence ratio across different PCR methodologies. 53
Figure 2.8: Global variation in the prevalence ratio and the relative size of the submicroscopic reservoir. 54
Figure 2.9: Tabulation of diagnostic properties by global region. 55
Figure 2.10: The effect of historical and current transmission intensity on the prevalence of submicroscopic malaria infection in Africa. 57
Figure 2.11: Tabulation of diagnostic properties by transmission archetype. 58
Figure 2.12: Sensitivity analysis to assess the robustness of the results surrounding historical and current patterns of transmission intensity. 59
Figure 2.13: The influence of age on submicroscopic malaria infection. 61
Figure 2.14: The potential contribution of submicroscopic infections to onwards transmission according to current and historical transmission intensity. 62
Figure 3.1 Map of Africa highlighting the geographic distribution of countries where the Demographic and Health Surveys (DHS) used in the analyses were conducted. 83
Figure 3.2 Malaria infection and the extent of household clustering according to detectability by different programmatic strategies. 84
Figure 3.3 Malaria infection and the extent of household clustering according to detectability by different programmatic strategies when comparing DHS surveys carried out in all age-groups, compared to those conducted in children under 7 only. 85
Figure 4.1 The Raw Mosquito Data Extracted During the Systematic Review Process. 107
Figure 4.2 Exploring Species Complex-Specific Patterns of Mosquito Population Dynamics. 108
Figure 4.3 Temporal Cluster Statistical Properties. 109
Figure 4.4 Characterisation and Clustering of Time-Series with Similar Temporal Properties. 110
Figure 4.5 Results of Clustering When Fitting Mosquito Catch Data Using An Uninformative Prior. 111
Figure 4.6 Comparison of Individual Study Sizes By Cluster. 112
Figure 4.7 Exploring the Cross-Correlation Between Rainfall and Mosquito Densities. 113
Figure 4.8 Exploring Drivers of Mosquito Population Dynamics Using Multinomial Logistic Regression. 116

Figure 4.9 Species-Cluster Coefficient Values and Hierarchical Clustering Results When Data Were Subsampled. 117
Figure 4.10 Cross-Cluster Correlations For Ecological Coefficients. 118
Figure 4.11 Predictive Maps of Mosquito Population Seasonality Across India. 119
Figure 5.1 Model schematic illustrating the different disease states and the possible transitions between them. 135
Figure 5.2 Flow diagram illustrating the probabilities associated with feeding outcomes. 147
Figure 5.3 Sources and Locations of Anopheles stephensi Time-Series Data and Examples for Each Country. 149
Figure 5.4 Results of model fitting to the longitudinal entomological data collated in this study. 150
Figure 5.5 Characterisation and Clustering to Identify Time-Series with Similar Temporal Properties. 151
Figure 5.6 Results of Clustering For 4 Clusters Instead of 2. 152
Figure 5.7 Random Forest Prediction of Temporal Cluster Membership. 154
Figure 5.8 Partial Dependence Plots for Covariates Used in the Random Forest Classification Modelling. 155
Figure 5.9 Random Forest Classification Results Without Upsampling Cluster 2. 156
Figure 5.10 Random Forest Classification Results With Held-Out Data. 157
Figure 5.11 Exploring Variation In Total Catch Size By Cluster. 158
Figure 5.12 Random Forest Classification Results Including Monthly Catch Size As A Model Covariate. 159
Figure 5.13 Random Forest Prediction of Percentage of Vector Density In Any 3 Month Period. 161
Figure 5.14 Predicting the Possible Seasonal Dynamics of Anopheles stephensi Across the Horn of Africa. 162
Figure 5.15 Modelling the Public-Health Impact of Indoor Residual Spraying (IRS) and How This Is Impacted by Anopheles stephensi Seasonality. 163
Figure 5.16 Sources and Locations of Anopheles stephensi Time-Series Data According to Urban/Rural Assignment. 164

List of Tables

Table 3.1 Overview of programmatic strategies for identifying households likely to have asymptomatic and/or subpatent infections 71
Table 3.2 Definition and Extraction of Treatment Seeking Variables 74
Table 3.3 Summary of the Collated Demography and Health Survey Data For Each Survey. 86
Table 4.1 Number of time series collated for each species 95
Table 4.2 Number of time series collated according to method of collection 95
Table 5.1 Number of time series collated according to method of collection 129

Acknowledgements

I struggled with writing this – reflecting on how far I'd come, how many people helped me get here, and the events of the past five years elicited far more emotion than I was anticipating. I also found it tricky doing justice to what I wanted to acknowledge; that is, to find the words to convey the magnitude and depth of my appreciation whilst avoiding straying into meaningless platitudes or cringey tropes. Still, the Acknowledgements section is always my favourite part when I read other PhD theses, so I had to try, and I hope you'll indulge me here. I hope I've got there with the words below, and I hope that you read this as how I intended it: a tribute to the innumerable acts of kindness, support and intellectual generosity that have been gifted to me across the course of the PhD and are responsible for making me the scientist I am today.

I'd like to begin with a huge thanks to my supervisors Azra Ghani, Pete Winskill and Samir Bhatt. It took us a little longer than anticipated to get here, given the pandemic, but your continued support and guidance and mentorship has been invaluable on this journey. I can't thank you enough for the trust you placed in me to take ownership of this work; and the guidance to help me through the bits I couldn't navigate on my own. It's been a privilege to work with you all and thank you for all that you've done in shaping me into the scientist I am today.

Being a part of the Department of Infectious Disease Epidemiology and MRC Centre for Global Infectious Disease Analysis has been in short, a joy. So many people within the department have helped me over the years – all of them have my thanks. However, I'd particularly like to extend the biggest, most heartfelt thanks I can muster to:

- The malaria modelling group – I've been overwhelmed by the generosity of its members and I couldn't have asked for a more friendly and supportive environment in which to do a PhD. Thanks in particular to Patrick Walker, Lucy Okell, Tom Churcher, Ellie Sherrard-Smith and Hannah Slater, all of whom have played a significant role in shaping the work presented here.
- Rich FitzJohn, who has put up with a simply extraordinary barrage of programming questions and who can take sole credit for helping me move away from the "bloop", laden code the early stages of my PhD were riddled with.
- To everyone who reviewed my Fellowship applications, often at last minute, and often only with a minimal idea what the work was about ahead of time – particularly to Patrick Walker, Jeff Eaton and Erik Volz for their contributions in this respect – thank you.

So much of the day-to-day of my PhD work has involved abstraction – of biological systems; and aggregation – of individuals into parse-able statistics. I don't doubt the importance of this, but it's so easy in this context to lose sight of the people these numbers represent, and the core motivations underpinning my work – to improve the health of real, living people around the world. In this context, another huge thanks goes to Natsuko Imai, Anne Cori and Thibaut Jombart who gifted me what I suspect will prove to be one of the privileges of my lifetime when they helped me deploy to the North Kivu Ebola Outbreak in 2019 in DRC as an analyst with the WHO. This opportunity only came about because of your tireless efforts and faith in me. The future work that I will (hopefully) get to do in outbreak response is because of you all. Thank you.

One of my favourite parts of science; one of the unquestionably wonderful things that can anchor you when code isn't working or the data's noisier than anticipated, is the opportunity it affords you to work with wonderful people from around the world. To that end, I'd also like to thank and recognise the following people who I've had the immense privilege of working with, and learning from, over the course of my PhD:

- To Noe, Richy, Xavier and the rest of the EpiCell analytics team in Goma who taught me a great deal about the challenges and practicalities of outbreak response. Your patience and kindness during the immense challenges posed by an ongoing outbreak meant the world to me.
- To Gillian Stresman, Jackie Cook, Teun Bousema and Chris Drakeley, whose collaboration shaped Chapters 2 and 3 of this thesis and whose knowledge of all things malaria epidemiology was invaluable, thank you.
- To Ashwani Kumar, Marianne Sinka and Sam Pironon for being indispensable encyclopedia of mosquito knowledge and whose insights and contributions have significantly shaped the work contained in Chapters 4 and 5 of this thesis – thank you.
- To Maria-Gloria Basanez and Martin Walker for giving me my first big break working on loiasis, and to all of the Montpellier crew – to Cedric Chesnais, Seb Pion, Michel Boussinez and Jeremie Campillo. I look forward one day to getting loiasis the recognition it deserves as a public health threat. Thank you.

PhDs are tough. Doing this PhD however has been made immeasurably easier and more enjoyable by the wonderful, kind and intelligent people I've had the privilege to call my friends whilst a student in DIDE (maybe the real doctorate was the friends we made along the way etc etc). To everyone who calls (or has called) VA5 home – to Kelly, Sarah, Andria, Nora and Hayley. It has been beyond an honour to walk this path with you and watch all of you develop into the incredible scientists you've become. And to those who don't work in VA5 but who occupy the VA5 of my heart – to OJ, Amy, Matt, Arran, Gina, Izzy, James, Helen and Andra – meeting all of you has been one of the biggest highlights of the PhD. Thanks for being so wonderful and kind. And to all of my friends outside of work, especially those I've had the joy of living with over the past 5 years – to Clare, Patrick, Sam, Alex, Dom and Cian. Your support, company and friendship have and continues to mean the world to me – it was a salve during the PhD's toughest moments and continues to be one of the things in my life I'm most grateful for.

I'd be remiss if I didn't mention the COVID-19 pandemic and its impact here. In February 2020 I paused my PhD and started working full-time as a member of the Imperial College COVID-19 Response Team. What was intended to be 2 months of acute response work somehow snowballed into 18 months of desperately challenging, rewarding and (at times) sad work. Words can't really do justice to that experience – to describe how proud I am of what we did during that time, and at how cripplingly sad I am about the tragedy that has since transpired globally. The work we did during that time was some of the most challenging (both scientifically and emotionally) I have ever been involved in and the one silver lining was the opportunity I got to work with some of the most talented and compassionate scientists I think I'll ever meet. To Patrick, OJ, Azra, Pete, Bob, Sam, Tom, Swapnil, Seth, Iwona, Olli, Natsuko, Katy, Lucy, Ilaria, Alexandra, Gina, Neil, Marc, Nick, Tim, Zulma, Ester, Nuno, Lewis, Carlos, Christine, Nikita, Avery and all the many, many others I've had the honour of working with during some of the most trying of circumstances over the past 2 years; thank you.

I finish with the most important thanks, which is to my parents, Ann and Andy. The motivations that drive my work – curiosity about the world and the desire to improve it – are motivations I inherited from them. I regard myself as immensely privileged to have been raised by people with such a strong sense of justice and fairness. Throughout my entire life, they have time and time again provided loving, unconditional support to all my academic (and non-academic) endeavours. Knowing they have my back no matter what gave me both the opportunity and confidence to switch from an ill-fated plant science PhD in molecular epigenetics to the Master's in Infectious Disease Epidemiology at Imperial that eventually led me to this thesis. I couldn't have done any of this without them. What I have produced here is because of them; and it is for them.

Chapter 1 Introduction

In this Chapter, I review the epidemiology, public health burden, and transmission dynamics of malaria, primarily the specific manifestation of the disease caused by the parasite *Plasmodium falciparum*. A specific focus of this Chapter is on the factors that underly the marked differences in the transmission dynamics and epidemiology of malaria across different contexts and ecologies. Developing a better understanding of these different factors, how they contribute to the diverse malaria dynamics observed across different settings, and what this variation means for the control of the disease, represents the main objective of this thesis work. I conclude this Chapter by describing the aims and objectives of this thesis, as well as a brief summary of the work contained in each of the chapters that follow this one.

Malaria Global Epidemiology and Burden

With over 620,000 estimated deaths in 2020 (World Health Organization, 2021b), malaria represents one of the most serious infectious diseases globally (Roth et al., 2018). The human form of the disease is caused by 5 members of the *Plasmodium* genus. These are the human parasites *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*, and also includes the more recently identified zoonotic malaria of macaques, *Plasmodium knowlesi* (Singh et al., 2004), though clear human to human transmission for this latter parasite is yet to be confirmed. The parasite is transmitted by mosquitoes belonging to the *Anopheles* genus, with parasite development and maturation occurring in a series of defined steps that span both the human and mosquito hosts. Of the roughly 460 recognised members of the *Anopheles* genus, approximately 70 possess the capacity to transmit human malaria parasites (Warrell and Gilles, 2017), with 41 having been identified as dominant vector species responsible for the majority of transmission that occurs (Sinka et al., 2012) **(Figure 1.1)**. Both parasites and vectors are heterogeneously distributed across the tropical and sub-tropical regions where malaria is endemic, a feature that results in marked differences in the transmission dynamics and epidemiology of malaria across different contexts and ecologies.

Globally, sub-Saharan Africa has the highest burden of morbidity and mortality associated with the disease, with nineteen countries in sub-Saharan Africa along with India accounting for almost 85% of the global burden **(Figure 1.2)**. 95% of all malaria cases and 96% of all malaria deaths between 2019 and 2020 were estimated to have occurred in the World Health Organization (WHO) African (AFRO) region (World Health Organization, 2021b), where the majority of malaria morbidity and mortality is caused by the parasite *Plasmodium falciparum*. Outside sub-Saharan Africa, the dominant malaria causing species is highly variable – across the Americas, *Plasmodium vivax* dominates, with over 75% of malaria cases in 2018 attributed

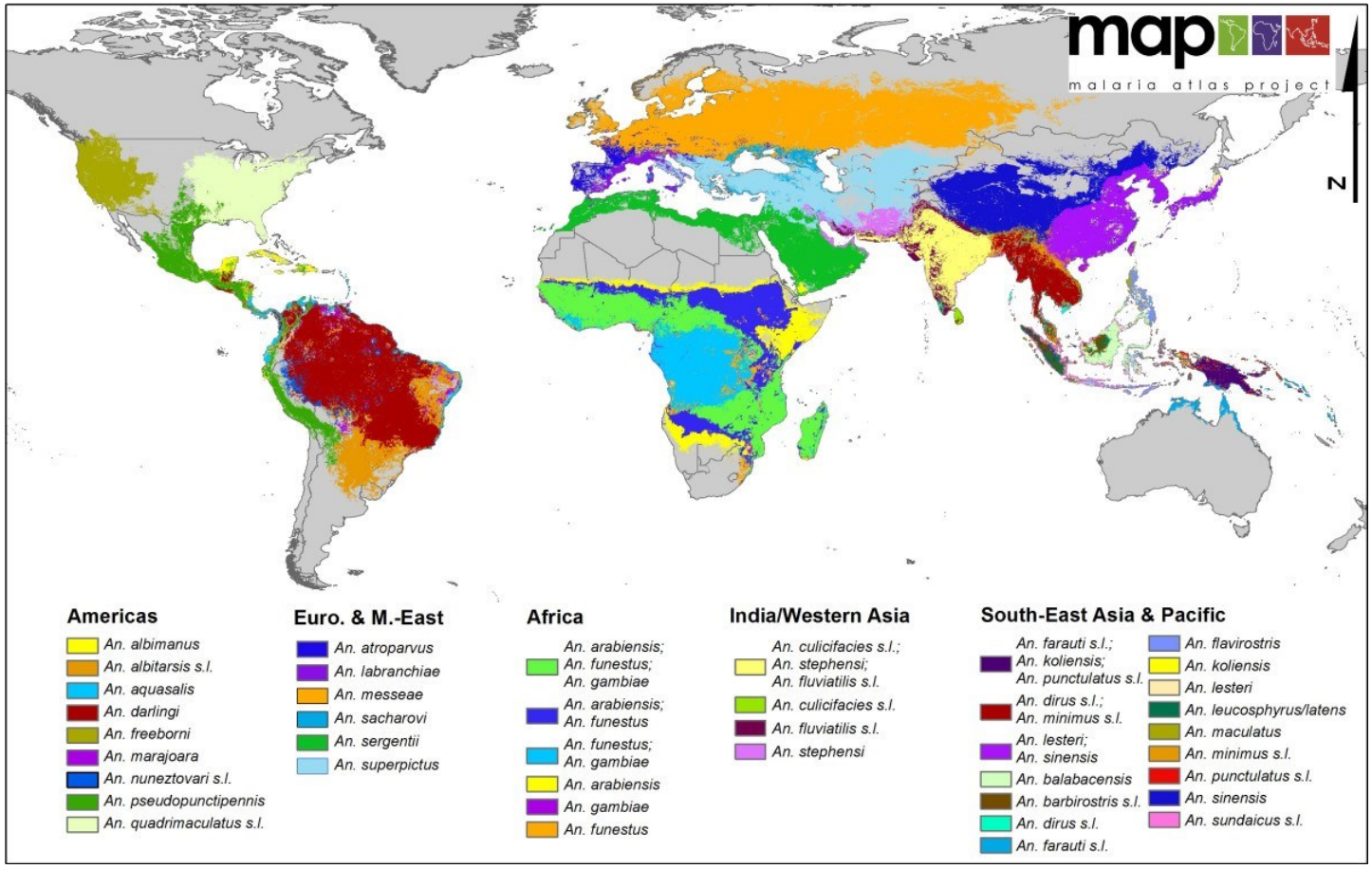


Figure 1.1 Global Map Indicating the Dominant Vector Species Responsible for Malaria Transmission. Sourced from Sinka et al 2012 (Sinka et al., 2012).

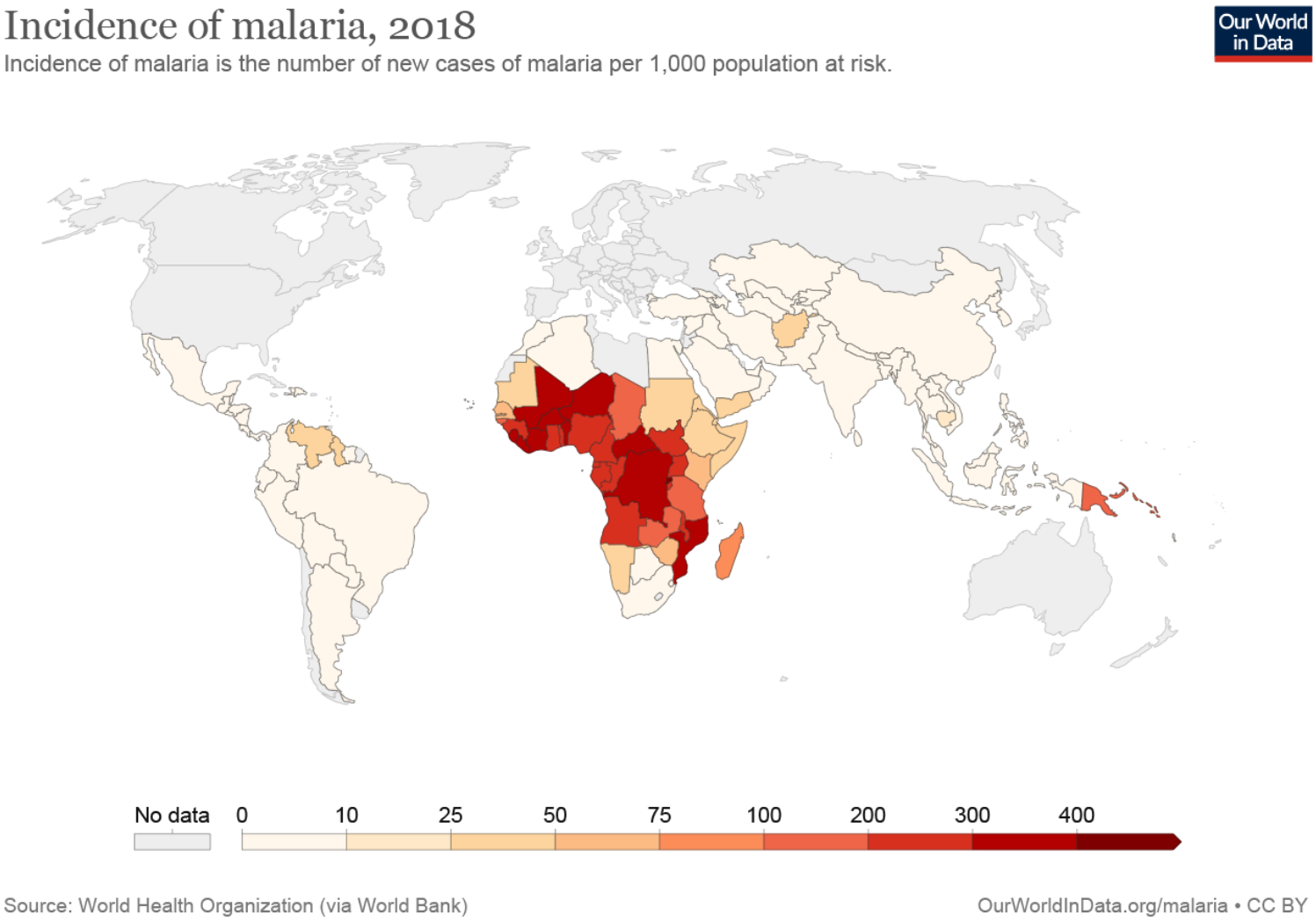


Figure 1.2 The Incidence of Malaria Globally. Sourced from Our World In Data (Roser and Ritchie, 2019). Colour indicates number of malaria cases per 1,000 population at risk.

to the parasite. Across South-East Asia, there was an almost equal split of estimated cases attributed to *Plasmodium falciparum* and *Plasmodium vivax*, with India alone accounting for almost 50% of the global *Plasmodium vivax* burden (Battle et al., 2019) **(Figure 1.3)**. Across much of southeast Asia, the simian malaria parasite *Plasmodium knowlesi* (which can cause severe and fatal diseases in humans) is also present (Moyes et al., 2014) and increasingly relevant public health threat. For example, across parts of Malaysia, *Plasmodium knowlesi* has become the most common cause of human malaria (and despite near elimination of other previously present human-only *Plasmodium* species (Cooper et al., 2019a)).

Despite this marked heterogeneity in the spatial distribution of parasites and cases, burden of disease, in particular mortality, is highly concentrated in sub-Saharan Africa, where an estimated 96% of malaria deaths in 2020 occurred – 80% of these in children under 5 years (World Health Organization, 2021b). Since the millennium, significant gains have been made in controlling and mitigating the public health impact of the disease. Between 2000 and 2015 an estimated 1.2 billion cases and 6 million deaths have been averted, with global incidence of malaria having fallen by an estimated 37% (Bhatt et al., 2015b), an achievement underpinned predominantly by significant scale-up of control interventions including insecticide-treated bednets (Bhatt et al., 2015a). Indeed, despite more limited recent progress, global trajectories in the decades since the year 2000 have overall been characterised by significant declines in morbidity and mortality.

Parasite Natural History and Lifecycle

The Lifecyle of the Malaria Causing Parasite *Plasmodium falciparum*

The lifecycle of the malaria causing parasite *Plasmodium falciparum* in the human host is structured into three key stages, spread across the human and mosquito hosts. The “exo-erythrocytic” stage begins with the injection of *Plasmodium falciparum* sporozoites (the infective motile stage of the parasite) into the human host via the bite of an infected female mosquito of the *Anopheles* genus. These sporozoites then migrate from the site of the bite, until they reach a blood vessel, whereupon they journey to the liver. In the liver they invade the host hepatocytes, develop into trophozoites and begin replicating through multiple asexual fissions (schizogony) to produce a significant number of merozoites (often tens of thousands), which are released into the bloodstream upon hepatocyte eruption (Meis et al., 1986). The process of infection, invasion, replication and release takes approximately 6-7 days in the case of *Plasmodium falciparum* (though can be many months or even years for hypnozoites in the case of *Plasmodium vivax*). Following release into the blood stream, the erythrocytic cycle commences. The released merozoites rapidly invade the host erythrocytes and develop into

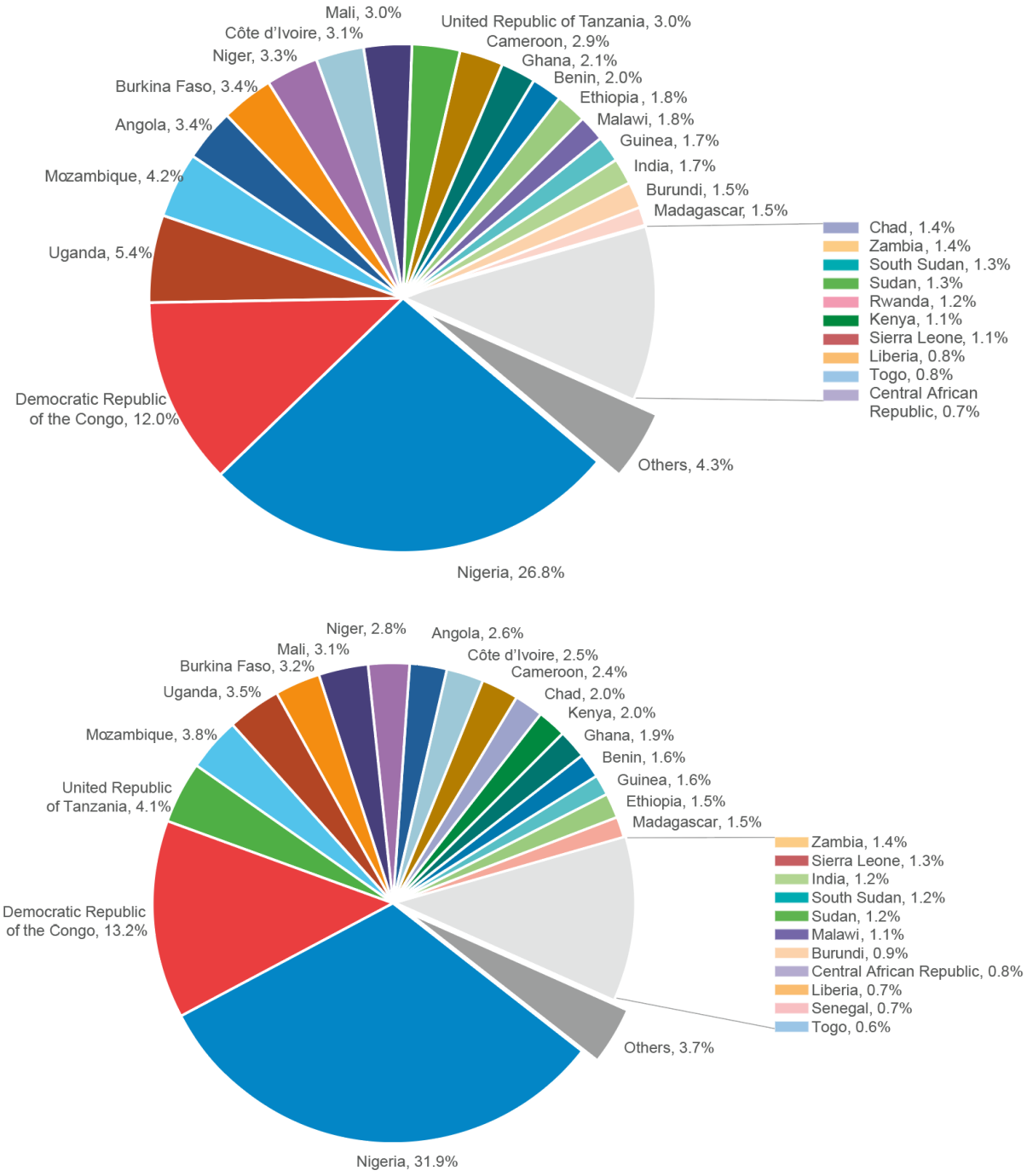


Figure 1.3 Estimated Country Share of Malaria In 2020. Results are displayed for **(A)** Total estimated malaria cases; and **(B)** *Plasmodium vivax* malaria cases. Sourced from the WHO World Malaria Report 2021.

ring stage trophozoites. Multiple rounds of asexual reproductive occur within the erythrocyte, eventually followed by rupture and release of the newly produced merozoites, ready to begin another round of the erythrocytic cycle. This asexual reproductive cycle takes approximately 1-2 days. Of the released merozoites, a number will differentiate into the sexual stage of the *Plasmodium falciparum* parasite, which are known as gametocytes and which are responsible for infecting the mosquito host when it takes a blood meal (Baker, 2010). Though there are both male and female gametocytes, gametocyte sex ratios are frequently female-biased (though see here for work highlighting plasticity of these ratios in response to a wide diversity of human-host related factors (Paul et al., 2000; Mitri et al., 2009)). **(Figure 1.4)**

The sporogenic stage of the *Plasmodium falciparum* life cycle takes place inside the anopheline mosquito host responsible for onwards transmission. Gametocytes concentrate in skin capillaries and are taken up by the mosquito during feeding. Upon feeding, the gametocytes are transferred to the mosquito's gut, whereupon they undergo a process of further development and maturation. Each male gametocyte produces eight microgametes through 3 rounds of mitosis, whilst the female gametocyte matures into a microgamete directly. The male and female gametocytes then fuse to form a diploid zygote, which elongates into the motile ookinete, which exits the gut through invasion and subsequent passage through the epithelium as an oocyst. Oocysts then undergo multiple rounds of replication to form sporozoites, which then migrate from the mosquito's abdomen to its salivary glands, whereupon they become available for transmission to humans during subsequent successful bloodmeal feeding by the mosquito. The time required for this process of sporogyny i.e. the time between the malaria causing parasite infecting a mosquito, reproducing and migrating to the salivary glands whereupon they can be transmitted, is known as the extrinsic incubation period (EIP). The length of the EIP is typically long (typically 10-14 days) relative to the lifespan of the mosquito (Smith and McKenzie, 2004), meaning it is predominantly older mosquitoes that pass on infection (and that malaria transmission responds acutely to changes in the longevity and survival of mosquitoes (Macdonald, 1956)). It is also highly plastic – parasite development rates are highly sensitive to a number of factors, including environmental temperature (Shapiro, Whitehead and Thomas, 2017), mosquito nutritional status (Hien et al., 2016) and innate immune responses (Clayton, Dong and Dimopoulos, 2014) amongst others. All of these factors are able to modulate parasite development rates (as well as various mosquito traits related to vector competency (Mordecai et al., 2019)) and in doing so, shift the timing and establishment of the *Plasmodium falciparum* parasite in the mosquito vector.

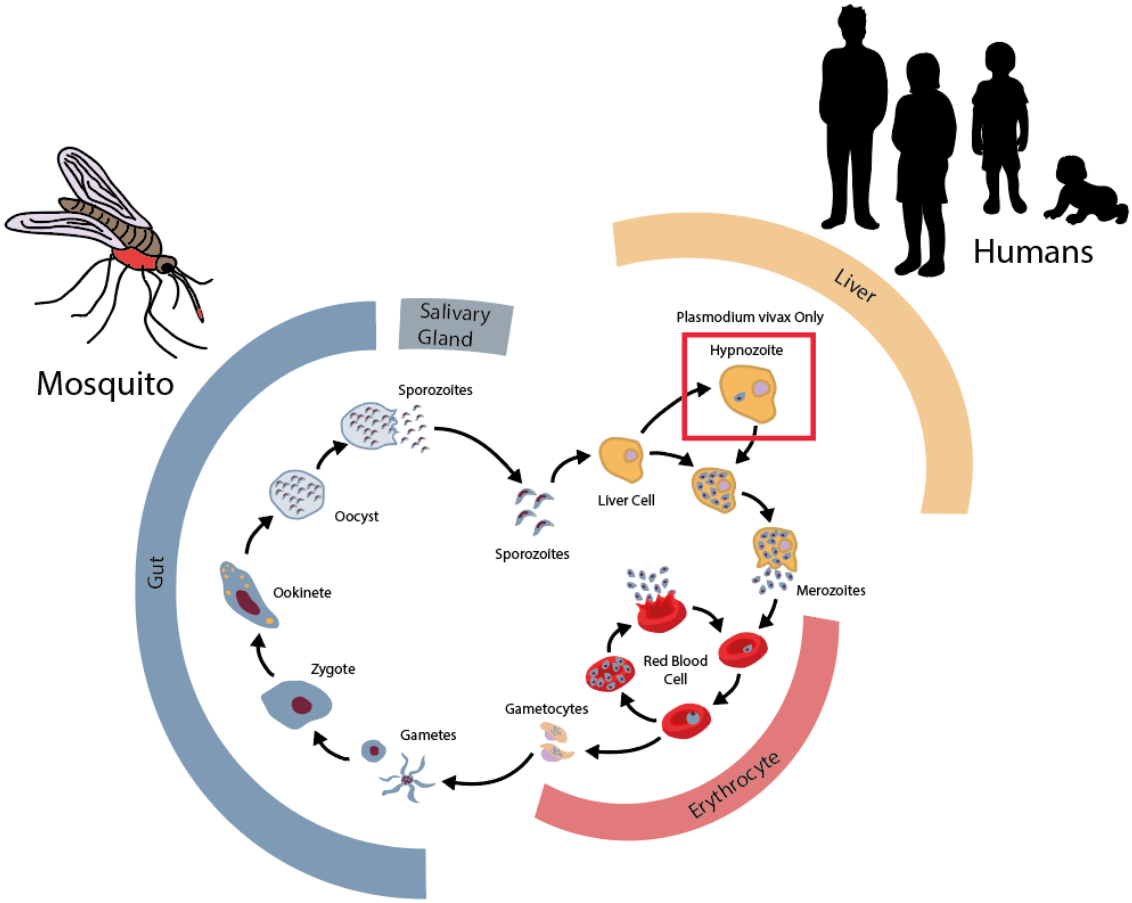


Figure 1.4 Lifecycle of *Plasmodium falciparum* and *Plasmodium vivax* across both the mosquito and human host. The life cycle of malaria involves both mosquito and human hosts. During feeding on a human, sporozoites present in the mosquito salivary gland enter the bloodstream. These then pass into the liver where they invade liver cells (hepatocytes) and multiply asexually over the next 7 to 10 days. Following this, the merozoites are released from liver cells into the bloodstream, where they invade red blood cells (erythrocytes). What follows is a cycle of multiplication, bursting of the erythrocyte and release of the new parasites into the bloodstream, followed by invasion of more erythrocytes. A small proportion of the infected red blood cells exit from this cycle of asexual multiplication and instead develop into sexual forms of the parasite, known as gametocytes. These circulate in the bloodstream where they can be picked up by mosquitoes during their feeding on humans. Ingestion of the gametocytes by a mosquito triggers further maturation into mature sex cells known as gametes. Fertilisation then occurs, with the fertilised female gamete developing into ookinetes that traverse the mosquito's midgut wall, forming oocysts on the exterior surface. Inside these oocysts, numerous sporozoites develop. Multiplication of these sporozoites leads to bursting of the oocyst, releasing the sporozoites into the body cavity and allowing passage to the mosquito's salivary glands. Upon reaching the salivary glands, the sporozoites are now ready for transmission to the human during the mosquito's next feed.

Plasmodium falciparum Malaria Natural History, Morbidity, Mortality and Immunity

The symptoms and consequences of infection with *Plasmodium* species vary in severity depending on both the parasite species and a number of host factors, including the level of host immunity (linked to past exposure to the parasite, which increases with age (Rodriguez-Barraquer et al., 2018)) and genetic factors. In individuals with limited prior exposure (and hence immunity), rapid division and proliferation of merozoites in the erythrocytic stage of the parasite life cycle leads to development of symptoms on average 7-15 days after inoculation by the mosquito vector. Initial symptoms typically include nausea, headaches and fever. As the infection progresses, continual depletion of erythrocytes due to parasite replication can result in acute anaemia (and the cyclical fever and chills characteristic of malaria infections (Lamikanra et al., 2007)). A proportion of these symptomatic cases will go on to develop severe malaria, which is associated with significantly elevated mortality and whose complications include severe anaemia, end-organ damage, cerebral malaria and numerous pulmonary complications (Phillips et al., 2017). The exact clinical manifestations of severe disease depend in part on the setting and its overarching patterns and intensity of malaria transmission – in high transmission areas where the average age at first infection is lower, severe disease frequently manifests as severe anaemia (which is more common in young children). By contrast, cerebral malaria is typically concentrated in older children, and so more likely to occur in areas of low to moderate transmission (Njuguna et al., 2019; Reyburn et al., 2005). Cerebral malaria typically has a higher case fatality rate, reaching up to 25% (Mockenhaupt et al., 2004; Seydel et al., 2015), though recent work has highlighted the substantially increased risk of mortality in the months following clinical discharge (in those who have received treatment) that severe anaemia results in (Kwambai et al., 2020).

The burden of severe malaria in endemic areas is typically concentrated in children who have little to no pre-existing immunity against the malaria causing *Plasmodium* parasites. Indeed, infection with malaria results in an immune response, with cumulative exposures resulting in gradual acquisition of immunity against symptomatic malaria. These dynamics are a key driver of the epidemiology of malaria in endemic settings, and result in dynamics that vary significantly with overall levels of transmission – specifically that acquisition of immunity is highest in high transmission settings (Griffin et al., 2015), with adults in these settings often not developing malaria symptoms during infection due to the high levels of immunity in the population (Langhorne et al., 2008). Gradual acquisition of adaptive immunity against symptomatic malaria is typically subdivided into anti-parasite immunity (i.e. an increased ability to control parasite densities during infection) and anti-disease immunity (an ability to tolerate higher parasite densities without developing the fever and other symptoms characteristic of a malaria infection) (Rodriguez-Barraquer et al., 2018). Both aspects of immunity build up in response to repeated

exposures and infections, though at substantially different rates. In highly endemic settings, immunity to severe disease typically develops rapidly, as evidenced by the drop in frequency of malaria deaths typically falling between the ages of 2-5 years (Doolan, Dobaño and Baird, 2009), and the majority of malaria morbidity and mortality occurring in children under the age of 5 (Murray et al., 2012). Anti-parasite immunity by contrast, appears to build up more slowly, developing over the course of multiple infections (Rodriguez-Barraquer et al., 2018).

Diagnosis and Treatment of Malaria

Diagnosis of malaria is complicated by non-specificity of symptoms and the frequent overlap and similarity in clinical presentation to other diseases that typically co-occur with malaria (Glennon et al., 2020). Based on this, malaria detection is frequently determined using either microscopy of collected blood to identify parasites circulating in the blood or via rapid diagnostic tests (RDTs) to detect circulating parasite antigens (typically lactate dehydrogenase or histidine-rich protein 2). More recently, there has been increased usage of more sensitive molecular methods (although such usage is still primarily restricted to research contexts). These techniques (typically utilising polymerase chain reaction based methodologies (Snounou et al., 1993)) have highlighted the presence of infections with parasite densities lower than the threshold of detection by routine methods such as microscopy (so-called “submicroscopic” infections (Okell et al., 2012)).

Treatment of malaria uses antimalarial drugs, with prompt access to treatment a key determinant of survival for severe malaria cases, and delays in the ability to access treatment are significantly associated with elevated mortality due to the disease (Mousa et al., 2020). The first anti-malaria widely used to treat malaria globally was chloroquine (CQ), though widespread usage as part of the Global Malaria Eradication Programme during 1955-1969 and the emergence of resistance (Wellems and Plowe, 2001) led to its replacement as the primary drug for treatment of malaria cases. CQ was subsequently replaced by sulfaxdoxine/pyrimethamine (SP). Resistance to SP has since emerged (Gatton, Martin and Cheng, 2004), although the geographical distribution of resistant *Plasmodium* genotypes is more limited than that of CQ. For example, across sub-Saharan Africa, SP resistance is primarily centred around East Africa, with malaria parasites across West Africa still largely susceptible to SP (Okell, Griffin and Roper, 2017), and indeed, the drug is still routinely used (alongside amodiaquine) in largescale seasonal malaria chemoprevention (SMC) campaigns in the region (Baba et al., 2020). The 1970s saw the discovery of the highly potent and effective anti-malarial drug artemisinin which has been crucial in the progress made in recent years to reduce the global burden of malaria – estimates suggest approximately 20% of the reduction in malaria incidence between 2000 and 2015 can be attributed to artemisinin and artemisinin-related treatments (Bhatt et al., 2015b)).

Contemporary usage of the drug is in the form of artemisinin combination therapies (ACTs), which represent formulations in which artemisinin (or an artemisinin derivative) are combined with a longer lasting partner drug. Such combination therapies have in part been motivated by a desire to mitigate and slow the spread of artemisinin resistance. This resistance is primarily restricted to South-East Asia (World Health Organization, 2018), although recent reports of parasite mutations associated with artemisinin resistance in Rwanda provide significant cause for concern (Uwimana et al., 2020).

Interventions and the Control of Malaria

Brief Historical Overview of Malaria Control and Elimination Policy

In 1955, the World Health Organization launched a global campaign aiming to eradicate malaria in all endemic settings around the world. This initiative (the Global Malaria Eradication Programme) primarily focussed on vector control efforts and included indoor residual spraying of households and other buildings with the insecticide dichloro-diphenly-trichloroethane (DDT) (The Lancet - Editorial, 2007). The programme was marked by several successes. Malaria was largely eradicated from southern Europe as well as parts of north Africa and the Middle East, with elimination achieved in a total of 37 countries. A number of these sustained this elimination in the decades following the decision to end the programme in 1969 when it was recognised that eradication was not achievable with the available means in many of the targeted areas (Nájera, González-Silva and Alonso, 2011). However, the end of the Global Malaria Eradication Programme also subsequently saw widespread resurgence of malaria in many of the areas that had previously been targeted with control measures (Cohen et al., 2012). The millennium saw the Roll Back Malaria Summit, which yielded a declaration aiming to halve malaria mortality by 2010. 2000 also saw the formation of the Bill and Melinda Gates Foundation, which made a significant financial and political commitment to elimination and eradication of malaria (Roberts and Enserink, 2007), as well The Global Fund to Fight AIDS, Tuberculosis and Malaria, whose disbursements against these 3 diseases regularly eclipses $4 billion annually. The renewed commitment evidenced by these events has since translated into significant reductions in the burden of malaria (as described in more detail above), with an estimated 41% decreases in global incidence of malaria cases between 2000 and 2015, and over 6 million deaths averted (Bhatt et al., 2015b). In 2016, the World Health Organization identified 21 countries with the potential to achieve zero indigenous cases of malaria by 2020 (the so-called “E-2020” initiative). Of these countries, 8 to-date reported zero indigenous cases of malaria in 2020 (World Health Organization, 2021) **(Figure 1.5)**. However, global estimated malaria incidence remained

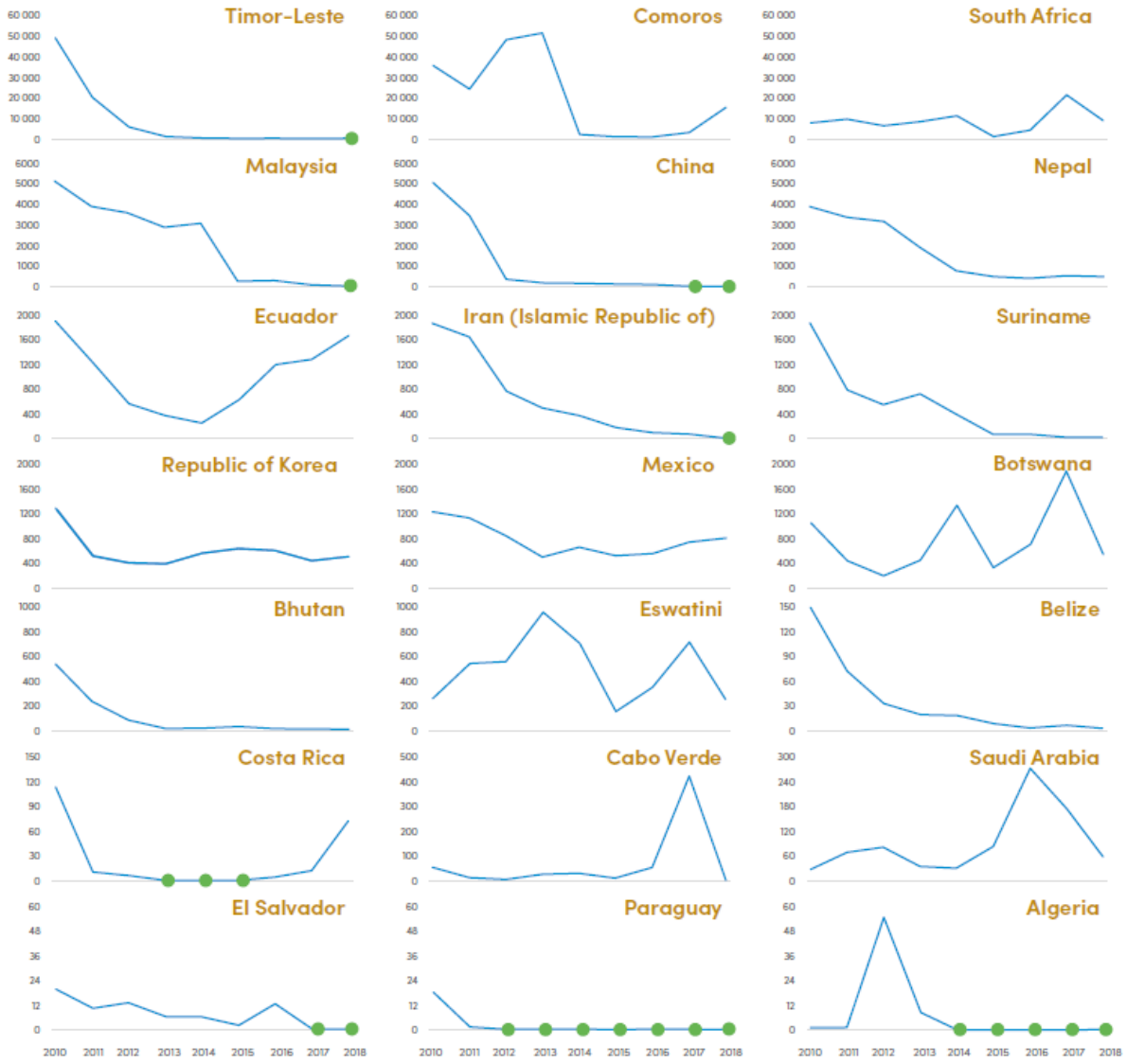


Figure 1.5 Trends In Indigenous Malaria Cases in E-2020 Countries, 2010-2018. Countries are presented from highest to lower number of malaria cases in 2010. Years with zero (indigenous) malaria cases are highlighted with green points. Sourced from the WHO World Malaria Report 2019

largely constant across 2016 and 2017, declining slightly between 2017 and 2019, before increasing in 2020 and 2021, a factor thought to be attributable to the impact of COVID-19 on disrupting provision of malaria prevention, diagnosis and treatment (Sherrard-Smith et al., 2020; Hogan et al., 2020; Weiss et al., 2021). Motivated by this stalled progress, recent years have seen the launch of the "High Burden to High Impact" programme, a targeted and country-led approach focussed on the 11 countries where approximately 70% of the world's malaria burden is concentrated – specifically Burkina Faso, Cameroon, Democratic Republic of the Congo, Ghana, Mali, Mozambique, Niger, Nigeria, Uganda, Tanzania and India.

Methods of Malaria Control i. Anti-Malarial Measures Targeting the Malaria Parasite

In addition to the direct treatment of disease described above, anti-malarial drugs are increasingly being used as preventative interventions. Examples of such measures include intermittent preventive therapy in infants (IPTi), which seeks to protect infants during the vulnerable first months of life through administration of anti-malarials such as SP (Aponte et al., 2009), as well as seasonal malaria chemoprevention (SMC), which aims to protect school age children in regions with high seasonality of malaria transmission by administering prophylactic anti-malarial drugs over the course of the rainy season (which previous work has indicated is likely to overlap with the period of highest transmission (Cairns et al., 2012)). SMC has had significant impact across the Sahelian region of West Africa, where implementation is now widespread (Baba et al., 2020). Additionally, intermittent preventive treatment in pregnancy (IPTp) involving regular administration of SP is recommend by the WHO (World Health Organization, 2014) and has been associated with reduced incidence of both neonatal mortality (Menéndez et al., 2010) and low birthweight (Desai et al., 2018). In addition to these interventions targeting specific age-groups, other preventive strategies treating entire populations include mass drug administration (MDA), which involves treatment of an entire population with anti-malarials in order to reduce prevalence of infection and transmission (Mwesigwa et al., 2019; Tripura et al., 2018; von Seidlein et al., 2019b), or mass screen and treat (MSAT), which involves screening a community for malaria infection using diagnostic tools or clinical markers such as presence of a fever, and then treating those meeting the criteria for likely malaria infection (Parker et al., 2017). Concerns around the logistical feasibility and costs of treating entire communities in this way has in some cases motivated the usage of reactive strategies, such as reactive case detection (RACD), which involves testing and (if positive) treating household members of each malaria case passively detected at health facilities (Stresman et al., 2020; Hustedt et al., 2016). Other research has investigated the usage of the gametocyte primaquine (the only currently available anti-malaria drug that clears mature *Plasmodium falciparum* gametocytes in infected humans) in conjunction with artemether-

lumafantrine as a way to simultaneously treat disease and limit onwards transmission of malaria to mosquitoes (Eziefula et al., 2014; Stone et al., 2022).

In the instances mentioned above, treatment has typically been with anti-malarial drugs that aim to either clear current infections or provide some temporary degree of prophylactic protection from infection in the near future. There is however increasing interest in the use of endectocidal compounds (i.e. those that possess activity against the anopheline mosquito vector as well as standard anti-parasitic activity). For example, previous work has highlighted the mosquitocidal activity of ivermectin on anopheline mosquitoes (Smit et al., 2019, 2018), as well as the potential population-level impact of delivering the drug during mass-drug administration campaigns as assessed through mathematical modelling (Slater et al., 2020), impact that has since been established in recent randomized trials evaluating the impact of adding ivermectin to MDA regimen (Dabira et al., 2022).

There has also been considerable interest in the development of vaccines for protection against malaria. Malaria vaccines can be split into three categories depending on the particular stage of the *Plasmodium* parasite lifecycle being targeted. These are pre-erythrocytic vaccines, which aim to prevent infection through targeting sporozoites inoculated into the skin, blood-stage vaccines, which aim to reduce the probability of clinical/severe disease through reducing parasite densities during infection, and transmission blocking vaccines, which aim to prevent onwards transmission of the *Plasmodium* parasite from humans to mosquito by targeting the sexual stage of the parasite in humans. Though many candidate vaccines have been explored to date, few have found widespread success. Perhaps the most notable exception is the vaccine RTS,S/AS01, developed by GlaxoSmithKline (GSK) in collaboration with the Walter Reed Army Institute of Research (WRAIR) and which has recently been recommended for widespread delivery to children across sub-Saharan Africa by the World Health Organization. Results from a Phase III randomised, controlled evaluation of the vaccine suggested efficacy against clinical malaria in the region of 30-50% (RTS,S Clinical Trials Partnership, 2015; RTS,S Clinical Trials Partnership et al., 2011). Concerns however over overall efficacy and waning protection over time (Olotu et al., 2016) have motivated considerations about potential usage of the vaccine as a seasonal vaccine in areas where malaria transmission is highly seasonal. Indeed, recent work has highlighted the substantial reduction in annual malaria burden that can be achieved by pairing delivery of RTS'S alongside SMC campaigns (Chandramohan et al., 2021). The search for more efficacious, longer lasting vaccinations remains however, with promising recent work including the reported 75% efficacy against clinical malaria for the low dose, pre-erythrocytic malaria vaccine R21 (Datoo et al., 2021).

Methods of Malaria Control ii. Anti-Vectorial Measures Targeting the Mosquito Vector

A number of interventions for the control of malaria are focussed on preventing successful blood-feeding of the anopheline mosquito vectors responsible for onwards transmission, or through direct killing of the mosquito (or both). Insecticide treated bednets (ITNs) provide both a chemical and physical barrier against mosquito feeding and provide a barrier during sleeping against night-time feeding by mosquitoes, which predominates patterns of feeding by many of the key malaria vectors in sub-Saharan Africa (Sherrard-Smith et al., 2019). In doing so, ITNs both provide direct protection to individuals sleeping under the net, as well as indirect protection through preventing uninfected mosquitoes becoming infected during feeding on infected humans. This indirect impact is achieved typically through the effects of mosquitoes coming into contact with the insecticide impregnated bednet material, which typically results in mosquito mortality or sub-lethal effects leading to reduced fitness and transmission potential (Viana et al., 2016). Indoor residual spraying (IRS) involves coating the walls and other surfaces of households with a residual insecticide that either kills mosquitoes that land and rest following feeding, or deters mosquitoes from entering the household (Tangena et al., 2020). The effects of IRS can last months, and in some cases, even years depending on the exact insecticide used (Sherrard-Smith et al., 2018), although difficulties surrounding achieving sufficient coverage and acceptability have meant the impact of IRS is comparatively smaller than those of ITNs (estimated to be responsible for approximately 13% of the reduction in malaria cases observed between 2000 and 2015, compared to 68% for ITNs (Bhatt et al., 2015b)).

Whilst interventions such as ITNs and IRS have successfully reduced malaria burden across sub-Saharan Africa, their effectiveness is increasingly being eroded by the spread of resistance to their killing effects on the mosquito vector (Ranson and Lissenden, 2016; Moyes et al., 2020; Hancock et al., 2018). This killing effect is thought to be one of the largest contributors to the overall impact of ITNs in particular on malaria burden (Killeen et al., 2007; Killeen, 2014). This has motivated development of new ITNs, either with different insecticides or a mixture of the most commonly and traditionally used pyrethroid alongside synergists such as piperonyl butoxide (Gleave et al., 2021). In randomised controlled trials to date, piperonyl butoxide nets have reduced malaria prevalence in the areas they have been deployed significantly more than equivalent pyrethroid-only nets (Protopopoff et al., 2018), and even more recent work evaluating a diverse range of dual-active ingredient ITNs has further underscored their utility over pyrethroid-only bednets (Mosha et al., 2022).

In addition to these more commonly and widely utilised vector control interventions, a number of others exist aimed at mitigating the burden of malaria, at various stages of development and degrees of implementation. These include larviciding, which aims to reduced malaria

transmission by targeting the immature stages (specifically larvae and pupae) of the anopheline mosquito, in doing so, reducing the proportion of mosquitoes that successfully reach adulthood and are available to contribute to malaria transmission (Leslie Choi, 2017). Another recently developed intervention is the attractive targeted sugar bait (ATSBs, (Fiorenzano, Koehler and Xue, 2017)), which involves placing supplies of readily accessible manufactured sugar alternative to the plant sugars mosquitoes typically consume around settlements. The addition of a toxin that rapidly kills mosquitoes upon ingestion or contact leads to death following feeding. ATSBs therefore both suppress the overall mosquito population and reduce the number of mosquitoes living long enough to pass the EIP and be capable of transmitting malaria onwards. Both field experiments (Müller et al., 2010; Traore et al., 2020) and modelling work (Marshall et al., 2013; Fraser et al., 2021) has highlighted the potential impact of ATSBs, which have the ability (unlike ITNs or IRS) to target outdoor-feeding mosquitoes, a source of substantial residual malaria transmission across sub-Saharan Africa (Sherrard-Smith et al., 2019; Musiime et al., 2019; Sougoufara, Ottih and Tripet, 2020). There is also increasing interest in the use of genetic control technologies primarily to either reduce population sizes or replace existing populations with vectors unable to transmit disease (Marshall and Taylor, 2009; Wang et al., 2021). These include approaches centred around the use of the *Wolbachia* bacterium or transgene-based approaches (often gene-drive based (Alphey et al., 2020)), typically either aiming to induce cytoplasmic incompatibility (Adams et al., 2021; Walker et al., 2021) or reduce vector susceptibility to infection (Gomes et al., 2017; Dong, Simões and Dimopoulos, 2020).

Heterogeneity and Variation In Epidemiology and Transmission Dynamics: Implications for Control and Elimination

Malaria epidemiology and its transmission dynamics is characterised by extensive heterogeneity across a range of spatial and temporal scales. This heterogeneity is underpinned by factors relating to the human host (e.g. genetic variation influencing susceptibility), the parasite (e.g. the predominance of different *Plasmodium* species and differences in their life-history), the mosquito vector (e.g. their bionomics and behaviours) and the environment (e.g. annual patterns of rainfall). Together, diversity and variation in these factors leads to marked differences in the epidemiology and transmission dynamics of malaria across different settings – differences that have material consequences for interventions aimed at combatting the disease. The next section of this thesis Chapter describes some of the main factors contributing to heterogeneity in malaria dynamics and epidemiology, as well as the implications of this heterogeneity on the control efforts required to effectively mitigate the disease's public health impact.

Sources of Heterogeneity i. Factors Relating to the Parasite

The predominant parasite responsible for malaria in a setting has a significant influence on the efforts and interventions required for control. Globally the distribution of the *Plasmodium* parasites responsible for causing malaria is highly heterogeneous. Whereas *Plasmodium falciparum* is the dominant parasite across sub-Saharan Africa, across Central and South America, the (more) limited malaria burden is typically associated with *Plasmodium vivax* infection. By contrast, the two parasites are similarly prevalent across South-East Asia, although there is marked heterogeneity within the region, with India accounting for over 50% of the global *Plasmodium vivax* burden and Papua New Guinea having the highest levels of *Plasmodium vivax* transmission in the world as measured by prevalence by light microscopy (Gething et al., 2012).

Plasmodium falciparum and *Plasmodium vivax* possess different lifecycles which pose specific challenges for their control – this is perhaps most notably seen with the formation of hypnozoites (dormant liver stages of the parasite formed during primary infection) during *Plasmodium vivax* infection. These hypnozoites are able to reactivate in the weeks to years after initial clearance of the primary blood stage infection (White, 2011; Battle et al., 2014), leading to relapse of the disease. This phenomenon makes control (and elimination) of *Plasmodium vivax* particularly challenging and indeed, trials exploring potential new anti-malarial formulations for treatment of the parasite have highlighted that as many as 80% of new blood-stage infections are attributable to relapses (Robinson et al., 2015). Whilst work has highlighted the significant impact on malaria transmission associated with both *Plasmodium falciparum* and *Plasmodium vivax* following increased coverage of ITNs (Koepfli et al., 2017), existence of hypnozoites necessitates the use of drugs belonging to the 8-aminoquinolines (8-AQ) such as Primaquine, and recent work has highlighted the likely insufficiency of control efforts that do not also involve targeting hypnozoites for treatment (White et al., 2014). Use of Primaquine involves a lengthy treatment regimen however and can cause severe haemolysis in glucose-6-phosphate-dehydrogenase (G6PD) deficient individuals (Ramos Júnior et al., 2010; Ashley, Recht and White, 2014), further complicating targeting and treatment of individuals compared to *Plasmodium falciparum* infections. Though sharing some common features, the epidemiology of *Plasmodum falciparum* and *Plasmodium vivax* malaria are therefore marked by some important differences. Previous work has highlighted extensive co-endemicity of *Plasmodium falciparum* and *Plasmodium vivax*, though significant heterogeneity in comparative proportions (Price et al., 2020) – given the differences in epidemiology and treatment described above, heterogeneity in the species composition of malaria-causing parasites across settings is likely to have material consequences for how best to control and combat the disease.

In addition to between species considerations, heterogeneity in the tools required for treatment and control of malaria can also arise from within-species genetic variation. Previous work has

documented variation across locations in the asexual blood stage multiplication rate of *Plasmodium falciparum* thought to be underpinned by genetic variation (Murray et al., 2017) as well as extensive fine-scale genetic variation in the genes encoding the parasite's surface coat and its role in partially evading acquired immunity (Day et al., 2017). However, the most notable examples of the impact of parasite genetic variation on the epidemiology of malaria has been the selection for and development of resistance against many of the most commonly used anti-malarials. To date, resistance has emerged against all commonly used anti-malarials, including to many of the frontline artemisinins and partner drugs, leading to the failure of ACTs against *Plasmodium falciparum* in several settings across South-East Asia (Haldar, Bhattacharjee and Safeukui, 2018). This resistance has for the most part been underpinned by variation in the *Plasmodium falciparum* K13 (PfKelch13) propeller domain (Ashley et al., 2014). Currently, the global distribution of ACT resistance is heterogeneously distributed globally, being concentrated in South-East Asia and comparatively less common elsewhere (Kagoro et al., 2022). Indeed, there has been limited evidence of its presence in sub-Saharan Africa to date, with slow clearing of infections following ACT treatment (a proxy for resistance) previously observed at frequencies of <1% prior to 2015 (WWARN Artemisinin based Combination Therapy (ACT) Africa Baseline Study Group, 2015), though more recent work has highlighted an increasing prevalence of such mutations in Rwanda (Uwimana et al., 2020). Patterns of resistance and variation in the degree and extent of resistance to different anti-malarials (including ACTs as described above, but also note the observed variation for SP resistance, which is common in East Africa and comparatively rarer in West Africa (Okell, Griffin and Roper, 2017)) therefore has important implications for what constitutes the most appropriate tools for the treatment of malaria in different settings.

Sources of Heterogeneity ii. Factors Relating to the Human Host

Variation in human genetic features also contributes to heterogeneity in malaria epidemiology across settings. Given the dependence on the erythrocyte for replication and proliferation, genetic disorders that affect erythrocyte biology can influence malaria susceptibility (Kwiatkowski, 2005) – these include sickle cell trait (which impairs parasite growth through disrupting and altering haemoglobin polymerisation (Archer et al., 2018)) as well as deficiency of the enzyme G6PD, which is thought to influence the susceptibility of erythrocytes that have been invaded by the parasite to breakdown by endogenous host factors (Ruwende and Hill, 1998). For *Plasmodium vivax*, the parasite's invasion of reticulocytes (a subset of erythrocytes) is dependent on interactions between the *Plasmodium vivax* Duffy Binding Protein (PvDBP) and the Duffy human antigen receptor for chemokines – absence of the receptor (i.e. being Duffy-negative) is thought to provide partial protection from disease (and possibly infection) (Golassa et al., 2020). These specific examples sit alongside mounting evidence of the role of a diversity

of common human host variants relating to erythrocyte function that influence malaria parasite fitness (Ebel et al., 2021) and the propensity for individuals to suffer from severe disease following infection (Timmann et al., 2012).

In addition to these genetic factors, a variety of other non-genetic factors relating to patterns of behaviour, exposure and immunity across populations, as well as comorbidities, coinfections and health inequities also leads to diversity in malaria transmission dynamics. Infectiousness and the degree to which individuals contribute to onwards transmission is highly heterogeneous, a phenomenon underpinned by extensive between-individual variation in both exposure to (infective) mosquito bites and their comparative infectiousness to mosquitoes. Previous work has highlighted the extensive heterogeneity in exposure to malaria that can occur even amongst individuals residing in the same community (where similar overall patterns of mosquito contact occur on average) (Rodriguez-Barraquer et al., 2016). Studies carried out across Burkina Faso and Kenya have highlighted that adults tend to receive more mosquito bites than children (Gonçalves et al., 2017), whilst other work conducted in Uganda has identified extensive overdispersion in the number of mosquito bites received within age-groups, with the degree of overdispersion negatively correlated with overall malaria transmission intensity (Cooper et al., 2019b). Infectiousness to mosquitoes is similarly highly heterogeneous between individuals – previous work has highlighted significant variation with age in both the likelihood of infections carrying gametocytes, and the gametocyte densities associated with infection (Coalson et al., 2016). Patterns of gametocyte density are strongly correlated with whether or not the infection is “submicroscopic” (Slater et al., 2019), with submicroscopic infections typically less infectious than infections which are microscopically detectable (Slater et al., 2019). There is extensive variation between settings in the size of the submicroscopic reservoir (i.e. what proportion of the infected population harbour submicroscopic infections). Previous work has highlighted that the size of submicroscopic reservoir is negatively correlated with overall levels of transmission, and that submicroscopic infections tend to predominate in low-transmission settings (Okell et al., 2012; Lin, Saunders and Meshnick, 2014), but also that much of the empirically observed variation in the prevalence of submicroscopic infection is unexplained by overall transmission levels (suggesting the existence of other factors influencing the size of the submicroscopic reservoir). Recent work has also highlighted extensive heterogeneity and individual-level variation in susceptibility to *Plasmodium falciparum* malaria, and by extension the rate of acquisition of clinical protection, as well as extensive variation in the rates at which individuals (particularly children) acquire immunity to malaria (Valletta et al., 2022). Together, these factors produce extensive individual-level variation within communities regarding who is infected, who is likely to be detected, and who is likely to be the most infectious. These are factors that have consequences for the viability and utility of different types of control efforts, such as those targeting whole communities (e.g. mass-drug administration) or more targeted approaches (e.g.

focussed on risk or age groups or based on symptoms such as mass-screen and treat based approaches).

In addition to this variation between individuals, malaria is also highly spatially heterogeneous, with this heterogeneity manifesting at a range of spatial scales. Within communities, extensive clustering of malaria infections around individual households has previously been documented (Stresman et al., 2020). This fine-scale variation in malaria risk is also significantly shaped by the patterns and sources of mosquito exposure that predominate in communities. In some instances, the primary site of contact with the mosquito vectors responsible for transmission occurs around the household – the majority of malaria transmission is this context is peri-domestic (Huho et al., 2013; Stresman, Bousema and Cook, 2019). This contrasts with other settings where the primary site of transmission lies away from the household – for example, in the case of forest workers across parts of Southeast Asia, where malaria is concentrated in groups of forest workers who are routinely exposed to bites from the highly efficient, exophagic vector *Anopheles dirus* during the course of their work (Dutta et al., 1996). This variation in who and how people are exposed to infectious bites leads to significant heterogeneity across settings in which populations are most at-risk and where the primary site of transmission is. In turn, this has implications for the appropriateness of different control measures. Where malaria transmission is peri-domestic, interventions targeting the household (such as reactive case detection (Stresman et al., 2020) or indoor residual spraying) are likely to be impactful. By contrast, in other settings where transmission that occurs away from the home predominates (e.g. in Cambodian forests where exposure is primarily related to occupational practices (Rossi et al., 2018b)), clustering of infections in those sharing similar occupational practices potentially necessitates other targeting strategies (such as long-lasting insecticide treated hammocks for forest workers (Thang et al., 2009)).

At a larger spatial scale, and in addition to the factors described above, malaria transmission dynamics are also extensively influenced by the patterns of spatial connectivity that link settings. These contribute to the transmission of malaria on spatial scales that exceed the limits of mosquito dispersal and influence rates of parasite importations between locations and shape patterns of geographical variation in malaria prevalence on regional scales (Wesolowski et al., 2012), and hold particular relevance in pre-elimination and elimination settings where prevalence is low and an increasing proportion of malaria cases are imported rather than locally acquired (Churcher et al., 2014; Guerra et al., 2019; Raman et al., 2020).

Sources of Heterogeneity iii. Factors Relating to the Mosquito Vector

In addition to factors relating to the parasite and the human host, another significant source of heterogeneity in the transmission dynamics and patterns of malaria transmission is variation in

the dominant vectors responsible for transmission. The global distribution of the approximately 70 *Anopheles* vectors able to transmit human malaria parasites is highly heterogeneous; a feature that results in marked differences in the transmission dynamics and epidemiology of malaria across different contexts and ecologies.

An area of particularly marked anopheline diversity and heterogeneity in the geographical distribution of the primary vectors responsible for malaria transmission is India (Dev and Sharma, 2013). In urban areas, transmission is dominated by *Anopheles stephensi* (Kumar and Thavaselvam, 1992) whereas in rural settings across central, eastern and northern India, transmission is predominantly attributed to *Anopheles culicifacies* (responsible for 65% of malaria cases in rural and peri-urban areas (Wangdi et al., 2016)) and *Anopheles fluviatilis* (Nanda et al., 2000). Across the northeastern region, bordering Bhutan, Bangladesh and Myanmar, the main vectors include *Anopheles minimus* (Dev, Sharma and Hojai, 2009) and *Anopheles dirus* (Prakash et al., 2001). The bionomic properties of these vectors are similarly diverse. *Anopheles culifacies* is primarily zoophilic (Joshi et al., 1988), whilst *Anopheles fluviatilis* displays a wide array of feeding preferences depending on the particular sibling species being considered (Dev and Sharma, 2013). Whilst both *Anopheles culicifacies* and *Anopheles fluviatilis* are predominantly nocturnal feeders and generally endophilic, other vectors, most notably *Anopheles dirus s.l*, can be highly exophilic and exophagic (Obsomer, Defourny and Coosemans, 2007).

This heterogeneity in vector species distributions and their respective bionomic properties and behaviours has important ramifications for malaria dynamics and transmission given differing vector behaviours (e.g. tendencies to bite indoors or outdoors), the impact of interventions aimed at controlling and combatting the disease (such as ITNs or IRS) will be highly variable depending on the ecological context and which particular vector species (or combinations of species) predominate. For example, where biting occurs predominantly outdoors interventions such as ITNs or IRS, which predominantly target indoor feeding and resting, are likely to have limited impact, and require different interventions (such as ATSBs or baited traps) to reduce malaria transmission (Sougoufara, Ottih and Tripet, 2020). Relatedly, in settings where the dominant malaria vectors also feed on animals in addition to humans, treating livestock or domestic animals with insecticidal drugs such as ivermectin are likely to be highly impactful (Chaccour et al., 2018). This bionomic diversity and its influence on intervention impact is in addition to extensive temporal heterogeneity in the profile of malaria risk (described in further detail below), itself underpinned by interactions between vector preferences for different larval habitats, the immediate hydrological environment (and its response to patterns of precipitation), and the broader ecological structure of the setting, which further influences the appropriateness and timing of key control interventions (such as IRS, IPTi or SMC).

This bionomic diversity is further underpinned by genetic variation in the vector, and in particular, its influence on the viability and utility of different insecticides used in malaria control. Across many settings, resistance against the widely utilised pyrethroid-based compounds typically present in ITNs and IRS has developed, with many mosquitoes (at least across sub-Saharan Africa) now at least partially resistant to the compounds (Hancock et al., 2018), findings which have motivated the development of new replacement compounds operating through different mechanisms to kill the mosquito (Sherrard-Smith et al., 2018). Development of this resistance is usually underpinned by selection for genetic variation resulting in upregulation of detoxification enzymes (e.g. esterases or P450 monooxygenases (Edi et al., 2014)), mutations to the site targeted by the insecticides (Kawada et al., 2011), or cuticular thickening (Wood et al., 2010). The degree of pyrethroid resistance (and to other insecticides more generally) present in the local mosquito population will have a significant impact on the expected effect of vector control interventions, given the significant influence on transmission that interventions affecting mosquito lifespan have, and the significant impact that specifically the killing effect of IRS and ITNs have on malaria burden (Protopopoff et al., 2018; Mosha et al., 2022).

Sources of Heterogeneity iv. Factors Relating to the Broader Environment

The epidemiology of malaria also varies extensively depending on the overall levels of local malaria transmission intensity. Patterns of clinical disease associated with malaria infection vary substantially with transmission intensity (Carneiro et al., 2010) – in highly endemic areas, disease burden is typically greatest in infants and young children. By contrast, in areas of lower transmission (where the average age of first infection is later and acquisition of immunity through cumulative exposure is slower), many cases also occur in older children and adults (Griffin, Ferguson and Ghani, 2014; Brasseur et al., 2011; Ceesay et al., 2008; Lalloo, Olukoya and Olliaro, 2006). Indeed, the age distribution of malaria infections appears to adapt relatively rapidly to reflect changes in transmission, with infection profiles typically shifting to older individuals as transmission declines. This phenomenon is observed in both infections in the wider community (Griffin, Ferguson and Ghani, 2014), as well as clinical cases (Brasseur et al., 2011). The resulting age shift has also been shown to influence the composition of disease sequalae, particularly in terms of severe malaria. Because youngest children are most at risk of severe anaemia, shifting the average age of first admission to hospital has resulted in an increase in the proportion of severe malaria cases presenting with cerebral malaria in some settings (O'Meara et al., 2008; Paton et al., 2021). Shifting case distributions depending on overall levels of transmission has important implications regarding the effectiveness of age-targeted control (such as a focus on protecting and preventing disease in children under 5, as in the now widespread SMC campaigns across the Sahel (Baba et al., 2020)), and the need to

consider potential alternative, less targeted options for burden reduction (such as mass-drug administration) or expansion of approaches like SMC to older age-groups (Cissé et al., 2016).

In addition to the overall levels of malaria transmission experienced by individuals in a given setting, there is also frequently extensive temporal heterogeneity in malaria risk over the course of a year, with the degree of this seasonality in malaria transmission varying significantly across locations (Cairns et al., 2012). This seasonality is frequently underpinned by changes to size of the populations of *Anopheles* mosquitoes responsible for malaria transmission, which are highly dynamic and frequently exhibit substantial temporal fluctuations in size that shape the profile of malaria risk. The extent and timing of these fluctuations has important implications for the control of malaria, given that the viability, efficacy and cost-effectiveness of a number of malaria interventions (e.g. SMC (Wilson and IPTc Taskforce, 2011; Ross et al., 2011) or IRS (Pluess et al., 2010)) depends on ensuring appropriate timing of their delivery relative to peaks in risk. Indeed, in settings where malaria transmission is marked by significant seasonal fluctuations in disease risk and transmission is highly seasonal, significant success has been achieved with SMC programmes. Perhaps the most notable example has been across the Sahelian region of West Africa, where administration of prophylactic anti-malarial drugs to children over the course of the rainy season has led to reductions in hospital malaria deaths in the region of 40-60% (Baba et al., 2020), with even more promising results from recent trials combining SMC campaigns with seasonal delivery of the RTS'S vaccine (Chandramohan et al., 2021). Achieving maximal impact from seasonally delivered interventions such as these however is intimately dependent on appropriate timing relative to the seasonal peak in disease risk – the impact and cost-effectiveness of such programs are lower in settings where malaria transmission is more perennial (Selvaraj, Wenger and Gerardin, 2018).

The drivers of these temporal fluctuations are complex and multifaceted, but largely driven by changes in the suitability and habitability of the environment. For example, mosquitos are highly sensitive to the ambient temperature of a setting, with local temperature strongly influencing the development rate of various stages of the vector life cycle (Kirby and Lindsay, 2009), as well as biting rates, adult mosquito and parasite mortality rates, and vector competence (Shapiro, Whitehead and Thomas, 2017). The role of temperature can also impact different mosquito species in different ways, with temperature shown to have a variable, and different impact on the developmental rates of *Anopheles arabiensis* and *Anopheles funestus* (Lyons, Coetzee and Chown, 2013). Rainfall is also an important determinant of transmission intensity in a region, due to the requirement of the early life cycle stages of the mosquito for an aquatic habitat in which to develop. Close relationships have been observed between the occurrence of rainfall and peaks in parasite prevalence (Odongo-Aginya et al., 2005) and disease incidence (Baird et al., 2002), though the influence of rainfall is not strictly linear, instead being determined by a

complex combination of the intensity, duration and level of precipitation an area receives (Fillinger et al., 2004; Koenraadt, Githeko and Takken, 2004). These rainfall dynamics strongly interact with the structure of the local environment to modulate the availability and extent of aquatic habitats suitable for egg laying and larval development. For example, the comparative composition of permanent (typically large bodies of water) and temporary (e.g. small puddles created by recent rainfall) aquatic habitats in a setting following rainfall varies dramatically depending on the local ecology and environment (Majambere et al., 2008). Variability in these factors as well as differences in the productivity (based on nutrient availability) and suitability (extent of pollution, water chemical composition etc) of aquatic habitats for breeding across different ecologies similarly structure the influence of rainfall on mosquito population dynamics and lead to noticeable differences, as has been observed across rural and urban settings with similar patterns of rainfall (Gimnig et al., 2001; Mattah et al., 2017).

The above factors in turn interact with species-specific breeding preferences to further generate and shape variation in transmission dynamics across settings – although all members of the *Anopheles* genus share a requirement for standing water in which to breed, the exact location and requirements for breeding vary extensively. A close correlation between rainfall and peaks in *Anopheles gambiae* populations (Appawu et al., 2004; Okello et al., 2006; White et al., 2011) across African settings has been observed (in-keeping with its preference for transient, rain-fed pools of water in which to breed (Gimnig et al., 2001)). By contrast, previous work has highlighted the preference *Anopheles fluviatilis* frequently displays for streams and surrounding stagnant water as breeding sites (Dasgupta et al., 2018) – such breeding sites are typically unsuitable for breeding periods of heavy rain and instead are most productive following the cessation of rains, during the dry season.

Together then, these climatic factors (as well many others such as humidity (Gray and Bradley, 2005) and patterns of land use) interact with the geographical distribution of different anopheline mosquito species and their specific preferences around breeding to yield extensive variation in the degree, extent and timing of seasonal fluctuations (if any) in malaria risk across settings, a phenomenon which has material consequences for control interventions aimed to malaria control, particularly those that are typically seasonally delivered. This seasonal variation is further shaped by inter-annual fluctuations in the suitability of the environment for malaria transmission (such as inter-annual variability in rainfall across East Africa (Pascual et al., 2008) or the size of monsoon experienced by northwest India (Cash et al., 2013)) to further modify and alter temporal variation in malaria risk across settings.

The above relates primarily to rural settings, but urban centres also represent sites of malaria transmission (in some contexts). Across sub-Saharan Africa, entomological inoculation rates and malaria transmission tends to be lower in urban settings compared to rural ones (Doumbe-

Belisse et al., 2021; Robert et al., 2003) – this phenomenon is thought to arise from differences between settings in the types of housing that predominate (Trape and Zoulani, 1987; Killeen et al., 2019) and the more limited availability of viable water sources for the vectors that dominate transmission across the continent (De Silva and Marshall, 2012; Awolola et al., 2007; Kasili et al., 2009). By contrast, urban malaria transmission is a common phenomenon across much of South Asia, due to the presence of the highly efficient urban malaria vector *Anopheles stephensi*, which is able to transmit both *Plasmodium falciparum* and *Plasmodium vivax* effectively, and whose ability to thrive in urban settings is suggested to be underpinned by an increased tolerance for breeding in polluted water sources (Batra et al., 2001), and a superior ability to utilise the manmade hydrological habitats present in urban settings (Thomas et al., 2016; Kumar and Thavaselvam, 1992). Concerningly, the vector has recently been imported to sub-Saharan Africa (having first been identified in Djibouti City in 2012 (Faulde, Rueda and Khaireh, 2014) and is now present across the Horn of Africa), with its further establishment and proliferation potentially threatening urban centres across the continent (Feachem et al., 2019).

Sources of Heterogeneity v. Factors Relating to Ongoing Malaria Control

The above primarily relates to how heterogeneity in various factors between settings can influence which malaria control interventions are most suitable and how they should be delivered to achieve maximum impact. It is however also important to note that the epidemiology and transmission dynamics of malaria in the same setting are highly dynamic and can change over longer time-periods as transmission declines due to control efforts (Cotter et al., 2013). Thus, what constitutes the most effective and impactful forms of control at a single location will be intimately shaped by current transmission levels and the history of control efforts that have preceded them. Though the exact way in which the epidemiology of malaria changes as transmission declines is often driven by setting-specific factors, previous work has highlighted a number of common patterns which have material consequences for intervention impact.

For example, previous work has highlighted the shifting composition of the infectious reservoir and a declining detectability of infections by conventional diagnostics like microscopy as transmission declines (Okell et al., 2012). Submicroscopic infections frequently predominate in low transmission settings and although they typically possess lower gametocyte densities than microscopically-detectable infections (Slater et al., 2019), can harbour gametocytes and so still contribute to onwards transmission. Given this, the comparative importance of control efforts explicitly targeting submicroscopically infected individuals is likely to vary significantly with current levels of transmission, and has material implications for the comparative impact of approaches targeting entire populations (such as mass-drug administration (Brady et al., 2017))

versus only targeting those with detectable malaria infections (such as mass-screen and treat (Kim et al., 2021)).

Another important way in which malaria epidemiology changes as settings progressively control and reduce transmission is related to imported cases. In settings approaching malaria elimination where transmission has significantly reduced, imported malaria cases become increasingly important and represent a significant threat to achievement and maintenance of elimination (Le Menach et al., 2011). Indeed, importation is thought to have contributed to resurgences of malaria in settings trying to achieve elimination such as Zanzibar (Smith et al., 2011), as well as those that have previously successfully achieved elimination such as with the reintroduction of *Plasmodium vivax* to Greece (Danis et al., 2011). Together, these highlight the changing approach to prioritising and targeting imported malaria cases that is required as transmission declines and settings approach and aim for elimination. Declines in malaria transmission also result in significantly increased spatial heterogeneity and clustering of cases (Bousema et al., 2010, 2012). This increased focality of malaria infections as transmission declines requires extensive adaptation by control programmes to effectively target remaining parasite reservoirs (potentially involving use of active-case detection methods to identify infected individuals (Sturrock et al., 2013b)) whilst recognising that overall reductions in malaria burden may lead to reduced prioritisation of the disease and motivate the use of potentially more cost-effective, less resource-intensive approaches to control such as reactive-case detection (Stresman et al., 2020).

Control efforts leading to reductions in burden can also perturb and alter the entomological dynamics that underpin malaria transmission. Previous work from Zambia has highlighted how suppression of transmission using vector control efforts (such as ITNs and IRS) targeting predominantly indoor feeding and resting by mosquitoes can lead to an increasing proportion of onwards transmission being driven by secondary anopheline vectors which are primarily exophilic and less affected by indoor interventions (Gebhardt et al., 2022). Relatedly, even amongst the primary vectors of malaria transmission in sub-Saharan Africa (*Anopheles gambiae, Anopheles funestus* and *Anopheles arabiensis*), outdoor biting and resting has recently been implicated as a significant source of residual malaria transmission that becomes progressively more important as malaria is controlled through traditional vector-control tools (Sherrard-Smith et al., 2019). This work has highlighted the crucial need for vector control approaches targeting outdoor biting anopheline malaria vectors such as host/odour baited traps, resting traps or ATSBs (Sougoufara, Ottih and Tripet, 2020), which are likely to prove particularly crucial in low transmission settings where effective control through ITNs and IRS has led to an increased proportion of transmission occurring through exposure outdoors (Russell et al., 2011).

These changes in patterns of exposure to infectious mosquito bites can also lead to the emergence of new risk groups (frequently adult men), as the comparative importance of peri-domestic (i.e. exposure occurring immediately occurring around the home) and occupational exposures (typically away from the home) to vectors changes (Chuquiyauri et al., 2012). Whilst peri-domestic exposure is common in the high transmission settings across sub-Saharan Africa, in numerous other (typically low-transmission settings), exposure frequently occurs away from the home (e.g. across the forests of the Peruvian Amazon (Saavedra et al., 2019) and Cambodia (Rossi et al., 2018b)). In these contexts, the utility of household-based ITNs and IRS is likely to be limited and previous research has highlighted the crucial need for other tools such as insecticide-treated hammock nets (Thang et al., 2009) or targeted chemoprophylaxis to forest-goers (von Seidlein et al., 2019a) in order to effectively control malaria in these settings.

Thesis Aims

The examples presented above highlight the diversity of different ways in which the epidemiology of malaria can vary – either between settings or as settings progressively control and reduce transmission. This variation drives heterogeneity across and between settings in where malaria is concentrated, how malaria is transmitted and who is most at-risk. In turn, these changes have material consequences for what constitutes the most impactful control interventions to be utilising. For example, the highest impact interventions in high transmission settings (where the aim is burden reduction) are likely to be significantly different to those required in low transmission settings (where goals may be centred around achieving elimination). Together, this underscores the crucial need to better understand and quantify the factors underlying heterogeneity in malaria epidemiology and transmission dynamics, in order to better tailor control efforts to the different and setting-specific eco-epidemiological contexts driving malaria burden.

Motivated by the above, the aim of this thesis is to use a combination of statistical and mathematical modelling to further our understanding of variation in the epidemiological and entomological determinants of malaria transmission and dynamics, and explore the implications of this variation for control efforts, particularly for the appropriateness of different interventions in different settings. In Chapter 2, I carry out a systematic literature review and meta-analysis exploring the prevalence of submicroscopic malaria infections across a diverse range of settings globally, with an emphasis on understanding how and why the size of the submicroscopic reservoir varies so much across different settings, and the implication this heterogeneity might have for malaria control efforts. In Chapter 3, I continue my focus on the human host and characterise the degree of fine-scale spatial clustering of malaria infections at the household level across a diverse range of sub-Saharan African settings. I explore how the extent of this

clustering varies depending on overall levels of malaria transmission, and what this heterogeneity means for the appropriateness of reactive vs proactive based infection detection strategies. Chapter 4 shifts the focus away from variation in the spatial heterogeneity focussed and human infections and turns instead to variation in the temporal dynamics of the anopheline vectors responsible for sustaining malaria transmission. This Chapter focuses on developing a better understanding of the annual fluctuations in patterns of mosquito abundance and develops a statistical framework to characterise the drivers underlying the diverse seasonal patterns often observed across populations of different mosquito species, with a particular focus on what a better understanding of these dynamics can facilitate with regards to the appropriate timing of seasonally delivered malaria control interventions such as SMC or IRS. Chapter 5 extends the framework developed in Chapter 4 and applies it to the vector *Anopheles stephensi* (a highly efficient urban malaria vector typically endemic to South Asia and the Middle East) in order to better understand the consequences of its recent establishment in the Horn of Africa. Integrating these results into a previously developed model of malaria transmission, I explore the potential public health impact and consequences of diversity in its temporal dynamics on the viability and suitability of different IRS compounds. Finally, in Chapter 6, I summarise and discuss the results presented across the course of the thesis, consider key findings and limitations, and suggest potential avenues and directions for future research.

Chapter 2 Global Patterns of Submicroscopic *Plasmodium falciparum* Malaria Infection: Insights from a Systematic Review and Meta-Analysis of Population Surveys

Accurate ascertainment of malaria infection represents a crucial component of malaria control, and yet mounting evidence has highlighted the often-substantial prevalence of infections with parasite densities lower than the threshold of detection by microscopy (so called "submicroscopic" infections). The drivers of these infections remain uncertain, despite their established relevance to onwards transmission. Using a Bayesian regression modelling approach in tandem with a systematic review of the literature, in this Chapter I explore a number of human host- and setting-specific factors and examine their association with submicroscopic malaria infection. Using these results, I explore how variation in these factors shapes heterogeneity globally in the size and extent of the submicroscopic reservoir and the implications of this heterogeneity on the utility of specifically targeting submicroscopic malaria infections in near-elimination settings.

Introduction

The ability to accurately detect malaria infection during population surveys is a cornerstone of effective surveillance and control of the *Plasmodium* parasite. Routinely, malaria detection is undertaken using microscopy of blood films or rapid diagnostic tests, although in recent years there has been an increase in the use of more sensitive molecular methods in research contexts. These techniques (typically PCR based) (Snounou et al., 1993) have revealed the widespread presence of infections with parasite densities lower than the threshold of detection by routine methods such as microscopy (Lamptey et al., 2018; Zhou et al., 2016; Mueller et al., 2009). Such submicroscopic infections are present across a range of different settings and populations (Tadesse et al., 2017; Steenkeste et al., 2010). Although rarely causative of severe symptoms, these infections have been associated with some adverse outcomes during pregnancy (Cottrell et al., 2015) and in children younger than 10 years (Katrak et al., 2018).

These infections are also relevant to public health because of their potential to be transmittable, despite being undetectable by conventional diagnostics. Although typically characterised by lower parasite densities and infectivity than microscopically detectable infections (Slater et al., 2019), individuals with submicroscopic infections frequently harbour gametocytes (the transmissible form of the parasite) and can contribute to onwards transmission of malaria. Individuals with submicroscopic infections have been shown to contribute to transmission across areas of high (Lin Ouédraogo et al., 2015) and low (Tadesse et al., 2018) transmission intensity, as well as seasonal (Ouédraogo et al., 2009) and perennial (Schneider et al.,

2007) settings, underscoring the potential relevance of this infection subgroup to malaria control efforts.

Despite this potential relevance to malaria transmission, our understanding of the factors influencing the size of the submicroscopic reservoir remains far from complete. Previous reviews have found that microscopy misses on average, half of all *Plasmodium falciparum* infections compared with PCR-based methods in cross-sectional surveys (Okell et al., 2009) and that adults are more likely to harbour submicroscopic infections than children (Okell et al., 2012). However, these reviews also identified extensive unexplained variation in the size of the submicroscopically infected population across settings, suggesting the existence of other important factors that determine the size of the reservoir. For example, although the extent of submicroscopic infection is highly heterogeneous across different locations (Tadesse et al., 2017; Idris et al., 2016), it remains unclear whether this represents systematic variation according to geographical location or is reflective of other underlying location-specific characteristics.

Resolving these gaps in our understanding of submicroscopic epidemiology has material consequences for the future of malaria control. Given that submicroscopic infections can contribute to malaria transmission, understanding when and where they are most prevalent is vital to the control of the disease. Whilst recent years have seen increases in transmission across a number of settings, elimination remains a target for a number of countries (World Health Organization, 2021). Low-transmission settings (such as those aiming for elimination) can have high proportions of submicroscopically infected individuals (Okell et al., 2012; Lin, Saunders and Meshnick, 2014). Understanding the prevalence, detectability, and infectiousness of low-density infections in these settings will be essential for planning for the elimination of malaria: is there benefit in detecting and treating such infections, or are resources better spent elsewhere? Improving our understanding of the drivers of submicroscopic infection in these areas is, therefore, crucial to better define when and where submicroscopic infection is likely to occur and how the size of the submicroscopic reservoir is likely to change as areas approach elimination.

Here I update previous reviews on submicroscopic malaria infection prevalence (Okell et al., 2009, 2012), leveraging the increase in the usage of molecular methods over the past decade to explore novel determinants of submicroscopic infection prevalence. These determinants include geographical location and historical patterns of transmission, diagnostic methodology, seasonality, and the role of age at a finer resolution than previously possible. These results are then integrated with estimates of the infectivity of submicroscopic individuals to mosquitoes to estimate their contribution to malaria transmission across a range of different settings.

Methods

Search strategy and selection criteria

A systematic review and meta-analysis of available data on submicroscopic malaria prevalence was carried out. Cross-sectional malaria prevalence data in which both microscopy and PCR based methods had been used to determine infection were compiled, updating a previous review published in 2012(Okell et al., 2012). PubMed and Web of Science were searched using the terms “PCR” OR “Polymerase Chain Reaction” AND “falciparum” from Jan 1, 2010, (i.e., the end date of the previous systematic review (Okell et al., 2012)) to Oct 11, 2020. Only studies in English were included. Studies reporting asexual *Plasmodium falciparum* prevalence by microscopy and PCR in the same population were included. See **Figure 2.1** for details of retained and excluded records at each stage of review. Surveys of pregnant women, in which participants were chosen on the basis of symptoms or treatment, or which did not involve a population from a defined location were excluded. Submicroscopic infections were defined as where infection was detectable by PCR but not by microscopy. Specificity of microscopy compared to PCR is high (average 98·4% (Okell et al., 2009)), and so it was assumed that microscopy-positive individuals are also PCR-positive.

Data extraction

From each study, I extracted information on the number of individuals tested by PCR and microscopy, as well as the number of tested individuals positive for *Plasmodium falciparum* malaria by each method. I also extracted data on the exact diagnostic methodologies used (specifically, the PCR method used and number of microscopy slides scanned) and characteristics of the survey location and timing (the global region, country, specific location, and sampling season). Where available, information on the age range of survey participants were also extracted.

ANOVA and Tukey's honest significant difference

Data were analysed using an ANOVA based approach to assess differences in the mean prevalence ratio (defined as the proportion of PCR positive infections also detectable by light microscopy) and the factors underlying these differences. Data were weighted according to the cross-sectional survey sample size and controlling for the PCR prevalence recorded in the survey. Tukey's honest significant difference (HSD) test was used to post-hoc examine pairs of

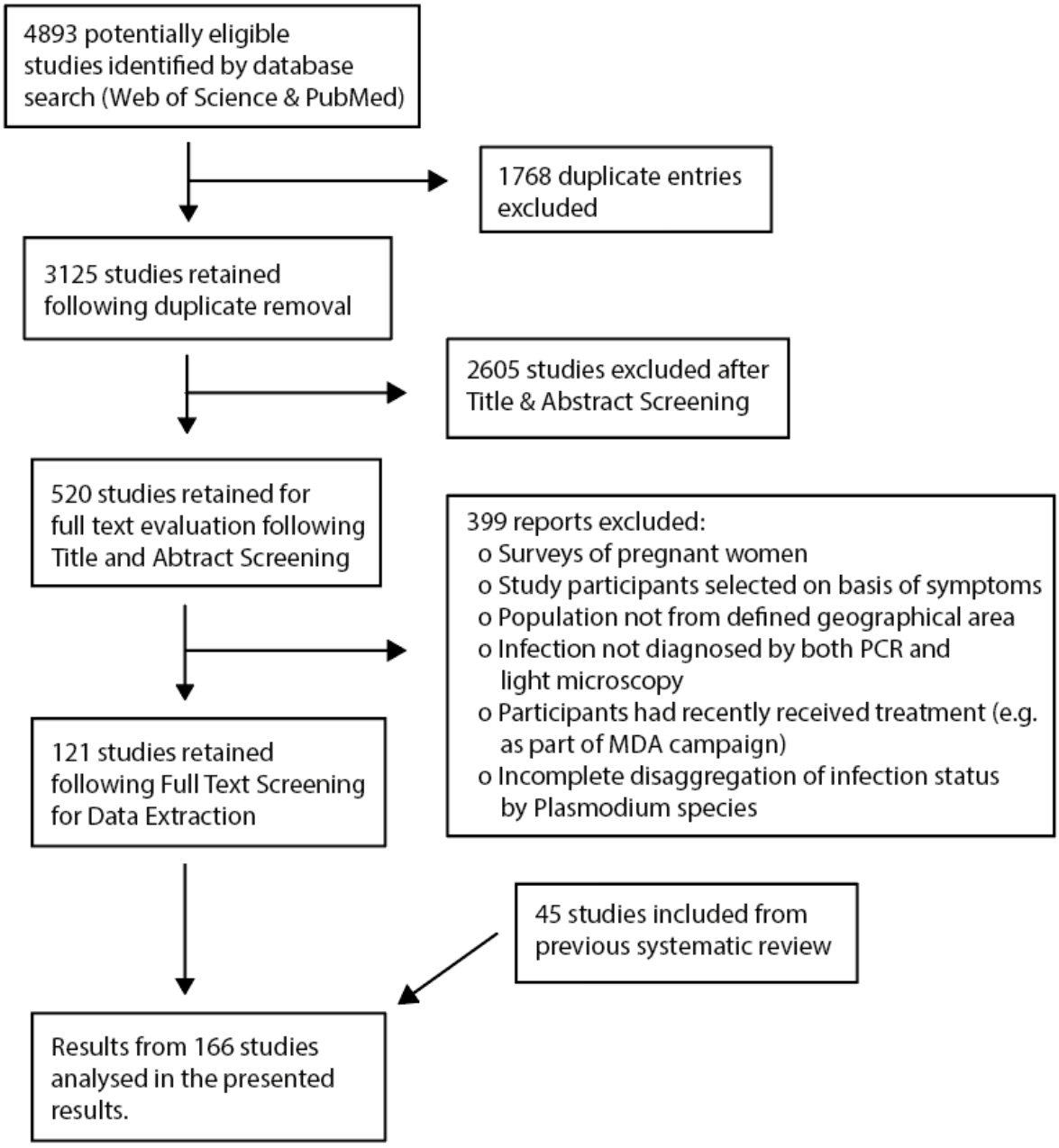


Figure 2.1: Systematic review overview, workflow and selection of eligible studies. Searches for malaria prevalence data where infection status had been determined using both microscopy and PCR based methods were carried out using a systematic review, updating a systematic review last conducted in 2017. A total of 121 studies were newly identified in the update. Alongside 45 studies identified in the previous systematic review, this gave a total of 166 studies included in the formal analyses presented here.

factors for significant differences in the mean prevalence ratio. These analyses were done using R statistical software, version 4.0.2.

Log-linear regression model formulation

In line with a previous review (Okell et al., 2012), data were also analysed using a Bayesian log-linear regression-based framework to estimate microscopy prevalence and the prevalence ratio as a function of PCR prevalence. We define the following $\mathrm{p_{LM,i}}$ as the underlying prevalence of microscopically detectable malaria in survey i, and $\mathrm{p}_{PCR,i}$ as the underlying prevalence of PCR-detectable malaria in survey i. We then define the following:

$$\mathrm{LM_i} = \ln\left(\frac{\mathrm{p_{LM,i}}}{1 - \mathrm{p_{LM,i}}}\right)$$

$$\mathrm{PCR_i} = \ln\left(\frac{\mathrm{p_{PCR,i}}}{1 - \mathrm{p_{PCR,i}}}\right)$$

i.e. where $\mathrm{LM_i}$ is the log odds of microscopy prevalence in survey *i*, and $\mathrm{PCR_i}$ is the log odds of PCR prevalence. Per the methodology described by Sharp and Thompson (Sharp and Thompson, 2000), we define the following equation to relate $\mathrm{LM_i}$ and $\mathrm{PCR_i}$:

$$\mathrm{LM_i} = \mathrm{PCR_i} + \delta_\mathrm{i}$$

where δ_i is the log odds ratio of microscopy to PCR prevalence. δ_i is defined as:

$$\delta_\mathrm{i} = \delta'_\mathrm{i} + \beta_0(\mathrm{PCR_i} - \overline{\mathrm{PCR}})$$

where $\overline{\mathrm{PCR}}$ is the log odds of the mean survey PCR prevalence, β_0 is a coefficient shared across all surveys describing how log odds of microscopy prevalence varies with log odds of PCR prevalence, and δ'_i is a survey-specific intercept coefficient describing the effect of PCR prevalence on microscopy prevalence. This model structure allows δ_i to vary between surveys, with β_0 controlling the extent of this variation. In addition to this logit-linear model, I also explored a range of different model structures representing relationships between microscopy and PCR prevalence of varying flexibility, to assess their capacity to fit to this newly collated and updated dataset. These were formulated as follows:

$$\mathrm{LM_i} = \mathrm{PCR_i} + \delta_\mathrm{i}$$

where:

$$\delta_\mathrm{i} = \delta'_\mathrm{i} \quad \textbf{(Basic)}$$

$$\delta_\mathrm{i} = \delta'_\mathrm{i} + \beta(\mathrm{PCR_i} - \overline{\mathrm{PCR}}) \quad \textbf{(Linear)}$$

$$\delta_\mathrm{i} = \delta'_\mathrm{i} + \beta(\mathrm{PCR_i} - \overline{\mathrm{PCR}}) + \gamma(\mathrm{PCR_i} - \overline{\mathrm{PCR}})^2 \quad \textbf{(Quadratic)}$$

$$\delta_\mathrm{i} = \delta'_\mathrm{i} + \beta(\mathrm{PCR_i} - \overline{\mathrm{PCR}}) + \gamma(\mathrm{PCR_i} - \overline{\mathrm{PCR}})^2 + \sigma(\mathrm{PCR_i} - \overline{\mathrm{PCR}})^3 \quad \textbf{(Cubic)}$$

Where, as above, $\mathrm{LM_i}$ is the log odds of microscopy prevalence in trial *i* and $\mathrm{PCR_i}$ is the log odds of PCR prevalence in trial *i*. $\overline{\mathrm{PCR}}$ is the log odds of the mean PCR prevalence across surveys. δ_i is the log odds ratio (OR) of microscopy to PCR prevalence, with δ_i' the expected log OR when the log odds of PCR prevalence is equal to the overall mean across trials. β, γ and σ are regression coefficients specifying the extent and nature of how δ_i varies between surveys.

Modelled prevalence values were assumed to be drawn from a binomial distribution with the sample size of the survey as the number of trials (people tested) and the underlying true prevalence as the probability of "success" (malaria positivity) in any given trials:

$$\mathrm{Positive}_{LM,\mathrm{i}} \sim \mathrm{Binomial}(\mathrm{p_{LM,i}}, \mathrm{N}_i)$$
$$\mathrm{Positive}_{PCR,\mathrm{i}} \sim \mathrm{Binomial}(\mathrm{p_{PCR,i}}, \mathrm{N}_i)$$

where $\mathrm{Positive}_{LM,\mathrm{i}}$ and $\mathrm{Positive}_{PCR,\mathrm{i}}$ are the observed number of malaria positive individuals by each diagnostic method, $\mathrm{p_{LM,i}}$ and $\mathrm{p_{PCR,i}}$ re the underlying prevalences according to each diagnostic, and N_i is the sample size for survey *i.*

Bayesian Gibbs Sampling MCMC Based Model fitting

The models described above were fitted within a Bayesian Markov chain Monte Carlo (MCMC) framework, implemented in the probabilistic programming language JAGS (Just Another Gibbs Sampler (Plummer and Others, 2003)) which uses a Gibb-sampling based approach to sample the posterior distribution of Bayesian models. The code implementing the analyses described below is openly available via GitHub at https://github.com/cwhittaker1000/submicroscopic_malaria. Each of the models described above were fitted to the entirety of the collated prevalence data (including both surveys identified in this review as well as those from previous reviews).

The goal of Bayesian model fitting schemes underpinned by MCMC-based algorithms such as these is for the algorithm to generate a set of samples that correspond to a random sample from the posterior distribution of interest – specifically to draw from the posterior density $p(x|y)$ where x represent the model parameters and y the observed data. Given a multivariate posterior distribution, and with $X = (x_1, \ldots, x_n)$ representing a single sample from the joint distribution $p(x_1, \ldots, x_n|y)$ (where y is our observed data and $(x_1, \ldots, x_n)$ are our model parameters), denote the ith sample by $X^i = \left(x_1^i, \ldots, x_n^i\right)$. The Gibbs Sampling algorithm then updates via the following steps:

1. The aim is to generate the next sample, which we will call X^{i+1}. $X^{i+1} = \left(x_1^{i+1}, \ldots, x_n^{i+1}\right)$. Starting from X^i, we sample the parameters in order from $1 \ldots n$ conditioning on the current values of all other model parameters. I.e. if j indexes the parameters, then for

x_j^{i+1} sample from the distribution of that parameter conditional on the current values of all the other parameters sampled so far i.e. $p\left(x_j^{i+1} | \, x_1^{i+1} \ldots, x_{j-1}^{i+1}, \; x_{j+1}^{i}, \ldots, x_n^{i}\right)$.

2. Repeat this step n times (i.e. once for each parameter).
3. Repeat steps 1. and 2. k times, where k is the pre-specified number of MCMC iterations.

Uninformative normal prior distributions were assigned to all parameters in the model. In each instance, 4 chains of 10,000 iterations were run for purposes of model fitting and parameter inference. 5,000 of these iterations were discarded as burn-in, leaving 5,000 iterations from each chain and therefore a total of 20,000 iterations available for inference. This sample was further thinned by selecting only every 10th element in order to minimise auto-correlation, leaving a sample of 2,000 values upon which inference was based. The Gelman-Rubin convergence statistic were monitored in all cases to assess convergence and were all consistently <1.02, indicating stability of the chains and likely convergence to the underlying true posterior distribution.

Model comparison

Model comparison was carried out using the deviance information criterion (DIC), which considers both the capacity of the model to fit the data, as well the model's underlying complexity (Spiegelhalter et al., 2014). It is formulated as follows:

$$\mathrm{DIC} = \mathrm{D}(\bar{\theta}) + 2p_D$$

where $\mathrm{D}(\bar{\theta})$ is the deviance evaluated at the expectation of θ (the vector of parameter values that together specify the model used, so δ_1', β, γ and σ for our purposes) and p_D is the variance of the deviance evaluated across all values of θ. This latter quantity can be considered to be a proxy measure of the effective number of parameters the model contains, and so reflects the underlying complexity of the model. Lower DIC values are preferred and so in doing so, the DIC trades off model fit (as indicated by $\mathrm{D}(\bar{\theta})$) and model complexity (p_D) to enable (in theory) selection of the model best able to extrapolate to new, unobserved data. For the data considered here, a model with a linear relationship linking PCR and microscopy prevalence on the logit scale (the model used in previous reviews) was found to have the lowest DIC and is therefore the preferred model. Based on this, this model was used for all subsequent analyses, though model fits to the data for each of the different formulations are provided in **Figure 2.9**.

Historical and current regional transmission intensity stratification

Surveys done in Africa were geolocated and prevalence estimates (aggregated to the administrative unit 1 level, which represents the highest level of officially delineated area within a specific country) from the Malaria Atlas Project (Bhatt et al., 2015b) (MAP) were used to characterise current and historical transmission intensity of the region that each survey

belonged to. Note that this distinguishes between local malaria transmission (defined by the prevalence recorded in each survey and hereafter referred to as survey PCR prevalence), and malaria transmission at the regional level (reflecting broader patterns of transmission and hereafter referred to as regional prevalence, which represents prevalence averaged at the administrative unit 1 level). This regional-level transmission represents the average of a heterogeneous mixture of higher and lower-transmission areas, and has relevance to local transmission because factors such as human movement patterns and circulating parasite genetic diversity are often similar across nearby settings in the same region, even if transmission levels differ markedly (Tatem and Smith, 2010; Tessema et al., 2019).

Regional transmission levels (both historical and current) were then used to stratify each study into one of three transmission archetypes.

(1) Historically high and currently high refers to areas with historically (defined as 15 years previous to the date of the survey) high transmission intensity (>15% slide prevalence in children aged 2–10 years) and remain so at the time of the survey;
(2) historically high and currently low refers to areas of historically high transmission intensity that have declined in the 15 years previous to the date of the survey to low levels (<15% slide prevalence in children aged 2–10 years); and
(3) historically low and currently low refers to areas with historical and current low transmission (<15% slide prevalence in children aged 2–10 years).

Where MAP estimates were unavailable (dates earlier than 2000), it was assumed that the year 2000 was reflective of historical transmission intensity because the substantial increase in international financing for malaria control occurred from 2000 onwards (approximately a twentyfold increase between 2000 and 2015) (WHO, 2014). Separate Bayesian log-linear regression models were then fitted to data from each transmission archetype to assess the effect of historical and current transmission intensity on the prevalence ratio.

Estimation of contributions to onwards transmission

The results of the regression modelling described earlier (which provides an estimate of the proportion of infections that are expected to be submicroscopic given an estimate of microsopically detectable malaria prevalence) were integrated with estimates of infectivity for submicroscopic and microsopically detectable infections to estimate the potential contribution of submicroscopic infections to transmission across different settings. Estimates of comparative infectivity of microscopically-detectable infections versus submicroscopic infections (hereafter referred to as the infectivity ratio) are variable, ranging from a 2× to a 20× difference (Coleman et al., 2004; Lin et al., 2016). Given this, three scenarios were considered in which

microscopically-detectable infections were 2×, 5×, or 20× more infectious to mosquitoes than submicroscopic infections. Proportional contribution to transmission by submicroscopic infections was calculated as:

$$\frac{(\mathrm{p_{PCR}} - \mathrm{p_{LM}})}{(\mathrm{p_{LM}} * Infectivity\ Ratio) + (\mathrm{p_{PCR}} - \mathrm{p_{LM}})}$$

Where $\mathrm{p_{PCR}}$ and $\mathrm{p_{LM}}$ describe the prevalence of malaria by PCR and microscopy respective; the term $(\mathrm{p_{PCR}} - \mathrm{p_{LM}})$ therefore describes the proportion of individuals positive by PCR and not by microscopy (i.e. the proportion of individuals with submicroscopic infections). The relative infectiousness of submicroscopic infections is set to 1, and the infectivity ratio (either 2, 5, or 20) is a multiplicative factor reflecting the fact that microscopically detectable infections are more infectious. The equation's denominator reflects total onwards transmission occurring within the population, the numerator the amount of transmission attributable to submicroscopic infections. These analyses assume that submicroscopic and microscopically infected populations do not differ in other factors that are likely to influence transmission (eg, age and mosquito exposure).

Results

4893 potentially eligible studies were identified in the systematic review update and 1768 duplicates were excluded, leaving 3125 studies for screening. After screening titles and abstracts for relevance, 520 studies were kept for full-text evaluation, of which 121 were included. These 121 studies, alongside 45 identified during previous systematic reviews (Okell et al., 2012, 2009), yielded 551 datapoints comprising distinct cross-sectional surveys in which surveyed individuals had malaria infection assessed by both PCR and microscopy **(Figure 2.2)**. The number of prevalence survey pairs is greater than the number of included studies because many studies presented results from multiple different locations. 164 of these 551 datapoints were from cross-sectional surveys done in a specific age-group (0–5 years, 6–15 years, and >15 years) and were analysed separately; 387 datapoints were from cross-sectional surveys done in populations that spanned more than one age-group. Across these data (n=387) included in the primary analyses, microscopy detected 44·9% (95% CI 42·0–47·8) of all PCR-detectable infections, although this varied across settings. In a small number of instances where the number of microscopically detected infections was higher than those identified by PCR (n=10), the prevalence ratio was adjusted to 1 (this adjustment does not qualitatively alter the results described here).

I fitted a Bayesian log-linear regression model to this collated data **(Figure 2.2)** and found that the prevalence ratio (defined as the proportion of PCR positive infections also detectable by

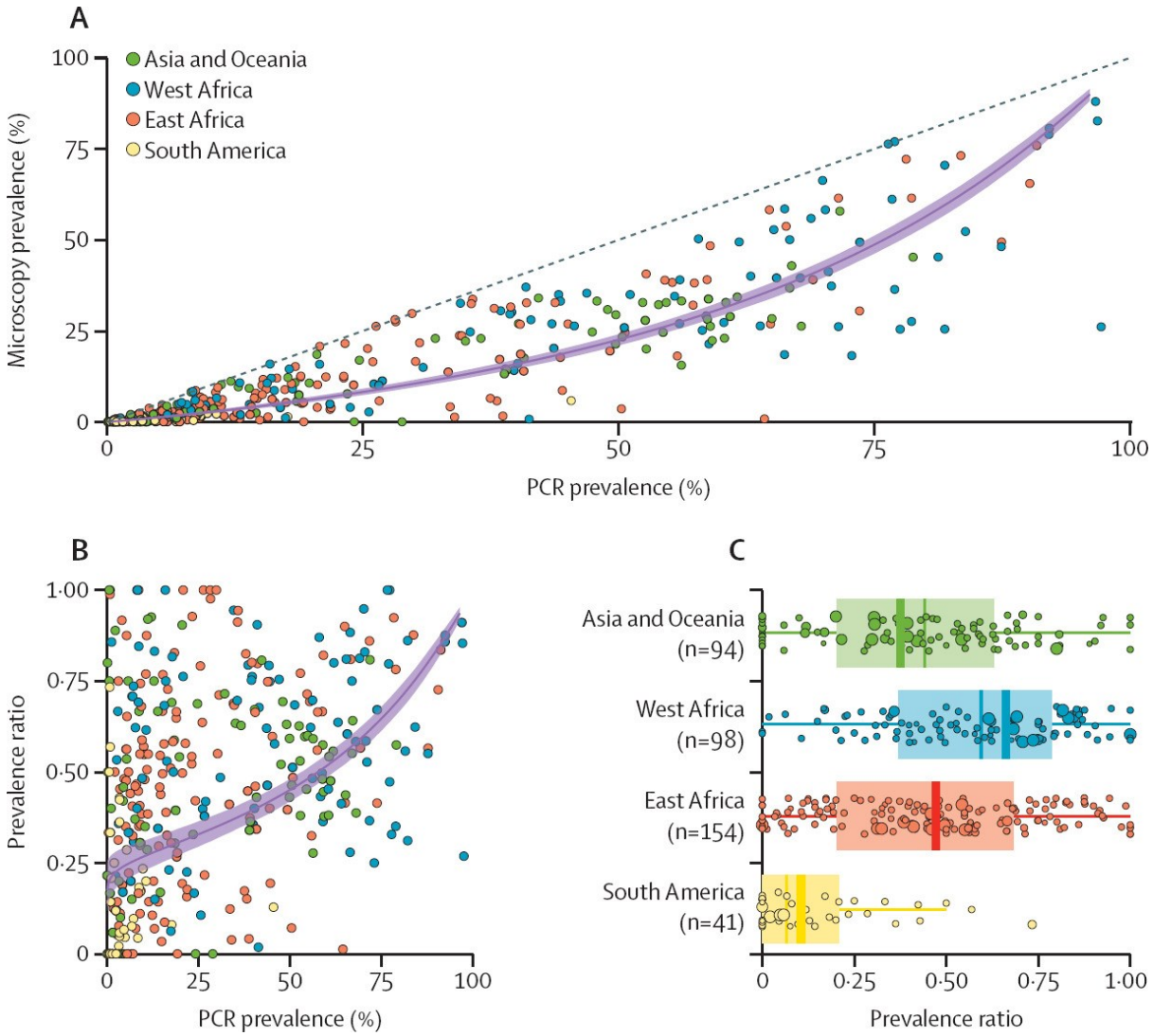


Figure 2.2: Prevalence of infection by PCR *vs* microscopy in 267 prevalence survey pairs and model fits. Bayesian Markov chain Monte Carlo methods were used to fit a linear relationship between PCR prevalence and microscopy prevalence on the log odds scale. (A) 387 microscopy and PCR prevalence surveys were identified in this study and previous systematic reviews. The fitted model relationship (purple line) and the 95% credible interval of the mean (light purple shaded area). (B) The prevalence ratio (ie, the proportion of PCR positive individuals also detectable by microscopy) according to underlying PCR prevalence for each of the 387 survey microscopy–PCR pairs (points) used to fit the full model. The estimated mean prevalence ratio (purple line) and 95% credible interval of the mean (light purple shaded area) are also shown. (C) Box plot of the prevalence ratio disaggregated by global region. For each region, the size of the point reflects the number of individuals tested by microscopy and PCR. Thick coloured bar on the box plot represents the weighted mean prevalence ratio for each global region. Thin line indicates the median, box indicates IQR, and whisker limits span 1·5× the IQR.

microscopy) increased as malaria transmission (measured by survey PCR prevalence) increased, indicating a declining proportion of submicroscopically infected individuals in the settings of highest malaria prevalence. An average of 60–70% of infections were submicroscopic in the areas of lowest PCR prevalence, but only 10–20% were submicroscopic in the highest prevalence areas. I also fitted the model separately to the data collected in previous reviews, and compared it to a model fitted to the data newly collated here. There was no difference in the modelled relationship between PCR and microscopy prevalence when fitted to the previously compiled dataset compared to the data newly collated here **(Figure 2.3)**, and model fit (as assessed by the correlation between the observed and model-predicted microscopy prevalence) was similar (R^2 of ~0.9 for models fitted to both previously collected and newly collated datsets) **(Figure 2.4)**. Some more flexible model structures incorporating non-linear relationships between PCR and microscopy prevalence were also fitted to the data. A log-linear model provided the best overall fit (as measured by the Deviance Information Criterion, DIC) **(Figure 2.5)**. There was a small but significant effect of sampling season after controlling for survey PCR prevalence (ANOVA, df=1, p=0·0017), with submicroscopic infections less common during the wet season than the dry season **(Figure 2.6**). There was also a significant effect of PCR methodology (ANOVA, df=4, p=0·038) **(Figure 2.7)** with the prevalence ratio marginally lower in surveys using quantitative PCR (qPCR) and RT-PCR to determine infection status. Scanning a higher number of microscopy fields to determine the presence or absence of infection was also significantly associated with the prevalence ratio increasing (ANOVA, df=1, p=0·0053).

Grouping surveys by global region (west Africa, east Africa, South America, and Asia and Oceania) revealed marked geographical variation in the prevalence ratio (ANOVA, p<0·0001, df=3), being lower in South American surveys than all other regions (Tukey's HSD, p<0·0001 for all pairwise comparisons) and higher in west African surveys than all other regions (Tukey's HSD, p<0·0001 for all pairwise comparisons) **(Figure 2.2)**. To examine these differences in more detail, a separate Bayesian log-linear model was fitted to the data from each global region and the modelled prevalence ratio across the range of transmission intensities found in each assessed **(Figure 2.8 A–D)**. These results revealed that the prevalence ratio in surveys from South America was lower (i.e., more infections were submicroscopic) than would be expected based on their respective transmission intensities alone, and consistently lower than all other global regions **(Figure 2.8 E)**. Across all settings, nested PCR predominated as the methodology used, although South America had higher levels of qPCR usage than other settings. No significant variation in which microscopy methodology was used between regions was observed **(Figure 2.9)**.

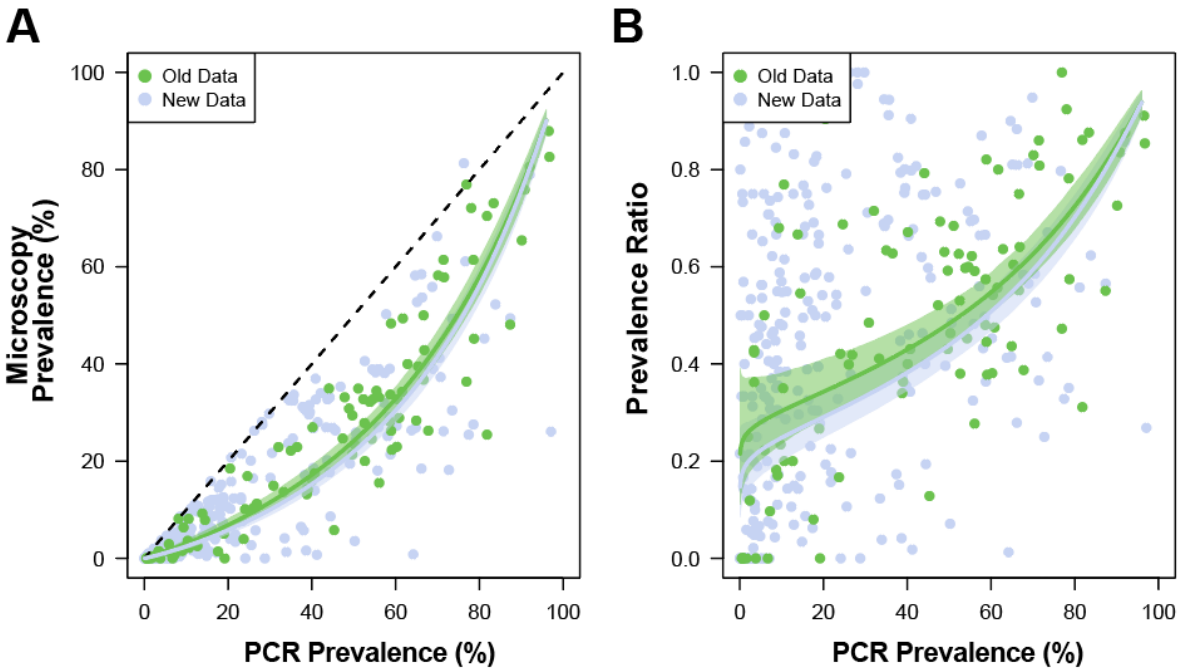


Figure 2.3: Prevalence of infection by PCR versus microscopy and model fits for previously collated data and data newly identified as part of this review. Bayesian Markov chain Monte Carlo methods were used to fit a log-linear relationship between PCR prevalence and microscopy prevalence separately to data collated during previous reviews on submicroscopic malaria infections (n = 100, green dots) and data newly identified as part of this review (n = 287, pale purple dots). **(A)** Microscopy and PCR Prevalence data from surveys, with the fitted model relationship (green and pale purple lines) and the 95% credible interval of the mean (shaded areas around each line). **(B)** The sensitivity of microscopy (defined here as the proportion of PCR positive individuals also detectable by microscopy) according to underlying PCR prevalence for each of the survey microscopy-PCR pairs used to fit the full model. For each dataset, the estimated average sensitivity (coloured line) and 95% credible interval of the mean (shaded area) also shown.

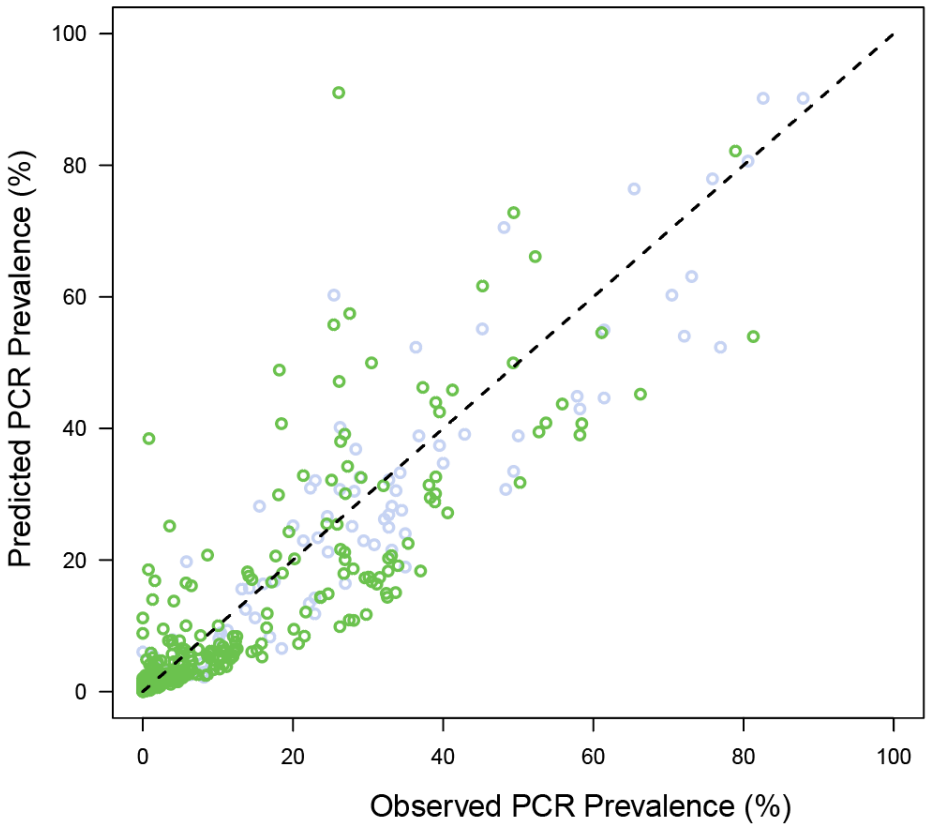
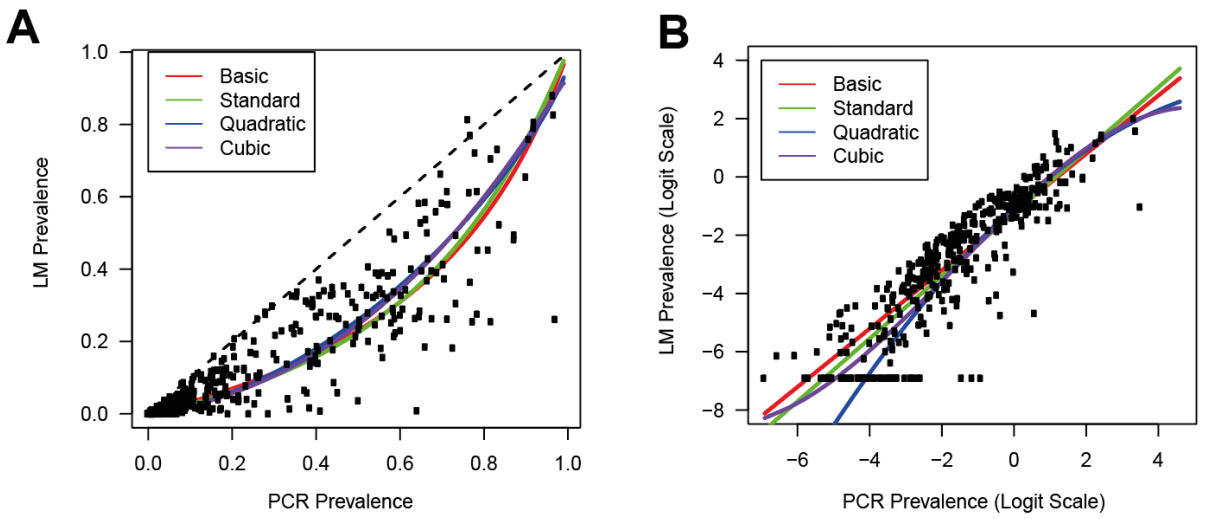


Figure 2.4: Comparison of Empirically Observed Microscopy Prevalence and Microscopy Prevalence Predicted by Bayesian Regression Modelling. Bayesian Markov chain Monte Carlo methods were used to fit a linear relationship between PCR prevalence and microscopy prevalence on the log odds scale. For each survey (green for newly identified surveys as part of this systematic review, grey for surveys identified in previous systematic reviews), the empirically observed and modelled microscopy prevalence were plotted and compared. The correlation between the observed and model-predicted microscopy prevalence across the surveys was 0.90 (measured via R^2, the correlation coefficient).



C

Model Structure	Model Equation	DIC
Basic	$\mathrm{LM_i} = \mathrm{PCR_i} + \delta_i$	5377
Standard (Linear)	$\mathrm{LM_i} = \mathrm{PCR_i} + \delta_i + \beta(\mathrm{PCR_i} - \overline{\mathrm{PCR}})$	**5361**
Quadratic	$\mathrm{LM_i} = \mathrm{PCR_i} + \delta_i + \beta(\mathrm{PCR_i} - \overline{\mathrm{PCR}}) + \gamma(\mathrm{PCR_i} - \overline{\mathrm{PCR}})^2$	5368
Cubic	$\mathrm{LM_i} = \mathrm{PCR_i} + \delta_i + \beta(\mathrm{PCR_i} - \overline{\mathrm{PCR}}) + \gamma(\mathrm{PCR_i} - \overline{\mathrm{PCR}})^2 + \sigma(\mathrm{PCR_i} - \overline{\mathrm{PCR}})^3$	5370

Figure 2.5: Comparison of different model structures and their capacity to fit the collated data. Bayesian Markov chain Monte Carlo methods were used to fit a number of different relationships, varying in flexibility, to PCR and microscopy prevalence on the logit scale. **(A)** Microscopy and PCR Prevalence data from surveys (black points), with the fitted model relationship (red, green, blue and purple lines, denoting Basic, Linear, Quadratic and Cubic relationships on the logit scale respectively) all plotted on the natural scale. **(B)** As for (A), but plotted on the logit scale. **(C)** Description of the different model structures considered, along with the corresponding deviance information criterion (DIC) for each. Lower DIC values indicate a more preferred model.

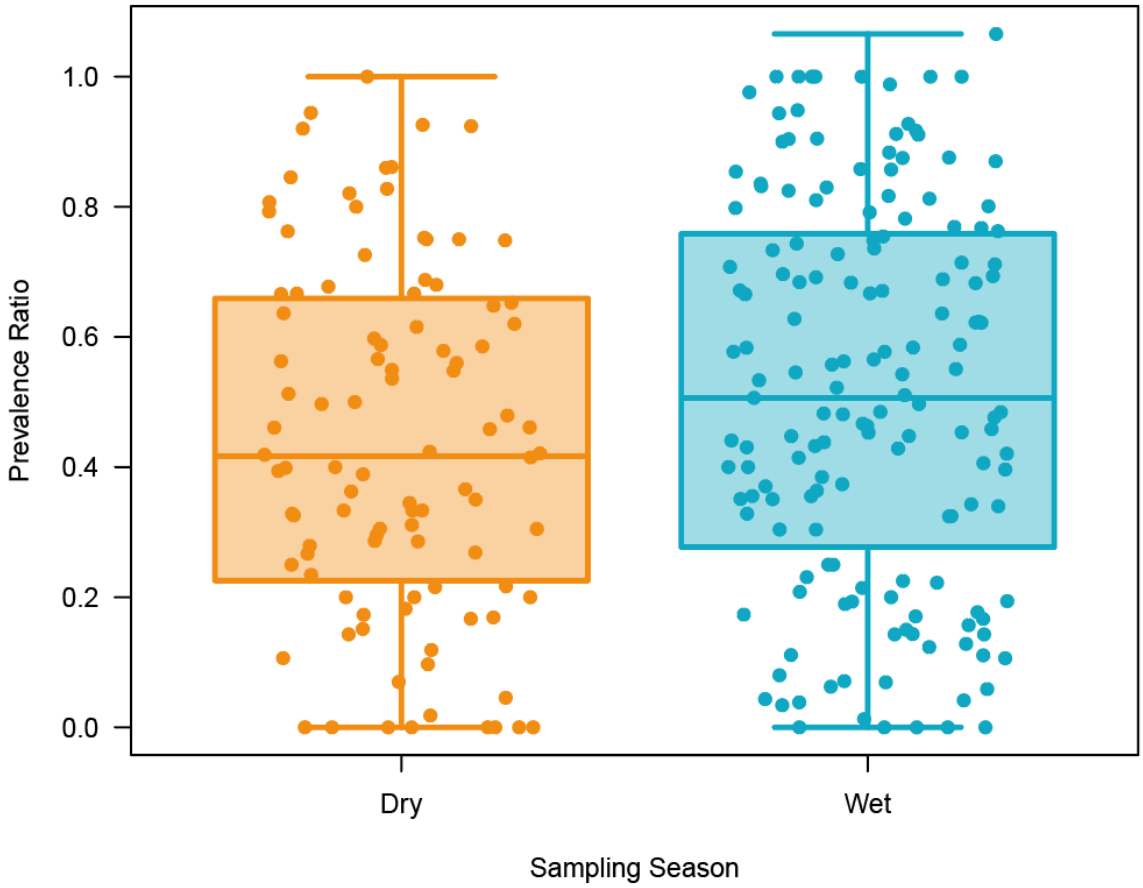


Figure 2.6: Comparing the prevalence ratio across different sampling seasons. Where available, information from references on which season sampling had occurred in was extracted and collated. Presented are boxplots of the prevalence ratio (defined as the ratio of microscopically detectable infections and PCR detectable infections, with a lower prevalence ratio indicating a higher proportion of individuals with submicroscopic infections) stratified by sampling season (n = 100 for dry season sampling, and n = 159 for wet season sampling), including also the raw datapoints (coloured circles). Note that an important caveat to these results is that I was unable to distinguish exact timing of sampling (e.g. early or late within season) in any more granular detail.

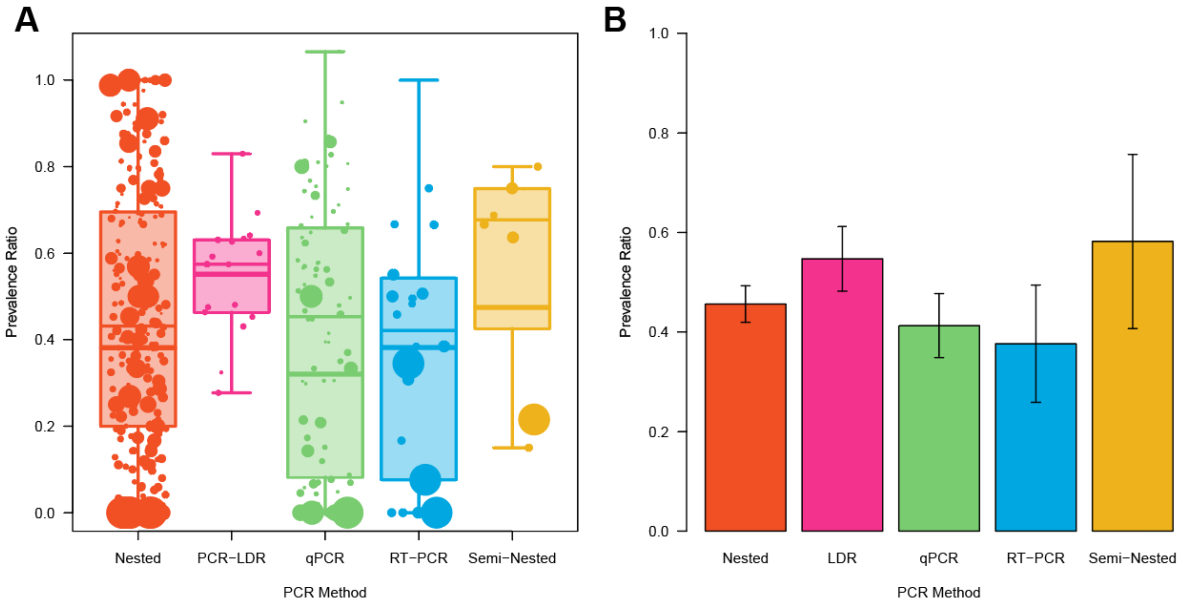


Figure 2.7: Comparing the prevalence ratio across different PCR methodologies. Where available, information from references on the seasonal timing of the sampling was extracted and collated. (A) Boxplots of the prevalence ratio (defined as the ratio of microscopically detectable infections and PCR detectable infections, with a lower prevalence ratio therefore indicating a higher proportion of individuals with submicroscopic infections) stratified by PCR methodology used to determine malaria infection in the survey, including also the raw datapoints (coloured circles, weighted by the inverse of their variance), and the weighted mean (thicker horizontal lines). There was a statistically significant difference in the mean prevalence ratio across PCR methodologies (p=0.038), with the prevalence ratio marginally lower in surveys using qPCR and RT-PCR to determine infection status (indicating that microscopy performs more poorly compared to qPCR than with other PCR methodologies). This significance remained even after explicitly accounting for PCR Prevalence in the underlying model. **(B)** Barplot of the mean prevalence ratio (and corresponding 95% confidence interval of the mean, as indicated by the error bars) for each different PCR diagnostic methodology used to ascertain malaria infection status.

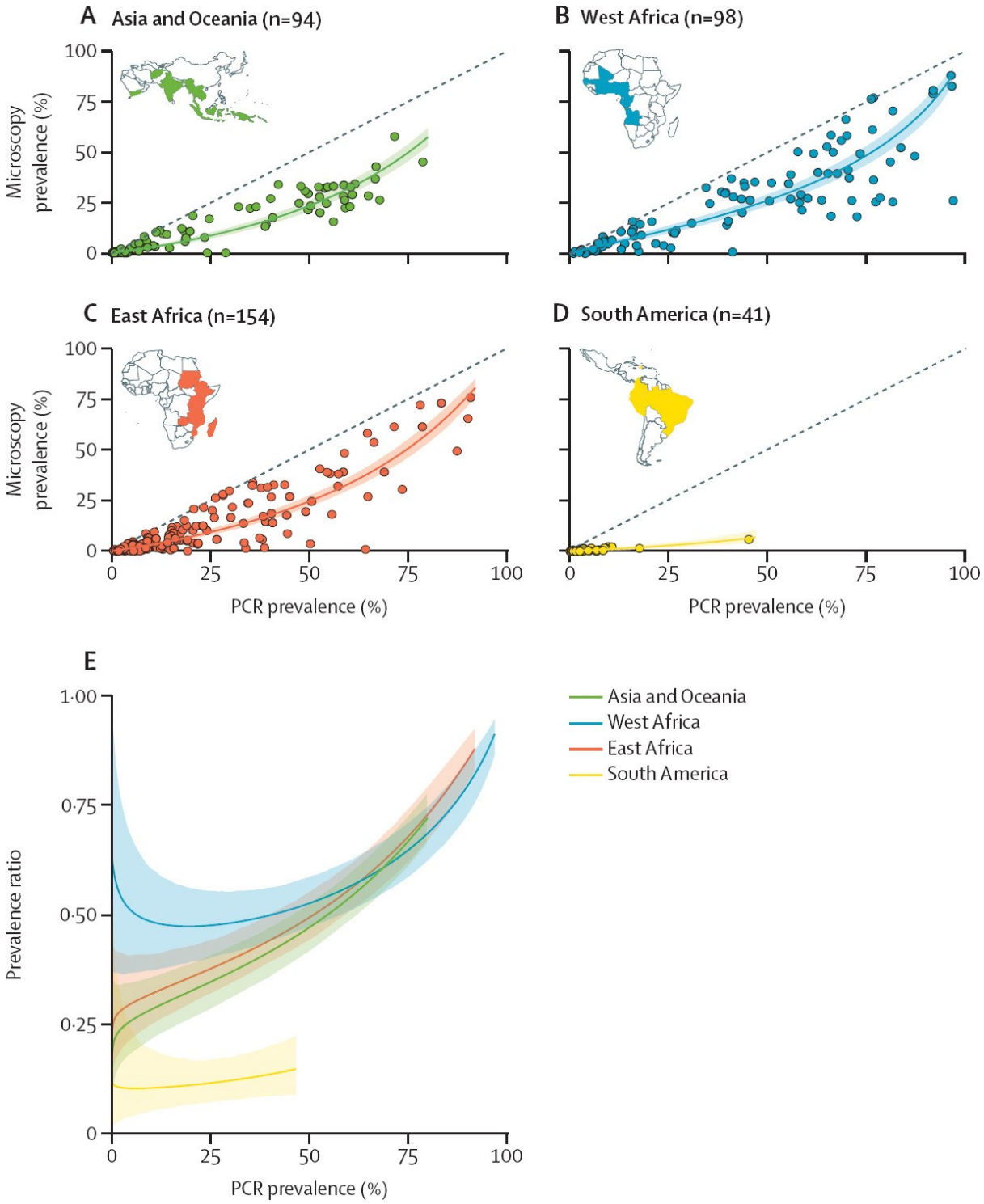


Figure 2.8: Global variation in the prevalence ratio and the relative size of the submicroscopic reservoir. Microscopy and PCR prevalence in included surveys (points), the model-fitted relationship (coloured line) and 95% credible interval (shaded area) for Asia and Oceania **(A)**, west Africa **(B)**, east Africa **(C)**, and South America **(D)**. **(E)** The model-fitted average microscopy: PCR prevalence ratio by PCR prevalence for each of the four regions (coloured line) and 95% credible interval (shaded area). Coloured countries on each regional map indicates countries for which studies were identified during the systematic review.

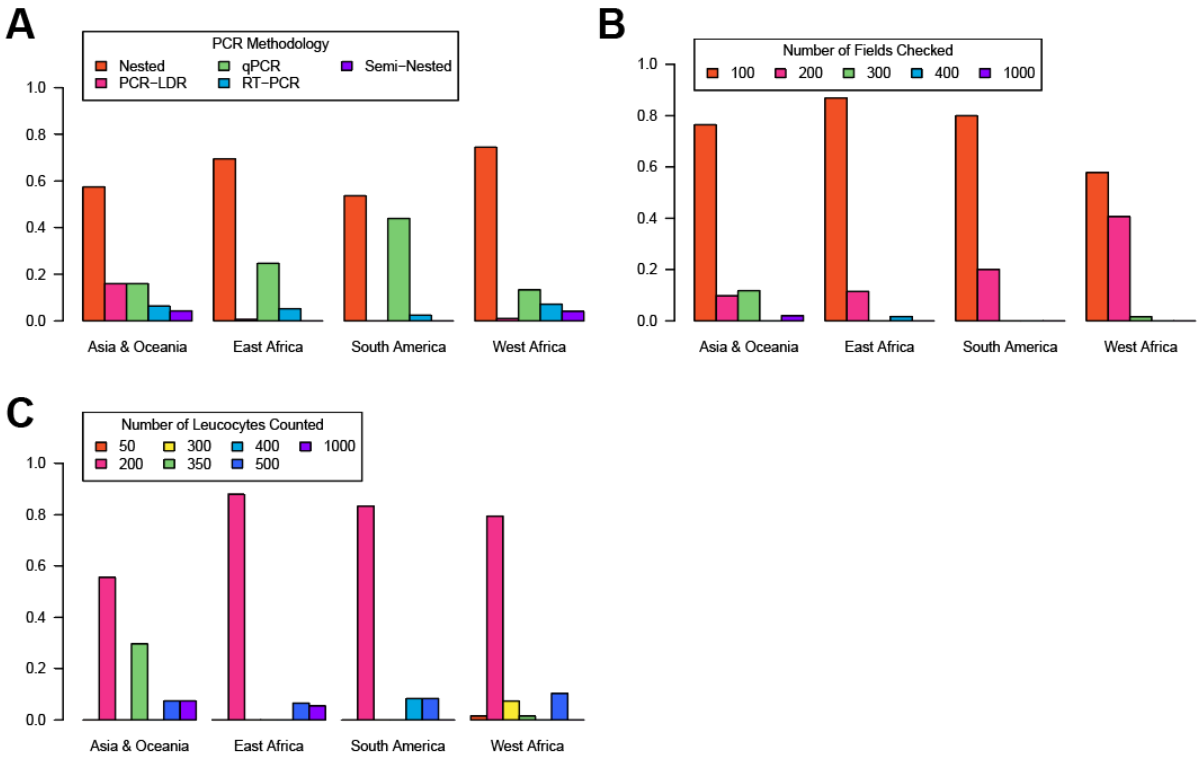


Figure 2.9: Tabulation of diagnostic properties by global region. Where available, information from each of the references were collated detailing some of the properties of the microscopy and PCR diagnostic methodologies used to determine infection status, and then the proportion of studies using each methodology disaggregated by global region. Specifically, these were **(A)** the type of PCR used to diagnose malaria infection, **(B)** the number microscopy fields checked when examining blood smears and **(C)** the number of leucocytes counted. Note that in some instances, the number of leucocytes counted was not used to determine infection status, but instead used to calculate parasite densities. The exact purpose of leucocyte counting was not consistently reported however, and so all instances in references that mention the number of leucocytes counted are tabulated here.

The majority of South American surveys had been done in areas marked by historically low transmission. I therefore investigated whether a high proportion of submicroscopic infections (low prevalence ratio) might be observed in areas with similarly historically low transmission in Africa. These results indicated that both regional historical prevalence (in the year 2000) and current prevalence (both averaged over the administrative unit 1 level) were significant predictors of the prevalence ratio when controlling for survey PCR prevalence (ANOVA, $p<0{\cdot}0001$ for regional historical prevalence at the administrative unit 1 level, $p=0{\cdot}042$ for regional current prevalence at the administrative unit 1 level), suggesting that historical transmission levels, in addition to current transmission levels, are an important determinant of the submicroscopic reservoir size.

I next classified each survey in my review from Africa into three transmission archetypes on the basis of the historical and current levels of transmission at the administrative unit 1 level **(Figure 2.10 A)** and fitted log-linear Bayesian regression models to each. The results were concordant with those from the ANOVA, with African surveys in regions with both historically and currently low transmission (n=40; Sudan, Ethiopia, and parts of Kenya and Tanzania) having on average a lower prevalence ratio (more submicroscopic infections) than other currently low endemicity areas in Africa where historical transmission was high (n=99) and settings where both historical and current transmission levels were high (n=90) **(Figure 2.10 B)**. There was no evidence of systematic differences in which PCR and microscopy methodologies were used across different transmission archetypes **(Figure 2.11)**.

In order to assess whether the threshold used to define transmission archetypes influenced the results, I repeated these analyses with a variety of different prevalence thresholds (specifically, 5, 10 and 20%) used to distinguish “high” from “low” transmission (with 15% used for the results presented in **Figure 2.10**). In all instances **(Figure 2.12 A – D)**, model fitting revealed an incremental decline in prevalence ratio (indicating a higher proportion of individuals who were submicroscopically infected) going from High High surveys to High Low and then Low Low surveys. This was most evident with the higher cutoffs of 20% or 15% **(Figure 2.12 A and D)** – as the cutoff threshold was lowered, an increasing number of surveys that had previously been classified as “High Low” were reclassified to “High High” and the difference in prevalence ratio between “High Low” and “Low Low” settings declined. Importantly however, the difference between “High High” and “Low Low” settings remained irrespective of the definitional threshold used, a finding that is also corroborated through the results of additional statistical analyses carried out **(Figure 2.12 E)**, highlighting that the observed results are robust to the choice of stratification threshold used to define high and low levels of transmission.

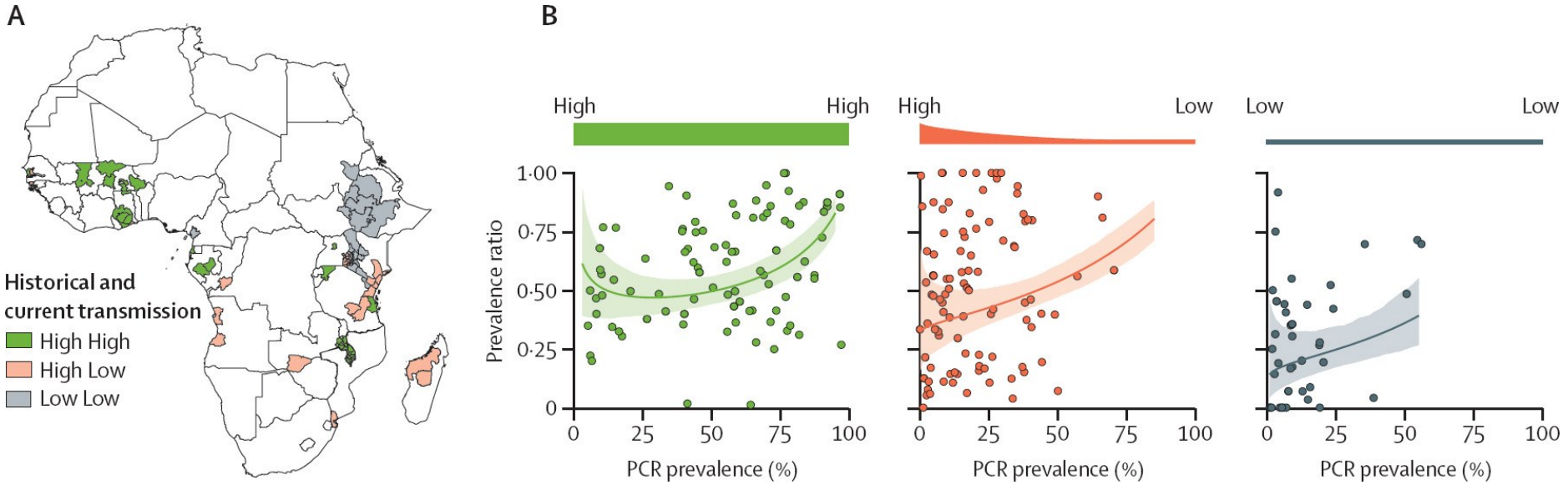


Figure 2.10: The effect of historical and current transmission intensity on the prevalence of submicroscopic malaria infection in Africa. **(A)** Map detailing the African countries and associated administrative unit 1 level regions for which prevalence surveys were identified, as well as their assigned transmission archetypes based on historical and current transmission intensity (high high=historically high and currently high; high low=historically high and currently low; low low=historically low and currently low). **(B)** The prevalence ratio of surveys in each transmission archetype (points; n=90 for high high, n=99 for high low, and n=40 for low low), and the modelled average prevalence ratio (coloured line) with 95% credible interval (shaded area).

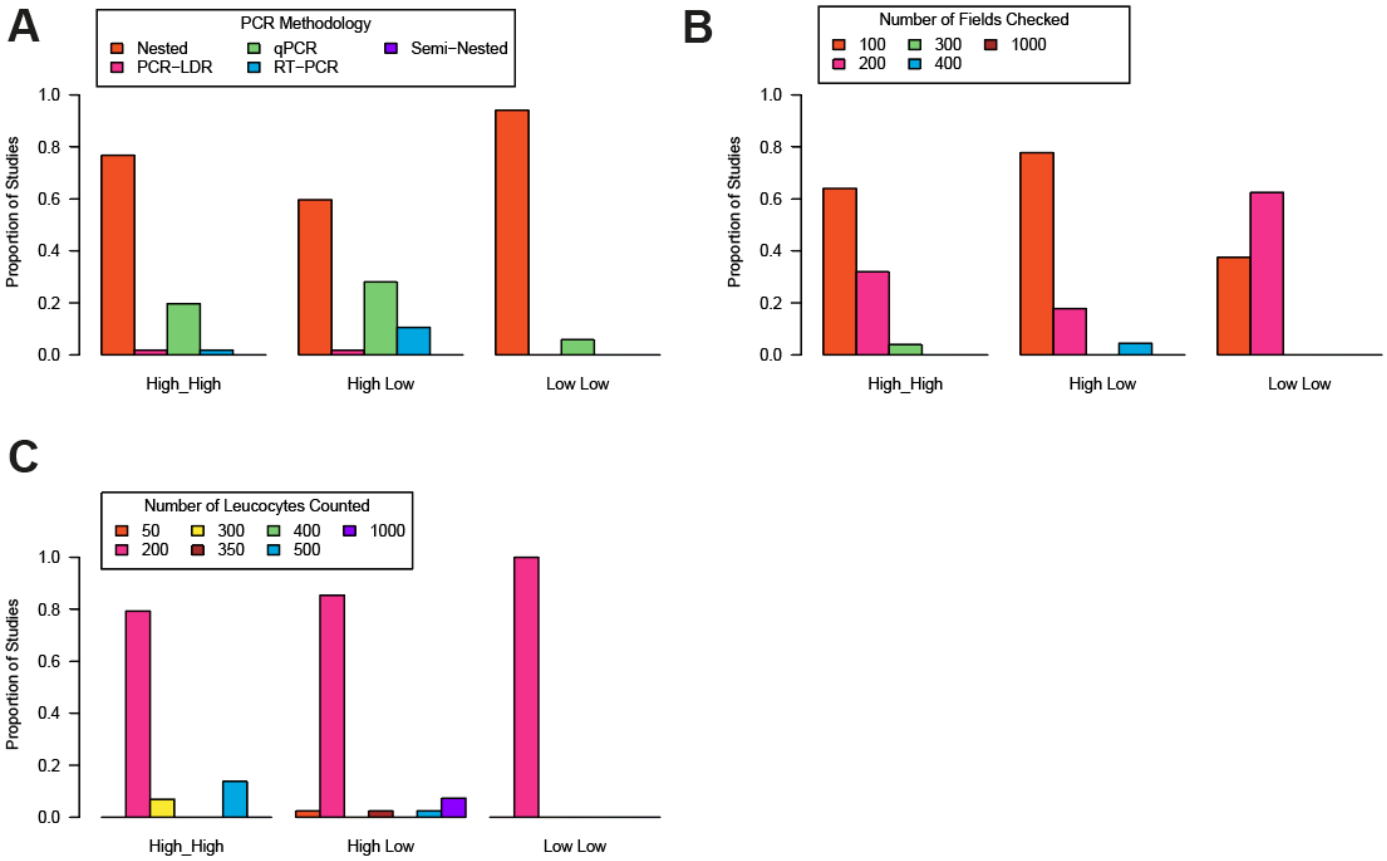
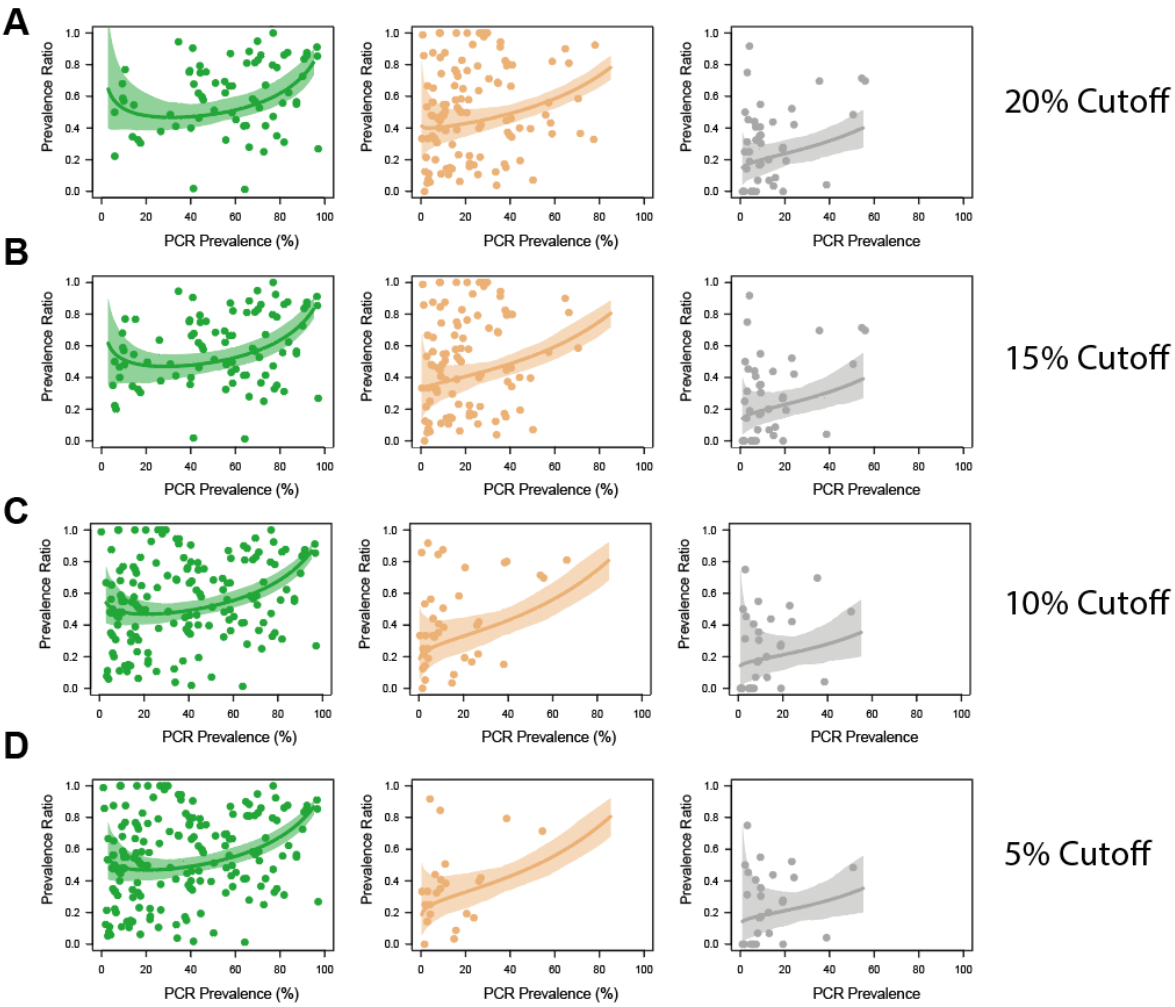


Figure 2.11: Tabulation of diagnostic properties by transmission archetype. Where available, information from each of the references were collated detailing some of the properties of the microscopy and PCR diagnostic methodologies used to determine infection status, and then the proportion of studies using each methodology disaggregated by transmission archetype (defined as described in the Methods section of the main text). Specifically, these were **(A)** the type of PCR used to diagnose malaria infection, **(B)** the number microscopy fields checked when examining blood smears and **(C)** the number of leucocytes counted. Note that in some instances, the number of leucocytes counted was not used to determine infection status, but instead used to calculate parasite densities. The exact purpose of leucocyte counting was not consistently reported however, and so all instances in references that mention the number of leucocytes counted are tabulated here.



Prevalence Cutoff	Overall Significance	High High vs High Low	High Low vs Low Low	High High vs Low Low
20%	Yes ($p < 0.001$)	No ($p = 0.42$)	Yes ($p < 0.001$)	Yes ($p < 0.001$)
15%	Yes ($p < 0.001$)	No ($p = 0.92$)	Yes ($p = 0.003$)	Yes ($p < 0.001$)
10%	Yes ($p < 0.001$)	Yes ($p < 0.001$)	Yes ($p = 0.003$)	Yes ($p < 0.001$)
5%	Yes ($p < 0.001$)	Yes ($p < 0.001$)	Marginal ($p = 0.07$)	Yes ($p < 0.001$)

Figure 2.12: Sensitivity analysis to assess the robustness of the results surrounding historical and current patterns of transmission intensity. The analyses presented in Figure 5 of the main text Results section were repeated using different thresholds (5, 10, 15 or 20%) for defining transmission archetypes. In each instance, the standard log-linear model was fitted and the results plotted for each definition (panels (A) – (D)), where the points represent a single survey, the line represents the model fit and the pale shaded area the 95% Credible Interval. Statistical analyses were also carried out (E), with ANOVA used to explore whether the mean prevalence ratio of the three transmission archetypes significantly differed (table 2nd column), with a post-hoc Tukey test carried out to assess pairwise differences (table columns 3-5).

Surveys carried out in specific age groups were also collated, and from these surveys, three age-based categories were defined: young children (0–5 years) giving 49 prevalence survey pairs, older children (6–15 years) giving 62 prevalence survey pairs, and adults (>15 years) giving 53 prevalence survey pairs. The prevalence ratio varied significantly between age groups (ANOVA, p<0·0001, df=2), and was significantly lower in adults (indicating a greater proportion of submicrosopic infections) than in young children (Tukey's HSD, p<0·0001) and older children (Tukey's HSD, p<0·0001) **(Figure 2.13 A)**. Fitting the Bayesian regression model separately to the data for each age group highlighted that the increased prevalence ratio observed in young children and older children compared with adults was less pronounced in higher-transmission settings. In high endemic areas with 70% overall PCR prevalence, the prevalence ratio for young children was predicted to be 1·42× that of adults, but 1·92× at low endemic areas with 10% overall PCR prevalence **(Figure 2.13 B)**. A similar result was observed for adults and older children, suggesting genuine differences in submicroscopic epidemiology both between age groups and across transmission settings.

I additionally explored how the contribution of submicroscopic infections to onwards transmission might vary across settings characterised by different historical transmission patterns, using a range of estimates for the comparative infectivity of submicroscopic and microscopically detectable infections. In transmission settings characterised by both historical and current low levels of transmission, submicroscopically infected individuals could account for an estimated 17·5% to 68·0% of onwards transmission **(Figure 2.14 C)**. By contrast, the results suggest the contribution of the submicroscopic reservoir to transmission is less important (although not negligible) in settings where transmission has only recently declined **(Figure 2.14 B)**, ranging from 7·8% to 46·0% depending on assumed comparative infectivity.

Discussion

Considerable debate surrounds the importance of the submicroscopic reservoir to malaria control efforts and whether it needs to be targeted by interventions (Lin, Saunders and Meshnick, 2014), particularly in areas of low transmission. Disaggregating the now greater quantity of available data (551 prevalence survey pairs from 44 countries) has given insight into the complex relationships underlying the global pattern of submicroscopic occurrence. This insight has facilitated a more refined evaluation of when and where submicroscopic infections are likely to be most prevalent and who is most likely to harbour them. This work suggests that some of the differences observed in the size of the submicroscopic reservoir can be explained by differences in historical patterns of transmission and the age profile of the infected population. Moreover, although previous work has generally noted the potential

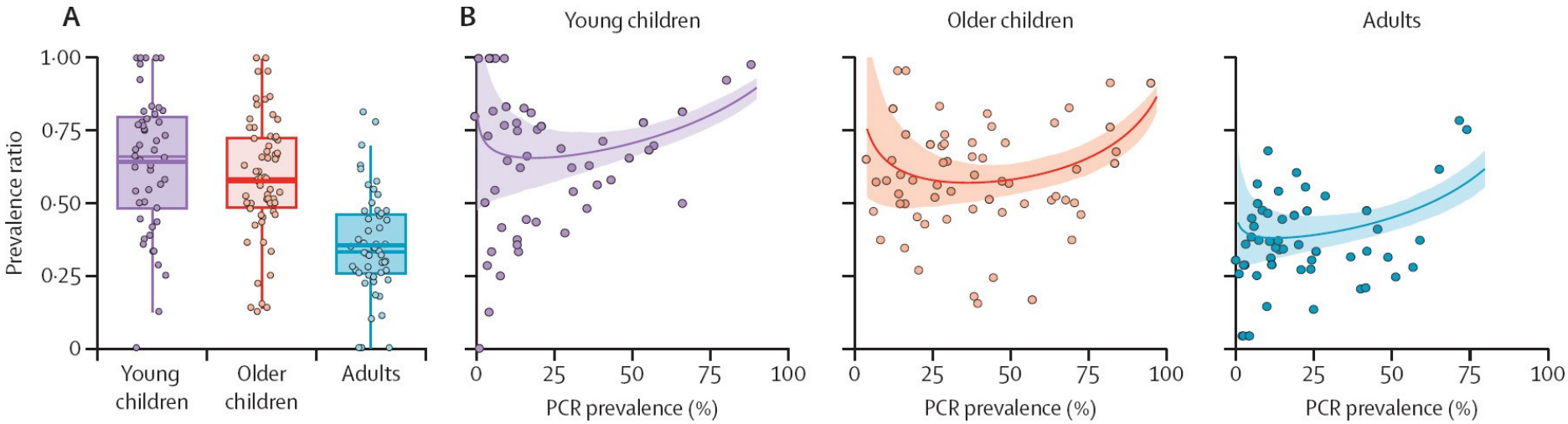


Figure 2.13: The influence of age on submicroscopic malaria infection. (A) Box plot of age disaggregated prevalence survey data for young children (0–5 years, purple points, n=49) older children (6–15 years, pink points, n=62), and adults (>15 years old, blue points, n=53). For each age group, the size of the point reflects the number of individuals tested by microscopy and PCR. Thick coloured bar on the boxplot represents the weighted mean prevalence ratio for each age group. Thin line indicates the median, box indicates IQR, and whisker limits span 1·5× the IQR. **(B)** The prevalence ratio in surveys where age-disaggregated data (points) were available by age group, showing the fitted model relationship (coloured lines) and the 95% credible interval (shaded areas).

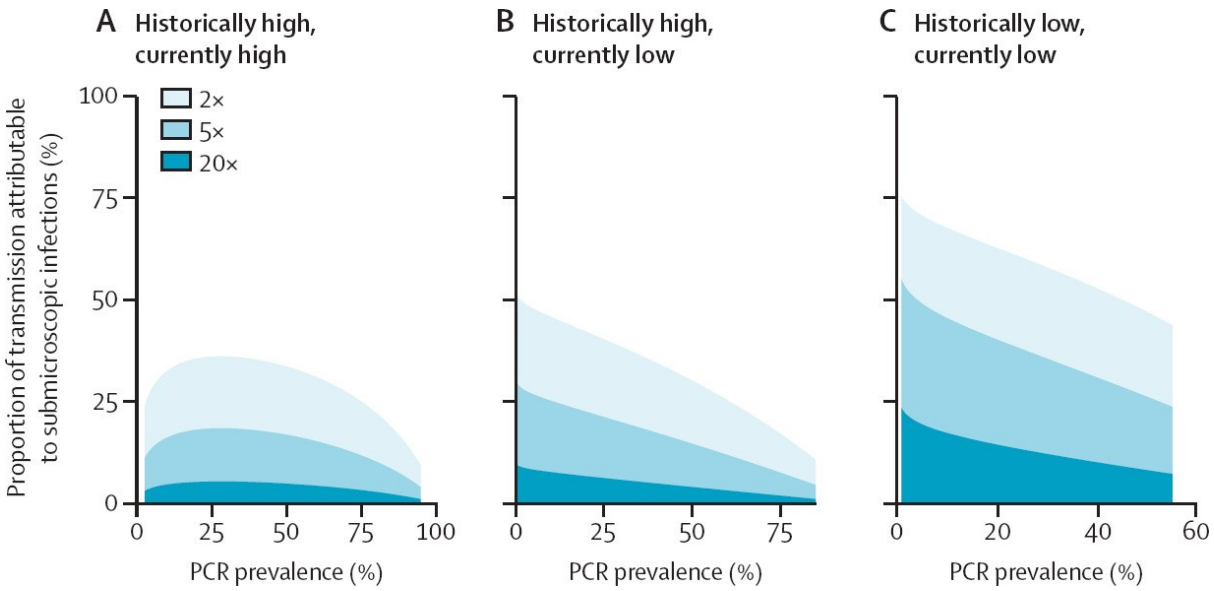


Figure 2.14: The potential contribution of submicroscopic infections to onwards transmission according to current and historical transmission intensity. Potential contribution of the submicroscopic reservoir to onwards transmission for each of the transmission archetypes: historically high and currently high (A), historically high and currently low (B), and historically low and currently low (C) if microscopic infections are either 2×, 5×, or 20× more infectious than submicroscopic infections.

relevance of submicroscopic infections in low-transmission settings (Cotter et al., 2013), these results suggest that this relevance is likely to be highly context dependent, potentially warranting different approaches to the control of submicroscopic infections in different locations. Both increasing age (independent of exposure) and increased immunity (due to previous exposure, which also increases with age) have been linked to lower parasite densities (Rodriguez-Barraquer et al., 2018). These results underscore the importance of age (and the demography of the population in general) as well as other setting-specific factors (such as historical patterns of transmission) in determining the size of the submicroscopic reservoir. However, an important caveat to these findings is that there were insufficient data to examine the role of these factors simultaneously. The average age at which an individual is infected is typically higher in low-transmission settings (Carneiro et al., 2010); therefore, a greater proportion of infected individuals in all-age surveys would be expected to be adults. It is also possible that systematic biases in the age of surveyed populations might exist between geographies or transmission archetypes. These results surrounding past transmission history might then be confounded by differences in the average age of infection across these different settings. However, the age distribution of malaria infection appears to adapt fairly rapidly to reflect changes in transmission, with infection profiles shifting to older individuals as transmission decline, a feature observed in both clinical cases (Brasseur et al., 2011) and infection in the wider community (Griffin, Ferguson and Ghani, 2014). Although this observation suggests that the difference in infected population age profile between surveys from historically low and currently low settings might not be substantial, I am unable to conclusively disentangle the potentially confounding role of age in the analyses of transmission archetypes (and global regions). There is also scope for residual confounding from variation in other factors such as microscopy or PCR methods across locations, although the results collating methodologies used across regions and archetypes suggest this might not be substantial. Although the global regional analyses showed that South American surveys were more likely to have used qPCR (a more sensitive diagnostic), the analysis of transmission archetypes revealed no systematic variation in the PCR methodologies used across archetypes. More broadly, it is also important to note that the data collated here represents cross-sectional surveys of locations that have not necessarily been sampled at random (and instead might be biased towards established research sites), which could introduce systematic bias into the findings.

Another potential limitation is the strong geographical bias in the transmission archetype stratification, which precludes exact determination of the extent to which variation in submicroscopic malaria is driven by geography compared to transmission patterns. The majority of surveys assigned to the historically low and currently low archetype are from east

Africa, while the majority of surveys in the historically high and currently high archetype are from west Africa. The observed results across transmission archetypes could, therefore, be reflecting geographical variation (i.e. the fact that a significant proportion of surveys belonging to a particular transmission archetype are also from the same broad geographical area) rather than variation driven by past transmission history. However, the fact that the proportion of infections that are submicroscopic in historically high and currently low settings (a strata also predominantly composed of studies from east Africa) was consistently lower than that observed for historically low and currently low settings, together with the results observed for South America, provide tentative support for an effect of transmission history on the size of the submicroscopic reservoir, independent of variation due to geographical location.

Several hypotheses other than geographical confounding could explain the observed results across the different transmission archetypes, including various haemoglobinopathies and human genetic traits that have been linked to lower average parasite densities (Kiwanuka, 2009). Parasite-related factors could also account for the results observed here, such as systematic variation across locations in asexual blood stage multiplication rate of *Plasmodium falciparum* (Murray et al., 2017) or selective pressures that vary with transmission intensity. Previous work has suggested that high-transmission settings might select for parasites with high replication rates and virulence (to outcompete other co-circulating *Plasmodium falciparum* strains), whereas low-transmission settings might select for non-virulent parasites with lower rates of replication better able to persist and avoid causing symptomatic infection (which would prevent drug exposure but be more likely to present submicroscopically) (Björkman and Morris, 2020). Lower genetic diversity (resulting in more rapidly acquired immunity to local parasite clones) might also contribute to the observed results. It is also not possible to definitively preclude a role for systematic variation in diagnostic quality across settings, although analyses have found that microscopy quality does not vary systematically with transmission intensity (Slater et al., 2019). Although these results highlight that PCR methodology does significantly influence the prevalence ratio, systematic variation in methodological quality across transmission archetype settings was not observed.

The analyses revealed a significant influence of seasonal effects on submicroscopic carriage, with submicroscopic infections more common in the dry season. This is in keeping with previous work showing that parasite densities rise slightly during the rainy season (even when prevalence does not change significantly) (Slater et al., 2019). It is important to note, however, that classification of sampling season was necessarily coarse due to limitations in available data (with information on timing within season being typically sparse). It is possible then that more granular disaggregations might reveal further variation that data limitations have

precluded exploring here. Seasonal effects have also previously been shown to play a role in shaping performance of rapid diagnostic tests for malaria (Watson et al., 2019b) and so future work exploring the factors driving the prevalence of false negatives in these diagnostics (which have shown similar relationships with overall transmission in previous reviews (Watson et al., 2019a)) would also likely be important given their increasing use over microscopy for surveillance and diagnosis of malaria infection.

This work suggests that the contribution of submicroscopic infections to onwards transmission is likely to be highly variable across settings. However, this analysis is based on the detection of asexual parasites and does not provide direct insight into gametocyte densities. Additionally, due to data constraints, I was unable to consider a range of relevant factors, such as age profile of the infected population and related skin surface area effects (whereby adults have larger skin surface areas available for biting by mosquitoes compared with children) and adjusting for these would likely increase the contribution to transmission from older children and adults (who are more likely to have submicroscopic infection). The relationship between asexual parasite and gametocyte density is highly non-linear and the distributions of parasite densities in the submicroscopic range can differ substantially between settings (Slater et al., 2019). The proportion of submicroscopic infections might, therefore, not linearly relate to their contribution to onwards transmission. For example, while a membrane feeding study done in Burkina Faso and Kenya (high-transmission settings) found that 45–75% of all mosquito infections were derived from submicroscopic infections (Gonçalves et al., 2017), only 4% of infections arose from submicroscopic individuals in a similar study carried out in Cambodia (a low-transmission setting) (Lin et al., 2016). Similarly, recent work from a setting with effective malaria control in eastern Uganda has suggested that the majority of onwards transmission (84%) arises from asymptomatic but microscopically detectable infections rather than submicroscopic infections (16%) (Andolina et al., 2021). These findings contrast with the predictions presented here and underscore the need to better resolve the relationship between submicroscopic parasite carriage, gametocyte densities, and mosquito infectivity.

Despite their potential relevance to maintenance of malaria transmission, our understanding of submicroscopic infections remains far from complete. Do submicroscopic infections represent a substantial source of transmission and threat to future progress? Do these infections need to be targeted to achieve malaria elimination? Although more work is required, these findings highlight important differences in submicroscopic epidemiology between settings and suggests the absence of a one-size-fits-all solution for malaria control targeting this infection subgroup. Such variation will probably warrant different approaches to malaria control if the infection is to be controlled most effectively in the effort towards elimination.

Conclusion

In this chapter, I have explored the factors underlying heterogeneity in the size and extent of the submicroscopic malaria reservoir across different settings and explored the implications such heterogeneity might have for the appropriateness of explicitly targeting this infected sub-group for treatment. In the next chapter, I examine another source of significant heterogeneity relating to malaria infections and the human host, specifically the fine-scale spatial heterogeneity and clustering of malaria infections at the household level. Using the same Bayesian regression modelling framework used in this Chapter, in the following Chapter I explore how the extent of spatial clustering of malaria infections varies with levels of transmission and explore how this variation in the degree and extent of clustering influences the viability of certain control measures centred around the household.

Chapter 3 Quantifying *Plasmodium falciparum* Infection Clustering Within Households to Inform Household-Based Intervention Strategies for Malaria Control Programs

In contexts where malaria transmission is peri-domestic, understanding spatial heterogeneity is key to optimising delivery of certain control strategies. This enables the most at-risk individuals to be identified and through this, more effective targeting of interventions. Using the Bayesian regression modelling approach applied in Chapter 2, in this Chapter I leverage malaria infection data from Demography and Health surveys from across sub-Saharan Africa to better understand the role of fine-scale spatial heterogeneity in the clustering of malaria infections at the household level. Specifically, I quantify the degree of spatial clustering, and how this varies with overall levels of malaria transmission. I explore the implications of this heterogeneity in the degree and extent of spatial clustering of infections on the viability of different household-based control measures and identify the settings in which they are likely to be more appropriate.

Introduction

The transmission and dynamics of malaria are highly heterogeneous between and within different populations, a phenomenon underpinned by a diversity of different factors including the vector species driving transmission (Sinka et al., 2010b; Massey et al., 2016), the parasite species causing disease (Phillips et al., 2017), patterns of exposure (Pollard et al., 2020; Sandfort et al., 2020; Kar et al., 2014) and the ecological structure of the location (Whittaker et al., 2021), amongst others. Spatial variation in infection is also a key factor underlying the heterogeneous nature of malaria transmission across settings. Variation in infection risk and disease burden is evident at a range of spatial scales (Grillet et al., 2010) and settings, but is thought to be particularly acute in the areas of lowest transmission (Stresman et al., 2018; Mogeni et al., 2017). In contexts where transmission of malaria is peri-domestic (that is, where contact with the vectors responsible for infection and onwards transmission occurs primarily around the main residence, in contrast to other settings where malaria exposure transmission is primarily occupational (Ekawati et al., 2020)), a better understanding of this spatial heterogeneity is key to optimising delivery of various control strategies, enabling the most at-risk populations to be identified and effective targeting of interventions.

The apparent clustering of malaria infections has led to usage in some cases of so called “reactive control strategies” – an example of this being Reactive Case Detection (RACD), which involves testing and treating all positive household members of confirmed malaria cases that have been passively detected at health facilities (Sturrock et al., 2013a; Aidoo et al., 2018). Such strategies are now widespread and most commonly employed in low transmission

settings (Rossi et al., 2018b; Sturrock et al., 2013a; Hustedt et al., 2016), in-keeping with WHO guidelines (World Health Organization, 2018c). Significant concerns remain however around how best to operationalise RACD, particularly in settings with low treatment-seeking rates (Rossi et al., 2018a) and where access to healthcare is limited (Weiss et al., 2020), which might lead to only a small fraction of the total infections present in a community being targeted.

These limitations have led to renewed interest in community-based approaches such as Mass Screen and Treat (MSAT) which involves testing all febrile members of a community for malaria and treating those positive for malaria (Millar, Toh and Valle, 2020); and Mass Test and Treat (MTAT), which involves testing the whole population for malaria using rapid diagnostic tests (RDTs) or light microscopy (LM), irrespective of symptoms, to identify and treat any infections that may be present (Conner et al., 2020). Whilst potentially able to reach a far larger fraction of the infected population (including asymptomatic or mild infections not leading to treatment seeking at health facilities), they are also far more operationally and resource intensive – potential modifications to these approaches that remain community-based but reduce programmatic costs have centred around focussing efforts on individual households, and treating entire households upon detection of a single malaria case, in turn reducing the number of individuals who must be screened (MSAT) or tested (MTAT).

Underpinning the viability of any of these strategies will be a better understanding of how and the degree to which malaria infections cluster within households; and in particular, whether the types of infection detected by each of the different programmes (RACD = symptomatic and presenting at health facility; MSAT = symptomatic; MTAT = detectable by RDT/LM) cluster within households. Significant uncertainties remain however surrounding the degree and generality of household clustering of malaria infections – empirical evidence is limited, particularly as to whether household clustering is consistently observed across settings characterised by differing patterns of vectors, ecologies, patterns of exposure and transmission intensities (Stresman, Bousema and Cook, 2019). By extension, it remains unclear in which settings household-based reactive strategies are likely to be the most appropriate and effective. Here I collate and analyse information on clustering of malaria infections diagnosed by rapid diagnostic tests (RDTs) or light microscopy (LM) in over 200,000 children from 57 Demographic and Health Surveys conducted in 23 countries across sub-Saharan Africa to answer the question of whether *Plasmodium falciparum* infections cluster within households of programmatically detectable infections. This work builds upon and complements recent systematic reviews of the literature exploring this question of clustering from the perspective of RACD (van Eijk et al., 2016), MSAT and MTAT campaigns (Stresman et al., 2020). In particular, utilisation of the Demographic and Health Surveys data allows

exploration of a wider diversity of settings than has previously been possible in analyses of existing published literature.

Utilising a Bayesian logistic regression-based framework, I systematically explore how the degree of household clustering of malaria infections changes with transmission intensity, whilst controlling for other potential confounders such as household size. The results highlight the operation relevance of malaria clustering to control programmes, but also substantial variation in the degree and extent of this household clustering across the range of transmission intensities spanned by the 23 countries considered. In doing so, these results highlight particular types of settings in which reactive control strategies for malaria are likely to have the most relevance and impact.

Methods

DHS Data: Overview of Data Collation, Survey Selection Criteria and Index Household Definition

The Demographic and Health Surveys (DHS) Program is an organisation responsible for collecting and disseminating accurate, nationally representative data on health and population in developing countries (Demographic and Health Surveys, 2022). These data are collected through conducting surveys involving in-depth interviews of households and individuals, following a set methodology and using a routine questionnaire consisting of core questions (collecting a suite of information on various economic, health-related and demographic variables) that are supplemented with a variety of different questions depending on the context and particular survey. The results of these surveys are then processed into different datasets (called “recodes”) characterised by different units of analysis (broadly, “who” or “what” is being studied). Depending on the recode, these units of analysis are typically either households, women, children or men. Each recode typically contains a core set of uniquely identifying information (Age, Sex, Region etc) that is present across all recodes, as well as information that is unique to that recode (and present in that recode only).

Individual-level data from 57 DHSs conducted across 23 African countries containing information on malaria infection status (as diagnosed by RDT or light microscopy) in children under 7 years of age were collated using the *rDHS* R package (Watson, FitzJohn and Eaton, 2019). From these surveys, I extracted a suite of demographic and household related information (described further below), as well as data enabling identification and definition of whether that individual resided in an *“Index Household”*. I describe this in more detail below, but briefly, an individual was defined as residing in an *“Index Household”* if the household contained at least one other person (i.e. other than the individual being considered) with a

malaria infection detectable through one of the programmatic strategies considered i.e. when not considering the infection status of the individual in question, are there any other infections present in the household that would have led to the considered individual having been identified and treated under the three programmatic strategies considered. The definitions for each strategy are as follows:

- **Mass Screen and Treat (MSAT):** Household contains an individual with a malaria infection detectable by either light microscopy or RDT (i.e. contains an individual testing positive for malaria at the time of the survey).
- **Mass Test and Treat (MTAT):** Household contains an individual with a malaria infection (as detected by light microscopy or RDT) and who was/is symptomatic i.e., the individual had had fever during the previous 2 weeks.
- **Reactive Case Detection (RACD):** Household contains an individual with a malaria infection, and treatment for that infection has recently (i.e. within the past 2 weeks) been sought at a health centre.

Defining the Index Households in this way allows me to mimic the detection strategies being assessed, and explore whether residency in an Index Household increases an individual's chances of being malaria positive - specifically, whether an individual who shares a household with someone who is currently malaria positive (MTAT), has recently had a symptomatic malaria infection (MSAT), or has recently sought treatment for malaria (RACD) is more likely to be malaria positive compared to households negative for each definition (see **Table 3.1**).

DHS Data: Detailed Information On Data Collation and Processing Workflow

Using the rDHS package (Watson, FitzJohn and Eaton, 2019), I identified all surveys that had assessed malaria infection status by either rapid diagnostic test (RDT, SurveyCharacteristicID 89 in the rDHS package) or light microscopy (SurveyCharacteristicID 90 in the rDHS package). In total, 57 surveys were identified in this way, as well as a single additional survey not reporting SurveyCharacteristicID 89 or 90 but where malaria infection status had been ascertained. Together, these 57 surveys spanned the years 2006 – 2018, and included over 2.5 million individuals (although the number included in my analyses is far lower as only a subset of these individuals had their malaria status tested, see below for further details) from 23 countries across sub-Saharan Africa.

Table 3.1 Overview of programmatic strategies for identifying households likely to have asymptomatic and/or subpatent infections. For the analyses presented here, I defined an index household, control and case population for each type of programmatic strategy. Infections in the index and control

Strategy	Definition of Index Household	Individual-Level Binary Indicator For Residence In Index Household
Reactive Case Detection (RACD)	Households with confirmed infections detected within health facilities. I.e., Households with any child who is malaria positive (by RDT/microscopy) and who has sought antimalarial treatment in the past 2 weeks.	For a given child, 1 if any other child in the household is positive for malaria and has sought anti-malarial treatment (based on the collated treatment seeking variables), and 0 otherwise (i.e., all other children in the household who are tested and provide responses are malaria negative and have not sought anti-malaria treatment respectively).
Mass Screen and Treat (MSAT)	Households with a symptomatic case (fever and RDT/light microscopy positive for malaria) which has been detected in the community as part of an active campaign. I.e., Households with a child who is RDT/microscopy positive and who has had a fever during the past 2 weeks.	For a given child, 1 if any other child in the household has had a malaria-related fever during the past two weeks (defined as that other child in the household being malaria positive and having had a fever during the past 2 weeks) and 0 otherwise (i.e. all other children in the household who were tested and provided responses were malaria negative and have not reported a fever during the past 2 weeks respectively).
Mass Test and Treat (MTAT)	Households with any infected individual (irrespective of symptoms, detected by RDT/light microscopy) detected in the community as part of an active campaign. I.e., Households with a child who is RDT/microscopy positive.	For a given child, 1 if any other child in the household is currently malaria positive and 0 otherwise (I.e. all other children in the household who are tested are malaria negative).

Note: To avoid index children making self-defining residency in an index household, when defining the binary indicator for a particular child, I only consider the other children who live in that house and their infection/symptoms/treatment seeking.

Of the surveys identified, 51 tested for malaria infection only in children under 7, whilst the remaining 6 surveys tested all individuals regardless of age. The outcome of interest in the analyses conducted here was the probability of an individual possessing a malaria infection (as determined by RDT or light microscopy) or not, and so individuals not tested for malaria were excluded. The primary dataset for consideration therefore consists of children under 7 whose malaria status had been determined by either RDT or microscopy and for whom information was available on recent fever occurrence and treatment seeking behaviour. This information (on malaria status, recent fever occurrence, treatment seeking status etc) is not contained in a single part of the survey and is instead across two datasets associated with each survey: the "People's Recode" (containing information on malaria infection status) and the "Children's Recode" (containing information on recent fever occurrence and treatment seeking behaviour), necessitating linkage of individuals across both datasets.

In order to do this linkage, I extracted uniquely identifying information that was present across both survey recodes. Specifically, these variables were Age, Sex, Cluster Number, Household Number and Line Number. Using these variables, I was able to link the vast majority of respondents (consistently >90%) across both datasets (see **Table 3.2** column *"Linked Across Recodes"*) allowing near complete collation of malaria infection status, recent fever occurrence and treatment seeking behaviour across surveyed individuals. Inability to link all individuals likely arises either from errors associated with individual enumeration of data entry or issues surrounding who is included in the "Children's Recode". Specifically, if a child in the household during the survey is without a mother (who would typically be interviewed to provide data for the "Children's Recode"), that child will not be present in the "Children's Recode" but will be present in the "Person's Recode". Irrespective, these instances represented a small fraction of the overall individuals surveyed and are therefore unlikely to affect the conclusions arising from subsequent analysis of the data.

Infection status was detected by either RDT (n = 11 surveys) or Light Microscopy and RDT (n = 47 surveys). There was one survey for which it was unclear which technique had been used – for the purposes of the results presented here, individuals were defined as positive if the individual tested positive for malaria by either method. Recent fever occurrence was defined based on the survey variable *h22* from the Children's Recode, which describes whether the child has experienced fever during the past 2 weeks. To characterise treatment seeking behaviour, I examined the full array of malaria treatment related variables detailed in each survey's Children's Recode, utilising them as a proxy for treatment seeking behaviour (i.e. I assume that an individual receiving a particular anti-malarial treatment recently is indicative of that individual having recently sought treatment). Although treatment related variables were largely consistent across surveys, there were a number of survey-specific treatment related

variables, asked as an additional, non-standard question during the survey. For the analyses presented here, I utilised all of the treatment-seeking variables available for a given survey, with an individual defined as having sought treatment if any one of these extracted variables was positive (see **Table 3.2** for the full list of variables utilised). In addition to these variables, I also extracted uniquely identifying information for each individual, including their region of residence, their household number (as enumerated by the survey) and unique line number, allowing linkage of children between the People's Recode (containing information of malaria status) and the Children's Recode (containing information on recent fever occurrence and treatment seeking behaviour) associated with each survey.

Using these collated variables, I defined a set of three binary indicators for each individual that describe whether that individual belongs to an Index household, as defined each of the programmatic strategies being considered. Briefly, an individual was defined as residing in an index household if the household contained at least one other infection detectable through one of the programmatic strategies considered i.e. when not considering the infection status of the child in question, are there any other infections present in the household that would have led to that household having been identified as an index household under the three programmatic strategies i.e. through i. clinical care seeking (RACD model), ii. through household surveys using routine diagnostics in symptomatic individuals (MSAT) or regardless of symptoms (iii. MTAT). For a given individual, these binary indicators describing whether that individual resides in an Index Household are defined in the following way:

- ***MTAT Index Household Status (MTAT)*:** 1 if any other child in the household is currently malaria positive at the time of the survey and 0 otherwise (i.e. all other children in the household who are tested are malaria negative).
- ***MSAT Index Household Status (MSAT)*:** 1 if any other child in the household has had a malaria-related fever during the past two weeks (defined as that other child in the household being malaria positive and having had a fever during the past 2 weeks) and 0 otherwise (i.e. all other children in the household who were tested and provide responses are malaria negative and have not reported a fever during the past 2 weeks respectively).
- ***RACD Index Household Status (RACD)*:** 1 if any other child in the household is positive for malaria and has sought anti-malarial treatment (based on the collated treatment seeking variables), and 0 otherwise (i.e. all other children in the household who are tested and provide responses are malaria negative and have not sought anti-malaria treatment respectively).

Table 3.2 Definition and Extraction of Treatment Seeking Variables

Variable Label	Description	Response	Survey Usage
h37a	Fansidar taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37aa	Artesunate rectal taken for fever	0 no, 1 yes, 9 missing	All Where Used
h37ab	Artesunate injection/IV taken for fever	0 no, 1 yes, 9 missing	All Where Used
h37b	Chloroquine taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37c	Amodiaquine taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37d	Quinine taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37da	Quinine injection/IV taken for fever	0 no, 1 yes, 9 missing	All Where Used
h37e	Combination with artemisinin taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37f	CS antimalarial taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37g	(CS) PRIMO taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37h	Other antimalarial taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
h37n	Artesunate mefloquine taken for fever	0 no, 1 yes, 9 missing	53 only
ml13a	Fansidar taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13aa	Artesunate taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13ab	Artesunate injection/IV taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13b	Chloroquine taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13c	Amodiaquine taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
ml13d	Quinine taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13da	Quinine injection/IV taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13e	Combination with artemisinin taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used

ml13f	Coartem taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
ml13g	AL/Artemether Lumefantrine taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13h	Other antimalarial taken for fever	0 no, 1 yes, 9 missing	All Where Used
ml13n	Artesunate (injection) taken for fever/cough	0 no, 1 yes, 9 missing	17, 53 Only
s623a	Medicine taken during fever: Antimalarial, combination with artemisinin (tca)	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623b	Medicine taken during fever: Antimalarial, Sp/fansidar	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623c	Medicine taken during fever: Antimalarial, Chloroquine	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623d	Medicine taken during fever: Antimalarial, Amodiaquine	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623e	Medicine taken during fever: Antimalarial, Quinine pills	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623f	Medicine taken during fever: Antimalarial, Quinino injection	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623g	Medicine taken during fever: Antimalarial, Coartem	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s623h	Medicine taken during fever: Antimalarial, Other	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s538c	Dihidroartemis- Piperaquine taken for fever	0 no, 1 yes, 9 missing	All Where Used
s538e	Artesunate-Amodiaquine taken for fever	0 no, 1 yes, 9 missing	All Where Used
s538g	Malaxin taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s538h	Malaritab taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s538i	Arinate taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s538j	Artesunate taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s538k	Mefloquine taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s125a	Antimalaria medicines prescribed or taken: SP/Sulphadoxine Pyrimethamine	0 no, 1 yes, 9 missing	All Where Used
s125b	Antimalaria medicines prescribed or taken: chloroquine	0 no, 1 yes, 9 missing	All Where Used
s125c	Antimalaria medicines prescribed or taken: DP/Dihydroartemisinin-Piperaquine	0 no, 1 yes, 9 missing	All Where Used

s125d	Antimalaria medicines prescribed or taken: quinine	0 no, 1 yes, 9 missing	All Where Used
s125e	Antimalaria medicines prescribed or taken: AA/Artesunate Amodiaquine	0 no, 1 yes, 9 missing	All Where Used
s125f	Antimalaria medicines prescribed or taken: artemisinin	0 no, 1 yes, 9 missing	All Where Used
s125g	Antimalaria medicines prescribed or taken: AL/Artemether-Lumefantrine	0 no, 1 yes, 9 missing	All Where Used
s412a	Combination with artemisinin taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412b	Fansidar taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412c	Chloroquine taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412d	Amodiaquine taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412e	Quinine pill taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412f	Quinine injection taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412g	Artesunate : by rectal taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412h	Artesunate : by injection taken for fever/cough	0 no, 1 yes, 9 missing	15, 28 Only
s412i	Fansidar and Amodiaquine (combined) taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s412j	Other antimalarial taken for fever/cough	0 no, 1 yes, 9 missing	28 Only
s326f	Larimal taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s326h	Arsumoon taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s326i	Falcimon taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s326j	Asaq Wintrop taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s326k	Artefan taken for fever/cough	0 no, 1 yes, 8 dk, 9 missing	All Where Used
s411a	Actipal taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411b	Larimal taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411c	Artemodi taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411d	Arsumoon taken for fever	0 no, 1 yes, 9 missing	All Where Used

s411e	Falcimon taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411f	Other ASAQ taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411g	Quinine (injection/IV) taken for fever/cough	0 no, 1 yes, 9 missing	All Where Used
s411h	Artefan taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411i	Lumartem taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411j	Other AL taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411asaq	Any ASAQ taken for fever	0 no, 1 yes, 9 missing	All Where Used
s411al	Any AL taken for fever	0 no, 1 yes, 9 missing	All Where Used
s124	Any antimalaria medicine prescribed to treat the malaria or took medicines without prescription	0 no, 1 yes, prescribed, 2, yes took without prescription	All Where Used
s311a	Fansidar taken for fever	0 no, 1 yes	All Where Used
s311b	Chloroquine taken for fever	0 no, 1 yes	All Where Used
s311c	Amodiaquine taken for fever	0 no, 1 yes	All Where Used
s311d	Quinine taken for fever	0 no, 1 yes	All Where Used
s311e	Combination with artemisinin taken for fever	0 no, 1 yes	All Where Used
s311f	Coartem taken for fever	0 no, 1 yes	All Where Used
s311g	Other antimalarial taken for fever	0 no, 1 yes	All Where Used
s623a	Medicine taken during fever: Antimalarial, combination with artemisinin (tca)	0 no, 1 yes, 8 dk	All Where Used
s623b	Medicine taken during fever: Antimalarial, Sp/fansidar	0 no, 1 yes, 8 dk	All Where Used
s623c	Medicine taken during fever: Antimalarial, Chloroquine	0 no, 1 yes, 8 dk	All Where Used
s623d	Medicine taken during fever: Antimalarial, Amodiaquine	0 no, 1 yes, 8 dk	All Where Used
s623e	Medicine taken during fever: Antimalarial, Quinine pills	0 no, 1 yes, 8 dk	All Where Used
s623f	Medicine taken during fever: Antimalarial, Quinino injection	0 no, 1 yes, 8 dk	All Where Used

s623g	Medicine taken during fever: Antimalarial, Coartem	0 no, 1 yes, 8 dk	All Where Used
s623h	Medicine taken during fever: Antimalarial, Other	0 no, 1 yes, 8 dk	All Where Used
Note: Sometimes the variable descriptions are different across surveys despite the same Variable Label. I went through all of the Variable Labels individually to assess. In instances where all different versions of the label still pertained to treatment seeking behaviour, the label was retained and used in all surveys (denote "All Where Used" in Survey Usage). Where only a subset of the label occurrences pertained to malaria treatment seeking, the variable was used only for a subset of surveys, with that information detailed in the "Survey Usage" column.			

This collation yielded 3 datasets containing identical information except for the binary indicator (which differs between datasets and are unique to the particular programmatic strategy being considered). This separation was done as many more individuals had been tested for malaria than reported recent fever occurrence or treatment seeking status and so defining the 3 separately enabled us to maximise the number of individuals included in each dataset. This means that each dataset contains all individuals for whom malaria status was determined and it was possible to define Index household status for. Definition of Index Household status in this way is contingent upon a household containing multiple individuals, and so I implicitly remove all children who are the sole respondent detailed in the surveys.

Analysis and Bayesian Regression Modelling of DHS Data

Model Formulation and Specification

To these collated DHS data, I fitted a logistic regression model relating malaria infection status to a suite of variables present in the dataset. The model was formulated as follows: I assume malaria infection status (a binary indicator taking either 0 or 1, indicating, respectively, absence and presence of a malaria infection in a given individual) for individual i, belonging to household j surveyed as part of survey k to be drawn from a Bernoulli distribution:

$$Malaria\ Infectious\ Status_{i,j,k} \sim\ \ Bernoulli(Prob\ Malaria\ Positive_{i,j,k})$$

where $Prob\ Malaria\ Positive_{i,j,k}$ represents the probability of that individual being malaria positive. I model this probability using logistic regression i.e. as being a linear function of a number of different covariates on the logit scale. It is specified in the following way:

$$logit\big(Prob\ Malaria\ Positive_{i,j,k}\big) = \alpha_k + \beta_1 Index\ Household_{i,j,k} + \beta_2 Prevalence_k + \beta_3 Household\ Size_{j,k} + + \beta_4 (Prevalence_k * Index\ Household_{i,j,k})$$

Where $Index\ Household_{i,j,k}$ is one of the three binary indicators (either 0 or 1, defined separately for each of the three programmatic strategies considered i.e. RACD, MTAT and MSAT) for individual i, belonging to household j surveyed as part of survey k describing one of three following scenarios:

- Whether anyone else in the individual's household is malaria positive (i.e. the individual resides in an Index Household as defined by the MTAT approach).
- Whether anyone else in the individual's household is malaria positive and has had fever in the past 2 weeks (i.e. the individual resides in an Index Household as defined by the MSAT approach).

- Whether anyone else in the individual's household is malaria positive and has sought treatment for malaria (i.e. the individual resides in an Index Household as defined by the RACD approach).

$Prevalence_k$ is the overall malaria prevalence recorded for survey k, $Household\ Size_{j,k}$ is the size of household j in survey k where individual i is resident, β_{1-4} are regression coefficients determining the magnitude of the influence of each variable on the probability of being malaria positive. Individually, they describe the odds ratio of being malaria positive vs malaria negative for a unit increase (for continuous variables) or an indicator is 1 (rather than 0, for binary variables), all other variables kept equal. α_k is an intercept within the regression framework used and $Prevalence_k * Index\ Household_{i,j,k}$ is an interaction term describing how the effect of $Index\ Household_{i,j,k}$ varies with the overall survey prevalence (a proxy for overall transmission intensity).

The primary interest is to assess how the influence of the three binary indicators for Index Household residency ($Index\ Household$) vary with transmission intensity (described by the survey prevalence $Prevalence$) and so I include an interaction term involving these two variables. This allows the influence of $Index\ Household$ to vary with survey prevalence, allowing me to formally test whether the odds ratio associated with $Index\ Household$ differs systematically according to transmission intensity – this formally tests the hypothesis that the association between being malaria positive and sharing a household with someone who satisfies one of the three programmatic criteria (Malaria Positivity for MTAT, Recent Malaria Fever for MSAT or Malaria and Treatment Seeking Behaviour for RACD) varies with malaria endemicity. It therefore explores the extent to which these infections (as detected by each of the different programmatic strategies) cluster and how this varies depending on the level of malaria endemicity.I fitted three separate instances of this model, one for each of the three programmatic strategies being considered (i.e. three different models utilising one of the three binary indicators).

Bayesian Hamiltonian Monte Carlo Model Fitting

This model was fitted within a Bayesian framework, with parameter inference carried out using a Hamiltonian Monte Carlo (HMC) based sampling scheme implemented in the probabilistic programming language Stan (Carpenter et al., 2017), with the model written using the R package RStan (Stan Development Team, 2022). Stan implements a Hamltonian Monte Carlo (HMC) based algorithm and a No-U-Turn Sampler (NUTS) based sampling scheme – the result of this is an adaptive HMC algorithm able to efficiently explore complex posterior densities (Hoffman and Gelman, 2011). As with other Bayesian model fitting schemes

underpinned by MCMC-based algorithms, the aim is for the sampling algorithm to generate a series of samples that correspond to a random, unbiased sample from the posterior distribution of interest – specifically to draw from the Bayesian posterior $p(\theta|y)$ where θ represent the model parameters and y the observed data. The simplest implementation of the HMC algorithm (so-called "static HMC" (Monnahan, Thorson and Branch, 2017)) works by augmenting the posterior distribution density with an independent momentum variable (λ) and then drawing from the joint density:

$$p(\lambda, \theta) = p(\lambda|\theta)\, p(\theta)$$

where θ represent the model parameters and $p(\lambda|\theta)$ is therefore the distribution of momenta conditional on a particular set of model parameters. This joint density $p(\lambda, \theta)$ defines the following Hamiltonian:

$$H(\lambda, \theta) = T(\lambda|\theta) + V(\theta)$$

where the term $T(\lambda|\theta) = -\log p(\lambda|\,\theta)$ is referred to as the "kinetic energy" (describing the distribution of momentum conditional on given parameter values) and $V(\theta) = -\log(p(\theta|y)p(\,\theta))$ is referred to as the "potential energy" (the negative log of the unnormalized posterior density) (Neal, 2011). This Hamiltonian is then used within the algorithm to generate new parameter proposals. Specifically, the HMC algorithm updates via the following steps:

1. Starting from the current vector of parameter values θ, sample a random momentum λ from a multivariate normal distribution, with $\lambda \sim MVN(0, M)$, where M is a covariance typically set to the identity matrix or estimated from warmup draws.
2. Evolve the joint system of momenta and parameters (λ, θ) (and the corresponding Hamiltonian specified by these equations) via Hamilton's equations (Betancourt, 2017). Specifically the current parameter and momentum values (λ, θ) are updated using the Leapfrog Integrator (specified by step size ϵ and number of steps L), which is used to simulate the system of ordinary differential equations describing these Hamiltonian dynamics.
3. The resulting state at the end of this simulation (λ', θ') (corresponding to applying L leapfrog steps of step size ϵ) is then used as a parameter proposal. These newly proposed parameters (λ', θ') are then either accepted or rejected via a metropolis acceptance step. Specifically, the newly proposed parameters are accepted with probability $\min\big(1, \exp\big(H(\lambda', \theta') - H(\lambda, \theta)\big)\big)$.

Stan extends this "static-HMC" algorithm to automatically optimise tuning of many of the different parameters described above. The software automatically tunes ϵ to match a targeted

metropolis acceptance rate, produces an estimate of M based on warmup samples, and via the No-U-Turn sampler (NUTS) algorithm is able to dynamically adapt L between HMC iterations (Hoffman and Gelman, 2011). Specifically, NUTS extends the static HMC procedure described above to automatically select the appropriate number of leapfrog steps through a tree-building algorithm through sequentially growing the number of leapfrog steps taken until the trajectory is detected to have turned back on itself (i.e. that a "U-Turn" has occurred). At this point, no more leapfrog steps are applied, and the resulting state is evaluated with the same metropolias acceptance step described above.

For each model, a total of 5 HMC chains, each 10,000 iterations in length, were run for purposes of model fitting and parameter inference. Half of each chain's iterations were discarded as burn-in/the adaptive phase of the sampling, leaving a total of 25,000 iterations available for inference. Non-informative normal priors were placed on all regression coefficient parameters. The Gelman-Rubin statistic was monitored to assess convergence of the MCMC chains, and in all cases, it was consistently <1.02, indicating stability of the chains and supporting the probability of convergence to the underlying true posterior distribution.

Results

From 57 Demographic and Health surveys spanning 23 African countries and the time-period 2006-2018, I extracted data for 72,498, 177,243, and 208,140 (depending on the availability of data to define the index household) children according to the RACD, MSAT, and MTAT strategies, respectively, representing 24,836, 50,590, and 59,050 patent infections **(Figure 3.1 and Table 3.1)**.The average transmission intensity recorded in these surveys varied widely, from 0.9% in the survey with the lowest recorded prevalence of patent malaria infection to 84.9% in the survey with the highest recorded prevalence.

Residing in an index household was consistently associated with increased odds of additional infections clustering within the same household (p<0.001 for all 3 strategies), and across all strategies used to define index households, a significant interaction between index household status and overall survey prevalence was observed ($P < 0.001$ in all instances), with clustering of infections becoming more prominent at low transmission levels **(Figure 3.2)**. For RACD, the odds ratio (OR) for index household residency ranged from 8.21 (7.47-9.02) at 10% overall malaria prevalence, to only 0.68 (0.60–0.77) at 80% survey prevalence of malaria. For MSAT, the OR for index household residency ranged from 15.8 (95% CI 15.2-16.5) when survey malaria prevalence was 10%, to only 2.67 (2.51-2.85) at a survey malaria prevalence of 80%. For MTAT, the OR for index household residency ranged from 10.3 (9.61-10.9) at 10% survey overall malaria prevalence, to only 1.05 (0.97-1.14) at a malaria survey prevalence of 80%.

Figure 3.1 Map of Africa highlighting the geographic distribution of countries where the Demographic and Health Surveys (DHS) used in the analyses were conducted. Data from 23 DHSs conducted across sub-Saharan Africa were utilised in the analyses presented here – countries where surveys were carried out are coloured in yellow above. The world map was obtained from the *rnaturalearth* R package (version 0.1).

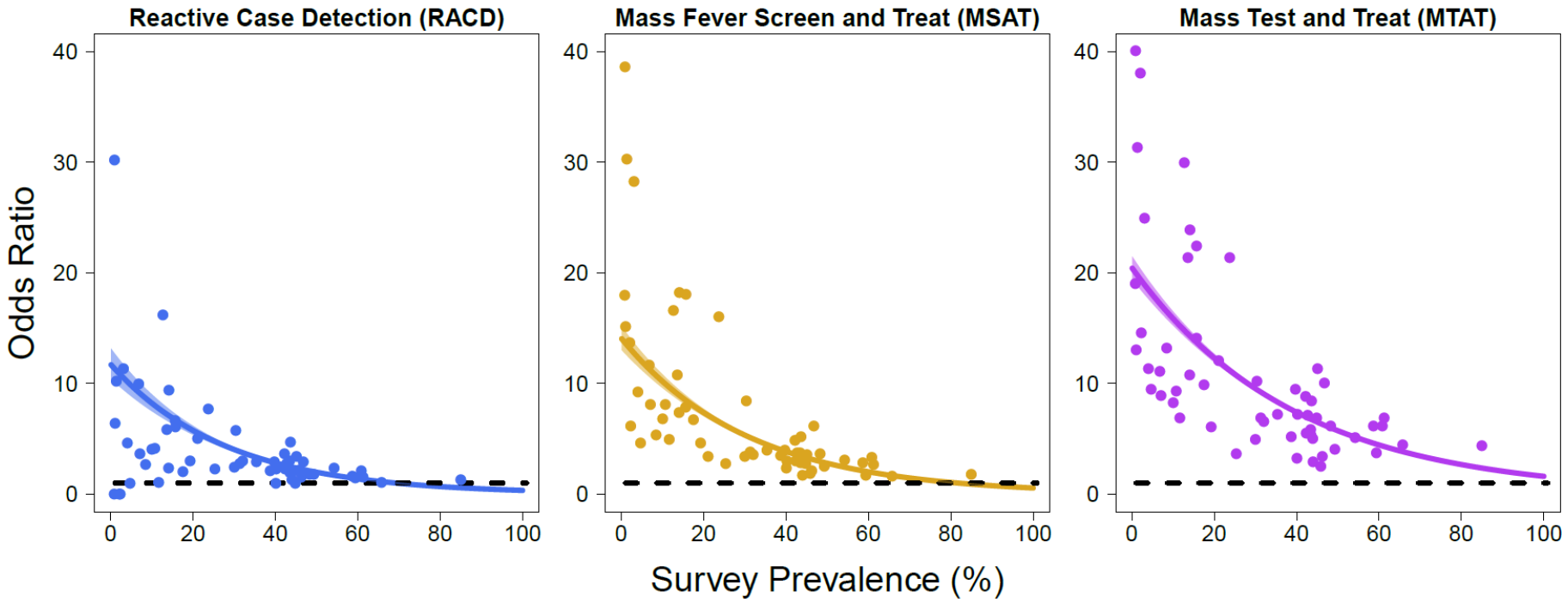


Figure 3.2 Malaria infection and the extent of household clustering according to detectability by different programmatic strategies. DHS data from 57 surveys and 23 countries detailing malaria infection status in children under 7 were collated, and the extent of infection clustering based on detectability by different programmatic strategies (RACD, MSAT or MTAT) was assessed. For each of these analyses, an individual was defined as residing in an index household if the household contained at least 1 other infection detectable through the programmatic strategy considered. The OR of being malaria positive in children who share a household with another child detectable through (A) RACD (clinical care seeking; population-level screening for infecton using routine diagnostics such as RDTs), (B) for the MSAT strategy (population-level recent fever screening), and (C) for the MTAT strategy (population-level screening for infecton using routine diagnostics such as RDTs) as compared to those who do not. Plots display the modelled OR (generated using a logistic-regression approach, coloured line) whilst points are the same ORs but calculated empirically for each survey. Pale shaded area represents the 95% Credible Interval.

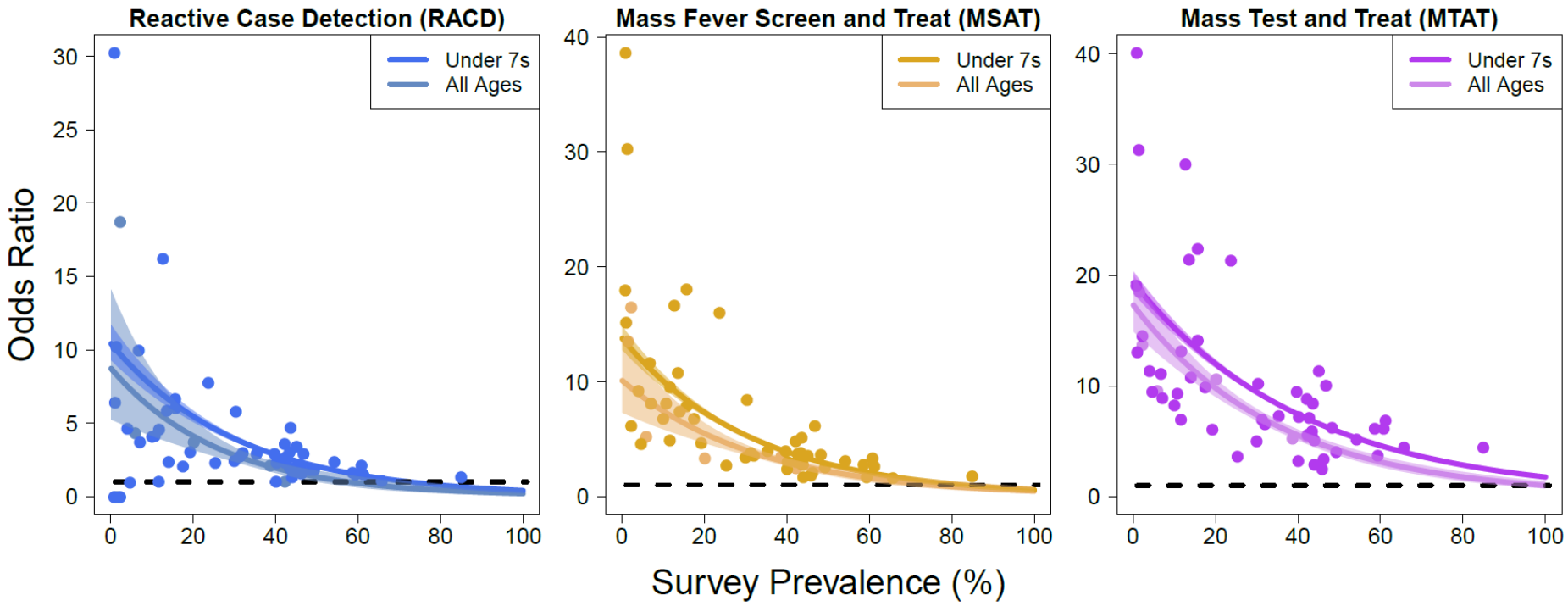


Figure 3.3 Malaria infection and the extent of household clustering according to detectability by different programmatic strategies when comparing DHS surveys carried out in all age-groups, compared to those conducted in children under 7 only. Data from the 51 DHS surveys in which only children under 7 had been tested for malaria infection (dark line with corresponding 95% credible interval in shaded area), and the 6 surveys in which all ages had been tested (lighter line with corresponding 95% credible interval in shaded area) were extracted and analysis repeated to compare the trends with the few all age surveys with the results from the larger dataset with children under 7 years of age. The dashed line represents an Odds Ratio (OR) of 1 with the dots representing the empirically estimated OR from each DHS study. Pale shaded area represents the 95% Credible Interval.

Table 3.3 Summary of the Collated Demography and Health Survey Data For Each Survey.

Survey	Country	Year	Diagnostic	Age Group	Survey Size (People's Recode)	Survey Size (Children's Recode)	Linked Across Recodes	# Included, MTAT	# Included, MSAT	# Included, RACD
1	Angola	2006	RDT	All Ages	14281	1698	1522	1436	690	150
2	Angola	2011	RDT	Under 7s	40083	8242	7621	2513	2195	891
3	Angola	2015	RDT	Under 7s	72879	14322	12710	6112	4634	958
4	Benin	2012	RDT	Under 7s	86432	13407	12259	3155	2466	441
5	Benin	2017	LM & RDT	Under 7s	73364	13589	12389	3972	3723	1127
6	Burkina Faso	2010	LM & RDT	Under 7s	81156	15044	13616	3668	3502	1156
7	Burkina Faso	2014	LM & RDT	Under 7s	39152	6841	6387	3799	3544	1790
8	Burkina Faso	2017	LM & RDT	Under 7s	36415	6061	5926	3052	2876	697
9	Burundi	2012	LM & RDT	Under 7s	23020	4267	4041	2087	2007	878
10	Cameroon	2011	RDT	Under 7s	70627	11732	10118	4822	3870	2247
11	DRC	2013	LM & RDT	Under 7s	94585	18716	16809	5413	4949	4964
12	Ivory Coast	2012	LM & RDT	All Ages	50010	7776	6744	2878	2227	932
13	Gambia	2013	LM & RDT	Under 7s	52176	8088	7488	2772	2518	821
14	Ghana	2014	LM & RDT	Under 7s	41202	5884	5463	1684	1332	339
15	Ghana	2016	LM & RDT	Under 7s	21874	3235	3051	1509	1377	706
16	Guinea	2012	LM & RDT	Under 7s	44769	7039	6262	1973	1784	825
17	Kenya	2015	LM & RDT	Under 7s	24271	3614	3389	8917	4536	2074
18	Liberia	2009	LM & RDT	Under 7s	22469	4193	3523	3702	2700	1455
19	Liberia	2011	LM & RDT	Under 7s	18970	3319	2906	1886	1544	941
20	Liberia	2016	RDT	Under 7s	21643	2956	2624	1457	1185	569
21	Madagascar	2011	LM & RDT	Under 7s	40160	6248	6015	4112	3318	577
22	Madagascar	2013	LM & RDT	Under 7s	38122	5477	5295	3575	2812	430
23	Madagascar	2016	LM & RDT	Under 7s	48349	6978	6797	3024	2582	430

24	Malawi	2012	LM & RDT	Under 7s	14022	2283	2194	941	862	279
25	Malawi	2014	LM & RDT	Under 7s	14026	2078	2020	744	681	209
26	Malawi	2017	LM & RDT	Under 7s	16495	2377	2273	1013	808	302
27	Mali	2012	LM & RDT	Under 7s	58004	10326	9290	4173	3412	577
28	Mali	2015	LM & RDT	Under 7s	39359	7749	7321	5954	5680	5680
29	Mali	2018	RDT	Under 7s	54270	9940	9034	2712	2502	588
30	Mozambique	2011	LM & RDT	Under 7s	61377	11102	10069	2673	2442	485
31	Mozambique	2015	LM & RDT	Under 7s	32399	5178	4903	2261	2006	958
32	Mozambique	2018	RDT	Under 7s	28516	4579	4326	2244	1981	633
33	Nigeria	2010	LM & RDT	Under 7s	29950	5978	5288	3565	3129	1847
34	Nigeria	2015	LM & RDT	Under 7s	37540	6524	6222	3893	3552	1777
35	Nigeria	2018	LM & RDT	Under 7s	181737	33924	29482	6776	6306	2204
36	Rwanda	2008	RDT	All Ages	32029	5489	5067	3701	3238	1356
37	Rwanda	2010	LM & RDT	All Ages	55718	9002	8361	2920	2328	1583
38	Rwanda	2015	LM & RDT	All Ages	53898	7856	7385	2208	1719	1697
39	Rwanda	2017	LM & RDT	All Ages	19484	2946	2797	2181	1562	547
40	Senegal	2008	LM & RDT	Under 7s	90586	15595	13920	3328	3137	1600
41	Senegal	2010	LM & RDT	Under 7s	77056	12326	11080	3944	3496	1464
42	Senegal	2012	LM & RDT	Under 7s	41400	6862	6275	6406	5785	1968
43	Senegal	2014	LM & RDT	Under 7s	40540	6842	6238	5652	5178	1299
44	Senegal	2015	LM & RDT	Under 7s	41681	6935	6327	5679	5191	1737
45	Senegal	2016	LM & RDT	Under 7s	41393	6725	6173	5492	5037	1425
46	Senegal	2017	LM & RDT	Under 7s	78492	12185	11112	8468	7991	3259
47	Sierra Leone	2016	LM & RDT	Under 7s	39836	6213	5753	5389	3990	1364
48	Tanzania	2007	Unclear	Under 7s	44656	7502	6746	3795	3461	739
49	Tanzania	2012	LM & RDT	Under 7s	53325	8648	7980	4699	4242	1164
50	Tanzania	2015	LM & RDT	Under 7s	63851	10233	9313	7128	5779	1453
51	Tanzania	2017	RDT	Under 7s	46950	7688	7162	4266	3858	997
52	Togo	2013	LM & RDT	Under 7s	45286	6979	6443	2525	2068	883

53	Togo	2017	LM & RDT	Under 7s	22340	3415	3281	1750	1632	473
54	Uganda	2009	LM & RDT	Under 7s	21121	4012	3560	2724	2416	1280
55	Uganda	2014	LM & RDT	Under 7s	26995	4728	4345	3323	2919	1091
56	Uganda	2016	RDT	Under 7s	88642	15522	13733	3555	2761	1425
57	Burundi	2016	LM & RDT	Under 7s	77425	13192	12231	4540	3723	2757

Whilst the majority (n = 51) of surveys had only assessed malaria infection status in children under 7, a small number of surveys (n = 6) had assessed malaria status in individuals of all ages. In these surveys, it was therefore possible to defined Index Household status for individuals older than 7. I therefore replicated the analyses retaining these individuals (spanning a wider range of ages) and compared the results of these analyses to the inferences drawn from the full range of surveys. Inclusion of these older individuals did not qualitatively affect the inferences presented here, and there was a significant effect of *Index Household* residency ($P<0.001$) and a significant interaction between *Index Household* and Survey Prevalence ($P<0.001$) for each of the three analyses of RACD, MTAT and MSAT respectively **(Figure 3.3)**.

Discussion

Collating data from over 200,000 individuals across 23 countries in sub-Saharan Africa, I explored the degree and extent of the household clustering of malaria infections across a range of different settings, and for a range of different programmatically relevant definitions of what constitutes a detectable malaria infection. The results highlight extensive clustering of malaria infections at the household level, for all programmatic definitions considered – whether defining malaria infection detectability based on that infection being patent and detectable by RDT/LM (as in the case of MTAT), based on both symptoms and detectability by RDT/LM (as in the case MSAT), or based on the detectable infection, fever and seeking treatment at health-centres (in the case of RACD). The results also highlight an important interaction between malaria transmission intensity and household clustering, with the degree of clustering becoming more pronounced as transmission intensity declines (i.e., a larger proportion of infected individuals within a population are clustered in fewer households). Particularly in these contexts, giving all residents of an *Index Household* (as defined by the programmatic strategy being employ) curative doses of antimalarial therapeutics may therefore provide malaria control programs with an option to easily target infections that may not otherwise be detected.

Importantly however, the appropriateness and viability of such strategies will depend intimately on the transmission intensity, and the associated resource and operational requirements this imposes. These results have highlighted that malaria infections tend to cluster (though to different degrees) in all but the settings of highest (>65%) prevalence, which in theory suggests that household-based strategies to identify and treat infections could be appropriate across a wide range of settings. The operational feasibility of such approaches is likely to be limited in the settings with the highest transmission however (Bannister-Tyrrell et al., 2019; Mlacha et al., 2017). In these settings, where the majority of the households are likely to have at least 1 detectable infection, uniformly applied interventions (such as

insecticide treated bednets) are typically considered more appropriate (Stresman et al., 2018). Moreover, because the absolute number of infections is higher, and the degree of clustering reduced, targeted strategies are likely to identify a smaller proportion of the overall infectious population compared to low transmission settings (Stresman et al., 2020). In settings with low transmission (e.g. between 1 and 10% prevalence of malaria infection) however, where such strategies become operationally feasible, the results presented here suggest that reactive strategies such as those considered here may well be able to identify and target a significant fraction of the infected population, and in doing so, materially impact onwards transmission (Stresman et al., 2020).

These analyses however are subject to a number of important limitations. Firstly, malaria infection in the DHS surveys was diagnosed using non-molecular methods i.e. either rapid diagnostic tests or light microscopy. These methods will therefore not detect individuals with submicroscopic malaria infections (i.e. those where parasite densities are below the limit of detection by RDT/LM and are undetectable with these methodologies, but which would be detectable with more sensitive molecular methods such as PCR). These infections have previously been shown to be infectious (Gonçalves et al., 2017; Slater et al., 2019) and my work in Chapter 2 has highlighted that they can constitute a significant fraction of the infectious reservoir in certain settings, particularly in the areas of lowest transmission which the results presented in this Chapter suggest are most amenable to reactive, household based control strategies. It is unclear therefore whether such clustering would be apparent when applied to the totality of malaria infections present, rather than just those detectable by typically employed diagnostics – and in turn, the comparative viability and utility of the different types of reactive based control measures.

Relatedly, in the vast majority of surveys conducted, malaria infection status (and by extension, occurrence of recent fever or treatment seeking behaviour) was limited to children, specifically those under 7. Because of this constraint, I am unable to consider the full complement of individuals residing in households in my analysis, and by extension, it remains unclear the degree to which the results presented here would generalise to mixed populations including both children and adults. However, the results presented here are consistent with a recently conducted systematic review of the literature that reviewed these programmatic strategies in a similar manner (and which was combined with the results of the analyses presented here and subsequently published (Stresman et al., 2020)) and that collated studies in which infection had been diagnosed using molecular methods in all resident members of households (Stresman et al., 2020). More generally, it is also important to note that reactive strategies such as RACD are mainly appropriate when peri-domestic transmission is the key

driver of malaria infection. Such modalities of transmission are common across sub-Saharan Africa, but other drivers of exposure dominate in other settings (such as Cambodian or Amazonian forests where much exposure is occupationally related (Saavedra et al., 2019; Rossi et al., 2018b)). In these settings, the appropriateness and viability of household-based approaches is far from clear, and ultimately reactive strategies need to be designed in a manner sensitive to the underlying local transmission context.

Despite these limitations however, these results highlight that where peri-domestic transmission dominates, the degree of household clustering of malaria infections is highly variable across the malaria endemicity spectrum, with the extent of this highest in the areas with the lowest transmission intensity. Together this work provides systematically collected, empirical evidence supporting the potential effectiveness of reactive, household-based approaches to controlling malaria transmission in the settings of lowest transmission, in-keeping with both the results of multiple trials (Rossi et al., 2018b) as well as recent meta-analyses comparing different approaches across the endemicity spectrum (Stresman et al., 2020). Such control strategies can potentially enhance elimination efforts in these settings particularly when combined with effective vector control and wider strengthening of existing health systems, which will be crucial in ensuring a sufficient proportion of all infections can be targeted.

Conclusion

In this Chapter, I have explored the spatial clustering of microscopically detectable malaria infections, how the extent of this clustering varies depending on overall levels of malaria transmission, and what this heterogeneity means for the appropriateness of different control measures. Both this Chapter and Chapter 2 have been focussed primarily on patterns of heterogeneity relating to the human host, specifically those surrounding spatial variation in risk of infection (this Chapter) and infectivity/detectability of these infections (Chapter 2). In Chapters 4 and 5, I will turn to instead focus on the anopheline mosquito populations that underpin transmission and explore entomological determinants of heterogeneity in the malaria parasite's transmission dynamics. Unlike Chapters 2 and 3, where work has primarily focussed on understanding spatial patterns of heterogeneity, these next chapters will be focussed on exploring the factors influencing the degree of seasonality in anopheline mosquito populations, to better understand the factors influencing temporal heterogeneity in the abundance of mosquito populations and its impact on the underlying temporal profile of malaria risk.

Chapter 4 The Ecological Structure of Mosquito Population Seasonal Dynamics

Populations of the anopheline mosquitoes responsible for malaria transmission are highly dynamic, frequently exhibiting substantial temporal (often seasonal) fluctuations in size that shape the profile of malaria risk. Understanding these dynamics is a crucial input to optimising certain control strategies (such as seasonal malaria chemoprevention or indoor-residual spraying), with impact depending on the ability to time delivery in relation to seasonal peaks in disease risk. In this Chapter, I develop a statistical framework enabling characterisation of the temporal patterns displayed by different mosquito species complexes and identification of "dynamical archetypes" sharing similar temporal properties. I apply this framework to temporally disaggregated anopheline catch data spanning 7 species and 117 unique locations across India collated via a systematic literature review. Integrating the collated data with this developed framework, I identify substantial heterogeneity in the extent and nature of seasonal dynamics both within and between different species and explore ecological factors driving this variation.

Introduction

Malaria transmission is underpinned by mosquito vectors belonging to the *Anopheles* genus – these vectors are heterogeneously distributed across the globe (Warrell and Gilles, 2017; Hay et al., 2010) and display marked variation in their vectorial capacity and bionomics (such as their propensity to bite humans vs other animals) (Massey et al., 2016). This results in marked differences in the transmission dynamics of malaria across different ecological contexts. To date, significant work has focussed on characterising the global spatial distribution (presence/absence) of these malaria vectors (and other mosquitos relevant to public health) (Sinka, 2013; Sinka et al., 2012). This work represents a vital input to surveillance and control programmes aimed at mitigating the impacts of vector-borne diseases worldwide. By contrast, less attention has been paid to understanding the temporal patterns of vector abundance, and how these dynamics are shaped by the local environment. Mosquito populations are highly temporally dynamic, exhibiting substantial annual fluctuations in size that drive the temporal profile of disease risk (Koenraadt, Githeko and Takken, 2004; Das et al., 2017). Understanding the determinants of these dynamics is important given that the efficacy of many malaria control interventions (such as seasonal malaria chemoprevention (Wilson and IPTc Taskforce, 2011; Ross et al., 2011) and indoor-residual spraying (Pluess et al., 2010)) depends on the timing of their delivery in relation to seasonal peaks in risk. Effective utilisation of these interventions will be vital for achieving the goals of the World Health

Organisation's "High Burden, High Impact" strategy, which aims to substantially reduce/eliminate malaria in India and the ten African nations with the highest global burden.

Despite their importance, many questions remain surrounding the drivers of mosquito population dynamics. A close relationship has been observed between rainfall occurrence, peaks in mosquito populations and malaria cases (Cairns et al., 2012) including for *Anopheles gambiae s.l.* (Appawu et al., 2004; Okello et al., 2006; White et al., 2011) across several African settings and *Anopheles dirus s.l.* across India and south-East Asia (Obsomer, Defourny and Coosemans, 2007); in-keeping with the aquatic breeding of mosquitoes and the preferences some species display for transient, rain-fed pools of water in which to breed (Gimnig et al., 2001). However, studies of *Anopheles funestus s.l.* populations have identified varying degrees of seasonality (Cohuet et al., 2004; Mendis et al., 2000) including population abundance peaking in the dry season (Matowo et al., 2021). Relatedly, for *Anopheles annularis s.l.,* a number of studies have demonstrated only limited seasonal peaks (despite highly seasonal rainfall), with the species detected in significant numbers over the course of the entire year even in periods when other major vectors (such as *Anopheles culicifacies s.l.*) are largely absent (Das et al., 2017; Singh et al., 2013; Das et al., 2011). This brings into question how generalisable relationships between rainfall and mosquito population dynamics are. The influence of other factors such as temperature (which has a marked influence on many mosquito traits including larval development (Bayoh and Lindsay, 2003), biting rates and mortality rates (Shapiro, Whitehead and Thomas, 2017)) remains similarly unclear. Recent field-based work has suggested that considerations of both rainfall and temperature are necessary to understand seasonal patterns of malaria incidence (Beck-Johnson et al., 2017). However, these analyses have been restricted to a small number of settings across sub-Saharan Africa; leaving the influence of temperature regimen on mosquito population dynamics largely unexplored in other ecological settings. Previous work has also suggested a potential role for numerous other ecological factors in shaping mosquito population dynamics, including land-use (such as irrigative practices (Sang et al., 2016) or structure of the built-environment in urban settings (Thomas et al., 2016)) or the local hydrological environment and presence of long-lived water bodies (Minakawa et al., 2012, 2008) (which potentially provide opportunities for breeding year-round).

Altogether, these results highlight outstanding questions surrounding the drivers of mosquito population dynamics. Whilst numerous entomological studies of *Anopheline* seasonality have been carried out, focus is typically on a single species and/or location –such studies are rarely gathered and synthesised together to identify generalisable patterns and facilitate systematic comparisons across key vector species. Using India as a case study, I collate a dataset of

temporally disaggregated mosquito catch data from across the country to better understand variation in mosquito population dynamics, the factors underlying this variation, and how dynamics vary across different species. I develop statistical methodologies enabling characterisation of the temporal patterns displayed by different mosquito species complexes and identification of "dynamical archetypes" sharing similar temporal properties. This work reveals pronounced heterogeneity in the extent and nature of seasonal dynamics, both between species complexes and across different locations. In doing so, these results highlight the importance of considering both species composition and ecological structure when implementing interventions aimed at controlling vector-borne diseases.

Methods

Systematic Review of Indian Entomological Literature

Web of Science and PubMed databases were searched on 17th October 2017 using the keywords "India" AND "Anophel*" to identify references with temporally disaggregated entomological data. I identified 1945 records with 1556 remaining after removing duplicates. References were selected for Inclusion/Exclusion according to the following criteria:

Inclusion Criteria:

- Reference contains temporally disaggregated adult mosquito catch data at a temporal resolution of monthly or higher.

Exclusion Criteria:

- Mosquito catch data is not temporally disaggregated to a sufficient extent (e.g. catches were done yearly or seasonally rather than monthly).
- Mosquito catch data was collected as part of a trial assessing a vector control intervention (which would perturb the natural dynamics of the vector, rendering the data unrepresentative of the population dynamics in the absence of control).
- Reference only contains information on immature/larval mosquito life cycle stages.
- Reference contained insufficient information to geolocate the area in which the study was conducted.

Following Title and Abstract screening 281 records were retained for full text evaluation. I included records containing temporally disaggregated adult mosquito catch data with monthly (or finer) temporal resolution spanning at least 12 months that had not been conducted as part of vector control intervention trials, and where sufficient information to geolocate the catch site was provided. 78 references were retained that yielded 117 geolocatable areas across India. These references contained 272 time-series spanning the malaria vectors *Anopheles*

annularis s.l., *culicifacies s.l.*, *dirus s.l.*, *fluviatlis s.l.*, *minimus s.l.*, *stephensi s.l.* and *subpictus s.l.* and spanning 5 collection methods.

Data Extraction, Collation and Initial Processing

Entomological Data Extraction

For each reference, I extracted all relevant entomological catch data detailed. I restricted extraction to 7 major *Anopheles* species known to be relevant to malaria transmission in India (although a number of others exist) and for which multiple catch data time series were available. These were *Anopheles annularis* (Dev and Sharma, 2013), *Anopheles culicifacies* (Singh et al., 1999), *Anopheles dirus* (Dutta et al., 1996; Prakash et al., 1997), *Anopheles fluviatilis* (Nanda et al., 2012; Tripathy et al., 2010), *Anopheles minimus* (Dev and Manguin, 2016; Dev, 1996), *Anopheles stephensi* (Korgaonkar et al., 2012) and *Anopheles subpictus* (Kumar et al., 2016). Where data were presented in the form of a table, data was copied directly from the table. Where graphs only were presented, estimates of the data were extracted using DataThief™ software. This yielded a total of 305 time series of monthly mosquito catch data, ranging in length from 5 – 46 months. I restricted subsequent analyses to time series that spanned a year (12 timepoints, monthly) or longer, a total of 272 time series. This yielded the following number of time series for each of the species considered:

Table 4.1 Number of time series collated for each species

	annularis	***culicifacies***	***dirus***	***fluviatilis***	***minimus***	***stephensi***	***subpictus***
# Time-Series	39	85	11	60	12	27	38

As the primary focus of this research was to explore annual and seasonal patterns of mosquito population dynamics, as well as the fact that variations in time series length are a factor known to affect their statistical properties (Fulcher, Little and Jones, 2013) (and which would therefore impact the comparability of the time series gathered and analysed here), all time series were standardised to be 12 months in length. For time series containing more than 12 time points (i.e. time series that spanned longer than a single year), I averaged the recorded catches for a given month. Where the study has been initiated in a month other than January, and concluded in a month other than December, the recorded counts were rearranged to yield a complete time series running from January to December. The studies analysed here employed a wide array of different sampling methodologies including Indoor and Outdoor Resting Collections, Human Landing Catches, Spray Catches and Trap Catches amongst others.

Table 4.2 Number of time series collated according to method of collection

	Landing Catch	Resting Collections	Pit Collections	Light Traps	Spray Catches
# Time Series	41	194	5	15	5

The majority of studies carried out were resting collections – within each of the different catch methods however, there was further variation according to the location the catch was carried out in (typically human dwellings or cattlesheds), the timing (daytime, night-time or overnight) and (where relevant) the bait used (typically either cattle or humans). Note that when summed (260), these values do not correspond to the number of time-series used (272) as in a small number of cases, multiple sampling methods were used, and the results not disaggregated (and have therefore not been counted for the purposes of the table above).

Results were typically, though not always, presented in the form of some sampling-effort standardised measure such as Man Hour Density (MHD). As such, though reflective of mosquito population dynamics, these measures do not represent the overall number of mosquitoes caught. To this end, where information on sampling effort (number of hours spent sampling, number of households/cattlesheds searched, number of human baits, number of traps set etc) was present, I used this information to convert MHD back to the raw counts. In the small number of instances where there was variable sampling effort across the time series (which would bias the conversion away from the true underlying population dynamics), I conservatively used the lowest sampling effort recorded across the time series in the conversion. Together, this allowed me to produce an estimate of the number of mosquitoes sampled (a raw count, based on equal sampling effort across the time series).

Environmental Covariate Assembly

For each of the 117 study locations I extracted a suite of environmental variables derived from satellite data that together describe the location's ecological structure. These include time-period and location specific rainfall data from The Climate Hazards Group Infrared Precipitation With Stations (CHIRPS) dataset (Funk et al., 2015), BioClimatic variables (a suite of biological relevant covariates defined from monthly rainfall and temperature satellite data (Fick and Hijmans, 2017)), various measures of aridity (Zomer et al., 2008), a number of covariates describing the seasonality and extent of water bodies (Lehner and Döll, 2004), landcover (Friedl et al., 2010) and a number of other variables previously used in defining the global distribution of anopheline vectors (Sinka et al., 2011b). The environmental covariates (i.e. the independent variables that, along with species, are used to predict the different seasonal patterns) used in this research consist of raster layers spanning all of India at a 2.5

arc-minute (~ 5km by 5km) spatial resolution. The covariates utilised here were initially selected from a set of 66 covariates derived from:

- Covariates previously used in other *Anopheles* mosquito mapping efforts (Sinka et al., 2010a), as well as in other mapping efforts looking at the spatial distribution of the malaria parasite, *Plasmodium falciparum* (Bhatt et al., 2015b).
- Consideration of some of the possible drivers of seasonal dynamics (primarily hydrological considerations surrounding the seasonality and availability of aquatic breeding sources, and how this might interact with environmental composition (Gimnig et al., 2001; Mattah et al., 2017) and species specific breeding preferences (Singh et al., 2014; Amerasinghe, Indrajith and Ariyasena, 1995) to structure population dynamics). From these considerations, a number of other raster layers were included that together describe further the underlying hydrological environment.

The majority of these covariates are derived from high temporal resolution satellite images that were initially gap-filled (Weiss et al., 2014) to eliminate missing data that typically arises from cloud cover. These images were then aggregated and summarised to produce a suite of synoptic environmental covariates for prediction. From these 66 covariates (a number of which are highly correlated with one another), a reduced subset of 25 covariates were selected. These were selected in the following way. Firstly, covariates were grouped into one of five categories based on the ecological features they were describing. These categories were Temperature, Rainfall, Aridity, Hydrological and Landcover. A subset of covariates were then selected in each category in order to minimise the correlation between covariates (based on correlation matrices of Spearman correlation coefficients) whilst also retaining measures of important quantities such as the mean, the dispersion etc for a given category. Based on this, the final covariates included in each category were the following:

- **Temperature:** Annual Mean Temperature, Temperature Seasonality & Mean Temperature in the Driest Quarter (3 covariates).
- **Rain:** Annual Rain, Rain Seasonality, CHIRPS Minimum and Rain in the Coldest Quarter (4 covariates).
- **Aridity:** Specific Humidity Standard Deviation, Tasseled Cap Wetness Standard Deviation, Tasseled Cap Brightness Standard Deviation (3 covariates).
- **Hydrological:** Water Areas Occurrence, Water Areas Recurrence and Flow Accumulation (3 covariates).
- **Landcover:** Dominant Landcover and City Accessibility (2 covariates).

together, comprising a total of 25 covariates (Dominant Landcover consists of 11 classes of landcover type). This reduction reduced the extent of multicollinearity in the covariates and reduced the scope for model overfitting. The association between each of these environmental covariates and membership of each dynamical archetype/cluster was then assessed using a Bayesian, regularised multinomial logistic regression-based framework (described in further detail below), that analyses the relationship linking each covariate to cluster/archetype membership whilst controlling for all other covariates included in the model.

Study Geolocation and Environmental Covariate Extraction

Geolocation of study areas was possible to a varying degree depending on the information available within the paper (and related literature). When villages names or the details of the administrative unit a study was carried out in were provided in the paper text, geolocation was carried out utilising a wide array of resources containing spatially explicit information on the location of Indian settlements and administrative units. These were Google Maps/Google Earth, Etrace, OneFiveNine, Veethi, Wikimapia, VillageInfo, MapsOfIndia, Geonames and AlipurduarTourism. Additionally, a number of the references identified in my review had previously been utilised as part of the Malaria Atlas Project (MAP) Presence/Absence mapping work and so had previously been geolocated (Sinka et al., 2011b). In these instances, the MAP location estimate was used. The precision of study location estimates varied greatly (due to the extent of spatial detail provided in the paper e.g. village vs district as well as the identifiability of villages/administrative units) – this uncertainty is explicitly incorporated into the analyses, with raster covariates extracted over the full area the study is believed to have been carried out in, and then the average of those raster values used.

In addition to the environmental covariates detailed above, for each of the 117 geolocated study locations, daily rainfall estimates specific to the location and time-period the study was conducted in were also collated. These data were taken from “The Climate Hazards Group Infrared Precipitation With Stations” (CHIRPS) dataset (Funk et al., 2015) and were subsequently aggregated up to the same temporal resolution as the mosquito catch data (i.e. monthly). Data from the CHIRPS dataset is only available from the year 1981, and so for locations where the sampling date predated this, daily rainfall data was extracted for the year 1981, and assumed to be representative of past rainfall. These rainfall data were used to calculate the cross-correlation coefficient between mosquito catches and rainfall.

Maps of Vector Presence/Absence

Extensive work has previously been undertaken mapping the distributions of key *Anopheline* vectors across Africa, the Middle East and Europe (Sinka et al., 2010a), the Americas (Sinka et al., 2011a) and the Asia and Pacific region (Sinka et al., 2011b). These maps describe the

probability of occurrence at a 5km by 5km resolution for many of the dominant vector species involved in malaria transmission. Here, I utilise updated versions of these maps that include presences up to the year 2016 as part of work conducted with the Humbug Project (http://humbug.ac.uk/), funded through a recent Google Impact Challenge grant. These maps describe the probability of occurrence for the species *An. annularis, An. culicifacies, An. dirus, An. fluviatilis, An. minimus, An. stephensi* and *An. subpictus*. These estimates of occurrence probability were then integrated with a multinomial logistic regression model of dynamics to generate estimates of the probability of a given location containing a particular temporal pattern/profile (see section **Penalised Multinomial Logistic Regression Modelling, Evaluation of Model Accuracy and Predictive Modelling** below for further technical details).

Time-Series Fitting and Interpolation

Negative Binomial Gaussian Process – Fitting and Inference:

I use a highly flexible stochastic process model, known as a Gaussian Process, to temporally interpolate between the monthly catch datapoints and integrate over uncertainties in the estimates of mosquito abundance (a product of both the catch methodology as well as generic random variation in the sampling of the mosquito population) spanning the entire year. Gaussian processes specify a distribution over functions such that any finite set of function values $f(x_1), f(x_2), \dots f(x_N)$ have a joint Gaussian distribution (Rasmussen, 2004). The Gaussian process is entirely specified by its mean function, defined as:

$$E[f(x)] = \mu(x)$$

and by its covariance function:

$$Cov[f(x), f(x')] = k(x, x')$$

also known as the kernel. This kernel is a positive-definite function of two inputs, x and x' that defines the covariance between any two points (and by extension the covariance matrix of our Gaussian Process when all pairwise combinations of points are considered). In doing so, the kernel encodes prior information about the extent to which I would expect two objects (x and x' in this instance) to be similar. A wide array of kernels exist that specify an equally wide array of similarity structures, such as the squared exponential (where similarity varies with the Euclidean distance separating x and x') and the linear kernel (which allows the relationship governing similarity to vary with not just the relative position of two inputs, i.e. $x - x'$, but with their absolute position, a property that makes this kernel "non-stationary"). Given the strong seasonality known to be present in mosquito catch time series and from the empirically observed patterns of abundance observed when examining the raw time series, I selected a

Periodic Kernel. This kernel defines similarity based on the distance between x and x' compared to some period p and so is able to accommodate patterns that broadly repeat themselves over time (such as seasonal or annual peaks in mosquito abundance).

$$k(x, x') = \alpha^2 \exp\left(-\frac{2}{l^2} sin^2\left(\frac{\pi|x - x'|}{p}\right)\right)$$

Where the p represents the period, α specifies the magnitude of the covariance given a certain period, and l represents a lengthscale parameter further constraining the extent to which two values separated by a given distance can co-vary with one another. Bayesian inference and fitting of Gaussian Processes typically utilises the following hierarchical formulation:

$$\theta \sim \pi(\theta)$$

$$f \sim GP(0, K_\theta(x))$$

$$y_i \sim MVN(f(x_i), \sigma^2)\ \forall i\ \in \{1, \dots, N\}$$

where θ represents a vector of hyperparameters involved in defining the kernel's properties, f is a distribution of functions from a zero-mean Gaussian Process with covariance function K_θ, f(x) are function evaluations at times x, and y our observed counts. However, mosquito catch data is rarely normally distributed (leaving aside limit theorems) and frequently displays high levels of overdispersion (Boussari et al., 2012), a common property of biological systems generally, but made more acute by the fact that for a number of the time series, the monthly catches reported represented the summed total of multiple catches made throughout the period (but that were not presented in the paper, where only monthly totals were presented); this process of summation also introduces overdispersion. Motivated by this, I adapted the above framework to accommodate a Negative Binomial likelihood, leading to the following inferential framework:

$$\theta \sim \pi(\theta)$$

$$f \sim GP(0, K_\theta(x))$$

$$y_i \sim Negative\ Binomial\left(e^{f(x_i)}, \sigma\right)\ \forall i\ \in \{1, \dots, N\}$$

where the exponential function e^x is used to reflect the fact that I use a log link between the observed counts and the underlying latent process reflecting the population dynamics, and σ represents the overdispersion parameter of the Negative Binomial distribution.

Prior Probability Specification

Prior distributions for the estimated parameters were defined as follows:

$$l \sim Normal(2, 1^2)$$

$$\alpha \sim Half - Normal(0, \sqrt{SD\ (y)})$$

$$p \sim Normal(12, 4^2)$$

$$\sigma \sim Half - Normal(0, 8^2)$$

Weakly informative priors were set on the scaling factor α, the period, p, and the overdispersion parameter, σ. The period prior was centred around 12 (a value which would represent annual variation) to reflect the fact that the majority of observed variation in mosquito abundance recorded has typically been observed to cycle annually due to annual variation in key ecological factors such as rainfall and ambient temperature. A wide standard deviation was used however in to allow the model to identify and accommodate instances of bimodality or periods operating across timescales longer than a year, although important to note is that the lower and upper bounds for the period were set to 4 and 18 months respectively, to avoid identifiability issues arising from the lack of data at temporal resolutions substantially below and above these bounds. A similarly wide prior was set over the overdispersion parameter σ and the scaling factor α. An informative prior was set for the lengthscale l, although the use of less informative priors, either for the lengthscale or for the period, did not significantly alter conclusions arising from the analysis (see **Figure 4.5**), highlighting the robustness of the results presented in the results section of this chapter.

Model Fitting and Parameter Inference

This Negative Binomial Gaussian Process were fitted using STAN a probabilistic programming language utilising a HMC-based sampling algorithm in conjunction with the No-U-Turn sampler for Bayesian inference (Hoffman and Gelman, 2011). The model specified above was implemented in R using the rStan package (Carpenter et al., 2017). For each time series, 4 chains of 5,000 iterations were run for purposes of model fitting and parameter inference. Half of each chain's iterations were discarded as burn-in/the adaptive phase of the sampling, leaving a total of 10,000 iterations available for inference. The Gelman-Rubin statistic was monitored for each model fitting to assess convergence and in all cases, estimates were consistently <1.02, indicating stability of the chains and probable convergence to the underlying true posterior distribution.

Fitted Time Series Normalisation and Von Mises Distribution Fitting:

Following this fitting process, and to establish comparability across the time series (which varied substantially in the absolute count numbers recorded and used a wide and highly

heterogeneous array of different sampling methods), I normalised each time series in the following way:

$$p_i = \frac{y_i}{\sum y_i}$$

where p_i is the normalised count for timepoint i and y_i is the un-normalised count for timepoint i as predicted from the Negative Binomial GP fitting described in the previous section.

To further characterise the periodic properties of these time series, I fit a Von Mises distribution, which is a continuous probability distribution on the circle with range from 0 to 2π. Broadly, it can be regarded as the circular analogue of the normal distribution on the line, with the probability density function for the angle x given by:

$$f(x|\mu,\kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}$$

where $I_0(\kappa)$ is the modified Bessel function of order 0, the parameter μ is a measure of location (analogous to the mean of the normal distribution, describing where on the circle the distribution is clustered around) and κ describes the concentration of density around μ (and thus its inverse is a measure of dispersion, analogous to σ^2 for the normal distribution.

I fit two sets of Von Mises densities to the normalised time series, the first containing a single component, specified as:

$$f(x|\mu_1,\kappa_1) = f_1(x|\mu_1,\kappa_1)$$

And the other possessing two components (sometimes called a mixture), formulated as:

$$f(x|\mu_1,\kappa_1,\mu_2,\kappa_2,w) = \omega f_1(x|\mu_1,\kappa_1) + (1-\omega) f_2(x|\mu_2,\kappa_2)$$

where x in both instances represents the normalised monthly count formulated as a random variable on the circle, i.e. by defining $x = \frac{2\pi p_i}{12}$. Fitting was carried out in R using the *optim* function and with the sum of squares as the loss function. The outputs arising from this fitting – the comparative suitability of the one and two component distributions, as well as the values of ω, μ and κ, were then explored to further characterise the temporal properties of the data.

Time-Series Characterisation and Clustering by Features

Motivated by previous work providing a framework to statistically characterise the empirical structure of time-series data (Fulcher, Little and Jones, 2013) and work characterising the

seasonality of malaria case incidence (Nguyen et al., 2020), I calculated several summary statistics for each smoothed time-series to characterise their temporal properties, specifically the following:

1. **Kullback-Leibler Divergence:** Also known as the relative entropy, the Kullback-Liebler divergence represents a measure of how different one probability distribution is from a second probability distribution (where a value of 0 indicates that the two distributions are identical). It is specified in the following manner:

$$E_i = p_i log_2\left(\frac{p_i}{q_i}\right)$$

$$E = \sum_{i=1}^{12} p_i log_2\left(\frac{p_i}{q_i}\right)$$

where p_i is the average value of the normalised time series for month i, and q_i = 1/12 for $i = 1, \ldots, 12$. This operation therefore measures the deviation of a normalised time series from a uniform distribution, in doing so, informing about the extent to which a seasonal peak (or peaks) is present in the time series.

2. **Periodic Kernel Median:** Fitting the Negative Binomial Gaussian Process with a periodic kernel allowed inference of the period, p, providing us with an estimate of the frequency of repeating patterns in the monthly abundance of mosquitoes. An estimate of p was calculated for each fitted time series using the MCMC samples, and the median value of p based on these samples used.
3. **Proportion of Points Greater Than 1.65x the Mean:** For each fitted, normalised time series, the proportion of points greater than 1.65x the mean of the time series was calculated. This informs about the extent to which the data is peaked, as well as the width of the peak.
4. **Peak Distance from January:** For each fitted, normalised time series, the maximum recorded value was noted and the distance of this value from January was calculated.
5. **Number of Peaks:** Estimates of the parameters governing the fitted two component Von Mises distribution were used to infer the number of peaks in each time series. Specifically, a time series was deemed to possess one peak if the value of the Von Mises component weighting was either < 0.3 or > 0.7 and the difference in means was $< \frac{2\pi}{3}$ or $> \frac{4\pi}{3}$, indicating that the majority of the density could be attributed to one of the two components, and that the two means identified during the fitting were temporally close to one another. Otherwise, a time series was judged to possess two peaks.

6. **Von Mises 1 Component Mean:** This operation is based on the number of peaks inferred from fitting of 1 and 2 component Von Mises distributions to the Negative Binomial GP fitted, normalised time series. If a 1 component Von Mises distribution was preferred, then the Von Mises mean corresponding to the maximum likelihood predicted value was used. If the 2 component Von Mises distribution was preferred, the value for this operation for that particular time series is set to -5.
7. **Von Mises Two Component Weight:** Estimates of the weight parameter governing the two component Von Mises distribution were also used to infer the bimodality of the time series. The weight specifies the proportion of each component that is used to fit the time series and thus a very high (or very low weight) indicates the dominance of a single component and the comparatively small contribution of the other.

I then applied a Principal Components Analysis (PCA) to these results to identify a lower-dimensional representation of the structure present in the data amenable to visualisation and implemented k-means clustering to identify clusters of time-series with similar temporal features – i.e. this clustering assigns each smoothed time-series to one cluster.

Statistical Modelling and Prediction of Seasonal Modality

Prediction of cluster membership was carried out using a multinomial logistic regression model. Multinomial logistic regression generalises logistic regression (which predicts a binary outcome) to instances with >2 possible outcomes and predicts the probabilities of all possible outcomes of a categorically distributed dependent variable (in this instance, the 4 clusters representing distinct temporal patterns) given a set of independent variables (in this instance, the species each time series belongs to and the previously mentioned suite of environmental covariates). Whereas logistic regression frameworks typically have a single coefficient per covariate (which describes the influence of that particular covariate on the outcome being 1 rather than 0), within a multinomial logistic framework, each category being predicted (each of the 4 clusters of temporal patterns in this instance) has a coefficient per covariate. Thus, a given covariate e.g. isothermality (defined as the mean diurnal temperature range divided by the overall average temperature, i.e. a measure of the variability of the temperature profile in a setting) will have 4 coefficients associated with it, with each of these 4 coefficients specifying the association between isothermality and membership of Clusters 1, 2, 3 and 4 respectively.

I employed an $\ell 2$ (ridge) penalty on all coefficients in order to reduce issues surrounding overfitting and partially mitigate multicollinearity across some of the environmental covariates: this regularised multinomial logistic regression model was then fitted within a Bayesian framework and implemented in Stan. Following fitting, the mean coefficient values were used to generate estimates of the probability that a given time series belongs to each of the 4

clusters. Time series were assigned to the cluster with the highest cluster probability and the misclassification rate computed based on the proportion of time series whose cluster membership was correctly predicted.

The results of these analyses were then integrated with recently produced maps of vector presence/absence (as part of work conducted with the Humbug Project (http://humbug.ac.uk/), funded through a Google Impact Challenge grant) to generate predictive maps of mosquito population dynamics across India, specifically, the probability of locations containing at least one mosquito species displaying a particular temporal pattern (1 of the 4 temporal patterns associated with the clusters). For each location, I individually calculated for each of the seven species the probability that the species was present (using the recently generated maps of vector presence/absence), and the probability that the species would display a particular temporal pattern, conditional on presence. For a given location (indexing suppressed for ease of notation), ror each temporal profile $TP\ i$ and the probability of a vector species j being present and displaying that particular temporal profile was calculated as follows:

$$p\left(TP_j^i\right) = p\left(TP_j^i \middle|\, VP_j\right)\, p\left(VP_j\right)$$

where $p(VP_j)$ describes the probability of the vector species i occurring in that location (taken from the Humbug vector occurrence probability maps) and $p\left(TP_j^i \middle|\, VP_j\right)$ describes the probability of vector species i displaying temporal profile j conditional on its occurrence in that location.

$1 - \ p\left(TP_j^i\right)$ is then the probability of the absence of vector species j displaying temporal profile i in the location being considered. Taking the product of these terms over all N vector species, we then have:

$$p(TP^{i'}) = \prod_{j=1}^{N}\left(1 - \ p\left(TP_j^i\right)\right)$$

Where $p(TP^{i'})$ is therefore the probability of temporal profile i being absent from the particular location being considered, across all species being considered. The probability of temporal profile i being present in the location being considered, in at least 1 vector species is then given as:

$$p\left(TP^i\right) = 1 - \ p(TP^{i'})$$

Results

A total of 272 time-series from 117 locations across India were identified through the systematic review, spanning seven species complexes that together represent the dominant malaria vectors in the country. These noisy time-series **(Fig. 4.1)** were then smoothed using a Negative Binomial Gaussian Process based framework **(Fig. 4.2)**. Substantial variation in temporal dynamics was observed between different species complexes in degree of seasonality and timing of seasonal peaks. Whilst *Anopheles dirus s.l.* populations tended to peak during the monsoon period (typically June to September), many *Anopheles fluviatilis s.l.* populations peaked between November and February (the dry season across most of India), reaching their lowest density during the monsoon. *Anopheles dirus s.l.* populations demonstrated the highest degree of seasonality with an average of 75% of the total annual catch being concentrated in a 4-month time-period **(Fig. 4.2C)**. This was in contrast to *Anopheles annularis s.l.*, where only 53% of the total annual catch on average was caught in any 4-month period. In addition to this variation between species complexes, I also observed extensive variation in temporal dynamics within a species complex. Across the 85 time-series collated for *Anopheles culicifacies s.l.*, populations varied substantially in both the extent and timing of their seasonal peaks; this ranged from sharp peaks in the monsoon season to less seasonal, more perennial characteristics similar to those observed for *Anopheles annularis s.l.*. A range of dynamics were also observed for time-series belonging to *Anopheles stephensi s.l.*, from peaks coincident with the monsoon season to bimodal dynamics displaying peaks both during and outside the rainy season.

An array of summary statistics were calculated for each time series in order to characterise their temporal properties (**Fig. 4.3**). This was followed by k-means clustering of the results, to assess whether the observed variation could be delineated into discrete groups, each characterised by distinct temporal patterns. I identified 4 groups (**Fig. 4.4A)** – these included time-series peaking during the monsoon season (Cluster 1), displaying bimodal characteristics (Cluster 2), peaking in the dry season (Cluster 3) or displaying perennial patterns of abundance (Cluster 4) **(Fig. 4.4C)**. Cluster assignment was robust to the choice of prior used in the time-series fitting and smoothing (**Fig. 4.5**). Average catch size varied between Clusters, ranging from a median catch size of 356 for Cluster 2 to 42 for Cluster 4 (see **Fig. 4.6**). The distinct patterns displayed by each group were not due to differences in the timing and extent of rainfall across India – I collated location and time-period specific rainfall data for each study (collated from the CHIRPS dataset (Funk et al., 2015)) and calculated the cross-correlation between mosquito density and rainfall. This varied between clusters –a high positive cross-correlation product between rainfall and mosquito density was observed for Cluster 1 (average

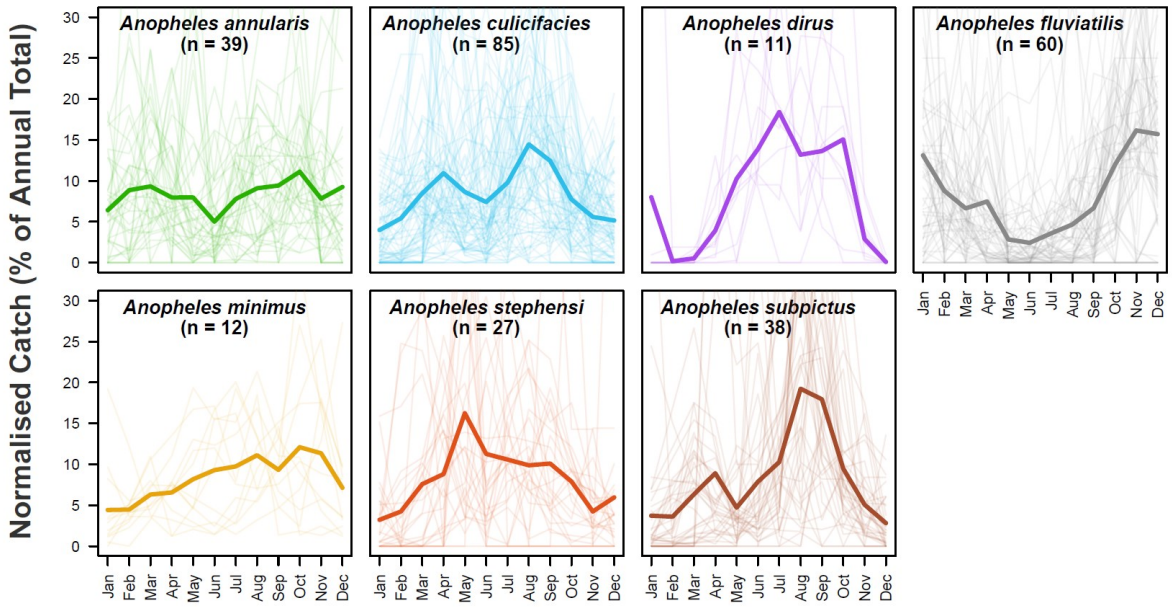


Figure 4.1 The Raw Mosquito Data Extracted During the Systematic Review Process. Through a systematic review, a total of 272 time series containing species-specific, monthly disaggregated mosquito catch data spanning at least 12 months were identified and extracted. Together, these time series span 118 locations across India and 7 major *Anopheline* species known to be involved in the transmission of malaria. For each panel presented here, pale lines represent a single normalised time series for that particular species, and the brighter line is the mean of all the time series belonging to that species, evaluated at that particular timepoint.

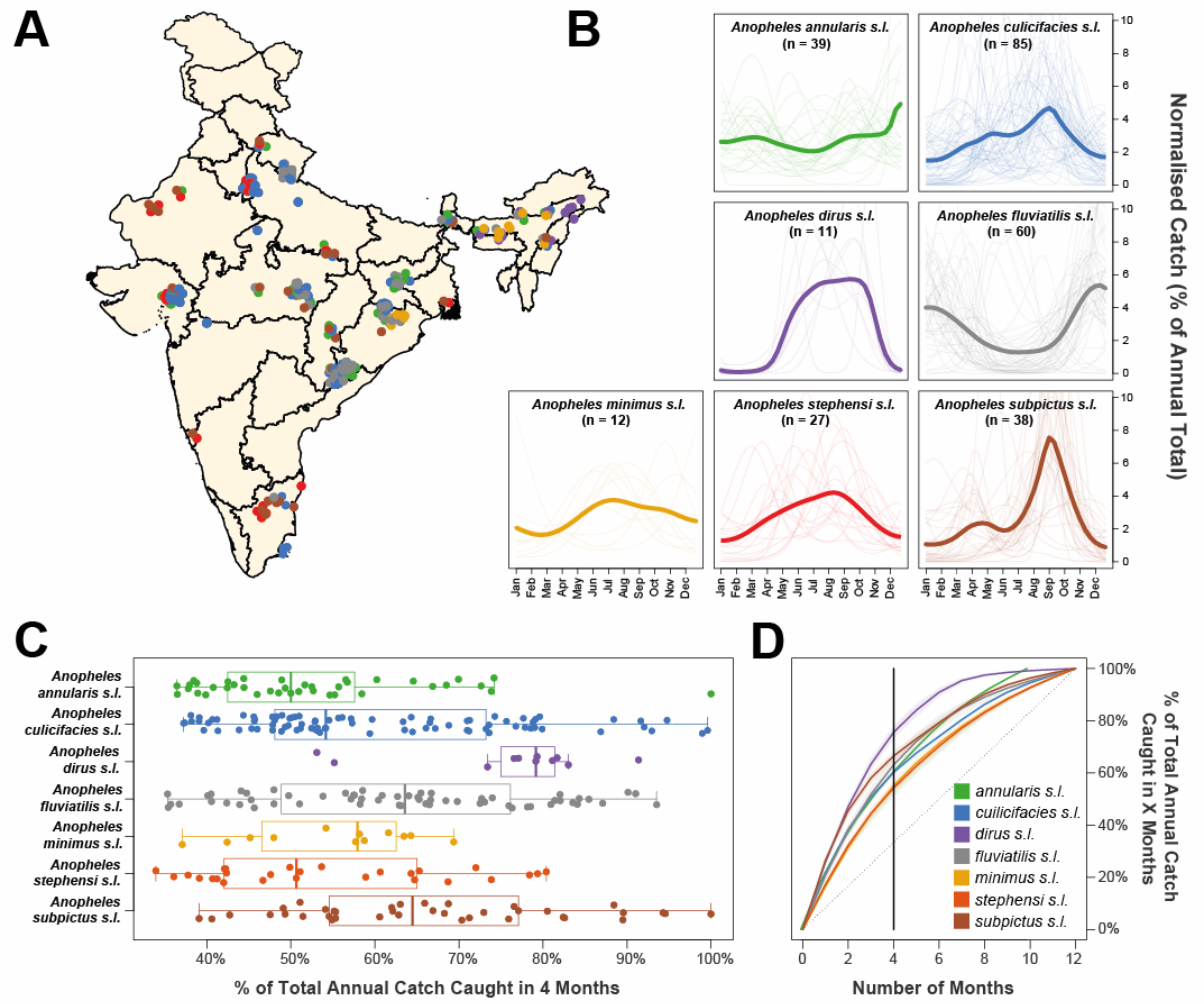


Figure 4.2 Exploring Species Complex-Specific Patterns of Mosquito Population Dynamics. Negative Binomial Gaussian Processes incorporating a periodic kernel were fitted to each of the 272 time-series collected from 118 locations across India and spanning the period 1979-2017 were collated as part of the systematic review. These fitted time-series (representing monthly catches over the course of a year) were then normalised and the results plotted here, disaggregated by species complex. **(A)** Map of India showing the different locations for which time-series data was available. Points represent a single collected time-series, coloured according to the species complex. **(B)** Normalised, Gaussian Process fitted time-series disaggregated by species complex. In all instances, pale lines represent a single time-series for that particular species complex, and the brighter line is the mean of all of the time-series belonging to that species complex, evaluated at that particular timepoint. **(C)** Boxplot of the maximum percentage of total annual study catch caught in any consecutive 4 month period (an approximate measure of seasonality, with a higher value implying greater seasonality). Each point is a study, coloured according to Anopheline species. **(D)** As for (C) but showing the mean percentage of the total annual study catch caught across a range of different months for each species.

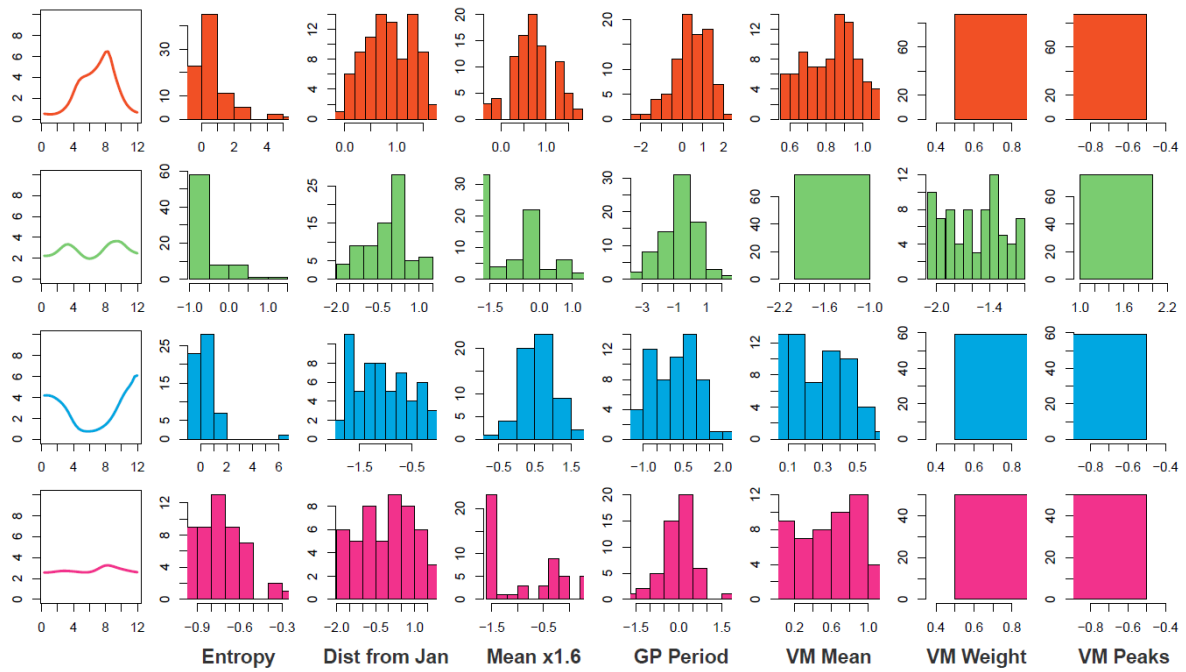


Figure 4.3 Temporal Cluster Statistical Properties. A series of mathematical operations were applied to the fitted time series in order to further characterise and explore their temporal properties. The results of this characterisation were then clustered using the k-means algorithm. For each cluster, the mean temporal profile is displayed, as well as the underlying distribution of each temporal property is displayed, namely the **Entropy**, the **Distance of the Highest Peak from January**, the **Proportion of Points > 1.6x the Mean,** the **Period of the Fitted Gaussian Process Kernel**, the **Mean of the Fitted Von Mises Distribution**, the **Weights of the Two Von Mises Components** and the **Optimal Number of Von Mises Components.** For further information on each of these operations, see Supplementary Information: Time Series Characterisation and Analysis.

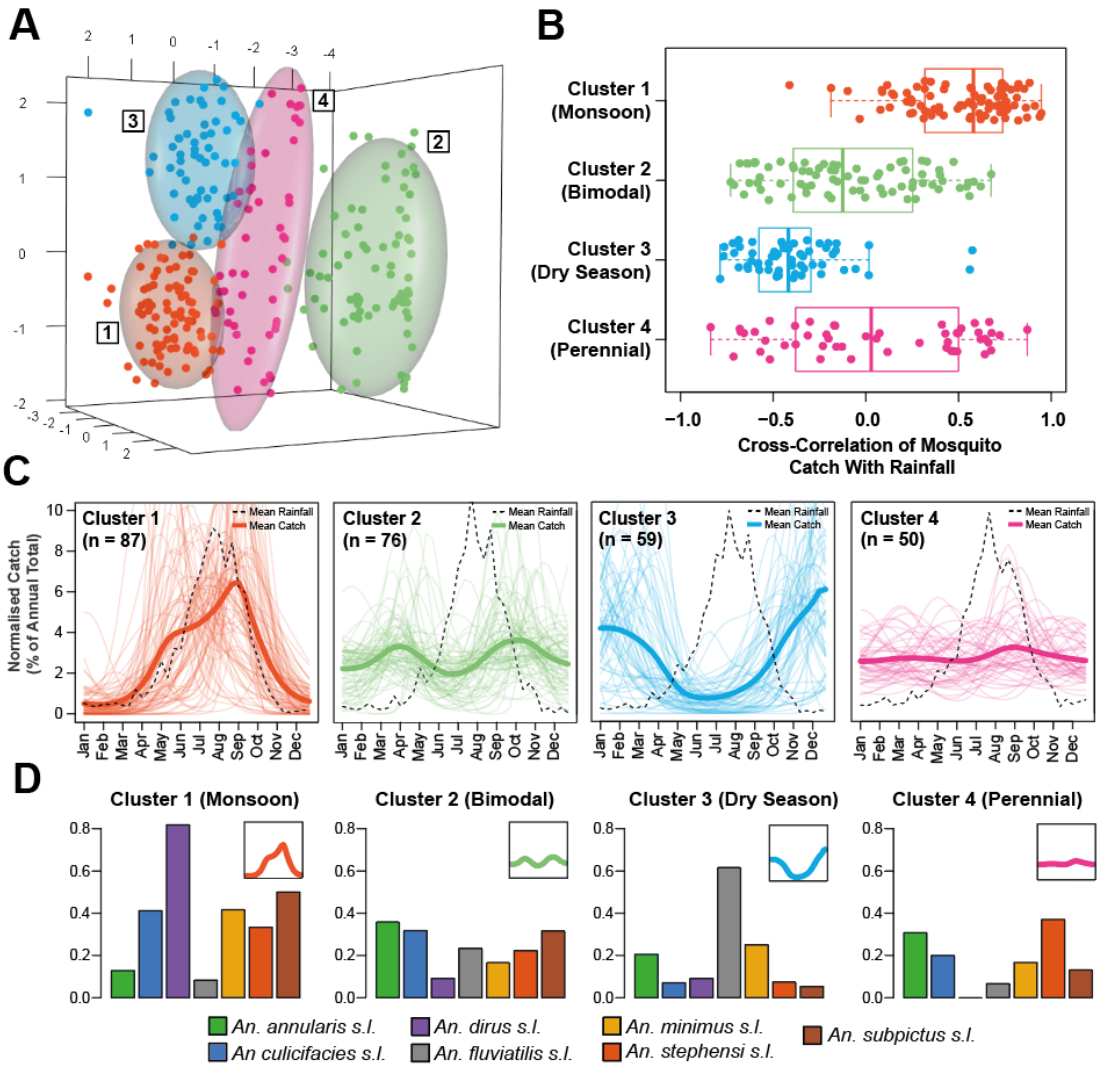
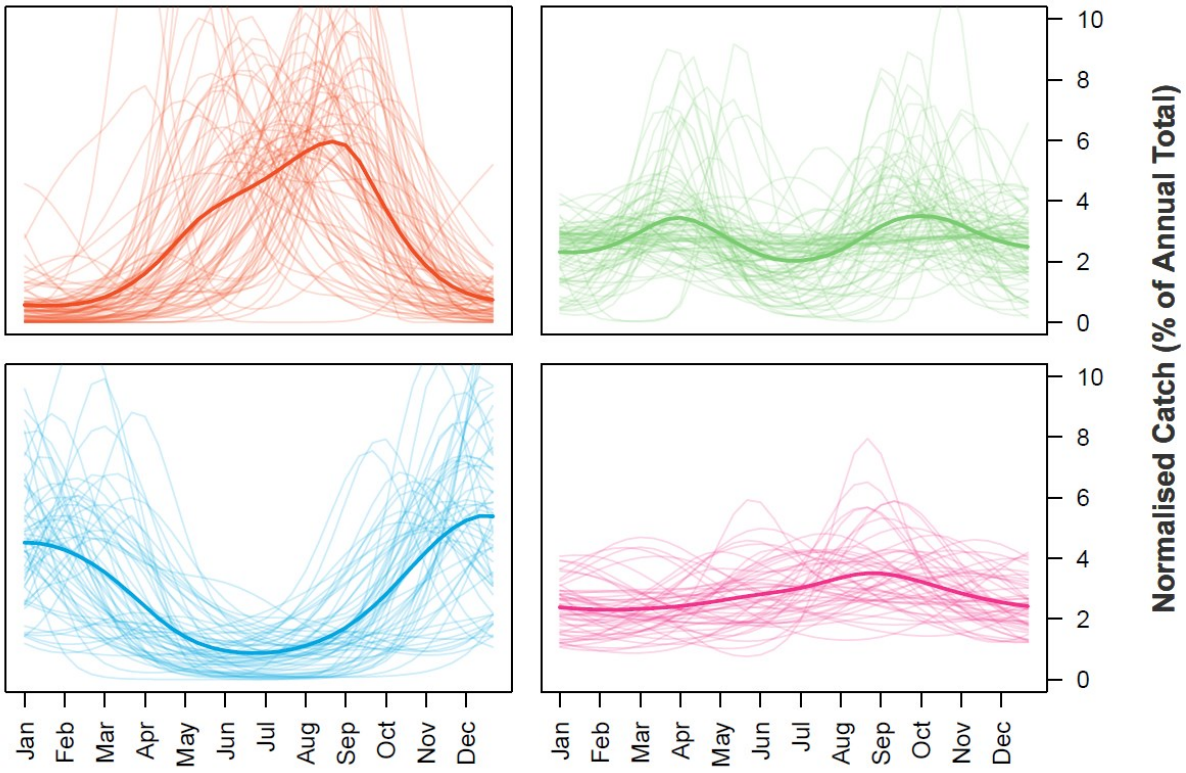


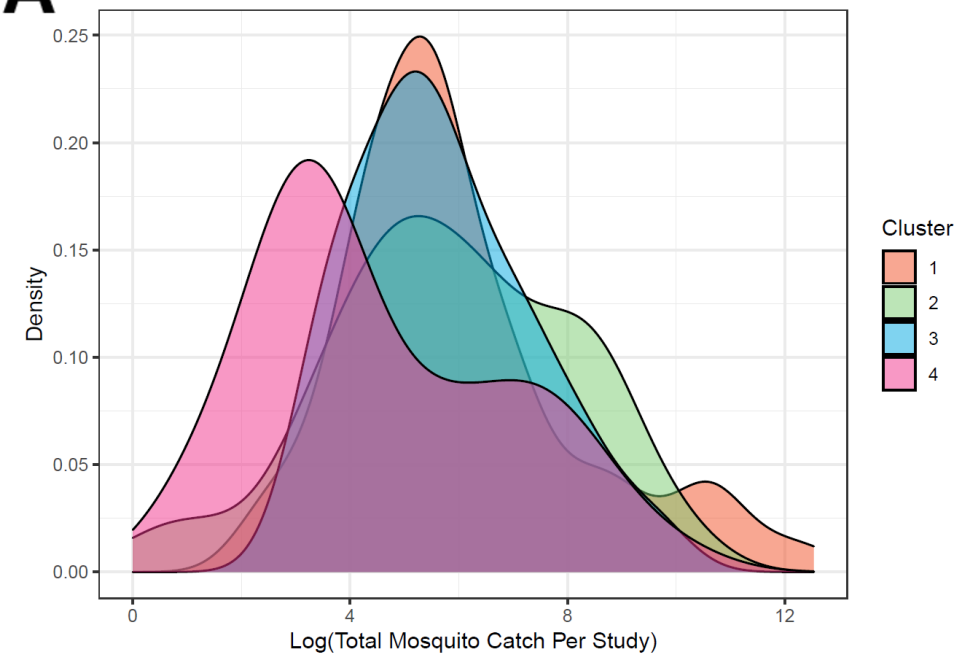
Figure 4.4 Characterisation and Clustering of Time-Series with Similar Temporal Properties. Statistical characterisation of time-series properties was followed by PCA and results clustered using the k-means algorithm. **(A)** Results of k-means clustering for 4 clusters, with a PCA applied for visualisation purposes. Point colour refers to cluster membership, ellipsoids demarcate the 75th quantile of the density associated with each cluster. First 3 principal components are plotted, explaining 82% of variation. **(B)** Boxplot of the cross-correlation between rainfall and mosquito catch for each location and time-series. Rainfall data is specific to study location and time-period and was extracted from the *The Climate Hazards Group Infrared Precipitation With Stations (CHIRPS) dataset*. Each point indicates an individual time-series, coloured according to cluster membership **(C)** Time-series belonging to each cluster. Pale lines represent individual time-series, brighter line represents the mean of all the time-series belonging to that cluster, evaluated at each timepoint. Dashed black line represents the mean rainfall across the time-series belonging to the cluster. Characterisation and clustering in this way revealed distinct groups of time-series that share similar temporal properties. **(D)** The proportion of time-series for each species complex belonging to each cluster - different coloured bars indicate different species complexes (see legend) and y axis corresponds to the proportion of time-series (for a given species complex) belonging to that cluster.



	Uninformative Prior			
Informative Prior	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Cluster 1	76	0	3	0
Cluster 2	2	72	0	11
Cluster 3	1	1	53	2
Cluster 4	8	3	3	37

Figure 4.5 Results of Clustering When Fitting Mosquito Catch Data Using An Uninformative Prior. In order to assess the sensitivity and robustness of the time series clustering, a less informative prior was used during the fitting process and the results displayed here. Top are the plots displaying the time series belonging to each cluster and which replicate the same 4 broad classes of temporal dynamics identified when clustering using the results from the fitting using an Informative Prior. Bottom table cross-tabulates Cluster assignments for individual time series across both sets of fitting – the majority (88%) of time series were consistently clustered across both sets of fitting, with the majority that displayed an incongruency being those belonging to Cluster 2 and Cluster 4 (from the Informative Prior fitting), the least peaked of the four temporal profiles. Predictive power based on the results of the multinomial logistic regression decreased when using the Uninformative Prior results but remained substantially above that of a random classifier (predictive accuracy was 0.51, compared to the 0.25 expected for a truly random classifier and 0.58 for the model constructed using time series fitted with Informative Priors).

A



B

Cluster	Median Catch Size	Mean Catch Size	Standard Deviation
1 (Monsoon)	252	8350	34300
2 (Bimodal)	356	2150	4520
3 (Dry Season)	242	1240	2940
4 (Perennial)	42	1170	3560
Overall	226	3760	19800

Comparison	t-test (Difference in Mean)	Mood's test (Difference in Median)
1 vs 2	p=0.10	p=0.24
1 vs 3	p=0.06	p=0.65
1 vs 4	p=0.06	p<0.001
2 vs 3	p=0.16	p=0.03
2 vs 4	p=0.18	p=0.15
3 vs 4	p=0.90	p<0.001

Figure 4.6 Comparison of Individual Study Sizes By Cluster. For each study the total number of mosquitoes caught over the year was calculated, and the results presented by cluster. **(A)** Density plot of the catch-sizes for the studies belonging to each of the archetype clusters. **(B)** Table showing the median, mean and standard deviation for the catch sizes across each of the clusters, as well as the results of hypothesis tests pairwise testing for either difference in means (t-test) or medians (Mood's test) between clusters.

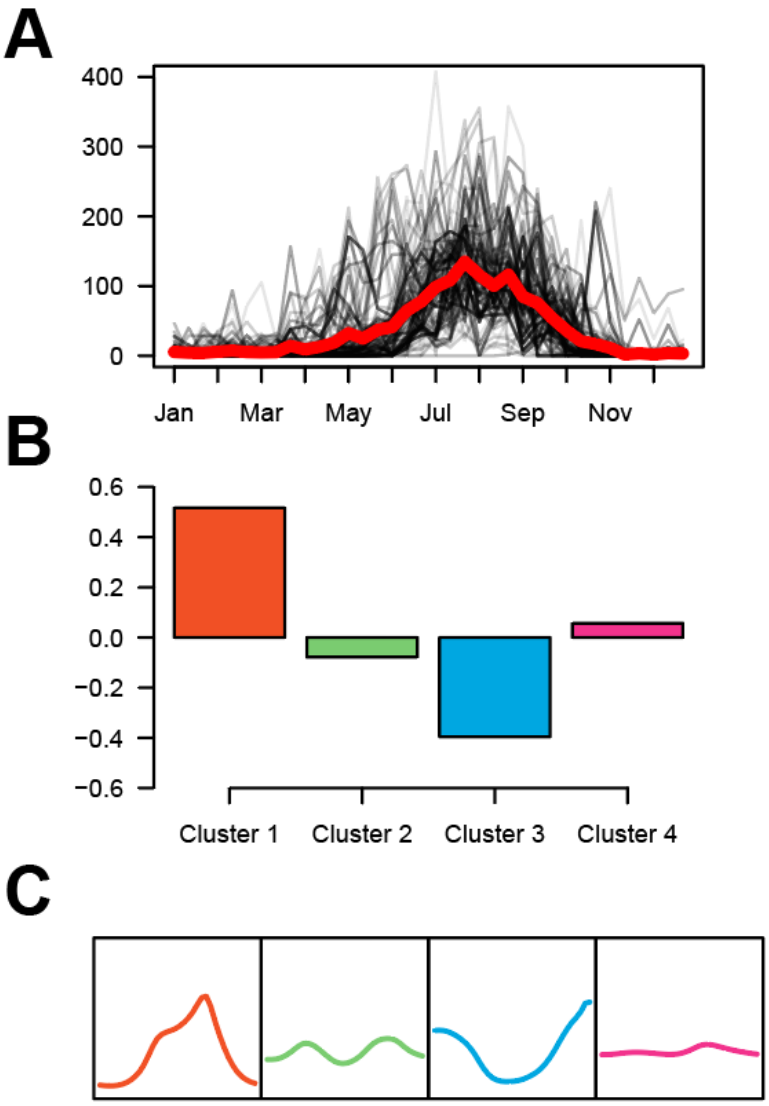


Figure 4.7 Exploring the Cross-Correlation Between Rainfall and Mosquito Densities. For the 117 locations across India where mosquito catch data had been collected, daily, year specific rainfall data was also extracted and collated. This rainfall data was then aggregated up to the same temporal and spatial scale as the collected mosquito data and the cross-correlation between the two quantities explored. **(A)** Rainfall dynamics in India across the course of a year. Each black line represents the rainfall in a given location, whilst the thicker red line represents the average rainfall profile across all 117 locations. **(B)** Cluster specific cross-correlations between rainfall and mosquito catch size. **(C)** Mean mosquito catch temporal profiles for each Cluster.

r=0.52), but a negative correlation for Cluster 3 (r=-0.41) and low correlation for Clusters 2 and 4 (r=-0.08 and 0.03 respectively, **Fig. 4.7**). This suggests that the observed patterns represent genuine differences between species and across locations in how mosquito populations respond to rainfall. For some species complexes, the majority of their time-series belonged to a single cluster (**Fig. 4.4D**) – *Anopheles dirus s.l* time-series were restricted primarily to Cluster 1 (monsoon season peaking) whilst *Anopheles fluviatilis s.l.* time-series were almost exclusively found in Cluster 3 (dry season peaking). Using binary indicators for species complex (seven total, indicating which species complex a particular time-series belongs to) and a suite of ecological variables (25 total) as predictors, I fitted a multinomial logistic regression model to the cluster labels (i.e. which cluster each time-series had been assigned to) to explore potential factors underlying the observed variation in temporal dynamics. This framework produces one coefficient estimate for each cluster and predictor (a total of 4 coefficients per cluster and predictor), with that coefficient defining the strength of the association between a predictor and a particular cluster. Across the species complex regression coefficients, *Anopheles culicifacies sl.* and *Anopheles subpictus s.l.* demonstrated positive associations with Cluster 1 (monsoon peaking dynamics), whereas for *Anopheles fluviatilis s.l.*, this relationship was negative (the species-complex associated with Cluster 3 instead) and *Anopheles annularis s.l.* was most strongly associated with Cluster 4 (perennial dynamics). To explore this variation more systematically, I employed a hierarchical clustering approach to identify groups of species with similar patterns of association with specific temporal dynamics **(Fig. 4.8A).** *Anopheles culicifacies s.l.* and *Anopheles subpictus s.l.* clustered together and showed a positive association with Cluster 1 and a negative association with Cluster 3). By contrast, *Anopheles fluviatilis s.l.* clustered on its own, positively associated with Cluster 3 and negatively associated with Cluster 1. There were significant disparities in the number of time-series available for each species (ranging from 85 for *Anopheles culicifacices s.l.* to only 11 for *Anopheles dirus s.l.*) and so I explored how robust the results of this clustering were to subsampling the data so that all species had the same number of time-series (as *Anopheles dirus s.l.*). Hierarchical clustering showed that these groupings were robust to subsampling, except in the case of *Anopheles dirus s.l.*, which instead clustered with *Anopheles culicifacies s.l.* and *Anopheles subpictus s.l.* (and showed positive associations with Cluster 1 dynamics, and a negative association with Cluster 3 dynamics, **Fig. 4.9**).

Both temperature seasonality and total annual rainfall were strongly associated with Cluster 1 (which possessed the dynamics most strongly correlated with rainfall) **(Fig. 4.8B)**. By contrast, perennial dynamics (Cluster 4) strongly associated with the continuous presence of water bodies and negatively associated with both temperature seasonality and rain seasonality. Strong associations with landcover were observed for Cluster 2 (strongly negative for

urbanicity) and Cluster 3 (strongly positive for woody savannas). In order to examine the broader patterns of association, I ranked the coefficients for each environmental variable within each cluster according to their magnitude, and selected the 15 with the strongest association in each cluster (positive or negative). The top 15 variables for each cluster were then compared to assess the extent of overlap, revealing that each cluster tended to associate with a unique set of ecological factors **(Fig. 4.8C)**. These mutually exclusive and cluster-specific patterns of association with environmental covariates were similarly borne out across an analysis of the correlation of all coefficients between clusters, which revealed them to be highly negatively correlated **(Fig. 4.10)**.

I next integrated these results with spatial predictions of mosquito species complex presence/absence to produce predictive maps of mosquito population dynamics across India; specifically, to generate estimates of the probability that a given location contains ≥1 mosquito species complex displaying a particular temporal pattern **(Fig. 4.11)**. These results predict that monsoon peaking dynamics (Cluster 1) are most likely in the North and Northeast **(Fig. 4.11A)**. This contrasts with the predicted spatial distribution of bimodal dynamics (Cluster 2), which are predicted to be more likely across central India and less likely in the Northeast. Dynamics involving peaks during the dry season tracks the predicted spatial distribution of *Anopheles fluviatilis s.l.* closely and are predicted to be most probable across central India **(Fig. 4.11C)** – a similar pattern was observed for spatial predictions of perennial dynamics **(Fig. 4.11D)**. Together these results suggest that spatial variability in both species complex occurrence and environmental factors together generate the complex patterns of mosquito temporal dynamics observed across India.

Discussion

Understanding the temporal dynamics of malaria transmission represents an important input to effective deployment of control interventions. Here I leverage a collection of temporally disaggregated mosquito time-series catch data from across India to explore these dynamics. These results reveal extensive variation in mosquito population dynamics between species complexes and across locations, ranging from highly seasonal and rainfall-concordant dynamics through to perennial and rainfall-discordant dynamics. Analysis of this variation has revealed a complex interplay between biotic (species complex-specific drivers) and abiotic (the broader ecological structure of the environment) factors in shaping these dynamics.

In a manner largely independent of the ecological setting, *Anopheles fluviatilis s.l.* populations typically peaked during the dry season. Whilst previous work has identified these dynamics (Gunasekaran et al., 1994; Sahu et al., 2017), this work highlights the consistency of this

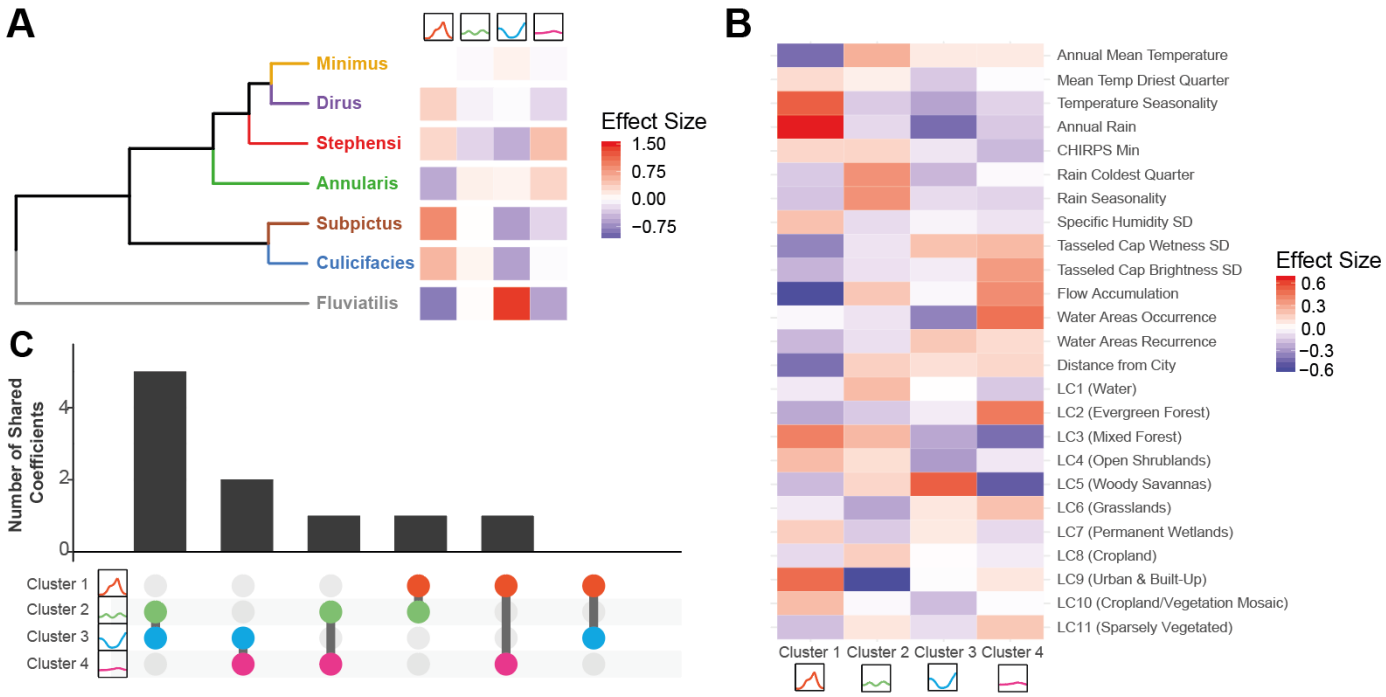


Figure 4.8 Exploring Drivers of Mosquito Population Dynamics Using Multinomial Logistic Regression. A multinomial logistic regression-based approach using both species complex and a suite of environmental variables was used to explore the factors associated with different mosquito population dynamics. The output of this regression is a single coefficient describing the strength of the association per variable and cluster. **(A)** Hierarchical clustering of the regression results for each species complex, as defined by the set of coefficient values describing the strength of the association between that species complex and the particular cluster. **(B)** The strength of the association between each of the 25 environmental covariates used and the relevant temporal cluster. **(C)** Upset plot summarising the environmental variable coefficients. For each cluster, the 15 environmental covariates with the strongest association were selected and the extent of overlap in this top 15 covariates compared across clusters; x-axis indicates the specific pairwise cluster comparison, y axis the number of shared top 15 covariates between the two clusters.

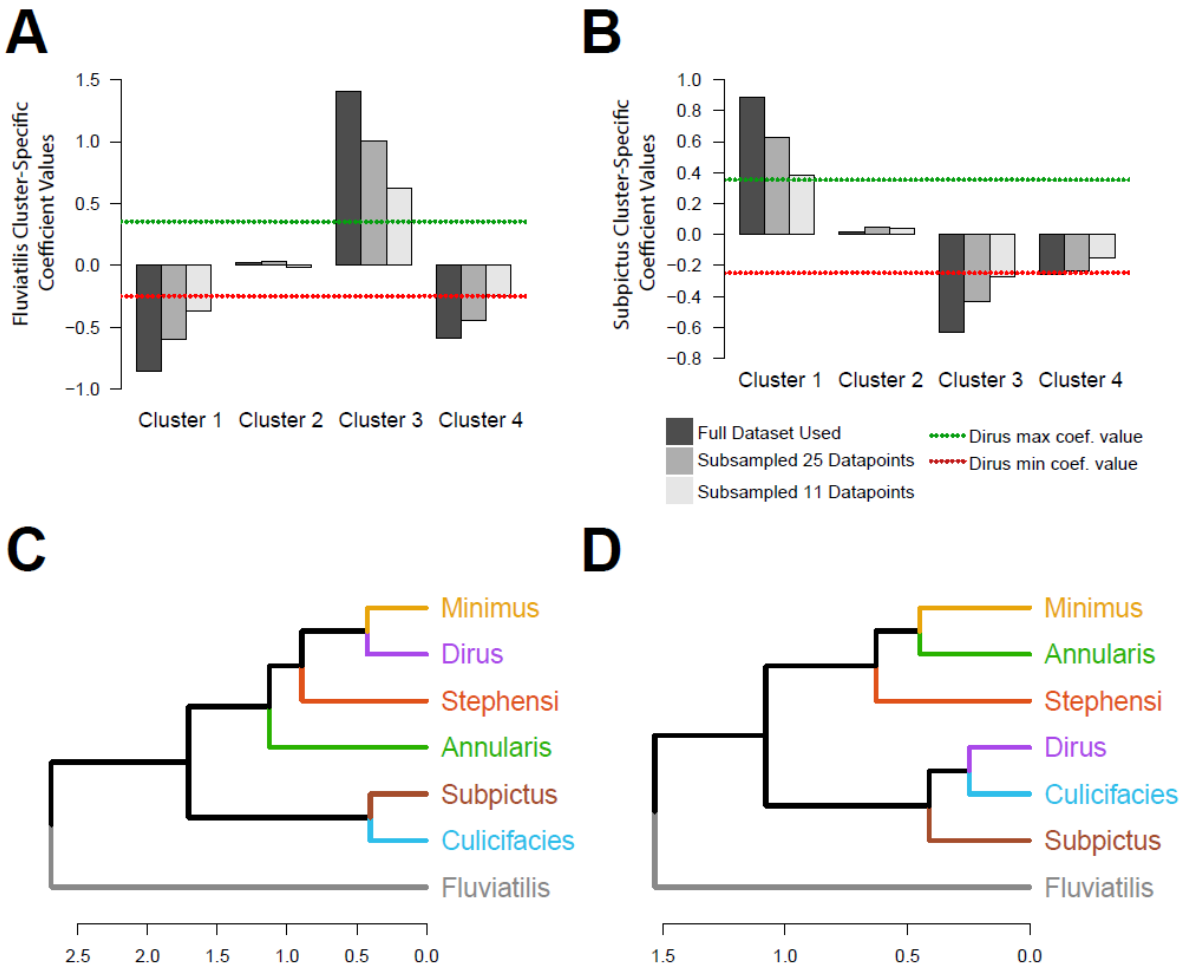


Figure 4.9 Species-Cluster Coefficient Values and Hierarchical Clustering Results When Data Were Subsampled. There were significant disparities in sample sizes for each species (highest being *Anopheles culicifacies s.l.* with 85 and lowest being *Anopheles dirus s.l.* with 11) and so a sensitivity analysis was conducted to assess whether differences in sample sizes were affecting the coefficient values inferred between species. Results presented above are derived using the mean coefficient values from 30 permutations, with each permutation involving a round of data subsampling and running of the multinomial logistic-regression based framework using this subsample. **(A)** Comparison of inferred cluster coefficient values for *Anopheles fluviatilis* when using the full dataset available (dark grey), subsampling so that each species had 25 (or less if fewer were available) datapoints available for inference, or subsampling to 11 datapoints per species (which matches the number of datapoints for the smallest sample – *Anopheles dirus s.l.*). Dotted red and green lines indicate the maximum and minimum coefficient values inferred for *Anopheles dirus s.l.* **(B)** As for (A), but for *Anopheles subpictus s.l.*. **(C)** Dendogram based on hierarchical clustering of the coefficient values inferred using the full dataset. **(D)** Dendogram based on hierarchical clustering of the coefficient values inferred using a dataset subsampled so that all species had 11 datapoints (to match the lowest sample size i.e. that of *Anopheles dirus s.l.*) only.

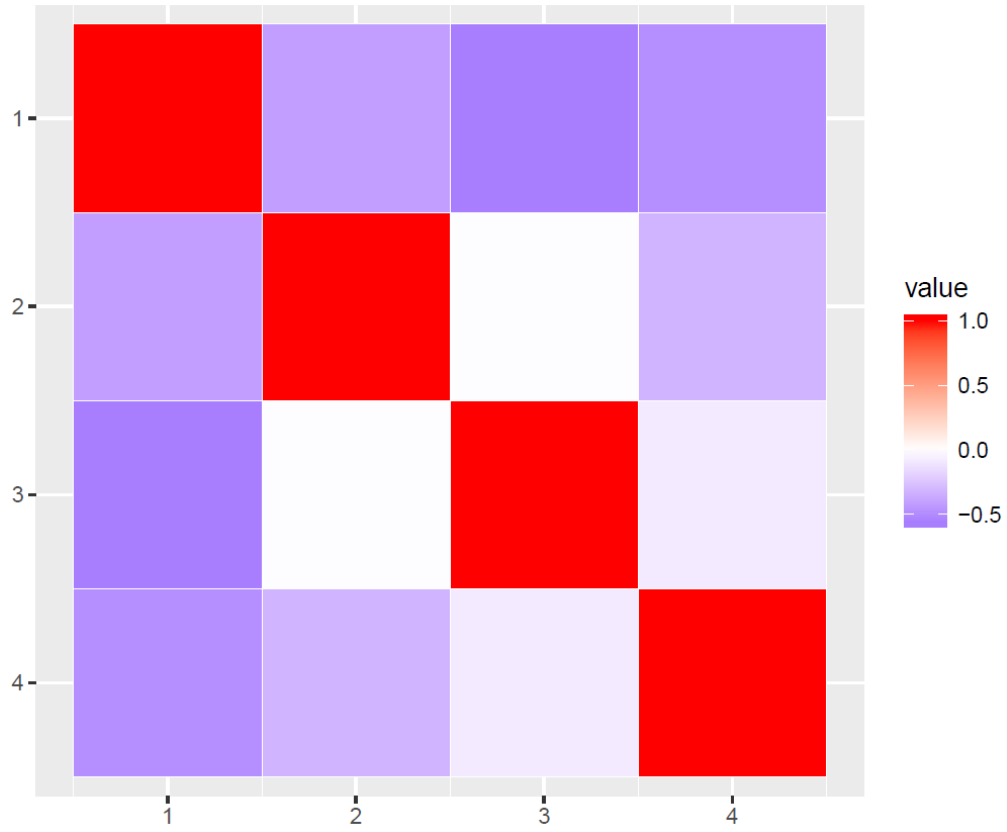


Figure 4.10 Cross-Cluster Correlations For Ecological Coefficients. The cross-correlation between the predicted ecological coefficients from the multinomial logistic regression for each Cluster.

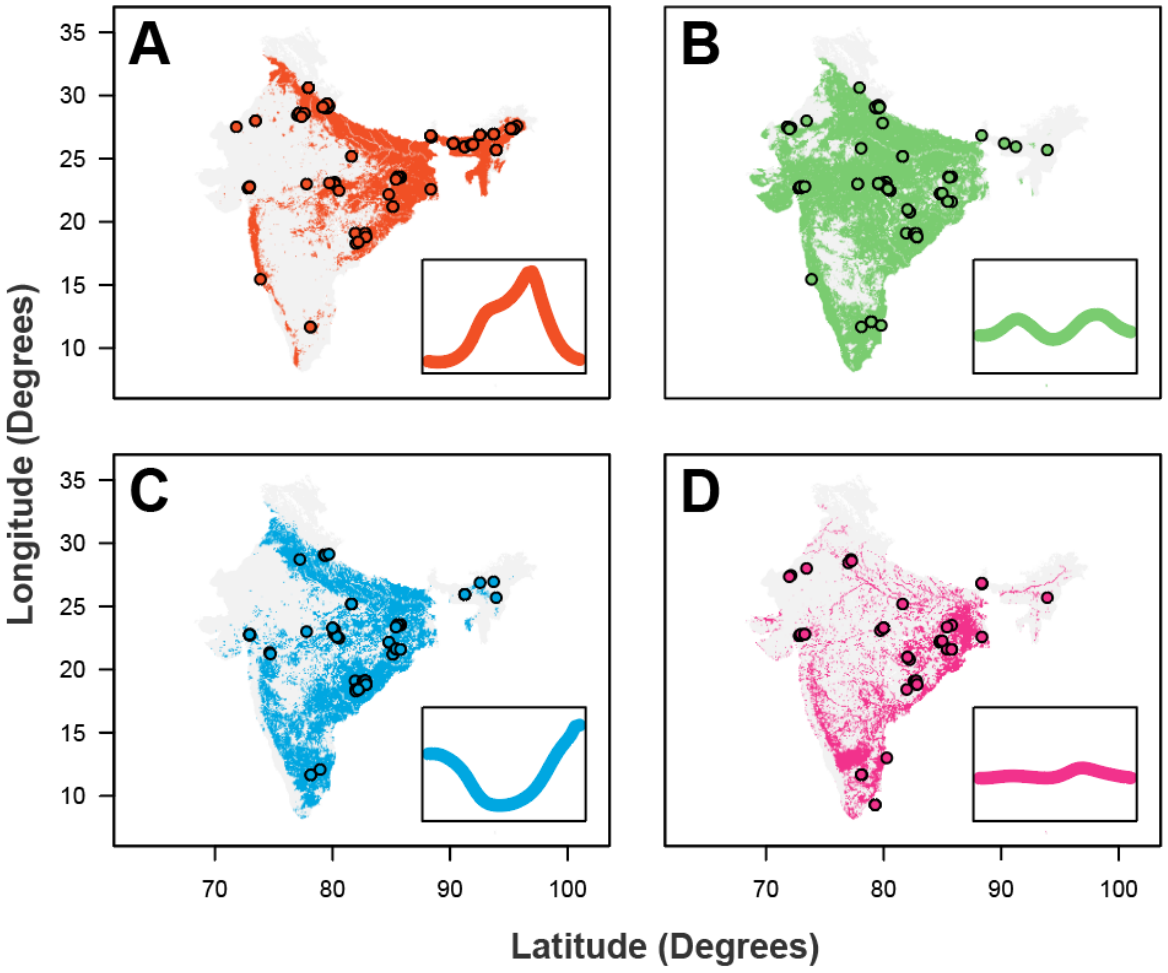


Figure 4.11 Predictive Maps of Mosquito Population Seasonality Across India. The results of the multinomial logistic regression were integrated with recently generated maps describing the probability of presence/absence for different anopheline species complexes (not shown). Together, these were used to generate estimates of a given area possessing at least one mosquito species complex with a particular temporal profile (as defined by the previously described clusters), with these probabilities then thresholder at 0.67 to produce a binary indicator (i.e. value 1 if the probability for a given pixel is > 0.66 and 0 otherwise). **(A)** Results of this analysis for Cluster 1 (the "monsoon peak" cluster) – red dots describe the locations in which a mosquito species complex with a temporal profile assigned to Cluster 1 were found. **(B)** As for A, but for the "bimodal" cluster (Cluster 2). **(C)** As for A, but for the "peak in dry season" cluster (Cluster 3). **(D)** As for A, but for the "perennial" cluster (Cluster 4). In all cases, the map colour describes the probability of a given area containing one or more mosquito species complex displaying that pattern of temporal dynamics. The coloured points indicate locations where a mosquito species complex displaying temporal dynamics belonging to that cluster were empirically observed.

observation across locations, showing that these dynamics are largely restricted to *Anopheles fluviatilis s.l.* and highlight the capacity for the population dynamics of a regionally important malaria vector to significantly depart from local patterns of rainfall. These results align with previous work that has indicated streams and surrounding stagnant water as breeding sites for this species complex (Dasgupta et al., 2018) – such breeding sites are typically unsuitable during the monsoon season when flooding occurs but become increasingly suitable as the dry season ensues. By contrast, *Anopheles culicifacies s.l.* displayed a wide array of temporal dynamics depending on the sampling site. These ranged from peaking during the monsoon to bimodal and even perennial behaviour – a finding consistent with documented variation in the species complex's breeding habits (Surendran and Ramasamy, 2005; Barik, Sahu and Swain, 2009; Jude et al., 2010). However, due to the inability to disaggregate time-series according to sibling species (which frequently show differences in preferred types of breeding sites (Barik, Sahu and Swain, 2009)), the drivers of this variation in temporal dynamics for *Anopheles culicifacies s.l.* remains unclear – specifically, whether is it due to sibling species displaying distinct temporal dynamics or because *Anopheles culicifacies s.l.* temporal dynamics are more plastic (and hence display different temporal dynamics depending on the particular ecological setting) than *Anopheles fluviatilis s.l.* (where the same dynamics were observed irrespective of the broader ecological structure).

My results highlight the limited utility of considering rainfall alone when trying to understand temporal patterns of mosquito abundance, with variable associations with rainfall observed across the populations studied here. Indeed, I identified a significant impact of temperature on population dynamics, with temperature seasonality strongly positively associated with the highly seasonal, monsoon peaking seasonal dynamics (Cluster 1) and both temperature seasonality and rainfall seasonality negatively associated with perennial (Cluster 4) dynamics. The role of temperature in shaping mosquito population dynamics is increasingly being recognised (Beck-Johnson et al., 2017; Mordecai et al., 2019), due in part to the significant influence it has on many individual mosquito life-history traits (Shapiro, Whitehead and Thomas, 2017; Johnson et al., 2015b), including biting rate, fecundity and mortality (amongst others); with its influence on these factors typically non-linear and unimodal with clear optima (Johnson et al., 2015a) and subject to interactions with other factors such as the demographic structure of the mosquito population (Miazgowicz et al., 2020). Together, this has significant consequences for mosquito population dynamics and, in turn, the range and dynamics of vector-borne diseases (such as malaria) they underpin (Ryan et al., 2015; Mordecai et al., 2013). These results therefore suggest a role for both rainfall and temperature in shaping annual patterns of mosquito abundance and underscores the importance of considering

seasonal fluctuations in a range of environmental variables when trying to understand seasonality in mosquito population dynamics.

In addition to temperature, I observed associations between temporal dynamics and a variety of other ecological covariates. The perennial patterns of abundance observed for Cluster 4 were strongly associated with flow accumulation and water area occurrence (acting as proxies for proximity to rivers and bodies of water). These factors were negatively associated with all other clusters. This is consistent with reports indicating that static water sources may provide sites available for oviposition and mosquito breeding year round (Minakawa et al., 2012; Kumar et al., 2016) and highlights the importance of the local hydrological environment (which in the cases of large bodies of water is only partially dependent on patterns of rainfall) in shaping mosquito population dynamics. Together with the results above, this highlights the importance of considering a broad array of ecological factors, including temperature, rainfall and the local hydrological environment more generally, when exploring the temporal dynamics of mosquito populations.

There was also a significant influence of landcover patterns, specifically urbanicity (measured by the two covariates Landcover and Distance to City) on temporal dynamics. Higher degrees of urbanicity were consistently and positively associated with monsoon peaking dynamics (Cluster 1). This is consistent with previous work which has extensively documented the differences in the nature of the hydrological environment utilised by mosquitoes in urban vs rural settings, and noted particularly the diverse array of physical features present in cities (e.g. tyres, wells, overhead tanks etc) that are able to hold water following rainfall and act as breeding sites for mosquitoes (Thomas et al., 2016; Lin et al., 2018). However, an important confounding factor is that the majority of the collated time-series from surveys that were carried out in urban settings were from *Anopheles stephensi s.l.*, a highly efficient urban vector able to sustain malaria transmission effectively in cities. Whether the observed association directly represents differences in how the hydrological environments associated with different types of settlement (e.g. rural vs urban settings) respond to rainfall, or is partly driven by the confounding described above remains unclear. Disentangling this is important however, particularly in light of recent work identifying *Anopheles stephensi s.l.* in the Horn of Africa (Faulde, Rueda and Khaireh, 2014). The species complex's range has historically been restricted to South Asia and the Middle East – mounting evidence of a range expansion (Sinka et al., 2020) and the possibility of established urban malaria transmission in the region would be detrimental to control efforts (particularly given cities are not typically foci of malaria transmission across sub-Saharan Africa (Doumbe-Belisse et al., 2021; Robert et al., 2003)). Understanding the temporal dynamics of this vector, particularly in urban centres, will be vital

in order to minimise spread and further proliferation of the species beyond the Horn of Africa. Further developing this understanding forms the basis for the following and final chapter of the PhD thesis.

It is important to note that factors other than mosquito dynamics are also involved in defining the temporal profile of malaria risk. Whilst an association between the size of mosquito populations and case numbers is well established (Bashar and Tuno, 2014; Galardo et al., 2009), the nature of this relationship remains less clear. Interactions between malaria endemicity (Churcher, Trape and Cohuet, 2015), mosquito abundance (Romeo-Aznar et al., 2018) and vector competence (Beck-Johnson et al., 2017) can lead to non-linear dynamics that can be further modified by human behavioural factors such as migration or occupational practices (Cohen et al., 2013). Due to limitations on the extent of entomological data describing relevant malaria metrics such as sporozoite positivity, I was unable to explore many of these factors. Similarly, the lack of disaggregation according to sibling species (which vary markedly in malaria vectorial efficiency) and accompanying epidemiological information (on malaria prevalence or incidence) precludes us from better resolving the comparative contributions of different mosquito species to transmission. This limits my ability to translate temporal patterns of mosquito populations into relevant metrics such as the Entomological Inoculation Rate (EIR). Whilst I mitigate this limitation somewhat by focussing the analyses specifically on dominant vector species-complexes previously established as relevant to malaria transmission in India (Dev and Sharma, 2013), it is not necessarily the case that each mosquito species analysed here is equally relevant to malaria transmission. Future work integrating these analyses with those exploring seasonality of case incidence (c.f. Nguyen et al. (Nguyen et al., 2020)) would therefore likely prove instructive.

There are a number of limitations to the work presented here – firstly, whilst location and time-period specific data were available for the collated rainfall, varying (often limited) degrees of geospatial information were present in each included study. The environmental covariates used in the multinomial-logistic regression were therefore spatially averaged over reported study area, and additionally often across multiple years due to the absence of time-period specific data. This spatio-temporal averaging may obscure relevant inter-annual variation in factors (e.g. rainfall) that affect population dynamics (Mendis et al., 2000), and may contribute to some of the more limited seasonality (e.g. Clusters 2 and 4) and timing of seasonal peaks (in e.g. Clusters 1 and 3) observed. I mitigate this somewhat by extracting time-period specific rainfall data for each study but cannot preclude some role of spatio-temporal averaging in the results presented here. Another limitation is the heterogeneity in mosquito sampling methods across the studies. Studies varied in the catch-method used (landing catch, resting collections,

pit collections, light traps and spray catches), as well as timing (dawn, dusk, night-time etc) and location (typically either human-dwellings or cattlesheds) of collections. This heterogeneity may interact with mosquito traits (such as timing (Sherrard-Smith et al., 2019) or degree of indoor/outdoor biting (Massey et al., 2016) and host preferences (Griffin et al., 2016)) that vary between species, and have implications for which species are sampled, and their comparative abundance (van de Straat et al., 2021). I partially mitigate this heterogeneity by normalising the catch data, but this incomplete accounting for differences in catch methodological characteristics might lead to biases in the presented inferences presented. There were also significant differences in the average number of mosquitoes caught between clusters, with Cluster 4 (perennial dynamics) having the lowest average catch size. Whilst differences in catch sizes between clusters were smaller than within cluster variation (where individual study counts ranged over several orders of magnitude and were highly overdispersed), it is possible that the lack of observed seasonality for Cluster 4 time-series might be an artefact of limited sampling effort and mosquitoes caught.

Overall, this work highlights that the substantial variation in temporal dynamics across mosquito populations can be clustered into a small number of dynamical archetypes, each characterised by distinct temporal properties and associated with distinct environmental factors. In doing so, this work underscores the crucial importance of integrating both species composition and ecological structure into our understanding of the temporal profile of malaria risk and provides a generically applicable framework to better identify and understand patterns of seasonal variation in vectors relevant to public health – a crucial and operationally relevant input for optimising the delivery of control interventions.

Conclusion

In this Chapter, I have explored the diversity in temporal dynamics displayed by key anopheline mosquito species implicated in the transmission of malaria across India. This exploration has been facilitated by development of a statistical framework enabling characterisation of the diversity and variation present in temporally disaggregated entomological catch data, and delineation of this heterogeneity into discrete temporal archetypes. In the following chapter of my thesis, I continue to focus on temporal variation in mosquito populations and leverage this same framework to investigate the temporal dynamics of *Anopheles stephensi*, a highly efficient malaria vector capable of sustaining transmission in urban settings and which has recently been introduced into the Horn of Africa. Specifically, I apply this same framework to characterise the species' temporal dynamics, explore the factors driving these dynamics, and integrate the collated temporal profiles of mosquito abundance

into an existing model of malaria transmission, in order to more directly and concretely link these dynamics to implications for malaria transmission and control.

Chapter 5 Seasonal Dynamics of an Emerging African Malaria Vector *Anopheles stephensi* and the Implications for Malaria Control

Across sub-Saharan Africa, urban centres often experience lower levels of malaria transmission compared to equivalent rural settings. Recent importation of the highly efficient urban malaria vector *Anopheles stephensi* to the Horn of Africa threatens this paradigm. In this Chapter, I apply the statistical framework developed in Chapter 4 to a dataset of *Anopheles stephensi* time-series catch data in order to better characterise the species' seasonal dynamics and explore the factors underpinning them. Integrating these results with an established model of malaria transmission, I highlight the crucial role that knowledge of timings of vector seasonality and peak densities will play in maximising the impact of vector control interventions such as indoor residual spraying (IRS) on mitigating and controlling the vector's further proliferation.

Introduction

There has been an estimated 40% reduction in the burden of malaria since 2000, predominantly due to significant scale-up of control interventions (Bhatt et al., 2015b). Alongside this expansion of control efforts, increasing urbanisation of Africa's populace (rising from 31% to 43% between 1990 and 2018, with >60% expected to live in urban areas by 2050 (United Nations, 2018)) is also thought to have indirectly contributed to reductions in disease burden. Previous work has found significantly lower annual Entomological Inoculation Rates (EIR) in urban compared to rural settings (Doumbe-Belisse et al., 2021; Robert et al., 2003). This is thought to be underpinned by factors including differences in the quality of housing (Trape and Zoulani, 1987; Killeen et al., 2019), reduced availability and suitability of habitats for anopheline breeding in urban settings (De Silva and Marshall, 2012; Awolola et al., 2007; Kasili et al., 2009), better access to treatment (Weiss et al., 2020), and higher population densities leading to higher human to mosquito ratios (and reduced transmission) (Romeo-Aznar et al., 2018). Whilst these trends are not always consistently identified (e.g. surveys where prevalence of malaria is higher in urban areas than surrounding locations (Mourou et al., 2012; Wang et al., 2006); or previous work highlighting that *Anopheles gambiae s.s.* can adapt to breeding in polluted water characteristic of urban environments (Klinkenberg et al., 2008)), increasing urbanicity across sub-Saharan Africa is likely to complement planned scale-up of malaria control interventions aimed at achieving the targets outlined in the World Health Organization's 2030 Global Technical Strategy for Malaria (World Health Organization, 2021a).

This impact of increasing urbanization on disease burden is contingent on these remaining areas of comparatively low malaria transmission. This phenomenon is currently under threat by the invasion and establishment of the highly efficient urban malaria vector *Anopheles stephensi*. Found throughout South Asia, there are three known forms of the species ("type", "intermediate" and "mysorensis"). The mysorensis form is predominantly found in rural settings, and typically possesses a low vectorial capacity (due to its zoophilic behaviour (Subbarao et al., 1987)). By contrast, the type and intermediate forms represent highly efficient vectors capable of transmitting both *Plasmodium falciparum* and *Plasmodium vivax*, with their ability to proliferate in urban locations distinguishing this species from other malaria vectors in the region. This efficiency as an urban vector is thought to be underpinned by an increased tolerance for breeding in polluted water sources (Batra et al., 2001), and a superior ability to utilise the anthropogenic hydrological habitats present in urban settings (Thomas et al., 2016; Kumar and Thavaselvam, 1992). The species was first identified in sub-Saharan Africa in Djibouti City in 2012 (Faulde, Rueda and Khaireh, 2014) and has since been recorded in both Ethiopia (Balkew et al., 2020; Tadesse et al., 2021) and Sudan (Ahmed et al., 2021b, 2021a), with recent work highlighting likely suitability for the species across some of the continent's largest population centres comprising over 100 million people (Sinka et al., 2020). Whilst causality has yet to be conclusively established, its emergence is thought to have contributed to the significant resurgence of malaria transmission in Djibouti (which experienced a 10-fold increase in cases between 2013 and 2019), highlighting the potential threat establishment of this vector poses to malaria control across the Horn of Africa (Hamlet et al., 2021) and the continent more generally (Feachem et al., 2019).

Experiences in Djibouti to date highlights the significant public-health threat this vector potentially poses. Despite this, substantial uncertainty remains regarding how its establishment might influence malaria dynamics in the region, particularly in the (predominantly urban) settings where the disease is currently largely absent. A key driver of this will be the vector's seasonal dynamics. As my work in Chapter 4 has highlighted, anopheline mosquito population dynamics are characterised by extensive diversity both between species and across locations; and frequently exhibit substantial annual fluctuations in size. Understanding the factors underlying these dynamics is crucial given that the effectiveness of many malaria control interventions (such as seasonal malaria chemoprevention (ACCESS-SMC Partnership, 2020) or indoor-residual spraying (Tukei, Beke and Lamadrid-Figueroa, 2017)) depends on the timing of their delivery relative to seasonal peaks in transmission. A better understanding of *Anopheles stephensi*'s seasonal dynamics is therefore likely to have material consequences for effective entomological monitoring and

surveillance of the vector's spread and establishment across the Horn of Africa, as well as effective control of malaria across the region.

Substantial uncertainty remains regarding *Anopheles stephensi*'s seasonal dynamics; studies carrying out longitudinal catches are present in the literature, but typically only focus on a single location, precluding systematic comparison and identification of generalisable patterns. Whilst the work presented in Chapter 4 began to answer some of these questions, the *Anopheles stephensi* data presented there was only from studies carried out in India. Here I extend the work presented in Chapter 4 and collate longitudinal mosquito catch data for *Anopheles stephensi* from across the extent of its geographical range, including India, the rest of South Asia and the Middle East in order to better understand the species' temporal dynamics. Leveraging the statistical framework developed in Chapter 4, I explore patterns of *Anopheles stephensi* temporal dynamics and the factors driving them. The results highlight pronounced variation in the extent and timing of seasonality, with distinct dynamics observed across rural and urban settings. Integrating these results with a model of malaria transmission highlights how this variation will influence the efficacy of malaria control efforts in parts of the Horn of Africa where the disease is currently (or has previously been) largely absent and underscores the need for rapid scaleup of entomological monitoring across the region.

Methods

Systematic Review of *Anopheles stephensi* Literature

I collated references from a previously published systematic review of the literature relating to *Anopheles stephensi* (focusing on its presence/absence across a wide geographical range (Sinka et al., 2020) and combined it with the systematic review carried out and presented in Chapter 4. I then updated these two previous searches (both conducted in 2017) by searching *Web of Science* and *PubMed* databases from January 2017 for further relevant references containing temporally disaggregated *Anopheles stephensi* catch data. Key words for this search were:

> (((anophel*) AND ((India) OR (BURMA) OR (MYANMAR) OR (BANGLADESH) OR (THAILAND) OR (ISLAMIC REPUBLIC OF IRAN) OR (ETHIOPIA) OR (DJIBOUTI) OR (SUDAN))) AND (("2017"[Date - Publication] : "3000"[Date - Publication])) OR ((anophel*) AND ((Pakistan) OR (Iran) OR (Afghanistan)) AND (("1990"[Date - Publication] : "3000"[Date - Publication]))

with references for Pakistan, Iran and Afghanistan searched for over an extended time-period (i.e. date range of 1990-2020 rather than 2017-2020) to ensure completeness of the collated

references, and fill in countries not included during previous reviews. Identified records were then screened according to the following Inclusion/Exclusion criteria:

Inclusion Criteria:

- Reference contains temporally disaggregated adult mosquito catch data for *Anopheles stephensi*, at a temporal resolution of monthly or higher.
- The time-period spanned by the survey must be at least 10 months in duration and have caught at least a total of 25 *Anopheles stephensi* (i.e. minimum monthly average catch of 2.5 mosquitoes) over the period for which catches were being carried out.

Exclusion Criteria:

- Mosquito catch data are not temporally disaggregated to a sufficient extent (e.g. catches were done yearly or seasonally rather than monthly).
- Mosquito catch data were collected as part of a trial assessing a vector control intervention (which would perturb the natural dynamics of the vector, rendering the data unrepresentative of the population dynamics in the absence of control).
- The reference only contained information on immature/larval mosquito life cycle stages rather than mature adults.

Overall, a total of 34 references were collated containing 65 time-series from catch surveys carried out in distinct locations from across Afghanistan (n=2), Djibouti (n=1), India (n=32, 27 of which are from the results presented in Chapter 4), Iran (n=17), Myanmar (n=5) and Pakistan (n=8). These were further supplemented with 2 references (from Pakistan and India respectively, yielding a total of 3 time-series) collated as part of a (currently unpublished) systematic review of the bionomics of secondary malaria (i.e., non-dominant) vectors across South Asia, yielding a total of 65 time-series from these 34 references.

Systematic Review Data Extraction, Collation and Initial Processing

Entomological Data Extraction

For each reference, I extracted all relevant entomological catch data provided that pertained specifically to *Anopheles stephensi*. Where data were presented in a table, data were copied directly from the table. Where the data were in a graph, data were extracted using the DataThief™ software. This yielded a total of 65 time series of monthly mosquito catch data (no reference presented data at a finer temporal resolution), ranging in length from 10 – 60 months, with a mean time-period of 15.6 months and a median time-period of 12 months, a mean catch size of 758 and a median catch size of 289.

Table 5.1 Number of time series collated according to method of collection

	Landing Catch	Resting Collections	Pit Collections	Light Traps	Spray Catches
# Time Series	3	33	2	4	14

Of the collated studies, the majority sampled mosquitoes via resting collections (n=33). As with the results presented in Chapter 4, there was also variation between surveys as to where mosquitoes had been sampled (e.g. human dwellings or cattlesheds), when sampling had been carried out (daytime, night-time or overnight) and for the small number of landing catch studies collated (n=3), which bait had been used (cattle or humans). Of the 65 collated time-series, 56 represented results arising from a survey carried out using 1 catch methodology (described in **Table 5.1** above). 9 time-series represented results which presented the total number of *Anopheles stephensi* mosquitoes caught across all methods of collection and could not be disaggregated by catch-type. They have not been counted in **Table 5.1**.

The primary focus of these analyses was to characterise annual and seasonal patterns of variation in *Anopheles stephensi* abundance. Given this, and also that variations in time-series length are a factor known to affect their statistical properties (Fulcher, Little and Jones, 2013) (and therefore limit the comparability of the time series gathered and analysed here), all time-series were standardised to be 12 months in length. For time series containing more than 12 time points (i.e. time series that spanned longer than a single year), I averaged the recorded catches for a given month. Where the study was initiated in a month other than January, and concluded in a month other than December, the recorded counts were rearranged to yield a complete time series running from January to December.

The results presented in the collated references were frequently presented in a form standardised by sampling effort, such as Man-Hour Density (MHD). They do not therefore represent the total number of mosquitoes caught each month (required for the statistical framework utilised to characterise temporal properties) and therefore, where information on sampling effort was present (e.g. number of hours spent sampling/catching *Anopheles stephensi*, number of households or cattlesheds searched, number of trap nights etc), I used this information to convert MHD back to the raw counts. In the small number of instances where there was variable sampling effort across the time series (which would bias the conversion away from the underlying population abundance), I conservatively used the lowest sampling effort recorded across the time series in the conversion. Together, this enabled production of an estimate of the number of mosquitoes sampled (a raw count, based on equal sampling effort across the time series).

Study Geolocation and Environmental Covariate Extraction

For each study where geolocation was possible, I recorded the location at both the administrative unit 1 and 2 level, based on information provided in the reference. A number of the references identified in my review had previously been utilised as part of previous reviews (i.e. the work presented in Chapter 4 and that of (Sinka et al., 2020)) – where these data were available, these descriptions of study location were used. For each location, I then extracted a suite of satellite-derived environmental covariates. These environmental covariates consist of raster layers spanning all of the countries in which studies had been conducted in (i.e. Afghanistan, Djibouti, India, Iran, Myanmar and Pakistan) at a 2.5 arc-minute (~5km by 5km, depending on the exact location and distance from the equator) spatial resolution and then averaged at the administrative unit 1 level. The covariates utilised here were derived from the *BioClimatic* variables (a suite of biological relevant covariates defined from monthly rainfall and temperature satellite data (Fick and Hijmans, 2017)) as well as measures of landcover and urbanicity (European Space Agency (ESA)), population density (Gaughan et al., 2013; Linard et al., 2012) and enhanced vegetation index (Justice et al., 2002). This provided a total of 51 covariates, many of which were highly correlated with one another. From these covariates, a reduced subset with minimal multi-collinearity were generated using the *tidymodels* collection of R packages (Kuhn and Wickham, 2020), which minimises the Spearman correlation coefficients between retained covariates via an optimisation procedure. I also excluded covariates where there was minimal variation for that (scaled and standardised to have mean 0 and unit variance) covariate across the full dataset, leaving 18 covariates in total. In addition to the environmental covariates described above, for each of the administrative units a survey had been carried out in, I also collated daily rainfall estimates for the time-period the survey had been conducted in, using the *"The Climate Hazards Group Infrared Precipitation With Stations"* (CHIRPS) dataset (Funk et al., 2015). These data were aggregated up to the same temporal resolution as the *Anopheles stephensi* catch data (i.e. monthly). These rainfall data were used to calculate the cross-correlation coefficient between mosquito catches and rainfall.

Time-Series Fitting and Interpolation

Negative Binomial Gaussian Process Fitting and Time-Series Characterisation

Using the methodology developed and presented in Chapter 4, I fitted the same Negative Binomial Gaussian Process-based model to smooth the collated mosquito count time-series. Model fitting was carried out within a Bayesian framework, using the probabilistic programming language STAN (Carpenter et al., 2017) – priors for each of the model parameters were the same as used in Chapter 4. I characterised time-series properties using the same summary

statistics presented in Chapter 4 as well as an additional metric motivated by previous work exploring malaria seasonality seeking to define operationally and programmatically relevant definitions of seasonality (Cairns et al., 2012). Specifically, I also included the maximum percentage of total annual catch occurring in any 3-month period as a summary statistic. To do this, I used a sliding 3-month window to calculate the maximum percentage of the total annual catch that was caught in any 3-month period across the course of the year. As with the work presented in Chapter 4, from these summary statistics, I obtain for each time-series a set of 8 real numbers providing a reduced representation of the temporal properties of each time-series. I then applied a PCA to identify a lower-dimensional representation of the structure present in the data and implemented k-means clustering to identify clusters of time-series with similar temporal properties.

Statistical Modelling and Prediction of Cluster Membership

Random Forest Prediction of Cluster Membership

The extracted suite of environmental variables derived from satellite data were then used as predictors within a Random-Forest based classification framework aimed at predicting the cluster membership (i.e. results of the k-means algorithm described above). Random Forests are a machine learning, ensemble-based statistical method that work by fitting a collection of decision trees to the data (where data are either a continuous outcome variable in the regression context, or a binary indicator in the classification context) (Breiman, 2001). The outputs of these decision trees are subsequently aggregated to produce a "forest" (or ensemble) of trees that together produce predictions for comparison with data. They have previous been shown to provide significant improvements in accuracy over traditional linear regression based approaches, particularly in contexts where non-linear relationships or interactions between covariates are likely present and to be relevant to prediction of an outcome (Biau, 2012).

Given a training set of covariates $X = x_1, \dots, x_n$ and responses $Y = y_1, \dots, y_n$, the algorithm underlying construction of the random forest takes the following steps:

1. Sample with replacement n training examples from Y yielding Y_S, and k covariates from X yielding X_S.
2. Train a classification tree f_S on X_S, Y_S by minimising the Gini impurity (a measure of how frequently a random chosen element from Y_S would be incorrectly classified if it was randomly labelled according to the distribution of labels in the relevant node of the tree).
3. Repeat steps 1 and 2 multiple times.

4. Having constructed a large number of decision trees (a “forest” of trees), aggregate and use this forest to make predictions given new data. In a classification context, this is typically carried out by taking the majority vote of the ensemble of decision trees.
    a. Note that evaluation of performance can either be carried out via cross-validation or by examining the out-of-bag error, which is the mean prediction error on each training sample x_i using only the trees that did not have x_i in their training subsample X_S.

I used a Random Forest based approach to either 1) classify time-series cluster membership (as defined via the PCA and k-means clustering analysis described above); or 2) predict *Anopheles stephensi* time-series seasonality (defined as the percentage of total annual vector density in any continuous 3-month period). These models were fitted using the software package *Ranger* (Wright and Ziegler, 2015), implemented in the *tidymodels* framework for R (Kuhn and Wickham, 2020), with 6-fold cross-validation utilised to optimise hyperparameter combinations. Presented results are based on averaging the results of 25 separate iterations of cross-validation and model fitting (to account for stochasticity in model fitting), and any predictions made using out-of-bag model estimates in all instances. Due to significant imbalances in class size across the time-series clusters (49 time-series in Cluster 1 compared to only 16 time-series in Cluster 2, I carried out upsampling using the SMOTE (synthetic minority over-sampling technique (Chawla et al., 2002)) algorithm. I also carried out model fitting without this upsampling, the results of which are presented in **Figure 5.8**.

In all instances, out-of-sample predictive accuracy was assessed using 6-fold cross-validation (CV) and used to optimise the hyperparameters associated with the Random Forest method algorithm. Random Forest models were fitted to the training dataset (i.e. the full dataset minus one of the CV folds) and then model accuracy assessed on the remaining fold of data not included in model training. In the case of the cluster classification example, the metric used to evaluate model performance was the area under the curve (AUC). In the case of the regression prediction of seasonality, the metric used to evaluate model performance was the root mean squared error (RMSE). The Random Forest hyperparameters providing the best out-of-sample AUC/RMSE (based on 6-fold CV) were then selected, and a final Random Forest model then fitted on the full set of data available. Predictive accuracy (assessed via AUC/RMSE) was then calculated for the entire dataset by using out-of-bag predictions for each sample i.e. predictions on each training sample using only the trees that did not have that training sample in their bootstrap sample. I also calculated both permutation variable importance and generated partial dependency plots (Molnar, 2020) for each model to assess the contribution of specific, individual environmental covariates to whether a time-series had a single seasonal peak or not. Together these methods allow evaluation of the importance of

each included covariate to model predictive accuracy, and in turn, allows “ranking” of covariates according to their contribution to the predictive performance of the model. This entire process was repeated 25 times in order to average over the stochasticity and variation inherent in the Random Forest fitting process.

I also carried out an additional sensitivity analysis where a set of the available data (n=12 time-series) was held-out at the onset, and the random forest model trained (using 6-fold cross-validation) on the remaining available data (n=53 time-series total, with 43 time-series used in model fitting and 10 time-series used for performance evaluation in each of the cross-validation folds). Optimal hyperparameters were selected in the same way as described above, and then a final model fitted to the full, non-held out data (n=53 time-series), and model predictive accuracy assessed by evaluating performance on the held-out data (n=12 time-series).

Calculating Probabilities of Catching *Anopheles stephensi* Given Seasonal Variation

I first normalised vector density over the course of the year such that the maximum value recorded (i.e. the highest recorded mosquito catch) was set to 1 and all other values were scaled proportionally (i.e. another catch at a different timepoint that was half the size of the peak catch size would be assigned 0.5). I then used these estimates as the probability of successfully sampling *Anopheles stephensi* if an entomological survey was carried out in that month. These estimates therefore represent conservative bounds on the likelihood of missing *Anopheles stephensi*, given I optimistically assume that the mosquito would successfully be detected if a survey is carried out in the month where its population is largest (i.e. the annual peak in abundance as reflected in the catch size). For each time-series, I then calculated the probability of not catching *Anopheles stephensi* in an entomological survey, given a certain number of consecutive months sampled, and averaged over all possible permutations of continuously sampled months possible (e.g. in the case of 11 consecutive months sampled, there are two possibilities: {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11} and {2, 3, 4, 5, 6, 7, 8, 9, 10, 11}). Note that the aim here is not to describe the exact probability of missing *Anopheles stephensi* in any given entomological survey, as this will depend on a wide array of other, poorly defined and heterogeneous factors (such as effort e.g. person-hours, type of catch methodology used etc). Instead, the aim is to highlight how variation in seasonal dynamics can influence the nature of surveillance required to successfully detect establishment of *Anopheles stephensi*.

Transmission Modelling of *Anopheles stephensi*-Driven Malaria Dynamics and Control

I integrated the temporal profiles of *Anopheles stephensi* abundance into a previously developed deterministic compartmental model of *Plasmodium falciparum* malaria transmission and disease (Griffin, Ferguson and Ghani, 2014; Challenger et al., 2021; Unwin et al., 2022) to explore the implications of the vector's establishment and seasonality on the dynamics of malaria transmission, with a particular focus on areas where malaria transmission is currently absent or only minimally present. Specifically, I use the modelling framework to understand how variation in seasonality of the mosquito might influence the impact of indoor residual spraying (IRS), a key vector control intervention. The deterministic, compartmental malaria model used here considers both human and mosquito populations. What follows is a description of the mathematical modelling framework, specifically a description of the human and vector models separately, followed by specific details about how exactly this framework was used to model malaria transmission underpinned by *Anopheles stephensi* in settings where malaria is currently absent or only minimally present.

Human-Component of the Transmission Model

The model groups humans within a population into discrete compartments based on their age, which is indexed by the subscript i. At each point in time, individuals of age i can exist in any one of six different infection states. Upon infection, individuals progress from an uninfected state (Susceptible, S_i) to either Asymptomatic (A_i) or clinical disease, with the comparative probability of these two outcomes depending on the degree of acquired natural immunity due to previous exposure to the parasite (a function of age, described in more detail below).

If an individual progresses to clinical disease, they enter either a Treated (T_i) or Untreated (D_i) Clinical Disease state that depends on the probability of receiving treatment. For those receiving treatment, individuals progress through a period of prophylactic protection following treatment (P_i), and then return to the susceptible compartment. For those developing clinical disease, they remain symptomatic for the duration of the disease, before moving to an

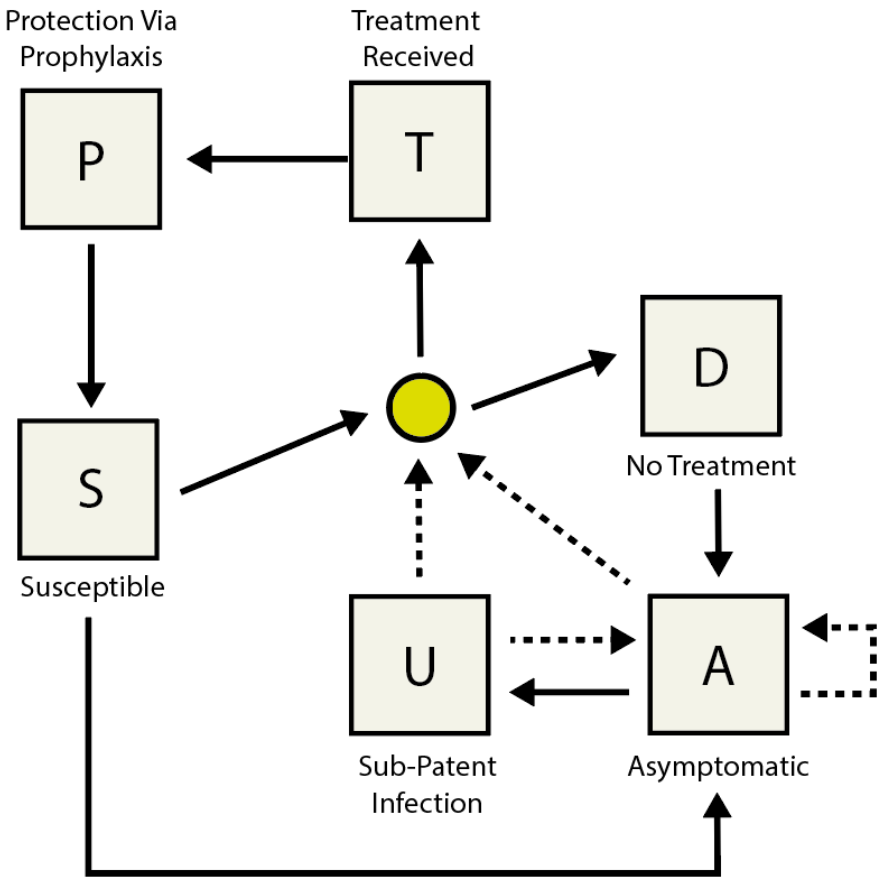


Figure 5.1 Model schematic illustrating the different disease states and the possible transitions between them. The states are shown in boxes with the transitions marked by arrows; dashed arrows indication superinfection. S = susceptible, D = clinical disease, T = successfully treated disease, P = prophylaxis from prior treatment, A = asymptomatic patent infection, U = subpatent infection.

asymptomatic state (A_i, detectable by light microscopy), before subsequently moving to a submicroscopically infected state (U_i, not detectable by light microscopy). Individuals who are currently asymptomatically infected (including individuals in both the $A_i$and $U_i$states) can be reinfected and develop clinical disease once again – if this does not occur, they subsequently clear the infection and return to the susceptible state (see **Figure 5.1** for schematic overview of the model and the transitions between states).

People are born into the first age compartment (i.e. i = 1) susceptible to infection (with new-borns possessing a degree of maternally-inherited immunity that gradually decays over the first 6 months of life) and progress through age-compartments via natural ageing. Within each age-compartment, susceptible individuals (i.e. those in S_i) are exposed to infectious mosquito bites according to a hazard of infection λ_i – this is a function of mosquito biting rate, overall mosquito population level (and degree of infection in the mosquito vector), as well as the pre-erthryocytic immunity that age-group i possesses (itself a function of age and prior exposure, described in more detail below). Infected individuals develop clinical disease or asymptomatic infection following a latent period τ, with the probability of progression to either of these two states depending on the probability of developing clinical disease Φ_i (itself dependent on the level of immunity against clinical disease, described in further detail below). The proportion developing clinical disease can be successfully treated with anti-malarials (with probability p_{treat}) and move to the state T_i. Individuals not receiving treatment (with probability $1 - p_{treat}$) move to infection state D_i. Treated individuals then recover from infection at rate r_{treat} and move to an uninfected state characterised by a partial degree of prophylactic protection from infection (i.e. P_i), before returning to S_i (after this protection wanes, at rate r_{drug}). Those not receiving treatment (i.e. D_i) recover to the asymptomatically infected state (i.e. A_i) at rate $r_{recover1}$, and then (as parasite densities are progressively controlled) to the submicroscopically infected state (i.e. U_i) at rate $r_{recover2}$. The possibility of superinfection (i.e. a new infection occurring whilst an individual is still infected) is included as a possibility, with individuals in states D_i, A_i and U_i all having the possibility for re-infection. All of the rates described above are constant and independent of age.

The model above is described by the following set of ordinary differential equations:

$$\frac{\partial S_i(t)}{\partial t}+\frac{\partial S_i(t)}{\partial a}=-\lambda_i(t-\tau)S_i(t)+r_pP_i(t)+r_uU_i(t),$$
$$\frac{\partial T_i(t)}{\partial t}+\frac{\partial T_i(t)}{\partial a}=\phi_i p_{Treat}\lambda_i(t-d_E)(S_i(t)+A_i(t)+U_i(t))-r_{treat}T_i(t),$$
$$\frac{\partial D_i(t)}{\partial t}+\frac{\partial D_i(t)}{\partial a}=\phi_i(1-p_{Treat})\lambda_i(t-d_E)(S_i(t)+A_i(t)+U_i(t))-r_{recover1}D_i(t),$$
$$\frac{\partial A_i(t)}{\partial t}+\frac{\partial A_i(t)}{\partial a}=(1-\phi_i)\lambda_i(t-d_E)(S_i(t)+U_i(t))+r_{recover1}D_i(t)-\phi_i\lambda_i(t-\tau)A_i(t)-r_{recover2}A_i(t),$$
$$\frac{\partial U_i(t)}{\partial t}+\frac{\partial U_i(t)}{\partial a}=r_{recover2}A_i(t)-r_{recover2}U_i(t)-\lambda_i(t-\tau)U_i(t),$$
$$\frac{\partial P_i(t)}{\partial t}+\frac{\partial P_i(t)}{\partial a}=r_{treat}T_i(t)-r_{drug}P_i(t).$$

In order to account for heterogeneity in biting rates both between and within age-groups, we further stratify each age-compartment (indexed by i) into separate heterogeneity compartments (indexed by j, suppressed in the above set of equations for clarity). Each age-compartment is assigned a unique biting rate (reflecting age-specific patterns in exposure relating to both lifestyle and skin-surface area):

$$\psi_i(a)=1-\rho\exp\left(-\frac{a_i}{a_0}\right)$$

for age-group i where ρ and a_0 are parameters that determine the relationship between age (i.e. body size) and biting rate. This age-specific biting rate is then modified by a relative biting rate within age-groups (ζ_j), which is specified by the following draw from a log-normal distribution with a mean of 1:

$$\log(\zeta_j)\sim N\left(\frac{-\sigma^2}{2},\sigma^2\right)$$

The entomological inoculation rate $\varepsilon_{i,j}$ (EIR, corresponding to the average number of infecting bites an individual receives) and the force of infection $\lambda_{i,j}$ experienced by age compartment i and within-age group heterogeneity group j at time t is then:

$$\varepsilon_{i,j} = \varepsilon_0(t)\zeta_j\psi_i(a)$$

$$\lambda_{i,j} = b_i(t)\varepsilon_{i,j}(a,t)$$

where $\varepsilon_0(t)$ is the average EIR experienced by the population at time t, $b_i(t)$ is the probability that an infectious bite leads to a patent infection (determined by the level of pre-erythrocytic immunity, described in further detail below).

Modelling Different Components of Immunity to *Plasmodium falciparum* Malaria

Acquisition (and loss over time) of naturally acquired immunity to malaria is dynamically modelled as a function of both age and exposure. Within the modelling framework described here, there are three key stages at which a host immune response is considered to act. Each of these are described in more detail below.

Pre-Erythrocytic Immunity (I_B) – which reduces the probability that an infection is able to establish following an infectious bite. In a population subject to an EIR of $\varepsilon_{i,j}(t)$, I_B is a function of age and time and follows the following partial differential equation:

$$\frac{\partial I_{B_{i,j}}}{\partial t} + \frac{\partial I_{B_{i,j}}}{\partial a} = \frac{\varepsilon_{i.j}}{\varepsilon_{i.j} u_B + 1} - \frac{I_{B_{i,j}}}{d_B}$$

with μ_B a parameter limiting the rate at which immunity to infection can be boosted at high exposure and $d_B$is the mean duration of immunity to infection. Given a level of pre-erythrocytic immunity at time $I_B(t)$ in age-group i and heterogeneity group j, the probability of an infectious bite resulting in successful infection is then given by the following Hill function:

$$b_{i,j}(t) = b_0 \left(b_1 + \frac{1 - b_1}{1 + \left(\frac{I_{B_{i,j}}(t)}{\mathrm{I}_{B0}} \right)^{\kappa_B}} \right)$$

where b_0 is the probability of infection with no immunity and I_{B0} and κ_B are scale and shape parameters for the Hill-Function.

Clinical Immunity (I_C) – which reduces the probability of developing clinical disease once infected, and that is made up of both clinical immunity accrued due to direct infectious exposure I_{CA} and clinical immunity that is maternally acquired (i.e. in infants in the months following birth, I_{CM}). In a population subject to an EIR of $\varepsilon_{i,j}(t)$, I_{CA} is a function of age and time and follows the following partial differential equation:

$$\frac{\partial I_{CA_{i,j}}}{\partial t} + \frac{\partial I_{CA_{i,j}}}{\partial a} = \frac{\lambda_{i,j}}{\lambda_{i,j} u_C + 1} - \frac{I_{CA_{i,j}}}{d_{CA}}$$

where μ_C limits the rate at which immunity to clinical disease can be boosted at high exposure and d_{CA} is the mean duration of clinical immunity. Maternally acquired immunity $I_{CM}(t)$ is assumed at birth to be a proportion (P_{CM}) of the level of immunity present in a 20-year old woman, $I_{C20,j}(t)$, living in the same location and which decays at a constant rate (1/d_M),

$$\frac{\partial I_{CM_{i,j}}}{\partial t}+\frac{\partial I_{CM_{i,j}}}{\partial a}=-\frac{I_{CM_{i,j}}}{d_M}, \qquad I_{CM_{i,j}}(t)=P_{CM}I_{C_{20,j}}(t).$$

The total clinical immunity by age and time is given by $I_C = I_{CA} + I_{CM}$. The probability of acquiring clinical disease upon infection by age is then given by a Hill function,

$$\phi_{i,j}(t)=\phi_0\left(\phi_1+\frac{1-\phi_1}{1+\left(\frac{(I_{CA_{i,j}}(t)+I_{CM_{i,j}}(t)}{I_{C0}}\right)^{\kappa_C}}\right)$$

where Φ_0 is the probability of disease with no immunity and I_{C0} and κ_C are scale and shape parameters for the Hill-Function.

Detection Immunity (I_D) – which reduces parasite load, reducing the probability of an infection being detectable diagnostically (and relatedly, the probability of onwards transmission to mosquitoes). This is given by the following partial differential equation:

$$\frac{\partial I_{D_{i,j}}}{\partial t}+\frac{\partial I_{D_{i,j}}}{\partial a}=\frac{\Lambda_{i,j}}{\Lambda_{i,j}u_D+1}-\frac{I_{D_{i,j}}}{d_D}$$

where μ_D limits the rate at which detection immunity can be boosted at high exposure and d_D is the mean duration of detection immunity. The detectability by microscopy of an asymptomatic infection in age compartment i and heterogeneity compartment j at time *t* is given by:

$$q_{i,j}(t)=d_1+\frac{(1-d_1)}{\left(\left(\frac{1+I_{D_{i,j}}(t)}{I_{D0}}\right)^{\kappa_D}f_{D_i}\right)}$$

where d_1 is the minimum probability of detection and I_{D0} and κ_D are scale and shape parameters for the Hill-Function. $f_{D,i}$ represents an age-dependent modifier of the detectability of infection taking the following form

$$f_{D_i} = 1 - \frac{(1 - f_{D0})}{\left(1 + \left(\frac{a_i}{a_D}\right)^{\gamma_D}\right)}$$

Vector Component of the Transmission Model

The vector model is based on a previously developed compartmental model of the anopheline mosquito life cycle and population dynamics (White et al., 2011). Within this framework, both adult mosquito populations and the earlier juvenile stages in the life-cycle are explicitly modelled. Immature mosquitoes start off as larvae, divided into early and late stage (E_L and L_L respectively), which then mature into pupae (P_L) before eventually maturing into adult mosquitoes (M). The average durations spent in each of these states are denoted by d_E, d_L and d_P respectively. Adult mosquitoes are stratified according to infection with *Plasmodium falciparum* status – they begin as susceptible (S_M) and upon infection, progress to an exposed (but non-infectious, E_M) state, and then onto the infectious state (I_M) following the extrinsic incubation period (EIP). Mosquitoes are infected through exposure to humans currently possessing transmissible infections i.e. the treated (T), clinical disease (D), asymptomatic (A) and submicroscopic (U) infection states. Once infected, mosquitoes pass through a latent period $\left(E_M\right)$ of fixed length τ_M and then they become infectious to humans (I_M). They are assumed to remain infectious until they die. The model is described by the equations below:

$$\frac{dE_L}{dt} = \beta_L M - \mu_{E_L}\left(1 + \frac{E_L + L_L}{K_L}\right)E_L - \frac{E_L}{d_{E_L}},$$

$$\frac{dL_L}{dt} = \frac{E_L}{d_{E_L}} - \mu_{L_L}\left(1 + \gamma_L\left(\frac{E_L + L_L}{K_L}\right)\right)L_L - \frac{L_L}{d_{L_L}},$$

$$\frac{dP_L}{dt} = \frac{L_L}{d_{L_L}} - \mu_{P_L}P_L - \frac{P_L}{d_{P_L}}.$$

$$\frac{dS_M}{dt} = \frac{P_L}{2d_{P_L}} - \lambda_M S_M - \mu_M S_M,$$

$$\frac{\partial E_M}{\partial t} = \lambda_M S_M - \lambda_M(t - \tau_M)S_M(t - \tau_M)P_M - \mu_M E_{M_i},$$

$$\frac{dI_M}{dt} = \lambda_M(t - \tau_M)S_M(t - \tau_M)P_M - \mu_M I_M,$$

Larval mortalities (μ_E and μ_L) are regulated by a density dependence, such that mortality increases with increasing population size. These mortalities are a function of both the overall larval population size and the (time-varying) carrying capacity K_t, which describes the capacity of the hydrological environment to support breeding sites. Solving the above

differential equations at equilibrium per (White et al., 2011) and rearranging allows derivation of an expression for the larval carrying capacity in terms of the number of observed mosquitoes at equilibrium:

$$K_L = M_0 \frac{2d_{L_L}\mu_0\left(1+d_{P_L}\mu_{P_L}\right)\gamma_L(x+1)}{\left(\frac{x}{\mu_{L_L}d_{E_L}}-\frac{1}{\mu_{L_L}d_{L_L}}-1\right)}$$

where M_0 is the initial female mosquito density (dependent on the EIR which is a tuneable model parameter), μ_0 is the baseline mosquito death rate and

$$x = -\frac{1}{2}\left(\gamma_L\frac{\mu_{L_L}}{\mu_{E_L}}-\frac{d_{E_L}}{d_{L_L}}+\left(\gamma_L-1\right)\mu_{L_L}d_{E_L}\right) + \sqrt{\frac{1}{4}\left(\gamma_L\frac{\mu_{L_L}}{\mu_{E_L}}-\frac{d_{E_L}}{d_{L_L}}+\left(\gamma_L-1\right)\mu_{L_L}d_{E_L}\right)^2+\gamma_L\frac{\nu\mu_{L_L}d_{E_L}}{2\mu_{E_L}\mu_0 d_{L_L}\left(1+d_{P_L}\mu_{P_L}\right)}},$$

where:

$$\nu = \frac{\beta_L\mu_M e^{\mu_M/f}}{\mu_M\left(e^{\mu_M/f}-1\right)\left(1-e^{-\mu_M/f}\right)}\ .$$

Here β_L is the maximum number of eggs per oviposition per mosquito and μ_M is the mosquito death rate defined as:

$$\mu_M = -f_R\log\left(p_1p_2\right),$$

where p_1 is the probability of a mosquito surviving one feeding cycle, p_2 is the probability of surviving one resting cycle and f_r is the feeding rate. The probability that a mosquito survives the extrinsic incubation period is therefore:

$$P_M = e^{-\mu_M\tau_M}$$

Seasonality is incorporated in the model by allowing a time-varying carrying capacity, which is represented in the framework as:

$$K(t) = K_L V\left(t\right)$$

Where K_L is the equilibrium carrying capacity and $V(t)$ a time-varying relative vector density derived from the collated *Anopheles stephensi* profiles that describes the density of the vector at time t relative to the mean annual vector density. Specifically, I took the fitted mosquito density temporal profile for each time-series, and then normalised it according to the following equation:

$$V(t) = \frac{D_t}{D_{Peak}}$$

where D_t is the fitted vector density (i.e. from the negative binomial gaussian process) at timepoint t, and D_{Peak} is the vector density at the point of highest density across the course of the year so that the peak density takes the value of 1 and all other values are <1 in proportion to the peak. $V(t)$ is then a multiplicative factor applied to the carrying capacity K_L to produce a time-varying carrying capacity ($K(t)$) that matches the temporal dynamics of each vector temporal profile. Note that this formulation makes the implicit assumption the seasonality profile inferred for adult densities can be directly applied to a larval carrying capacity (K_L) with any adjustment. Such an adjustment would require the carrying capacity to be lagged by a period specifying the time between larval emergence and development into mature adult mosquitoes. Given the primary aim of this work is to explore how the degree of seasonality (rather than exact timing) influences the impact of seasonally delivered interventions, adjusting for this delay was deemed unnecessary.

It is assumed that 50% of the emergent adult mosquitoes are female and all enter the susceptible state S_M Susceptible mosquitoes get infected at a rate that depends on the prevalence of infection in the human population (and includes a time lag to account for the time taken for infected humans to produce the transmissible gametocyte form of the parasite). These mosquitoes become infected at a rate that depends on the infectiousness of the human population including an appropriate time-lag τ_G to account for the time taken for parasites to become infectious gametocytes. The force of infection on mosquitoes, λ_M is the sum of the contributions from the different human infection state compartments:

$$\lambda_M(t) = \frac{\theta}{\omega} \sum_i \sum_j \zeta_j \psi_i \left(c_D D_{i,j}(t-\tau_G) + c_T T_{i,j}(t-\tau_G) + c_A A_{i,j}(t-\tau_G) + c_U U_{i,j}(t-\tau_G) \right)$$

where the parameter θ is the rate at which individuals are bitten and depends on the patterns and coverage of vector control in a setting (and so is described below in the section on the way in which IRS is modelled). C_U, C_A, C_D and C_T are parameters specifying the relative

infectiousness to mosquitoes from of humans belonging to the sub-patent infection, asymptomatic, clinical disease and successfully treated compartments respectively. The mean EIR experienced by adult humans at time t is defined to be:

$$\varepsilon_0(t) = \frac{1}{\omega} I_M(t).$$

where the parameter ω represents a normalising constant for the biting rate over all ages and adjusts exposure based on the comparative proportion of the population each age-group corresponds to:

$$\omega = \int_0^\infty \psi(a)\mathrm{g(a)da}$$

Modelling the Impact of Indoor Residual Spraying (IRS)

The mode of action and impact of IRS is modelled following a previously established intervention model (Winskill et al., 2017; Griffin et al., 2010; Sherrard-Smith et al., 2018) that probabilistically models the likelihood of the different outcomes that can arise during and following attempted feeding by mosquitoes. Specifically, it generates the probability of different events occurring following an attempt by a blood-seeking mosquito to feed on human. Within the modelling framework, we consider 5 different outcomes of a mosquito attempting to feed (and use the model described below to calculate the probabilities of each of these outcomes):

1) It bites a non-human host
2) It is killed by IRS before it bites
3) It is killed by IRS after it bites
4) It successfully feeds and survives that feeding attempt
5) It is repelled without feeding, either through the actions of IRS.

Define the probability that a mosquito of a given species successfully bites host i during a single attempt as y_i; the probability that a mosquito bites a host and survives the feeding attempt as w_i, and the probability that the mosquito is successfully repelled without feeding as z_i. Together, these probabilities exclude natural vector mortality so that for an individual i, in the absence of any protection or vector control interventions in place, both $y_i = 1$ and $w_i = 1$, and $z_i = 0$.

During a single feeding attempt the probability that a mosquito successfully feeds is:

$$W = (1 - Q_0) + Q_0 \sum_i \pi_i w_i \ ,$$

and is repelled without feeding with probability:

$$Z = Q_0 \sum_i \pi_i z_i ,$$

where in both equations Q_0 is the proportion of bites taken on humans in the absence of any intervention and π_i is the proportion of the total bites occurring on humans that person i receives in the absence of any intervention. Two periods of time relating to feeding are also defined here – the length of time the mosquito spends looking for a blood meal (δ_1); and the length of time resting following ingestion of a blood meal (δ_2). Given these two durations, the mosquito rate of feeding f_R is then given by:

$$f_R = \frac{1}{\delta_1 + \delta_2}.$$

It is assumed that that δ_2 is unaffected by vector control interventions, but that in the presence of interventions, the fact that mosquitoes can be repelled without feeding (i.e. $Z > 0$) leads to δ_1 increasing to:

$$\delta_1 = \frac{\delta_{10}}{1 - Z}$$

where δ_{10} is the value of δ_1 in the absence of any vector control interventions. Assuming a constant natural death rate μ_0 during each of the periods δ_1 and δ_2, the probabilities of surviving the periods of feeding (p_1) and resting (p_2) in the absence of any interventions is given by:

$$p_{10} = \exp(-\mu_0 \delta_{10}),\ p_2 = \exp(-\mu_0 \delta_2) ,$$

With interventions p_1 increases to:

$$p_1 = \frac{p_{10} W}{1 - Z\ p_{10}}$$

The probability of surviving one entire feeding and resting cycle is $p_1 p_2$. The mosquito death rate μ_M can therefore be calculated as:

$$p_1 p_2 = \exp(-\mu_M / f_R)$$

$$\mu_M = -f_R \log(p_1 p_2).$$

Given this, in the presence of vector control interventions that alter p_1 or p_2, the probability of surviving the extrinsic incubation period, P_M therefore also changes. The probability that a feeding cycle ends with a successful bite on person i, q_i is:

$$q_i = p_{10}\left(Q_0 \pi_i w_i + Z\ q_i\right)$$
$$q_i = \frac{p_{10} Q_0 \pi_i w_i}{1 - Z\ p_{10}}$$

The probability that a feeding cycle ends with a bite on an animal, q_A:

$$q_A = p_{10}\left(1 - Q_0 + Z\ q_A\right)$$
$$q_A = \frac{p_{10}(1 - Q_0)}{1 - Z\ p_{10}}$$

The proportion of successful mosquito bites which occurs on humans is therefore given by:

$$Q = 1 - \frac{q_A}{q_A + \sum_i q_i} = 1 - \frac{(1 - Q_0)}{(1 - Q_0) + Q_0 \sum_i \pi_i w_i}$$
$$= 1 - \frac{(1 - Q_0)}{W}$$

and the associated biting rate on humans is:

$$\alpha = Q\ f_R \quad .$$

The rate at which person *i* is bitten by a given mosquito species is therefore:

$$\theta_i = \frac{\alpha\ \pi_i w_i}{\sum_i \pi_i w_i}, \tag{1}$$

When IRS is in use, a proportion of biting mosquitoes some mosquitoes may die following biting a person due to picking up a lethal insecticide dose whilst resting on the walls of the sprayed home. Calculating the force of infection on humans therefore requires the biting rate on each person needs to be inflated by a factor of $\frac{y_i}{w_i}$ giving an effective biting rate of:

$$\tilde{\theta}_i = \frac{\alpha\ \pi_i y_i}{\sum_i \pi_i w_i} \tag{2}$$

The EIR then experienced by person i due to a given mosquito species is therefore $\tilde{\theta}_i I_M$. This formulation of the EIR is shaped by the values of y_i, w_i and z_i, which will vary according to the vector control interventions in place. Below, I describe the form these parameters take and how they are modified in the presence of IRS.

Define the rate at which an individual who is currently indoors is bitten at hour t as $\theta_I(t)$, and the corresponding figure for someone outdoors as $\theta_O(t)$. Knowing the proportion of human hosts indoors $p_I(t)$ at time t then enables calculation of the proportion of bites taken on humans whilst indoors:

$$\Phi_I = \frac{\sum_t p_I(t)\theta_I(t)}{\sum_t \left((1-p_I(t))\theta_O(t) + p_I(t)\theta_I(t)\right)} ,$$

Once a mosquito enters a house to feed, one of three things can happen: it can feed successfully, die or make a repeat attempt at feeding (if the initial attempt was unsuccessful). **Figure 5.2** shows the order in which the different processes operate within the mosquito feeding model used in this framework, specifically the order when a mosquito attempts to feed on an individual protected by IRS. This intervention model yields the following probabilities for the different outcomes of vector feeding in the presence of IRS vector control:

Successful feed (w_i): $1-\Phi_I+\Phi_I(1-r_I)(1-r_{IW}-d_{IW})(1-d_{IF})$

Biting (y_i): $1-\Phi_I+\Phi_I(1-r_I)(1-r_{IW}-d_{IW})$

Repulsion (z_i): $\Phi_I\left(r_I+(1-r_I)r_{IW}\right)$

where Φ_I is the probability (with respect to the time of day) of feeding indoors, r_I is the probability of being repelled before entering the house due to IRS. The parameters r_{IW}, d_{IW} and d_{IF} are the (time-varying) probabilities of being repelled (conditional on having entered the house) before feeding, killed before feeding or killed after feeding following resting on the wall of an IRS sprayed house. The time varying nature of these parameters arises from the chemical decay in the active ingredients over time. These parameter estimates – including both initial efficacy and the rate of decay – are all taken from a previously published systematic review of IRS compounds and their efficacy (Sherrard-Smith et al., 2018). Within the context of the work presented here, I explore the potential impact of 3 different IRS compounds – bendiocarb, clothiandin and pirimiphos methyl. I modelled the impact of a single round of IRS, timed to achieve the maximum reduction in malaria burden (as measured by total annual incidence in the 12-month period following spraying compared to a counterfactual of no IRS). For the purposes of the results presented here, parameter values for the model parameters

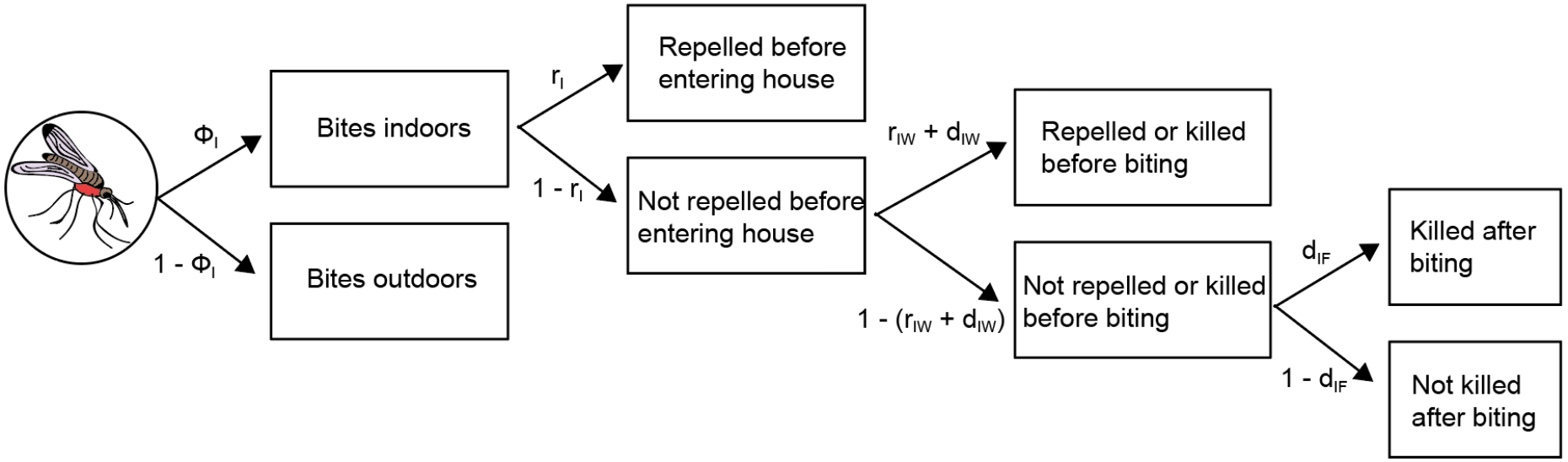


Figure 5.2 Flow diagram illustrating the probabilities associated with feeding outcomes. $\boldsymbol{\Phi_I}$ indicates the proportion of bites taken indoors, $\boldsymbol{r_I}$ the probability of being repelled from entering the house, $\boldsymbol{r_{IW}}$ the probability of being repelled from feeding (conditional on having entered the house) by the IRS, $\boldsymbol{d_{IW}}$ the probability of being killed before feeding and $\boldsymbol{d_{IF}}$ the probability of being killed after feeding.

described above are all taken from (Unwin et al., 2022), except for estimates of *Anopheles stephensi*'s binomics, which are taken from (Hamlet et al., 2022) and the estimates of IRS efficacy over time, which as described above are taken from (Sherrard-Smith et al., 2018).

Results

A total of 65 time-series from studies across Afghanistan, Djibouti, India, Iran, Myanmar and Pakistan were identified (**Figure 5.3A)**. These noisy time-series were then smoothed using a negative binomial gaussian process-based framework (see **Figure 5.3B** for example time-series from each country and **Figure 5.4** for the results of fitting for all of the collated time-series). Substantial variation in temporal dynamics was observed across the collated time-series in terms of the degree and timing of seasonality – ranging from highly seasonal dynamics with a single seasonal peak (e.g. Afghanistan example in **Figure 5.3B**) to more perennial patterns of abundance (Pakistan example in **Figure 5.3B**) and bimodal population dynamics with two peaks observed during a single year (Iran example in **Figure 5.3B**).

To these collated time-series, I then calculated a series of summary statistics that describe and characterise different aspects of their temporal properties and then applied PCA and k-means clustering to the results to cluster the time-series into discrete groups sharing similar temporal patterns. The results presented here highlight two distinct clusters, with each cluster characterised by distinct temporal patterns and degree of their seasonality (defined as the percentage of total vector density that occurs across any continuous 3-month period, **Figure 5.5**). Cluster 1 time-series typically had clear, single seasonal peaks and were more seasonal on average (57% of total vector density in a 3-month period) than Cluster 2 time-series, which had less seasonal (more perennial) patterns of annual abundance (average 36% vector density in any consecutive 3-month period), including time-series with two peaks across the course of a single year. Despite differing significantly in vector abundance seasonality **(Figure 5.5C**, top panel, $p<0.001$), there was no significant difference across Cluster 1 and Cluster 2 time-series in terms of rainfall seasonality **(Figure 5.5C**, bottom panel, $p=0.59$).

In order to further investigate the different patterns of temporal dynamics present in the collated dataset, I re-ran the k-means clustering algorithm this time specifying 4 clusters **(Figure 5.6)**. The less seasonal cluster from the 2 cluster analysis was retained (here Cluster 3), and Cluster 1 from the 2 cluster analysis was further disaggregated into 3 different clusters (here, Clusters 1, 2 and 4), each defined by differences in the timing of their seasonal peak relative to the timing of peaks in monthly rainfall (mean timing of vector density peak 7, 8.25 and 5.86 months after January for Clusters 1, 2 and 4 respectively) and the timing of the vector peak relative to peaks in rainfall (rainfall peak on average 1.03 and 2.32 months before vector

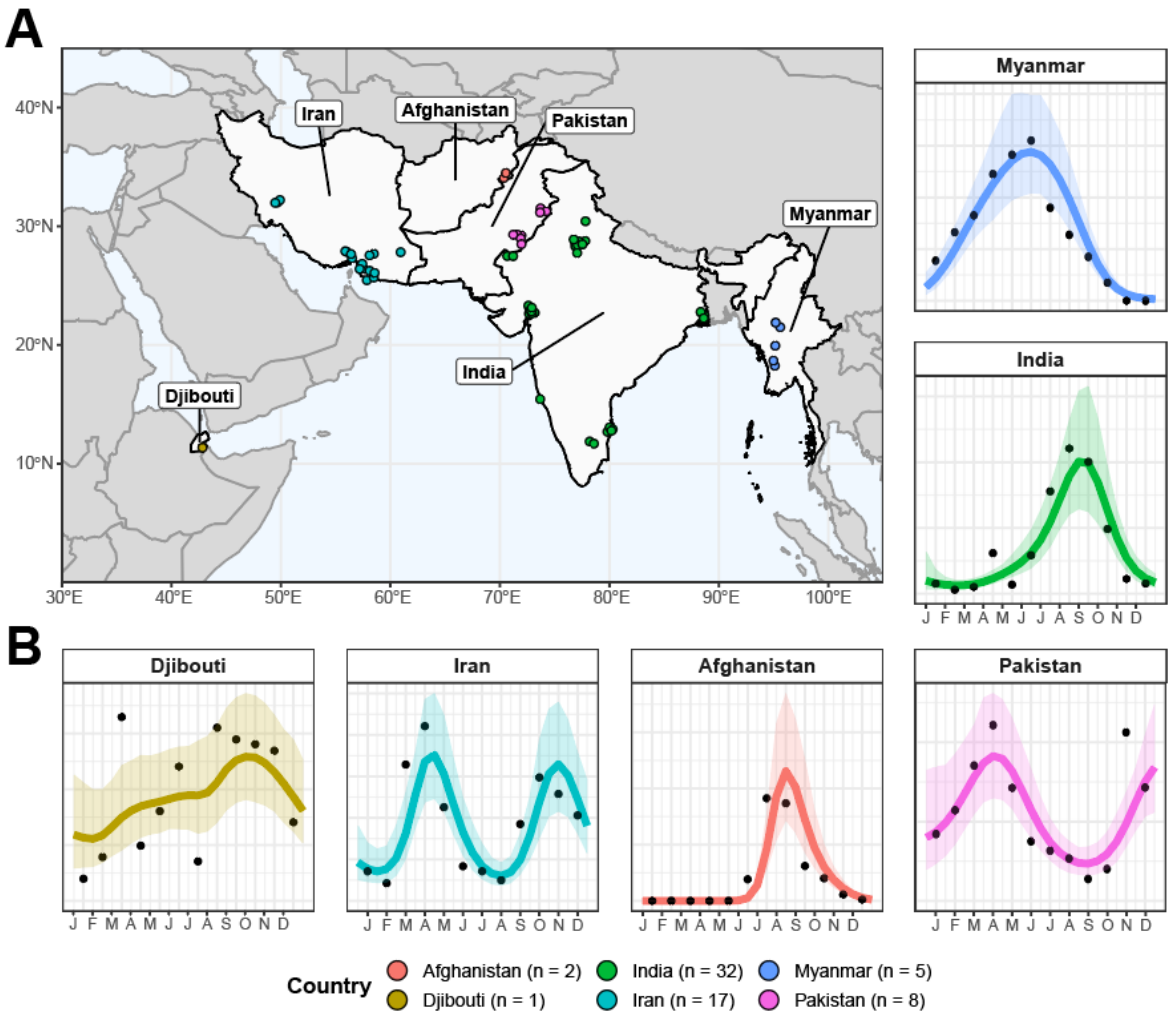


Figure 5.3 Sources and Locations of Anopheles stephensi Time-Series Data and Examples for Each Country.(A) Map of the geographical range over which collated time-series had been carried out in, with countries where studies had been carried out in highlighted in light grey, and the locations of individual studies indicated by points, coloured according to the country they were carried out in (Afghanistan = red, Djibouti = yellow, India = green, Iran = turquoise, Myanmar = blue and Pakistan = pink). **(B)** Example *Anopheles stephensi* time-series from each country, with the empirical monthly mosquito catch (black points) and fitted gaussian process curves (mean = coloured line, ribbon = 95% Bayesian Credible Interval) for each, coloured according to country. The x-axis indicates the month of sampling, the y-axis normalised annual vector density (i.e. arbitrary units).

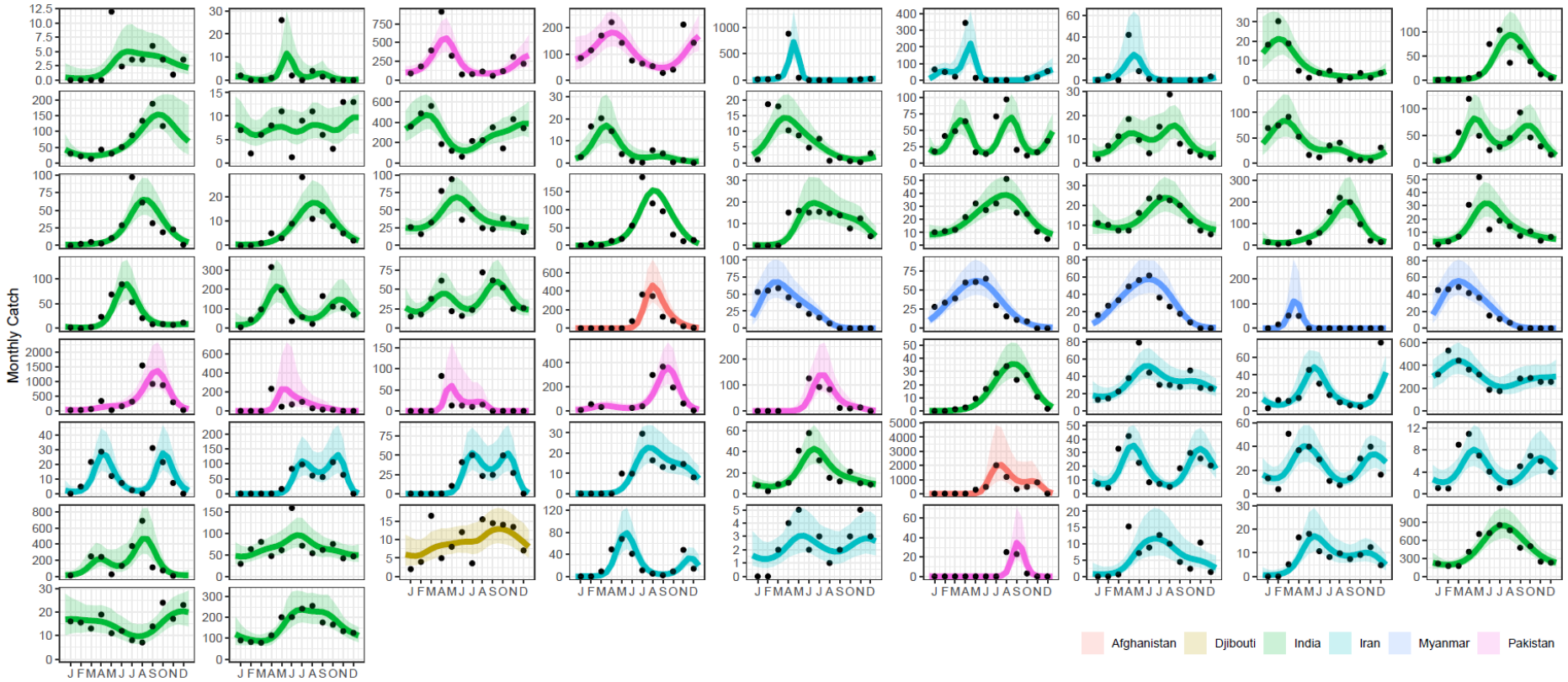


Figure 5.4 Results of model fitting to the longitudinal entomological data collated in this study. Reviews of the literature in tandem with previously published databases of entomological data identified 65 *Anopheles stephensi* time-series matching the inclusion criteria (>10 months of catch data at monthly temporal resolution or finer), and a negative binomial gaussian process with period kernel fitted to each time-series. For the results presented above, black points are the data, and the lines represent the model output, coloured according to the country in which the study was conducted. Line indicates the mean model output, with the shaded ribbon delineating the 95% credible interval (CI).

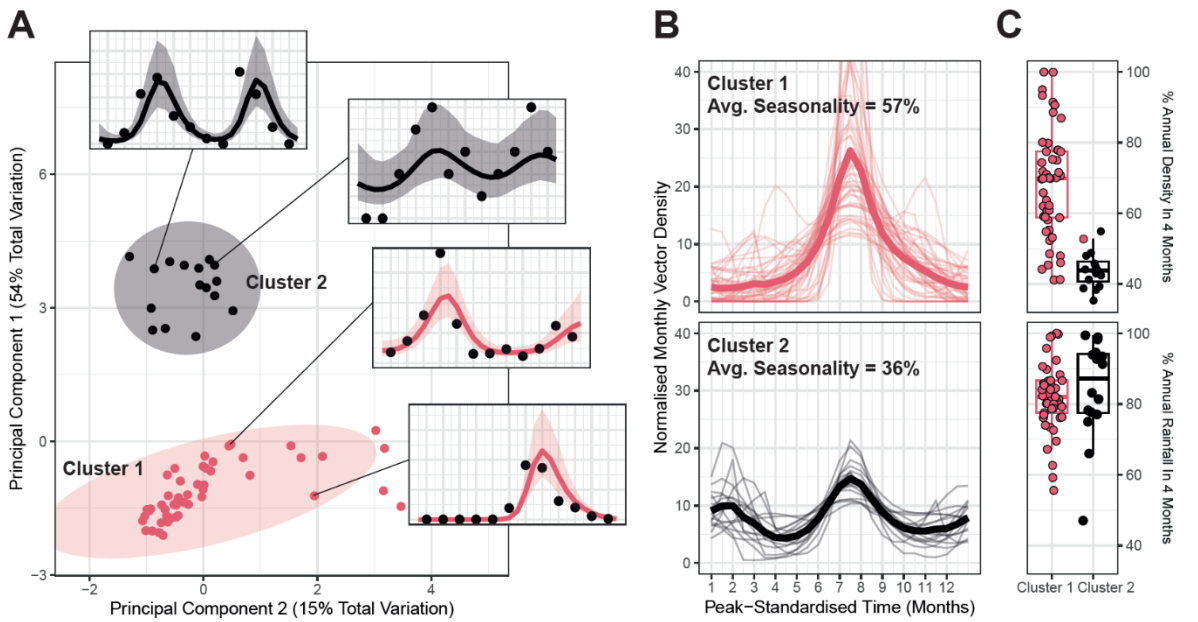


Figure 5.5 Characterisation and Clustering to Identify Time-Series with Similar Temporal Properties. **(A)** Results of principal components analysis (PCA) and k-means clustering for 2 clusters. Points indicate individual time-series, with point colour indicating cluster membership. Ellipsoids demarcate the 75th quantile of the density associated with each cluster. Principal components 1 and 2 are plotted, together explaining 69% of the total variation in temporal properties across the time-series. **(B)** Time-series belonging to each cluster. Pale lines represent individual time-series, brighter line the mean of all the time-series belonging to that cluster – in all cases vector density is normalised to sum to 1 over the course of the year, and time-standardised so that the highest vector density for each time-series is arbitrarily set to occur at month 7. **(C)** Boxplots of the percentage of annual total mosquito catch (top) and annual total rainfall (bottom) for each time-series. Rainfall data comes from the *CHIRPS* dataset and is specific to study location and time-period. Each point indicates an individual time-series.

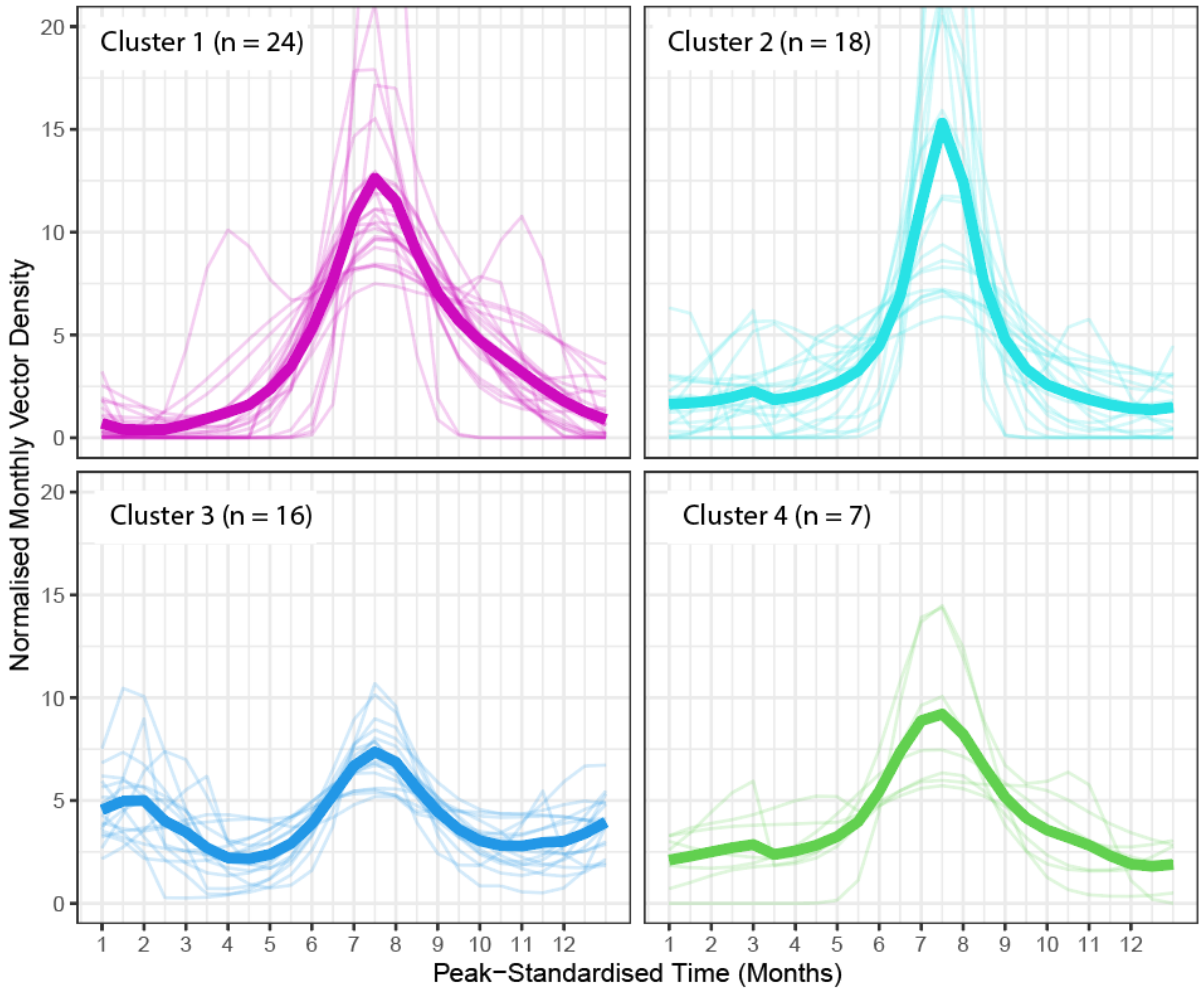


Figure 5.6 Results of Clustering For 4 Clusters Instead of 2. In order to further investigate the different patterns of temporal dynamics present in the collated dataset, I re-ran the k-means clustering algorithm this time specifying 4 clusters. The less seasonal cluster from the 2 cluster analysis in the main text (Cluster 2 in the main text results) was retained (here Cluster 3), and Cluster 1 from the main text was further disaggregated into 3 different clusters (here, Clusters 1, 2 and 4), each defined by different peak timings (mean timing of vector density peak 7, 8.25 and 5.86 months after January for Clusters 1, 2 and 4 respectively) and the timing of the vector peak relative to peaks in rainfall (rainfall peak on average 1.03 and 2.32 months before vector density peak for Clusters 1 and 2, 1.09 months after vector density peak on average for Cluster 4).

density peak for Clusters 1 and 2, 1.09 months after vector density peak on average for Cluster 4) **(Figure 5.6)**.

In order to explore in more depth potential drivers of the observed variation in seasonal dynamics across different time-series, I next fitted a random forest-based classification framework to satellite-derived environmental covariates in order to predict cluster membership (either Cluster 1 or Cluster 2, as defined in **Figure 5.5**). Due to the significant class size imbalance between Cluster 1 (n=49) and Cluster 2 (n=16), I upsampled the Cluster 2 data using the SMOTE algorithm to generate balanced classes. Across the 25 iterations of random forest model fitting, the mean AUC was 0.89 (indicating good predictive performance, **Figure 5.7A**) and on average, the model was able to correctly classify Cluster 1 and Cluster 2 time-series similarly well (83% and 85% of the time respectively). Both population per km^2 and temperature seasonality, as well as a number of measures of landcover (specifically LC30 which corresponds to mosaic cropland/natural vegetation, and LC20 which corresponds to irrigated or post-flooding cropland) were all highly predictive of Cluster membership (**Figure 5.7B)**. Time-series from surveys in locations with lower population density were more likely to belong to Cluster 2 (i.e. less seasonal), as were areas with high values of LC20 (i.e. land predominantly occupied by irrigated or post-flooding cropland). By contrast, areas with high values of LC10 and LC30 (i.e. substantial fraction of land covered either rainfed cropland or a mosaic of cropland and natural vegetation) were more likely to belong to Cluster 1 (i.e. more seasonal), as were areas in which rainfall was strongly seasonal. **Figure 5.8** presents the partial dependence covariate plots for all included covariates and their relationships with cluster membership in full. Examining the association between rurality/urbanicity (as defined by the authors of each study) and temporal dynamics (**Figure 5.7C**), there was an indication of an association with rurality/urbanicity and cluster membership, though this was not statistically significant (chi-squared test, p=0.07). 88% (n=22/25) time-series from urban settings were assigned to Cluster 1, compared to only 65% (n=24/37) from rural settings. Only 12% (3/25) time-series from urban settings were assigned to Cluster 1, compared to 35% (n=13/37) from rural settings.

In order to assess the sensitivity of the results to the upsampling procedure applied, I also carried out model fitting with no upsampling to balance the classes. Model predictive performance and variable importance rankings were similar when no up-sampling was applied to the dataset (AUC=0.81, **Figure 5.9**), though average predictive accuracy on Cluster 2 (50%) was substantially lower than for Cluster 1 time-series (94%), likely because of the significant disparities in class size. I also assessed model performance and variable importance ordering when fitting the model and explicitly holding out a small subset of the data to evaluate model performance (n=7 time-series, see **Figure 5.10**). Across the 25 iterations of random forest

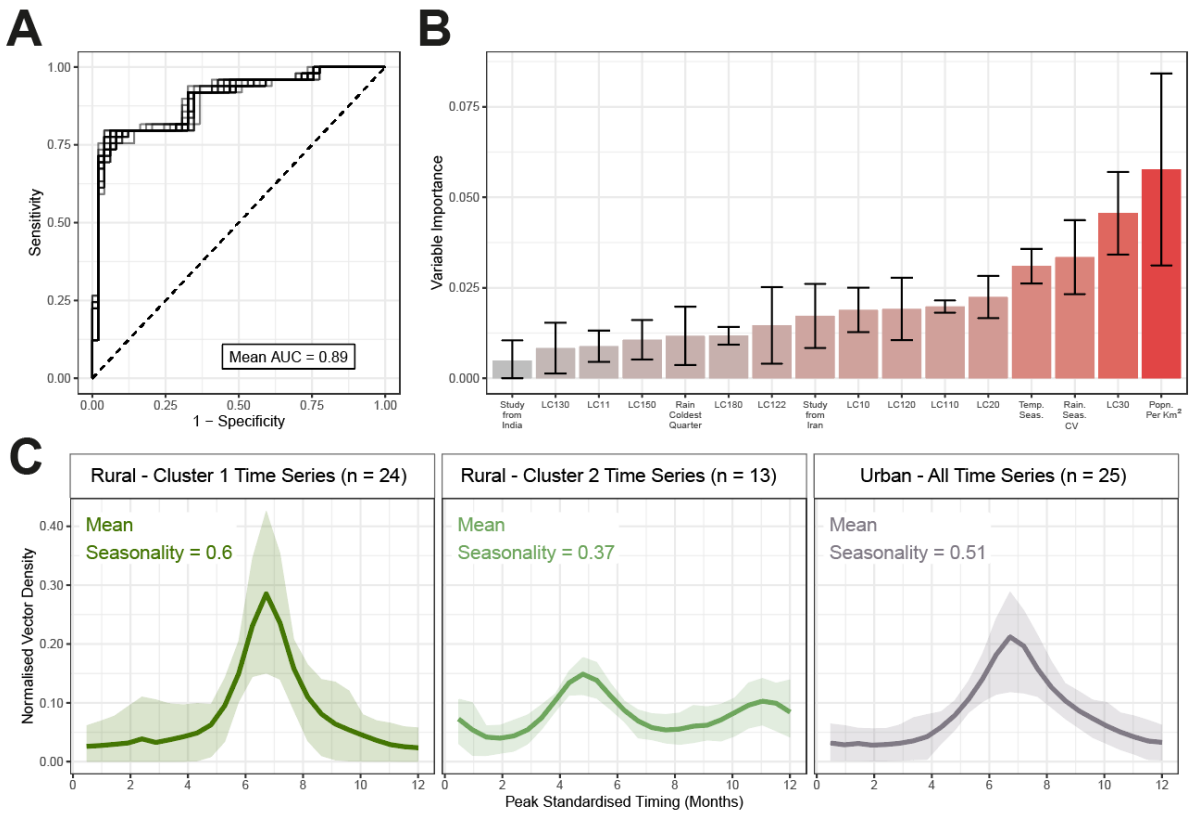


Figure 5.7 Random Forest Prediction of Temporal Cluster Membership. A random forest-based classification modelling framework was used to predict membership of the temporal cluster (either Cluster 1 or Cluster 2, as defined in **Fig 2)** and explore the ecological factors underpinning variation in *Anopheles stephensi* seasonality. **(A)** Receiver-operator curve (ROC) for each of the 25 individual iterations of random forest model fitting carried out, with results for each displayed as grey lines. The mean AUC across these 25 iterations was 0.89. **(B)** Variable importance plot for the covariates included in the random forest model – bar height indicates the mean variable importance across the 25 individual iterations of random forest fitting, with error bars representing the 95% confidence interval. **(C)** Collated *Anopheles stephensi* time-series, disaggregated according to urbanicity and cluster membership. Cluster 1 and Cluster 2 time-series from rural locations are plotted separately; all time series for time-series carried out in urban locations (22 belonging to Cluster 1 and only 3 belonging to Cluster 2) are plotted together. Coloured line indicates the mean and ribbon indicates the 90% range spanned by the group of time series belonging to each displayed grouping. For landcover (LC) variables, LC10 = rainfed cropland, LC11 = herbaceous cover, LC20 = irrigated or post-flooding cropland, LC30 = mosaic of cropland and natural vegetation, LC110 = mosaic of herbaceous cover and tree/shrub cover, LC120 = shrubland, LC122 = Deciduous shrubland, LC130 = grassland, LC150 = sparse vegetation (<15% cover of any type), LC180 = mixture of herbaceous cover and flooded with water.

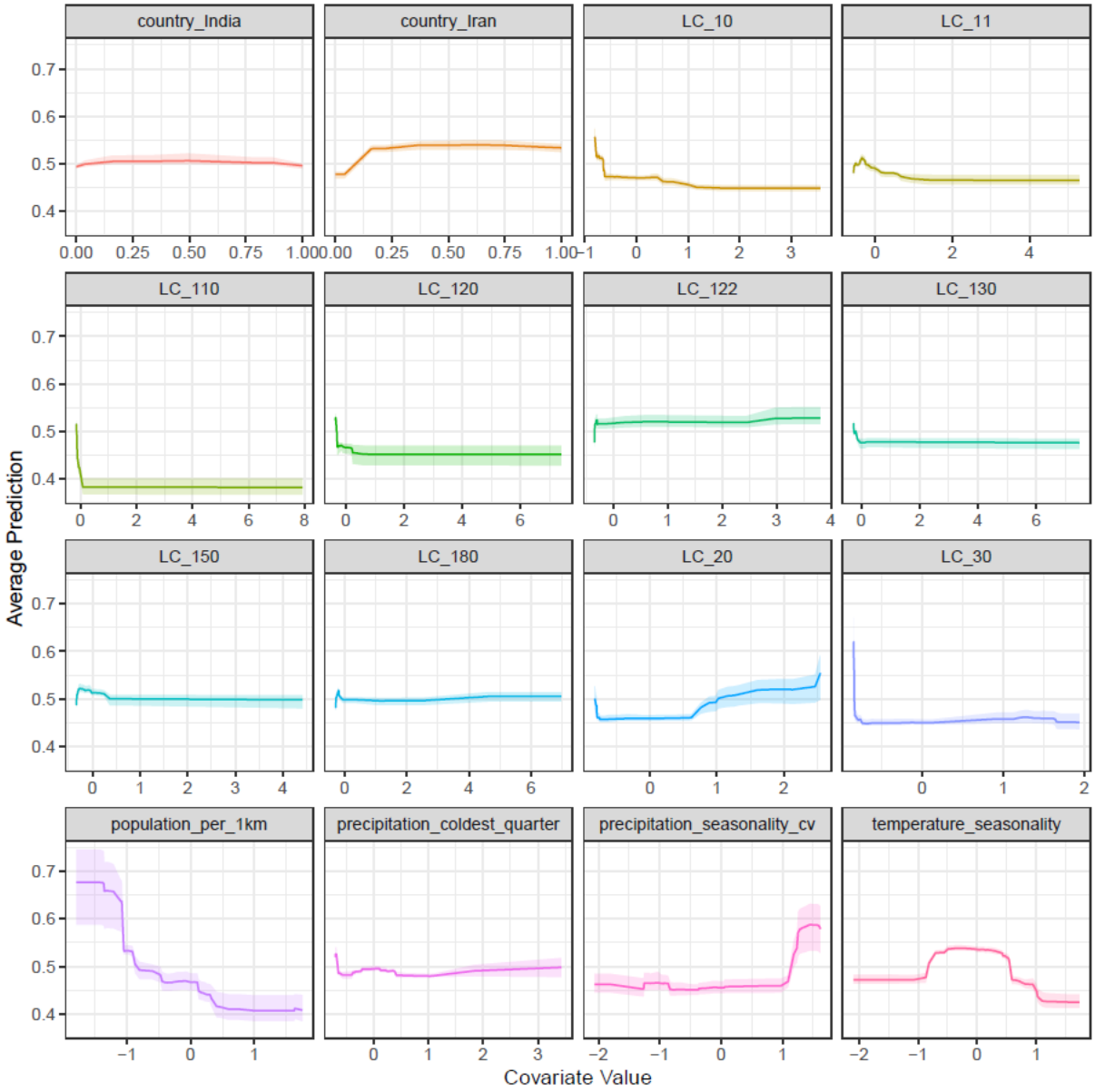


Figure 5.8 Partial Dependence Plots for Covariates Used in the Random Forest Classification Modelling. The y-axis on the left shows the probability of the time-series belonging to Cluster 2 (i.e. a high probability indicates the time-series is predicted to likely belong to Cluster 2, a low probability indicates the time-series likely belongs to Cluster 1). The x-axis describes the value of the (scaled, normalised) covariate. For landcover (LC) variables, LC10 = rainfed cropland, LC11 = herbaceous cover, LC20 = irrigated or post-flooding cropland, LC30 = mosaic of cropland and natural vegetation, LC110 = mosaic of herbaceous cover and tree/shrub cover, LC120 = shrubland, LC122 = Deciduous shrubland, LC130 = grassland, LC150 = sparse vegetation (<15% cover of any type), LC180 = mixture of herbaceous cover and flooded with water.

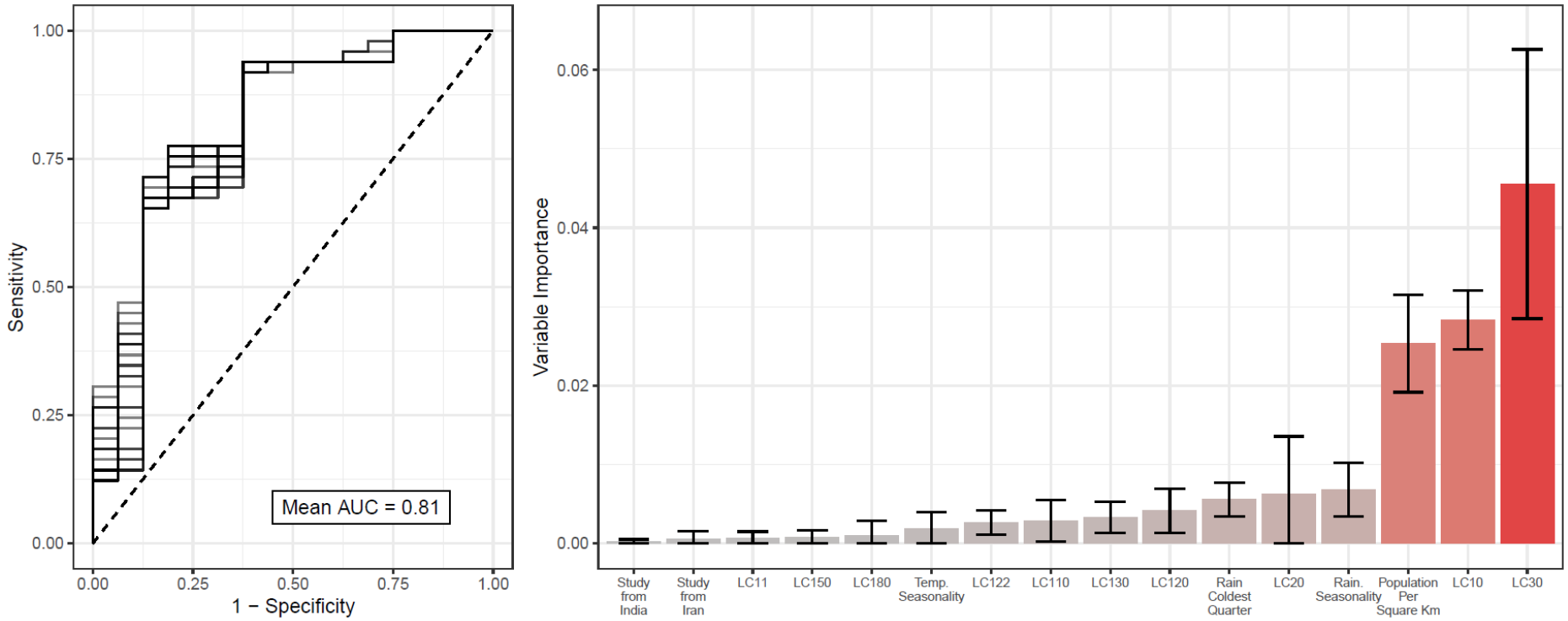


Figure 5.9 Random Forest Classification Results Without Upsampling Cluster 2. Due to the extreme class-imbalance of Clusters 1 and 2 (49 vs 16 time-series respectively), the results presented in the main text are following upsampling of the Cluster 2 time-series to create a dataset with equal numbers of time-series belonging to each cluster. As a sensitivity analysis, I also carried out the random forest fitting without upsampling and assessed both model fit (as measured by AUC) and variable importance. Model performance was somewhat reduced compared to the upsampled data (mean AUC of 0.81 vs mean AUC >0.9 for the upsampled dataset), whilst variable importance results were broadly consistent across both analyses, with population per square kilometre and various land-cover measures all emerging as important predictive variables. For landcover (LC) variables, LC10 = rainfed cropland, LC11 = herbaceous cover, LC20 = irrigated or post-flooding cropland, LC30 = mosaic of cropland and natural vegetation, LC110 = mosaic of herbaceous cover and tree/shrub cover, LC120 = shrubland, LC122 = Deciduous shrubland, LC130 = grassland, LC150 = sparse vegetation (<15% cover of any type), LC180 = mixture of herbaceous cover and flooded with water.

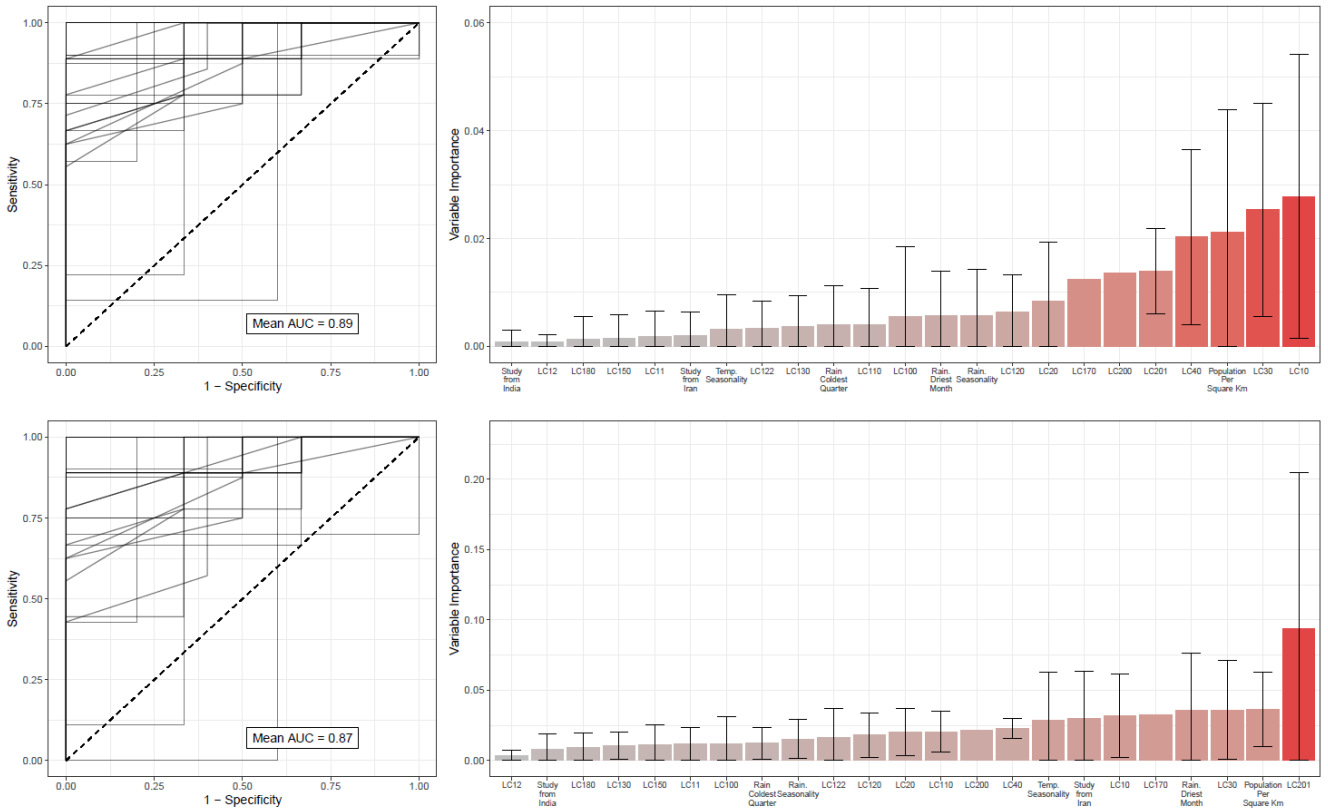


Figure 5.10 Random Forest Classification Results With Held-Out Data. Due to the overall sample size (n = 65 time-series), the results presented in the main text were generated using a random forest-based workflow where final model fitting (using hyperparameters tuned using 6-fold cross-validation) utilised the entirety of the dataset. As a sensitivity analysis, I also carried out the random forest fitting holding out a small portion of the dataset (n = 9) during model fitting, with model performance subsequently evaluated on this held-out data. Results presented above are in the case where data was upsampled to address class imbalance (top) and where no upsampling was carried out (bottom). For landcover (LC) variables, LC10 = rainfed cropland, LC11 = herbaceous cover, LC20 = irrigated or post-flooding cropland, LC30 = mosaic of cropland and natural vegetation, LC110 = mosaic of herbaceous cover and tree/shrub cover, LC120 = shrubland, LC122 = Deciduous shrubland, LC130 = grassland, LC150 = sparse vegetation (<15% cover of any type), LC180 = mixture of herbaceous cover and flooded with water.

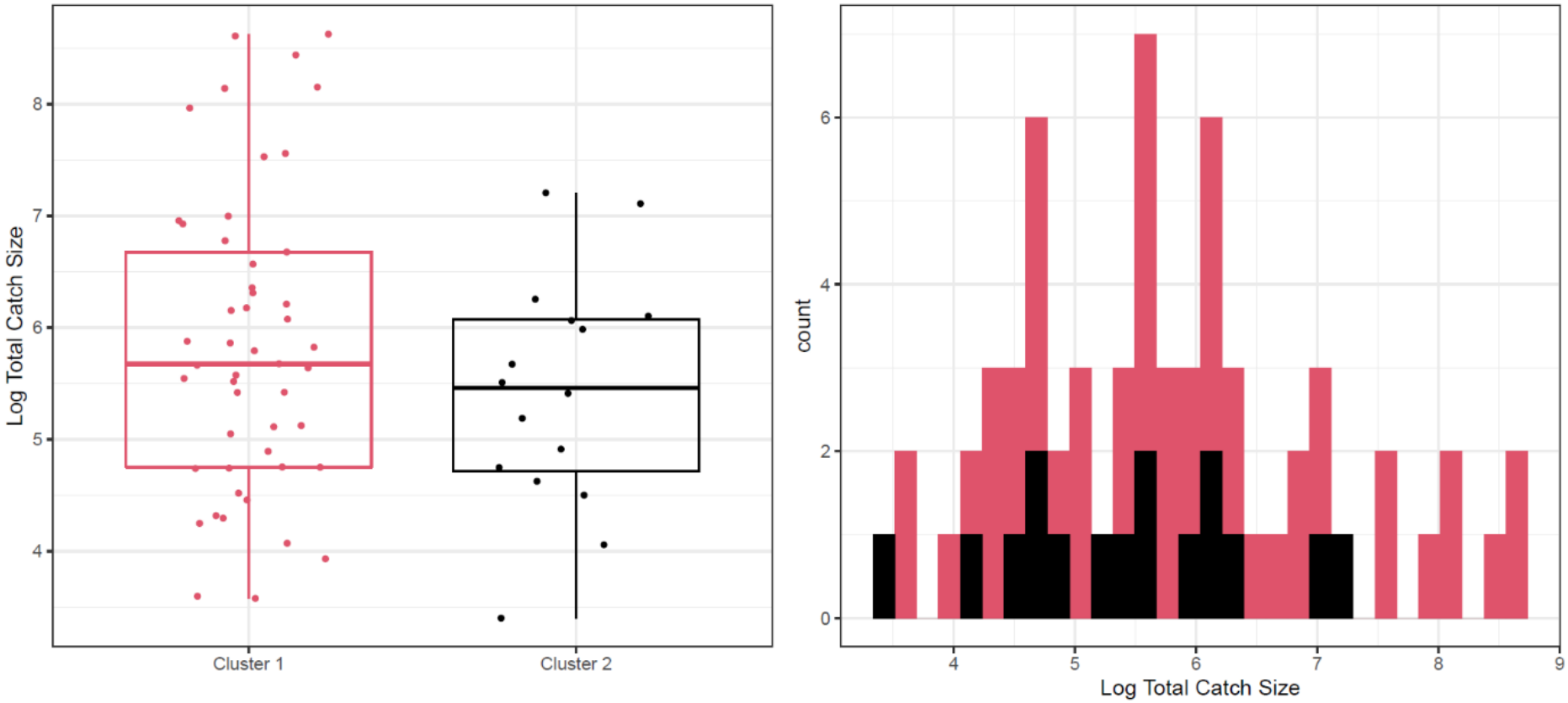


Figure 5.11 Exploring Variation In Total Catch Size By Cluster. Boxplot and histogram of the total number of *Anopheles stephensi* mosquitoes caught over the duration of each study, coloured according to cluster membership. Total catch size was highly overdispersed, varying over several orders of magnitude across both clusters.

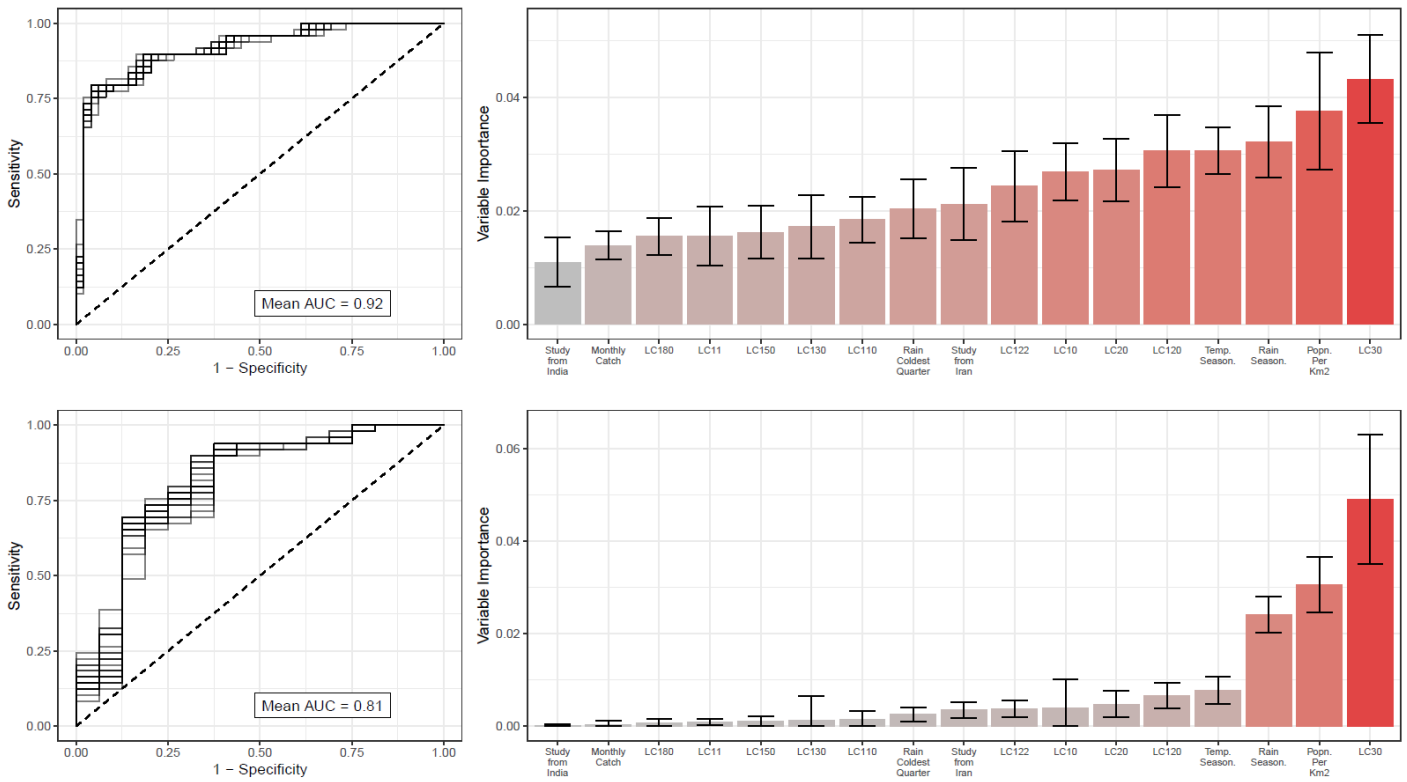


Figure 5.12 Random Forest Classification Results Including Monthly Catch Size As A Model Covariate. As a sensitivity analysis, I also carried out the random forest fitting including average monthly catch size for each time-series as a predictive covariate, to assess whether any of the cluster assignments might be due to the model spuriously not detecting seasonal peaks in studies with low catch sizes. Results presented above are in the case where data was upsampled to address class imbalance (top) and where no upsampling was carried out (bottom), with predictive performance and variable importance largely unchanged compared to results presented in the main text (which do not include average monthly catch size as a covariate). For landcover (LC) variables, LC10 = rainfed cropland, LC11 = herbaceous cover, LC20 = irrigated or post-flooding cropland, LC30 = mosaic of cropland and natural vegetation, LC110 = mosaic of herbaceous cover and tree/shrub cover, LC120 = shrubland, LC122 = Deciduous shrubland, LC130 = grassland, LC150 = sparse vegetation (<15% cover of any type), LC180 = mixture of herbaceous cover and flooded with water.

model fitting and performance evaluation carried out, model performance and variable importance ordering remained similar.

Study catch size was highly variable between the studies (**Figure 5.11**), and I therefore assessed whether any of the results presented here might be due to systematic differences in study catch size between clusters. Whilst median catch size across the two clusters did not differ (Moody's Median Test, p=0.47), the mean catch size did differ (t-test, p=0.025). I therefore carried out a sensitivity analysis including study average monthly catch as a covariate in the random forest model (**Figure 5.12**) – model results were largely unchanged compared with the other analyses that lacked monthly catch size as a covariate – predictive performance and results around the comparative importance of the different included variables remained similar, with monthly catch size the 2nd least most important variable in both instances of model fitting with and without upsampling. Via a random-forest based regression framework, I also assessed the capacity of the collated ecological covariates to predict seasonality (defined as the largest percentage of vector catch in any continuous 3-month period). Model predictive performance was more modest than in the classification setting, although model estimates and empirical values were positively correlated (r=0.43, see **Figure 5.13**) and measures of landcover including LC10, LC20 and LC30, as well as temperature and rainfall seasonality were all strongly associated predictors of seasonality, as in the classification modelling setting.

I next explored the potential implications of these results on the seasonal dynamics of *Anopheles stephensi* across the Horn of Africa. I collated the same satellite-derived environmental covariates for countries across the Horn of Africa where *Anopheles stephensi* has been reported and integrated them with the trained random forest classification model to predict potential temporal dynamics (as denoted by Cluster membership) of *Anopheles stephensi* across the region (**Figure 5.14**). The results highlight distinct geographical areas across the region considered more likely by the model to belong to Cluster 1 (more seasonal, **Fig. 5.14A**) and Cluster 2 (less seasonal), as well as substantial areas of significant uncertainty. Using the collated seasonal profiles belonging to each cluster, I next asked what consequences this uncertainty about the degree and timing of *Anopheles stephensi* seasonality might have on entomological surveillance of the vector. Specifically, I asked what the probability of missing *Anopheles stephensi* in entomological surveys might be as a function of the number of consecutive months sampled (with start month selected randomly i.e., assuming no knowledge of *Anopheles stephensi's* temporal dynamics). In instances where sites are sampled for a limited number of months, there is a significant risk of missing *Anopheles stephensi*, with the exact value dependent on the specific time-series (i.e. *Anopheles stephensi* temporal profile) being considered (**Figure 5.14B**). In the absence of a

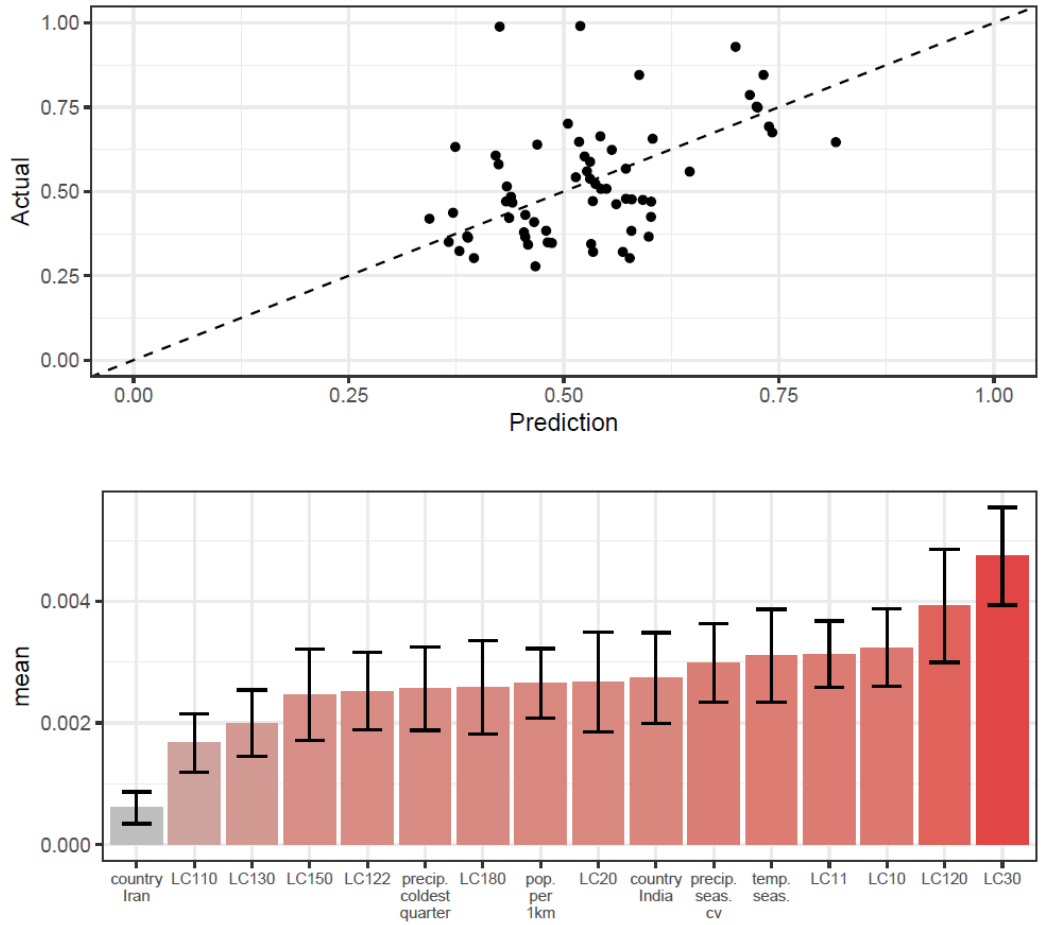


Figure 5.13 Random Forest Prediction of Percentage of Vector Density In Any 3 Month Period. As a further sensitivity analysis, I used a random forest modelling approach to predict the percentage of vector density occurring in a single continuous 3-month period. Results presented above are the average of 25 independents random forest model fittings, with no upsampling of the data carried out, and the final model fitted (using hyperparameters tuned using 6-fold cross-validation) to the full dataset. Model predictive power was moderate, with correlation between predicted and actual values = 0.43. For landcover (LC) variables, LC10 = rainfed cropland, LC11 = herbaceous cover, LC20 = irrigated or post-flooding cropland, LC30 = mosaic of cropland and natural vegetation, LC110 = mosaic of herbaceous cover and tree/shrub cover, LC120 = shrubland, LC122 = Deciduous shrubland, LC130 = grassland, LC150 = sparse vegetation (<15% cover of any type), LC180 = mixture of herbaceous cover and flooded with water.

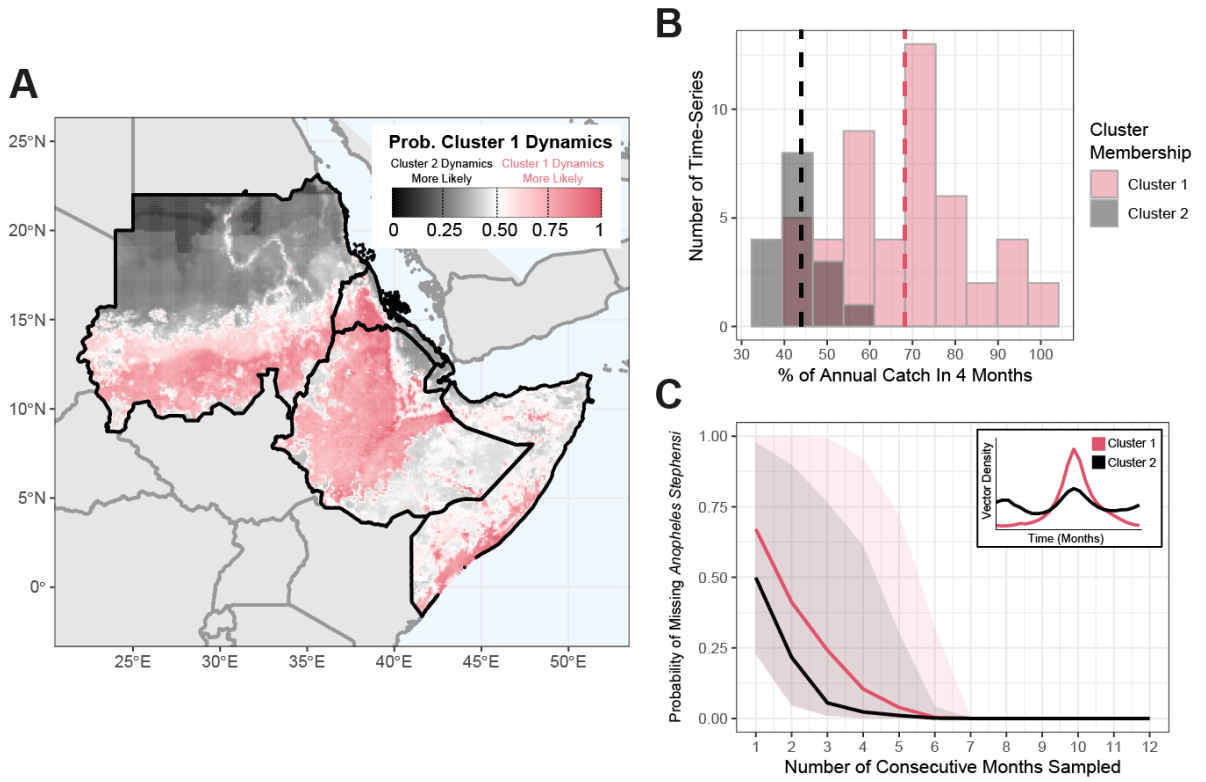


Figure 5.14 Predicting the Possible Seasonal Dynamics of Anopheles stephensi Across the Horn of Africa. (A) Environmental covariates were collated across countries in the Horn of Africa where *Anopheles stephensi* has been found, and the fitted random forest classification model from **Fig 3** used to predict potential temporal dynamics. Map shows the probability of temporal dynamics belonging to Cluster 1, with pink corresponding to Cluster 1 dynamics being more likely than Cluster 2, black indicating Cluster 2 dynamics are more likely than those for Cluster 1, and white indicating both are equally likely. **(B)** Histogram of the percentage of total annual catch in any continuous 4-month period for each time-series, coloured according to the cluster the time-series belong to. **(C)** The probability of missing *Anopheles stephensi* in an entomological catch survey (y-axis) as a function of the number of consecutive months randomly sampled, assuming the start-month is picked at random. Coloured lines are the mean results across all time-series belonging to each cluster, with the shaded area indicating the range spanned by all time-series belonging to each cluster.

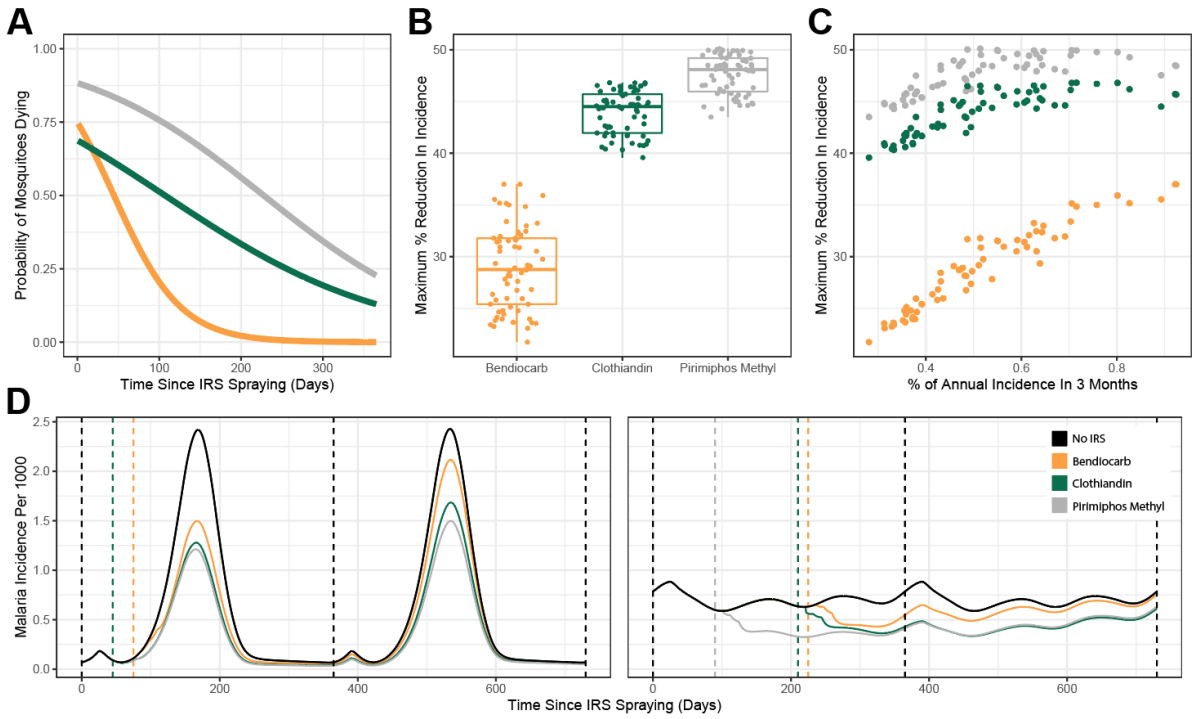


Figure 5.15 Modelling the Public-Health Impact of Indoor Residual Spraying (IRS) and How This Is Impacted by Anopheles stephensi Seasonality. (A) Probability of mosquitoes dying upon exposure to each IRS compound – yellow indicates bendiocarb, green indicates clothiandin and grey indicates pirimiphos methyl. **(B)** Percentage reduction in annual incidence (with optimal timing of IRS delivery), for each of the IRS compounds considered. Individual points correspond to specific time-series. **(C)** The relationship between percentage reduction in annual malaria incidences and the overall seasonality of malaria incidence in the setting (as modelled and implied by each *Anopheles stephensi* temporal profile). **(D)** Malaria incidence modelling results for a highly seasonal (left) and less seasonal (right) setting. Black lines indicate endemic dynamics in the absence of any IRS, coloured solid lines indicate incidence following a single IRS round (with timing of the round indicated by the coloured dashed lines).

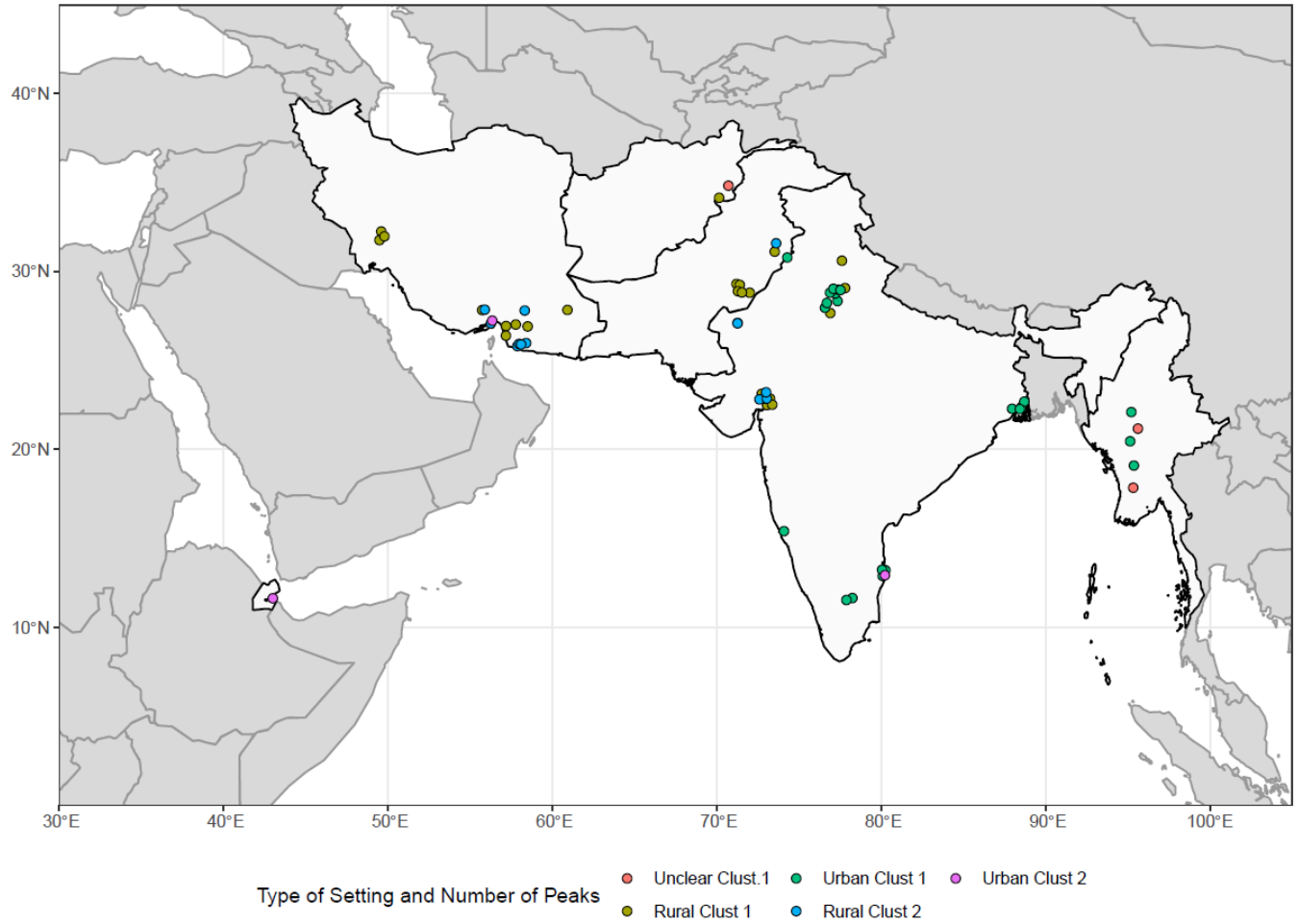


Figure 5.16 Sources and Locations of Anopheles stephensi Time-Series Data According to Urban/Rural Assignment. Collated time-series are displayed above coloured according to 1) whether or not the study was carried out in an urban or rural location; and 2) which cluster they were assigned to.

detailed understanding of *Anopheles stephensi*'s anticipated temporal dynamics, sampling for a limited number of months therefore appears to pose a significant risk of missing *Anopheles stephensi* and erroneously concluding it is absent.

I next integrated the collated temporal profiles of *Anopheles stephensi* abundance with a model of malaria transmission to model the impact of *Anopheles stephensi* seasonality on vector control measures. Specifically, I explored how variation in *Anopheles stephensi* temporal dynamics influences the impact of indoor residual spraying (IRS), a key malaria control intervention. I explore three different IRS compounds – bendiocarb, clothiandin and pirimihpos methyl (each commonly used throughout the Horn of Africa) that span a range of different functional half-lives and lethalities (**Figure 5.15A**) and model the impact of a single annual round of IRS, assuming optimal timing (as defined by subsequent impact on disease burden) of delivery is possible. Across the *Anopheles stephensi* temporal profiles considered, pirimiphos methyl consistently had the highest impact, reducing annual malaria incidence by 47.6% on average in the 12 months following spraying, compared to only 43.9% and 28.9% on average for clothiandin and bendiocarb (**Figure 5.15B**). The overall impact on disease burden was highly dependent on the degree of seasonality, with IRS having the most impact in highly seasonal settings, and the least impact in settings where the degree of seasonality was minimal (**Figure 5.15C** and **Figure 5.15D**). This influence of seasonality was most significant for bendiocarb (the insecticide with the shortest half-life), with reduction in malaria burden ranging from only 21.7% in the lowest seasonality setting, through to as high as 37.0% in the most seasonal setting, a 1.7 fold difference. I assume in all instances optimal timing of IRS delivery relative to seasonal peaks, which in turn requires sufficient understanding of the vector's dynamics and when peaks in density are going to occur.

Discussion

Invasion and establishment of *Anopheles stephensi* across the Horn of Africa represents an urgent threat to malaria control in the region. Understanding the temporal profile of malaria risk this vector might lead to will represent a crucial input to effective deployment of surveillance, monitoring and control interventions aimed at mitigating the potential impact of *Anopheles stephensi*'s arrival, particularly in urban settings across the region where malaria has historically been largely absent or only minimally present. Collating data from across South Asia and the Middle East, I identify extensive diversity across *Anopheles stephensi* populations in the extent and nature of their seasonal dynamics. This variation is associated with a wide array of ecological factors, including patterns of land-use and temporal fluctuations in rainfall and temperature. Perhaps most crucially, I find evidence of distinct temporal dynamics across rural

and urban settings – this variation has material consequences for the efficacy of interventions aimed at controlling the threat this vector poses to urban areas across the Horn of Africa.

The analyses have identified several ecological factors associated with *Anopheles stephensi* seasonal dynamics. Patterns of land-use (e.g. whether the land is rain-fed or irrigated), as well as the seasonality of temperature were identified as key drivers of the extent and nature of *Anopheles stephensi* seasonality. This is consistent with the work presented in Chapter 4 that highlighted features of the local hydrological environment and its interaction with land-use (particularly whether it is predominantly rain-fed or static, perennially available bodies of water) as a key factor shaping the seasonality of diverse anopheline species across India. Similarly, previous work has identified temperature as a key driver of mosquito population dynamics, due to its impact on an array of mosquito life-history traits including biting rate, lifespan and fecundity (amongst several others) (Beck-Johnson et al., 2017; Mordecai et al., 2019). These results therefore highlight the importance of considering both the hydrological environment (including patterns of land use and their interaction with rainfall and other hydrological features), as well seasonal fluctuations in temperature when trying to understand seasonal patterns of mosquito abundance.

Perhaps most notably, the analyses identified population per km^2 as a key predictor of cluster membership, with high population density being strongly associated with Cluster 1 dynamics (i.e. more seasonal patterns of abundance), a finding that was also observed when stratifying surveys according to whether they had been carried out in urban or rural settings. This potential disparity in temporal dynamics across rural and urban settings will likely have implications for the public health impact of different control interventions. These results suggest that urban *Anopheles stephensi* populations are likely to display seasonal dynamics, supporting the utility of interventions like IRS in these settings – the same is not necessarily true in rural settings, where a range of seasonal profiles including more perennial patterns of abundance were observed. Implementing these measures and achieving sufficient IRS (and ITN) coverage in urban settings is likely to prove challenging, given the historical absence of large-scale vector control campaigns. If these barriers can be surmounted, such measures are likely to be impactful however, given the results presented here as well as previous modelling work that has identified low altitude urban areas with minimal levels of pre-existing transmission as the areas likely to experience the largest increase in disease burden (due to high population densities, absence of existing vector control and minimal human population immunity (Hamlet et al., 2021)).

This work has also highlighted the paucity of currently available entomological data from the Horn of Africa region, particularly the absence of longitudinal studies surveying the same site over multiple months and the risk this poses for erroneously concluding *Anopheles stephensi*'s

absence. Indeed, these analyses demonstrate that a lack of a detailed understanding of the vector's dynamics (precluding targeting of entomological surveys to periods of highest vector density) risks missing *Anopheles stephensi*'s presence. Longitudinal surveys elucidating these dynamics would therefore be useful in enabling subsequent refinement and timing of shorter surveys aimed at detecting presence only, even before considering the additional information these surveys would provide on temporal dynamics that can facilitate the effective targeting and timing of control interventions such as IRS. Indeed, in my assessment of IRS impact, I assume implementation teams would have sufficient understanding of the vector's dynamics to time delivery of IRS optimally in relation to any seasonal peaks in vector abundance – something likely not possible for a newly establishing vector. Such surveillance is therefore vital – both for optimising control efforts, but also for enabling better awareness of the establishment during its earliest phases. Unnoticed proliferation of an invasive anopheline vector has previously been observed with *Anopheles arabiensis* in North-Eastern Brazil, where extensive spread prior to detection subsequently led to a significant malaria epidemic across the region (Killeen et al., 2002).

There are a number of important limitations to the work presented here. Firstly, I do not formally include considerations of insecticide resistance in the model of malaria transmission. Insecticide resistance is well-documented for *Anopheles stephensi*, including populations across Afghanistan (Safi et al., 2019), Iran (Vatandoost and Hanafi-Bojd, 2012), Pakistan (Ali Khan, Akram and Lee, 2018) and India (Tiwari et al., 2010). Recent populations assayed in Ethiopia showed resistance to insecticides of all four major insecticide classes (Yared et al., 2020; Balkew et al., 2021), suggesting that pyrethroid-only ITNs and IRS (both already in use across the country) might have limited impact at controlling malaria transmitted by *Anopheles stephensi*. Relatedly, I do not consider uncertainty in *Anopheles stephensi* bionomic properties (such as timing of biting, or whether resting occurs predominantly indoors or outdoors), which will further modulate the impact of interventions such as IRS whose killing is mediated primarily through indoors resting following feeding. Significant variation in *Anopheles stephensi*'s bionomic properties across settings has previously been identified (Massey et al., 2016), including a propensity for crepuscular biting and resting outside of houses compared to African anopheline species (Sinka et al., 2020). Previous work has identified these factors as key drivers of malaria intervention impact (Sherrard-Smith et al., 2019), and more specifically, the impact of control interventions aimed at mitigating the public health threat this vector poses (Hamlet et al., 2021). Whilst the aim of the work here is not to provide specific estimates of intervention impact, and instead highlight how seasonality modulates impact and underscore an understanding of seasonal dynamics as a crucial input to optimising control interventions, these limitations highlight the urgent need for a more detailed characterisation of *Anopheles stephensi* in the settings across the Horn of Africa where it is now present, in order to more precisely

quantify how its bionomic properties and insecticide resistance profile might erode intervention impact.

I also assume that the inferred relationships linking environmental features to temporal dynamics will translate from settings in South Asia and the Middle East to the Horn of Africa. Indeed, these results highlight significant plasticity and variation in *Anopheles stephensi*'s seasonal dynamics depending on the setting, and therefore the extent to which the results will extrapolate to new settings beyond *Anopheles stephensi*'s historical range remains unclear. Relatedly, due to the limited amount of data available and the wide geographical range over which the collated studies were conducted, I cannot rule out possible spatial confounding in shaping the associations inferred – though exploratory analysis of the distribution of locations stratified by rural/urban status and cluster assignment did not reveal obvious patterns of spatial confounding (**Figure 5.16**). One factor that I was unable to consider is the possibility of variation between *Anopheles stephensi* forms ("type", "intermediate" and "mysorensis") in their geographical range and temporal dynamics. Identification of the particular *Anopheles stephensi* form is challenging, often requiring close visual examination (Nagpal et al., 2003) or molecular methods (Chavshin et al., 2014). Availability of this data was limited, and I therefore lack the ability to disaggregate time-series by the specific *Anopheles stephensi* form caught. It therefore remains unclear whether the variation in temporal dynamics observed across the time-series is due to inherently flexible dynamics that are shaped by distinct environmental drivers, or instead maybe arising from different *Anopheles stephensi* forms.

Despite these limitations, this work highlights significant variation in temporal dynamics across surveyed *Anopheles stephensi* populations; variation that is shaped by distinct ecological factors, can differ pronouncedly between urban and rural settings, and which has material consequences for the potential effectiveness of vector control interventions. This work also underscores the need to better understand the vector's dynamics in settings where it has newly established. Indeed, the trajectory of *Anopheles stephensi*'s establishment and subsequent dynamics in Horn of Africa remains deeply unclear and the paucity of published studies from the region underscores the need for studies longitudinally surveying locations where *Anopheles stephensi* has recently arrived. This will be vital to understanding the patterns of seasonal variation the vector displays, a crucial and operationally relevant input for optimising the delivery of malaria control interventions aiming to mitigate the impact of this invasive vector.

Conclusion

With the work presented here, I have extended the framework developed in Chapter 4 and applied it to *Anopheles stephensi* data spanning South Asia and the Middle East in order to inform our understanding of how the seasonal dynamics of this efficient urban malaria vector

are likely to play out given its recent introduction to the Horn of Africa. This work has provided insight into the factors underlying variation and heterogeneity in the species' temporal dynamics, particularly highlighting the potential for discordant, asynchronous dynamics across urban and rural settings. Across this thesis, I have sought to quantify and explore some of the factors (both relating to the human host and the mosquito vector) that drive heterogeneity in malaria transmission and burden across a diversity of different settings. In the following, final chapter of this thesis, I synthesise these different strands of work together and reflect on how these different sources of heterogeneity contribute to the diverse transmission dynamics of malaria observed globally, and what they imply for the most effective control of the disease and its mitigation.

Chapter 6 Discussion

Thesis Aims and Objectives

In this thesis, my aim was to use a combination of statistical and mathematical modelling to further our understanding of some key factors (relating to the parasite, the human host, the mosquito vector and the broader environment) underlying variation in the transmission dynamics and burden of malaria between settings. In particular, the aim was to explore how heterogeneity in these key determinants of transmission might influence the appropriateness and efficacy of different malaria control interventions.

Summary of Findings

In Chapter 2, I carried out a systematic literature review and meta-analysis to explore how the prevalence of submicroscopic *Plasmodium falciparum* malaria infection varies across a diverse range of settings globally, using a Bayesian regression-based approach to understand the factors influencing the size of the submicroscopic reservoir, and the implications this might have for malaria control efforts. The results of this analysis highlighted extensive variation between settings in the size and extent of the submicroscopic reservoir, with the proportion of infections that are submicrosopic highest in low transmission settings. Crucially however, I showed that significant variation in the size of the submicroscopic reservoir exists even across settings characterised by similar current levels of transmission, and that this variation can be in-part explained by historical patterns of control, namely whether transmission has recently declined or has been low for many years. Integrating these results with estimates of infectivity in relation to parasite density suggests the contribution of submicroscopic infections to transmission across different settings is likely to be highly variable and dependent on the historical patterns of control and transmission characterising a setting. Variation in the extent to which the infectious reservoir is detectable by commonly used diagnostics such as microscopy has material implications for the efficacy of interventions only targeting detectable infections compared to those treating the whole community (or specifically targeting submicroscopic infections). My results better quantify the types of settings in which these different approaches are likely to be most relevant and impactful.

In Chapter 3, I continued the thesis' focus on the dynamics and characteristics of malaria infections in the human host and explored the spatial distribution of malaria infections within communities, specifically the degree of clustering within households and the factors affecting the extent of this phenomenon. A number of malaria control interventions specifically target individual households (e.g. reactive case detection) and are routinely implemented across a wide range of settings, but the appropriateness of these approaches (which depend on the

degree of household clustering of infections) remained largely unclear. Applying a Bayesian regression-based approached to data from the Demographic and Health Surveys spanning 23 African countries, I show that malaria infections in children consistently cluster together at the household level, though the degree of clustering is highest in areas of lowest transmission. This variation has material consequences for the appropriateness of malaria control strategies focussed on the household. In the settings of highest transmission where the degree of household clustering is most limited and overall prevalence of infection is higher, reactive household-based approaches are likely to offer limited benefit (in terms of resource saving) over community-based approaches such as mass-drug administration. By contrast, in the settings of lowest transmission where there are overall fewer infections and where the degree of clustering is most substantial, household targeted strategies may offer malaria control programs the ability to target infections in a highly cost-effective way, and also catch infections that may not otherwise be detected (such as asymptomatic infections). Crucially however, this will depend on the extent to which malaria transmission continues to be peri-domestic (i.e. occurring predominantly in and around the household).

Together, Chapters 2 and 3 explored variation in the detectability, prevalence and spatial clustering of malaria infections across communities, and the implications this variation has for effective control of the disease across different settings. In Chapter 4, my work shifted to focus on the anopheline mosquitoes that underpin malaria transmission. The aim with this work was to better understand the factors influencing the population dynamics of these mosquitoes, and their influence on the temporal patterns and seasonality of malaria transmission. I developed a statistical framework enabling characterisation and comparison of the temporal properties of entomological time-series catch data. Applying this framework to data collated from across India, I showed how this framework can identify the dominant temporal patterns present in the data and explore the ecological factors underpinning the empirically observed variation in the degree and extent of seasonality. This work highlights that even within the same location, different mosquito species can display substantially different temporal dynamics. This phenomenon, likely underpinned by species-specific factors and their interaction with the ecological structure of the surrounding environment (such as species-specific preferences for different types of breeding water sources and the availability of those given the setting's hydrological characteristics) will influence the most effective way to controlling the disease. Understanding these dynamics is a crucial input to optimising timing and delivery of vector control interventions such as IRS, and importantly, the efficacy of these interventions will depend intimately on both the timing and degree of seasonality displayed by the mosquito populations responsible for malaria transmission.

In the final Chapter of this thesis, I apply this same framework to a collection of time-series data assembled for the malaria vector *Anopheles stephensi*, a highly efficient urban malaria vector that has recently been introduced in the Horn of Africa and is thought to be driving a resurgence in malaria transmission in parts of the region. The results of my analyses revealed pronounced diversity in the extent and degree of seasonality displayed by different *Anopheles stephensi* populations. Using a random-forest based modelling approach, I showed that this variation is driven by a variety of ecological factors including temperature seasonality and land-use patterns. Importantly, this work revealed clear structuring of temporal dynamics across rural and urban settings, suggesting a key role of the built environment in modulating the vector's temporal dynamics. This finding suggests a potential need for different interventional strategies aimed at controlling the vector in cities compared to other non-urban locations. Extending the framework presented in Chapter 4, I then integrated the collated temporal profiles of *Anopheles stephensi* abundance into a previously published model of malaria transmission to link variation in vector temporal dynamics more directly to their implications for malaria transmission and control. These results highlight that maximising the impact of annually delivered vector control interventions such as IRS is intimately dependent on both the degree of vector seasonality in settings, and the ability to time these interventions appropriately relative to seasonal peaks in abundance.

The results presented here highlight some of the factors that can shape malaria epidemiology and whose variation across settings contributes to the diverse malaria ecologies observed globally. The comparative importance of these factors in shaping intervention impact is variable, however. For example, the impact of interventions such as IRS or SMC depends on their timing relative to seasonal peaks in transmission, and so their impact is modulated substantially by the degree and extent of seasonality in a given setting. By contrast, the overall impact of other vector control interventions such as the delivery and distribution of ITNs is likely less sensitive to timing relative to seasonal peaks in transmission. Similarly, whilst the impact of spatially targeted interventions such as reactive case detection intimately depends on the spatial distribution and clustering of malaria infections, approaches targeting the entire community simultaneously (such as mass drug administration campaigns) are less dependent on the underlying spatial distribution of infections with a community. The results presented here underscore the material implications different malarial ecologies have for the impact of different control interventions, and the need to tailor interventions to context if impact is to be maximised, and the most achieved given limited resources.

Limitations and Future Directions

There are a number of important limitations to the results presented here. Firstly, significant work remains to be done integrating all the different factors explored in this thesis into

transmission models to quantify their influence on the public health impact of interventions more formally. Indeed, it is only in Chapter 5 where I integrate the results of my analyses (in that case relating to vector temporal dynamics) into a mathematical model of transmission that allows the impact of the variation (in vector temporal dynamics) on the burden of malaria and the efficacy of different control strategies (specifically IRS) to be quantified. Mathematical models such as the one utilised in Chapter 5 represent useful platforms for integrating and synthesising a range of different characteristics of settings. Drivers of heterogeneity explored in this thesis are currently considered standalone – exploration of multiple different sources of heterogeneity simultaneously would also likely prove useful, given that real-world settings are characterised by a complex combination of these different factors, which may well interact to further modulate the impact of different control interventions.

Relatedly, future work expanding the range of interventions considered would likely prove instructive. In this thesis, results exploring intervention impact are limited to IRS. There are of course numerous other interventions that can and should be explored using the developed framework. For example, comparisons of treatment-based approaches to control that include both targeted (such as reactive case detection) and untargeted (such as mass-drug administration) strategies would be useful in quantifying just how much variation in the spatial clustering and detectability of malaria infections affect impact and would provide a more direct link from the results presented in Chapters 2 and 3 to variation in public health impact. Additionally, considering new interventions such as vaccines (e.g. RTS,S/ASO1 (RTS,S Clinical Trials Partnership, 2015; RTS,S Clinical Trials Partnership et al., 2011) or R21 (Datoo et al., 2021)), larviciding (Leslie Choi, 2017), attractive toxic sugar baits (ATSBs) (Müller et al., 2010) or the utilisation of genetically modified mosquitoes (Marshall and Taylor, 2009) and how this interacts with some of the factors considered in this thesis would also likely prove instructive. Given these interventions in some cases act at different parts of the malaria transmission cycle compared to the most commonly used interventions (e.g. larviciding and targeting of the immature developmental stages compared to IRS's impact on mature adults mosquitoes), quantifying their impact and how it is modulated by differences in malaria epidemiology across different settings will be vital, though has not been explored here.

Perhaps most crucially, there are a number of different factors not considered here that will similarly affect the appropriateness, viability and efficacy of different control strategies. Notable examples include insecticide resistance (Hancock et al., 2018) and drug resistance (Okell, Griffin and Roper, 2017), which will shape and influence the most appropriate strategy for control of malaria across all levels of transmission and endemicity. Previous work has also identified extensive variation within-species across different sites in vectorial factors such as mosquito bionomic properties (e.g. degree of endophagy, the human blood index etc) (Sherrard-

Smith et al., 2019), all of which have material implications for the success of vector control efforts predicated on indoor feeding and indoor resting, which includes ITNs and IRS, and which are not considered here. Other factors absent from consideration in this thesis include a number which are directly and specifically relevant to settings approaching elimination. During the transition to low endemic levels of transmission and possible elimination, there are often marked shifts in the nature and dynamics of malaria (Cotter et al., 2013) – whilst some have been considered here (such as the increasing proportion of submicroscopic infections or greater extent of spatial clustering), there are numerous changes to the disease's epidemiology that have not been considered here and that likely have important consequences for the optimal control strategy as settings shift from focuses on burden reduction to elimination of malaria. These include increasing fractions of imported malaria cases (Martens and Hall, 2000; Churcher et al., 2014), the emergence of new risk groups as the comparative importance of peri-domestic and occupational exposures to vectors changes (Chuquiyauri et al., 2012), and the persistence of transmission in hard-to-reach groups leading to even finer spatial heterogeneity in the distribution of burden (Bejon et al., 2014). All have material consequences for which interventions are likely to be most impactful, and thus their absence from consideration in this thesis is an important limitation to the results presented here.

Beyond these factors, it is also important to note that the focus of this thesis has been on the transmission dynamics and burden of *Plasmodium falciparum* malaria. For countries that have co-endemic *Plasmodium falciparum* and *Plasmodium vivax* malaria, there have been repeated observations of the fraction of cases attributable to *Plasmodium vivax* increasing as the total burden of malaria decreases (World Health Organization, 2020), and indeed, there is mounting recognition that *Plasmodium vivax* malaria is likely to pose a significant challenge to malaria elimination efforts (Cotter et al., 2013). Whilst many of the factors considered in this thesis will also be relevant to the transmission and control of *Plasmodium vivax* malaria, there are other factors that represent unique challenges to *Plasmodium vivax* malaria transmission and control. Previous work has highlighted extensive variation in the relapse frequency of *Plasmodium vivax* infection (White et al., 2016), itself underpinned by systematic variation between parasite genotypes between temperature and tropical settings as well as intense overdispersion in the number of hypnozoites individuals are host to (White et al., 2014). Recent work has seen the development of mathematical modelling approaches explicitly focussing on the transmission dynamics and burden of disease due to *Plasmodium vivax* (White et al., 2018), providing a framework in which to systematically explore the impacts of these different factors. Furthering our understanding of the factors influencing variation and heterogeneity in *Plasmodium vivax* transmission dynamics across settings, the degree to which these factors overlap or differ to the factors relevant to transmission of *Plasmodium falciparum* malaria, and how this affects control of the disease would therefore represent a likely instructive avenue of future inquiry.

Finally, the work in this thesis has made the assumption that the primary motivating factor for what constitutes the most appropriate intervention(s) to deploy in settings will centre around how to achieve the most impact. The reality is that the viability of different control interventions are shaped by a wealth of additional considerations including cost-effectiveness, technical or operational feasibility, and equity (amongst many others). All of these factors would further shape and constrain the feasibility of different control strategies in different settings. Indeed, whilst the results of this thesis highlight the likely wide variation in public health impact that can be achieved with different interventions and how this varies according to the setting, whether or not that intervention is viable will intimately depend on its cost, and how much impact is achieved given a certain amount of expenditure. Considerations of cost-effectiveness are therefore a vital input to considerations around the viability and practicality of different malaria control strategies; a crucial component that is absent from the analyses presented here. Previous work has utilised this same mathematical modelling framework to explore the cost-effectiveness of different malaria intervention packages (Walker et al., 2016), demonstrating the cost-benefits of carefully tailoring malaria intervention to the ecological context in which malaria transmission is situated and how cost-effectiveness of the same intervention can vary widely depending on the setting. Given the important insight into allocative decisions that incorporating the cost of malaria interventions can provide (Conteh et al., 2021), future work explicitly incorporating considerations of cost-effectiveness are likely to prove instructive.

Conclusions

Though the years since 2000 saw significant reductions globally in the burden of morbidity and mortality attributable to malaria, more recent progress has been limited. Progress stalled and malaria incidence remained constant in 2016 and 2017, before increasing in 2020 and 2021 following the advent of the COVID-19 pandemic. Reversing this trajectory and achieving the goals laid out in the World Health Organization's Global Technical Strategy for Malaria 2016-2030 (World Health Organization, 2015) will be contingent on maximising the impact of limited resources aimed at control of the disease. In turn, maximising impact will require both focussing of efforts on the settings where burden is currently greatest (e.g. the WHO's "High Burden to High Impact" strategy (World Health Organization, 2018b) and efficient tailoring of interventions to match the specific eco-epidemiological context of each setting. In this thesis, I have attempted to in-part address this latter need. Malaria's transmission dynamics and epidemiology is marked by extensive diversity and heterogeneity across settings – heterogeneity that materially influences the impact of control interventions. Across the work presented in this thesis, I have explored both entomological and epidemiological factors that give rise to this diversity, quantified their influence on malaria transmission dynamics, and began to develop a framework to assess the material consequences they have for the efficacy of the different malaria control

tools currently available. In doing so, my work has underscored the crucial importance of considering the setting-specific ecological and epidemiological context of malaria transmission when designing control strategies, and the need for appropriate targeting of efforts if impact is to be maximised.

References

ACCESS-SMC Partnership. (2020). Effectiveness of seasonal malaria chemoprevention at scale in west and central Africa: an observational study. *Lancet*, 396 (10265), pp.1829–1840.

Adams, K. L. et al. (2021). Wolbachia cifB induces cytoplasmic incompatibility in the malaria mosquito vector. *Nature microbiology*, 6 (12), pp.1575–1582.

Ahmed, A. et al. (2021a). Emergence of the invasive malaria vector Anopheles stephensi in Khartoum State, Central Sudan. *Parasites & vectors*, 14 (1), p.511.

Ahmed, A. et al. (2021b). Invasive malaria vector anopheles stephensi mosquitoes in Sudan, 2016-2018. *Emerging infectious diseases*, 27 (11), pp.2952–2954.

Aidoo, E. K. et al. (2018). Reactive case detection of Plasmodium falciparum in western Kenya highlands: effective in identifying additional cases, yet limited effect on transmission. *Malaria journal*, 17 (1), p.111.

Ali Khan, H. A., Akram, W. and Lee, S. (2018). Resistance to Selected Pyrethroid Insecticides in the Malaria Mosquito, Anopheles stephensi (Diptera: Culicidae), From Punjab, Pakistan. *Journal of medical entomology*, 55 (3), pp.735–738.

Alphey, L. S. et al. (2020). Opinion: Standardizing the definition of gene drive. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (49), pp.30864–30867.

Amerasinghe, F., Indrajith, N. and Ariyasena, T. (1995). Physico-chemical characteristics of mosquito breeding habitats in an irrigation development area in Sri Lanka. *Ceylon journal of science: Biological sciences*, 24, pp.13–29.

Andolina, C. et al. (2021). Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: a longitudinal, observational cohort study. *The Lancet infectious diseases*, 21 (11), pp.1568–1578.

Aponte, J. J. et al. (2009). Efficacy and safety of intermittent preventive treatment with sulfadoxine-pyrimethamine for malaria in African infants: a pooled analysis of six randomised, placebo-controlled trials. *Lancet*, 374 (9700), pp.1533–1542.

Appawu, M. et al. (2004). Malaria transmission dynamics at a site in northern Ghana proposed for testing malaria vaccines. *Tropical medicine & international health: TM & IH*, 9 (1), pp.164–170.

Archer, N. M. et al. (2018). Resistance to Plasmodium falciparum in sickle cell trait erythrocytes is driven by oxygen-dependent growth inhibition. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (28), pp.7350–7355.

Ashley, E. A. et al. (2014). Spread of Artemisinin Resistance in Plasmodium falciparum Malaria. *The New England journal of medicine*, 371 (5), pp.411–423.

Ashley, E. A., Recht, J. and White, N. J. (2014). Primaquine: the risks and the benefits. *Malaria journal*, 13, p.418.

Awolola, T. S. et al. (2007). Anopheles gambiae s.s. breeding in polluted water bodies in urban Lagos, southwestern Nigeria. *Journal of vector borne diseases*, 44 (4), pp.241–244.

Baba, E. et al. (2020). Effectiveness of seasonal malaria chemoprevention at scale in west and central Africa: an observational study. *The Lancet*, 396 (10265), pp.1829–1840.

Baird, J. K. et al. (2002). Seasonal malaria attack rates in infants and young children in northern Ghana. *The American journal of tropical medicine and hygiene*, 66 (3), pp.280–286.

Baker, D. A. (2010). Malaria gametocytogenesis. *Molecular and biochemical parasitology*, 172 (2), pp.57–65.

Balkew, M. et al. (2020). Geographical distribution of Anopheles stephensi in eastern Ethiopia. *Parasites & vectors*, 13 (1), p.35.

Balkew, M. et al. (2021). An update on the distribution, bionomics, and insecticide susceptibility of Anopheles stephensi in Ethiopia, 2018-2020. *Malaria journal*, 20 (1), p.263.

Bannister-Tyrrell, M. et al. (2019). Households or Hotspots? Defining Intervention Targets for Malaria Elimination in Ratanakiri Province, Eastern Cambodia. *The Journal of infectious diseases*, 220 (6), pp.1034–1043.

Barik, T. K., Sahu, B. and Swain, V. (2009). A review on Anopheles culicifacies: from bionomics to control with special reference to Indian subcontinent. *Acta tropica*, 109 (2), pp.87–97.

Bashar, K. and Tuno, N. (2014). Seasonal abundance of Anopheles mosquitoes and their association with meteorological factors and malaria incidence in Bangladesh. *Parasites & vectors*, 7, p.442.

Batra, C. P. et al. (2001). Impact of urbanization on bionomics of An. culicifacies and An. stephensi in Delhi. *Indian journal of malariology*, 38 (3–4), pp.61–75.

Battle, K. E. et al. (2014). Geographical variation in Plasmodium vivax relapse. *Malaria journal*, 13 (1), p.144.

Battle, K. E. et al. (2019). Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study. *The Lancet*, 394 (10195), pp.332–343.

Bayoh, M. N. and Lindsay, S. W. (2003). Effect of temperature on the development of the aquatic stages of Anopheles gambiae sensu stricto (Diptera: Culicidae). *Bulletin of entomological research*, 93 (5), pp.375–381.

Beck-Johnson, L. M. et al. (2017). The importance of temperature fluctuations in understanding mosquito population dynamics and malaria risk. *Royal Society open science*, 4 (3), p.160969.

Bejon, P. et al. (2014). A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. *eLife*, 3, p.e02130.

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv [stat.ME]. arXiv* [Online]. Available at: http://arxiv.org/abs/1701.02434.

Bhatt, S. et al. (2015a). Coverage and system efficiencies of insecticide-treated nets in Africa from 2000 to 2017. *eLife*, 4. [Online]. Available at: doi:10.7554/eLife.09672.

Bhatt, S. et al. (2015b). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature*, 526 (7572), pp.207–211.

Biau, G. (2012). Analysis of a random forests model. *Journal of machine learning research: JMLR*, 13 (1), pp.1063–1095.

Björkman, A. and Morris, U. (2020). Why Asymptomatic Plasmodium falciparum Infections Are Common in Low-Transmission Settings. *Trends in parasitology*, 36 (11), pp.898–905.

Bousema, T. et al. (2010). Identification of hot spots of malaria transmission for targeted malaria control. *The journal of infectious diseases*, 201 (11), pp.1764–1774.

Bousema, T. et al. (2012). Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS medicine*, 9 (1), p.e1001165.

Boussari, O. et al. (2012). Use of a mixture statistical model in studying malaria vectors density. *PloS one*, 7 (11), p.e50452.

Brady, O. J. et al. (2017). Role of mass drug administration in elimination of Plasmodium falciparum malaria: a consensus modelling study. *The Lancet. Global health*, 5 (7), pp.e680–e687.

Brasseur, P. et al. (2011). Changing patterns of malaria during 1996-2010 in an area of moderate transmission in southern Senegal. *Malaria journal*, 10, p.203.

Breiman, L. (2001). Random Forests. *Machine learning*, 45 (1), pp.5–32.

Cairns, M. et al. (2012). Estimating the potential public health impact of seasonal malaria chemoprevention in African children. *Nature communications*, 3, p.881.

Carneiro, I. et al. (2010). Age-patterns of malaria vary with severity, transmission intensity and seasonality in sub-Saharan Africa: a systematic review and pooled analysis. *PloS one*, 5 (2), p.e8988.

Carpenter, B. et al. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76 (1), pp.1–32.

Cash, B. A. et al. (2013). Malaria epidemics and the influence of the tropical South Atlantic on the Indian monsoon. *Nature climate change*, 3 (5), pp.502–507.

Ceesay, S. J. et al. (2008). Changes in malaria indices between 1999 and 2007 in The Gambia: a retrospective analysis. *Lancet*, 372 (9649), pp.1545–1554.

Chaccour, C. J. et al. (2018). Targeting cattle for malaria elimination: marked reduction of Anopheles arabiensis survival for over six months using a slow-release ivermectin implant formulation. *Parasites & vectors*, 11 (1). [Online]. Available at: doi:10.1186/s13071-018-2872-y.

Challenger, J. D. et al. (2021). Predicting the public health impact of a malaria transmission-blocking vaccine. *Nature communications*, 12 (1), p.1494.

Chandramohan, D. et al. (2021). Seasonal malaria vaccination with or without seasonal malaria chemoprevention. *The New England journal of medicine*, 385 (11), pp.1005–1017.

Chavshin, A. R. et al. (2014). Molecular characterization, biological forms and sporozoite rate of Anopheles stephensi in southern Iran. *Asian Pacific journal of tropical biomedicine*, 4 (1), pp.47–51.

Chawla, N. V. et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *The journal of artificial intelligence research*, 16, pp.321–357. [Accessed 1 April 2022].

Chuquiyauri, R. et al. (2012). Socio-demographics and the development of malaria elimination strategies in the low transmission setting. *Acta tropica*, 121 (3), pp.292–302.

Churcher, T. S. et al. (2014). Public health. Measuring the path toward malaria elimination. *Science (New York, N.Y.)*, 344 (6189), pp.1230–1232.

Churcher, T. S., Trape, J.-F. and Cohuet, A. (2015). Human-to-mosquito transmission efficiency increases as malaria is controlled. *Nature communications*, 6, p.6054.

Cissé, B. et al. (2016). Effectiveness of seasonal Malaria Chemoprevention in children under ten years of age in Senegal: A stepped-wedge cluster-randomised trial. *PLoS medicine*, 13 (11), p.e1002175.

Clayton, A. M., Dong, Y. and Dimopoulos, G. (2014). The Anopheles innate immune system in the defense against malaria infection. *Journal of innate immunity*, 6 (2), pp.169–181.

Coalson, J. E. et al. (2016). High prevalence of Plasmodium falciparum gametocyte infections in school-age children using molecular detection: patterns and predictors of risk from a cross-sectional study in southern Malawi. *Malaria journal*, 15 (1), p.527.

Cohen, J. M. et al. (2012). Malaria resurgence: a systematic review and assessment of its causes. *Malaria journal*, 11, p.122.

Cohen, J. M. et al. (2013). Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malaria journal*, 12, p.61.

Cohuet, A. et al. (2004). High Malaria Transmission Intensity Due to Anopheles funestus (Diptera: Culicidae) in a Village of Savannah–Forest Transition Area in Cameroon. *Journal of medical entomology*, 41 (5), pp.901–905. [Accessed 16 October 2021].

Coleman, R. E. et al. (2004). Infectivity of asymptomatic Plasmodium-infected human populations to Anopheles dirus mosquitoes in western Thailand. *Journal of medical entomology*, 41 (2), pp.201–208.

Conner, R. O. et al. (2020). Mass testing and treatment for malaria followed by weekly fever screening, testing and treatment in Northern Senegal: feasibility, cost and impact. *Malaria journal*, 19 (1), p.252.

Conteh, L. et al. (2021). Costs and cost-effectiveness of malaria control interventions: A systematic literature review. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 24 (8), pp.1213–1222.

Cooper, D. et al. (2019a). Plasmodium knowlesi Malaria in Sabah, Malaysia, 2015–2017: Ongoing Increase in Incidence Despite Near-elimination of the Human-only Plasmodium Species. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 70 (3), pp.361–367. [Accessed 12 April 2022].

Cooper, L. et al. (2019b). Pareto rules for malaria super-spreaders and super-spreading. *Nature communications*, 10 (1). [Online]. Available at: doi:10.1038/s41467-019-11861-y.

Cotter, C. et al. (2013). The changing epidemiology of malaria elimination: new strategies for new challenges. *The Lancet*, 382 (9895), pp.900–911.

Cottrell, G. et al. (2015). Submicroscopic Plasmodium falciparum Infections Are Associated With Maternal Anemia, Premature Births, and Low Birth Weight. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 60 (10), pp.1481–1488.

Dabira, E. D. et al. (2022). Mass drug administration of ivermectin and dihydroartemisinin-piperaquine against malaria in settings with high coverage of standard control interventions: a cluster-randomised controlled trial in The Gambia. *The Lancet infectious diseases*, 22 (4), pp.519–528.

Danis, K. et al. (2011). Autochthonous Plasmodium vivax malaria in Greece, 2011. *Euro surveillance : bulletin Europeen sur les maladies transmissibles [Euro surveillance : European communicable disease bulletin]*, 16 (42). [Online]. Available at: doi:10.2807/ese.16.42.19993-en.

Das, M. K. et al. (2017). Malaria epidemiology in an area of stable transmission in tribal population of Jharkhand, India. *Malaria journal*, 16 (1), p.181.

Das, N. G. et al. (2011). Diversity and seasonal densities of vector anophelines in relation to forest fringe malaria in district Sonitpur, Assam (India). *Journal of parasitic diseases: official organ of the Indian Society for Parasitology*, 35 (2), pp.123–128.

Dasgupta, S. et al. (2018). *Preferential breeding habitats of sibling species complexes of Anopheles fluviatilis in east-central India*. [Online]. Available at: https://www.dipterajournal.com/pdf/2018/vol5issue1/PartB/5-1-10-753.pdf.

Datoo, M. S. et al. (2021). Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. *Lancet*, 397 (10287), pp.1809–1818.

Day, K. P. et al. (2017). Evidence of strain structure in Plasmodium falciparum var gene repertoires in children from Gabon, West Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (20), pp.E4103–E4111.

De Silva, P. M. and Marshall, J. M. (2012). Factors contributing to urban malaria transmission in sub-saharan Africa: a systematic review. *Journal of tropical medicine*, 2012, p.819563.

Demographic and Health Surveys. (2022). *ICF: The DHS Program. Funded by USAID*. [Online]. Available at: https://dhsprogram.com/ [Accessed 2 May 2022].

Desai, M. et al. (2018). Prevention of malaria in pregnancy. *The Lancet infectious diseases*, 18 (4), pp.e119–e132.

Dev, V. (1996). Anopheles minimus: its bionomics and role in the transmission of malaria in Assam, India. *Bulletin of the World Health Organization*, 74 (1), pp.61–66.

Dev, V. and Manguin, S. (2016). Biology, distribution and control of Anopheles (Cellia) minimus in the context of malaria transmission in northeastern India. *Parasites & vectors*, 9 (1), p.585.

Dev, V. and Sharma, V. P. (2013). The Dominant Mosquito Vectors of Human Malaria in India. In: Manguin, S. (Ed). *Anopheles mosquitoes*. Rijeka: IntechOpen.

Dev, V., Sharma, V. P. and Hojai, D. (2009). Malaria transmission and disease burden in Assam: challenges and opportunities. *Journal of parasitic diseases: official organ of the Indian Society for Parasitology*, 33 (1–2), pp.13–22.

Dong, Y., Simões, M. L. and Dimopoulos, G. (2020). Versatile transgenic multistage effector-gene combinations for Plasmodium falciparum suppression in Anopheles. *Science advances*, 6 (20), p.eaay5898.

Doolan, D. L., Dobaño, C. and Baird, J. K. (2009). Acquired immunity to malaria. *Clinical microbiology reviews*, 22 (1), pp.13–36, Table of Contents.

Doumbe-Belisse, P. et al. (2021). Urban malaria in sub-Saharan Africa: dynamic of the vectorial system and the entomological inoculation rate. *Malaria journal*, 20 (1), p.364.

Dutta, P. et al. (1996). Feeding patterns of Anopheles dirus, the major vector of forest malaria in north east India. *The Southeast Asian journal of tropical medicine and public health*, 27 (2), pp.378–381.

Ebel, E. R. et al. (2021). Common host variation drives malaria parasite fitness in healthy human red cells. *eLife*, 10. [Online]. Available at: doi:10.7554/eLife.69808.

Edi, C. V. et al. (2014). CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito Anopheles gambiae. *PLoS genetics*, 10 (3), p.e1004236.

van Eijk, A. M. et al. (2016). What is the value of reactive case detection in malaria control? A case-study in India and a systematic review. *Malaria journal*, 15, p.67.

Ekawati, L. L. et al. (2020). Defining malaria risks among forest workers in Aceh, Indonesia: a formative assessment. *Malaria journal*, 19 (1), p.441.

European Space Agency (ESA). *ESA CCI Land Cover time-series v2.0.7 (1992 - 2015)*.

Eziefula, A. C. et al. (2014). Single dose primaquine for clearance of Plasmodium falciparum gametocytes in children with uncomplicated malaria in Uganda: a randomised, controlled, double-blind, dose-ranging trial. *The Lancet infectious diseases*, 14 (2), pp.130–139.

Faulde, M. K., Rueda, L. M. and Khaireh, B. A. (2014). First record of the Asian malaria vector Anopheles stephensi and its possible role in the resurgence of malaria in Djibouti, Horn of Africa. *Acta tropica*, 139, pp.39–43.

Feachem, R. G. A. et al. (2019). Malaria eradication within a generation: ambitious, achievable, and necessary. *The Lancet*, 394 (10203), pp.1056–1112.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology: a journal of the Royal Meteorological Society*, 37 (12), pp.4302–4315.

Fillinger, U. et al. (2004). The practical importance of permanent and semipermanent habitats for controlling aquatic stages of Anopheles gambiae sensu lato mosquitoes: operational observations from a rural town in western Kenya. *Tropical medicine & international health: TM & IH*, 9 (12), pp.1274–1289.

Fiorenzano, J. M., Koehler, P. G. and Xue, R.-D. (2017). Attractive toxic sugar bait (ATSB) for control of mosquitoes and its impact on non-target organisms: A review. *International journal of environmental research and public health*, 14 (4), p.398.

Fraser, K. J. et al. (2021). Estimating the potential impact of Attractive Targeted Sugar Baits (ATSBs) as a new vector control tool for Plasmodium falciparum malaria. *Malaria journal*, 20 (1), p.151.

Friedl, M. A. et al. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of environment*, 114 (1), pp.168–182.

Fulcher, B. D., Little, M. A. and Jones, N. S. (2013). Highly comparative time-series analysis: the empirical structure of time series and their methods. *Journal of the Royal Society, Interface / the Royal Society*, 10 (83), p.20130048.

Funk, C. et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2 (1), pp.1–21. [Accessed 16 October 2021].

Galardo, A. K. R. et al. (2009). Seasonal abundance of anopheline mosquitoes and their association with rainfall and malaria along the Matapí River, Amapí, Brazil. *Medical and veterinary entomology*, 23 (4), pp.335–349.

Gatton, M. L., Martin, L. B. and Cheng, Q. (2004). Evolution of resistance to sulfadoxine-pyrimethamine in Plasmodium falciparum. *Antimicrobial agents and chemotherapy*, 48 (6), pp.2116–2123.

Gaughan, A. E. et al. (2013). High resolution population distribution maps for Southeast Asia in 2010 and 2015. *PloS one*, 8 (2), p.e55882.

Gebhardt, M. E. et al. (2022). Understudied Anophelines Contribute to Malaria Transmission in a Low-Transmission Setting in the Choma District, Southern Province, Zambia. *The American journal of tropical medicine and hygiene*. [Online]. Available at: doi:10.4269/ajtmh.21-0989.

Gething, P. W. et al. (2012). A long neglected world malaria map: Plasmodium vivax endemicity in 2010. *PLoS neglected tropical diseases*, 6 (9), p.e1814.

Gimnig, J. E. et al. (2001). Characteristics of larval anopheline (Diptera: Culicidae) habitats in Western Kenya. *Journal of medical entomology*, 38 (2), pp.282–288.

Gleave, K. et al. (2021). Piperonyl butoxide (PBO) combined with pyrethroids in insecticide-treated nets to prevent malaria in Africa. *Cochrane database of systematic reviews* , (5). [Online]. Available at: doi:10.1002/14651858.CD012776.pub3 [Accessed 2 May 2022].

Glennon, E. E. et al. (2020). Syndromic detectability of haemorrhagic fever outbreaks. *bioRxiv*, medRxiv. [Online]. Available at: doi:10.1101/2020.03.28.20019463.

Golassa, L. et al. (2020). The biology of unconventional invasion of Duffy-negative reticulocytes by Plasmodium vivax and its implication in malaria epidemiology and public health. *Malaria journal*, 19 (1), p.299.

Gomes, F. M. et al. (2017). Effect of naturally occurring Wolbachia in Anopheles gambiae s.l. mosquitoes from Mali on Plasmodium falciparum malaria transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (47), pp.12566–12571.

Gonçalves, B. P. et al. (2017). Examining the human infectious reservoir for Plasmodium falciparum malaria in areas of differing transmission intensity. *Nature communications*, 8 (1), p.1133.

Gray, E. M. and Bradley, T. J. (2005). Physiology of desiccation resistance in Anopheles gambiae and Anopheles arabiensis. *The American journal of tropical medicine and hygiene*, 73 (3), pp.553–559.

Griffin, J. T. et al. (2010). Reducing Plasmodium falciparum malaria transmission in Africa: a model-based evaluation of intervention strategies. *PLoS medicine*, 7 (8), p.e1000324.

Griffin, J. T. et al. (2015). Gradual acquisition of immunity to severe malaria with increasing exposure. *Proceedings. Biological sciences*, 282 (1801), p.20142657.

Griffin, J. T. et al. (2016). Potential for reduction of burden and local elimination of malaria by reducing Plasmodium falciparum malaria transmission: a mathematical modelling study. *The Lancet infectious diseases*, 16 (4), pp.465–472.

Griffin, J. T., Ferguson, N. M. and Ghani, A. C. (2014). Estimates of the changing age-burden of Plasmodium falciparum malaria disease in sub-Saharan Africa. *Nature communications*, 5, p.3136.

Grillet, M.-E. et al. (2010). Disentangling the effect of local and global spatial variation on a mosquito-borne infection in a neotropical heterogeneous environment. *The American journal of tropical medicine and hygiene*, 82 (2), pp.194–201.

Guerra, C. A. et al. (2019). Human mobility patterns and malaria importation on Bioko Island. *Nature communications*, 10 (1), p.2332.

Gunasekaran, K. et al. (1994). Reliability of light trap sampling for Anopheles fluviatilis, a vector of malaria. *Acta tropica*, 58 (1), pp.1–11.

Haldar, K., Bhattacharjee, S. and Safeukui, I. (2018). Drug resistance in Plasmodium. *Nature reviews. Microbiology*, 16 (3), pp.156–170.

Hamlet, A. et al. (2021). The potential impact of Anopheles stephensi establishment on the transmission of Plasmodium falciparum in Ethiopia and prospective control measures. *bioRxiv*. [Online]. Available at: doi:10.1101/2021.08.19.21262272.

Hamlet, A. et al. (2022). The potential impact of Anopheles stephensi establishment on the transmission of Plasmodium falciparum in Ethiopia and prospective control measures. *BMC medicine*, 20 (1), p.135.

Hancock, P. A. et al. (2018). Associated patterns of insecticide resistance in field populations of malaria vectors across Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (23), pp.5938–5943.

Hay, S. I. et al. (2010). Developing global maps of the dominant anopheles vectors of human malaria. *PLoS medicine*, 7 (2), p.e1000209.

Hien, D. F. D. S. et al. (2016). Plant-Mediated Effects on Mosquito Capacity to Transmit Human Malaria. *PLoS pathogens*, 12 (8), p.e1005773.

Hoffman, M. D. and Gelman, A. (2011). The no-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv [stat.CO]*. *arXiv* [Online]. Available at: http://arxiv.org/abs/1111.4246.

Hogan, A. B. et al. (2020). Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: a modelling study. *The Lancet. Global health*, 8 (9), pp.e1132–e1141.

Huho, B. et al. (2013). Consistently high estimates for the proportion of human exposure to malaria vector populations occurring indoors in rural Africa. *International journal of epidemiology*, 42 (1), pp.235–247.

Hustedt, J. et al. (2016). Reactive case-detection of malaria in Pailin Province, Western Cambodia: lessons from a year-long evaluation in a pre-elimination setting. *Malaria journal*, 15 (1), p.132.

Idris, Z. M. et al. (2016). High and Heterogeneous Prevalence of Asymptomatic and Sub-microscopic Malaria Infections on Islands in Lake Victoria, Kenya. *Scientific reports*, 6, p.36958.

Johnson, L. R. et al. (2015a). Mapping the distribution of malaria: Current approaches and future directions. In: *Wiley Series in Probability and Statistics*. Hoboken, NJ, USA: John Wiley & Sons, Inc. pp.189–209.

Johnson, L. R. et al. (2015b). Understanding uncertainty in temperature effects on vector-borne disease: a Bayesian approach. *Ecology*, 96 (1), pp.203–213.

Joshi, H. et al. (1988). Host feeding patterns of Anopheles culicifacies species A and B. *Journal of the American Mosquito Control Association*, 4 (3), pp.248–251.

Jude, P. J. et al. (2010). Anopheles culicifacies breeding in brackish waters in Sri Lanka and implications for malaria control. *Malaria journal*, 9, p.106.

Justice, C. O. et al. (2002). An overview of MODIS Land data processing and product status. *Remote sensing of environment*, 83 (1–2), pp.3–15.

Kagoro, F. M. et al. (2022). Mapping genetic markers of artemisinin resistance in Plasmodium falciparum malaria in Asia: a systematic review and spatiotemporal analysis. *The Lancet. Microbe*, 3 (3), pp.e184–e192.

Kar, N. P. et al. (2014). A review of malaria transmission dynamics in forest ecosystems. *Parasites & vectors*, 7, p.265.

Kasili, S. et al. (2009). Entomological assessment of the potential for malaria transmission in Kibera slum of Nairobi, Kenya. *Journal of vector borne diseases*, 46 (4), pp.273–279.

Katrak, S. et al. (2018). Clinical consequences of submicroscopic malaria parasitaemia in Uganda. *Malaria journal*, 17 (1), p.67.

Kawada, H. et al. (2011). Multimodal pyrethroid resistance in malaria vectors, Anopheles gambiae s.s., Anopheles arabiensis, and Anopheles funestus s.s. in western Kenya. *PloS one*, 6 (8), p.e22574.

Killeen, G. (2014). Characterizing, controlling and eliminating residual malaria transmission. *Malaria journal*, 13 (S1), p.P53.

Killeen, G. F. et al. (2002). Eradication of Anopheles gambiae from Brazil: lessons for malaria control in Africa? *The Lancet infectious diseases*, 2 (10), pp.618–627.

Killeen, G. F. et al. (2007). Preventing childhood malaria in Africa by protecting adults from mosquitoes with insecticide-treated nets. *PLoS medicine*, 4 (7), p.e229.

Killeen, G. F. et al. (2019). Suppression of malaria vector densities and human infection prevalence associated with scale-up of mosquito-proofed housing in Dar es Salaam, Tanzania: re-analysis of an observational series of parasitological and entomological surveys. *The Lancet. Planetary health*, 3 (3), pp.e132–e143.

Kim, S. et al. (2021). A systematic review of the evidence on the effectiveness and cost-effectiveness of mass screen-and-treat interventions for malaria control. *The American journal of tropical medicine and hygiene*, 105 (6), pp.1722–1731.

Kirby, M. J. and Lindsay, S. W. (2009). Effect of temperature and inter-specific competition on the development and survival of Anopheles gambiae sensu stricto and An. arabiensis larvae. *Acta tropica*, 109 (2), pp.118–123.

Kiwanuka, G. N. (2009). Genetic diversity in Plasmodium falciparum merozoite surface protein 1 and 2 coding genes and its implications in malaria epidemiology: a review of published studies from 1997-2007. *Journal of vector borne diseases*, 46 (1), p.1.

Klinkenberg, E. et al. (2008). Impact of urban agriculture on malaria vectors in Accra, Ghana. *Malaria journal*, 7, p.151.

Koenraadt, C. J. M., Githeko, A. K. and Takken, W. (2004). The effects of rainfall and evapotranspiration on the temporal dynamics of Anopheles gambiae s.s. and Anopheles arabiensis in a Kenyan village. *Acta tropica*, 90 (2), pp.141–153.

Koepfli, C. et al. (2017). Sustained malaria control over an 8-year period in Papua New Guinea: The challenge of low-density asymptomatic Plasmodium infections. *The journal of infectious diseases*, 216 (11), pp.1434–1443.

Korgaonkar, N. S. et al. (2012). Mosquito biting activity on humans & detection of Plasmodium falciparum infection in Anopheles stephensi in Goa, India. *The Indian journal of medical research*, 135, pp.120–126.

Kuhn, M. and Wickham, H. (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. *Boston, MA, USA. [(accessed on 10 December 2020)]*.

Kumar, A. et al. (2016). Anopheles subpictus carry human malaria parasites in an urban area of Western India and may facilitate perennial malaria transmission. *Malaria journal*, 15, p.124.

Kumar, A. and Thavaselvam, D. (1992). Breeding habitats and their contribution to Anopheles stephensi in Panaji. *Indian journal of malariology*, 29 (1), pp.35–40.

Kwambai, T. K. et al. (2020). Malaria Chemoprevention in the Postdischarge Management of Severe Anemia. *The New England journal of medicine*, 383 (23), pp.2242–2254.

Kwiatkowski, D. P. (2005). How malaria has affected the human genome and what human genetics can teach us about malaria. *The American Journal of Human Genetics*, 77 (2), pp.171–192.

Lalloo, D. G., Olukoya, P. and Olliaro, P. (2006). Malaria in adolescence: burden of disease, consequences, and opportunities for intervention. *The Lancet infectious diseases*, 6 (12), pp.780–793.

Lamikanra, A. A. et al. (2007). Malarial anemia: of mice and men. *Blood*, 110 (1), pp.18–28.

Lamptey, H. et al. (2018). The prevalence of submicroscopic Plasmodium falciparum gametocyte carriage and multiplicity of infection in children, pregnant women and adults in a low malaria transmission area in Southern Ghana. *Malaria journal*, 17 (1), p.331.

Langhorne, J. et al. (2008). Immunity to malaria: more questions than answers. *Nature immunology*, 9 (7), pp.725–732.

Le Menach, A. et al. (2011). Travel risk, malaria importation and malaria transmission in Zanzibar. *Scientific reports*, 1, p.93.

Lehner, B. and Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, 296 (1), pp.1–22.

Leslie Choi, A. W. (2017). Larviciding to control malaria. *Cochrane database of systematic reviews* , 2017 (7). [Online]. Available at: doi:10.1002/14651858.CD012736 [Accessed 12 April 2022].

Lin, C.-H. et al. (2018). Location, seasonal, and functional characteristics of water holding containers with juvenile and pupal Aedes aegypti in Southern Taiwan: A cross-sectional study using hurdle model analyses. *PLoS neglected tropical diseases*, 12 (10), p.e0006882.

Lin, J. T. et al. (2016). Microscopic Plasmodium falciparum Gametocytemia and Infectivity to Mosquitoes in Cambodia. *The Journal of infectious diseases*, 213 (9), pp.1491–1494.

Lin, J. T., Saunders, D. L. and Meshnick, S. R. (2014). The role of submicroscopic parasitemia in malaria transmission: what is the evidence? *Trends in parasitology*, 30 (4), pp.183–190.

Lin Ouédraogo, A. et al. (2015). Dynamics of the Human Infectious Reservoir for Malaria Determined by Mosquito Feeding Assays and Ultrasensitive Malaria Diagnosis in Burkina Faso. *The Journal of infectious diseases*, 213 (1), pp.90–99. [Accessed 15 October 2021].

Linard, C. et al. (2012). Population distribution, settlement patterns and accessibility across Africa in 2010. *PloS one*, 7 (2), p.e31743.

Lyons, C. L., Coetzee, M. and Chown, S. L. (2013). Stable and fluctuating temperature effects on the development rate and survival of two malaria vectors, Anopheles arabiensis and Anopheles funestus. *Parasites & vectors*, 6 (1), p.104.

Macdonald, G. (1956). Epidemiological basis of malaria control. *Bulletin of the World Health Organization*, 15 (3–5), pp.613–626.

Majambere, S. et al. (2008). Spatial distribution of mosquito larvae and the potential for targeted larval control in The Gambia. *The American journal of tropical medicine and hygiene*, 79 (1), pp.19–27.

Marshall, J. M. et al. (2013). Quantifying the mosquito's sweet tooth: modelling the effectiveness of attractive toxic sugar baits (ATSB) for malaria vector control. *Malaria journal*, 12, p.291.

Marshall, J. M. and Taylor, C. E. (2009). Malaria control with transgenic mosquitoes. *PLoS medicine*, 6 (2), p.e20.

Martens, P. and Hall, L. (2000). Malaria on the move: human population movement and malaria transmission. *Emerging infectious diseases*, 6 (2), pp.103–109.

Massey, N. C. et al. (2016). A global bionomic database for the dominant vectors of human malaria. *Scientific data*, 3 (1), p.160014.

Matowo, N. S. et al. (2021). An increasing role of pyrethroid-resistant Anopheles funestus in malaria transmission in the Lake Zone, Tanzania. *Scientific reports*, 11 (1), p.13457.

Mattah, P. A. D. et al. (2017). Diversity in breeding sites and distribution of Anopheles mosquitoes in selected urban areas of southern Ghana. *Parasites & vectors*, 10 (1), p.25.

Meis, J. F. et al. (1986). Fine structure of the malaria parasite Plasmodium falciparum in human hepatocytes in vitro. *Cell and tissue research*, 244 (2), pp.345–350.

Mendis, C. et al. (2000). Anopheles arabiensis and An. funestus are equally important vectors of malaria in Matola coastal suburb of Maputo, southern Mozambique. *Medical and veterinary entomology*, 14 (2), pp.171–180.

Menéndez, C. et al. (2010). Malaria prevention with IPTp during pregnancy reduces neonatal mortality. *PloS one*, 5 (2), p.e9438.

Miazgowicz, K. L. et al. (2020). Age influences the thermal suitability of Plasmodium falciparum transmission in the Asian malaria vector Anopheles stephensi. *Proceedings. Biological sciences*, 287 (1931), p.20201093.

Millar, J., Toh, K. B. and Valle, D. (2020). To screen or not to screen: an interactive framework for comparing costs of mass malaria treatment interventions. *BMC medicine*, 18 (1), p.149.

Minakawa, N. et al. (2008). Recent reduction in the water level of Lake Victoria has created more habitats for Anopheles funestus. *Malaria journal*, 7 (1), p.119.

Minakawa, N. et al. (2012). Malaria vectors in Lake Victoria and adjacent habitats in western Kenya. *PloS one*, 7 (3), p.e32725.

Mitri, C. et al. (2009). Density-dependent impact of the human malaria parasite Plasmodium falciparum gametocyte sex ratio on mosquito infection rates. *Proceedings. Biological sciences / The Royal Society*, 276 (1673), pp.3721–3726.

Mlacha, Y. P. et al. (2017). Fine scale mapping of malaria infection clusters by using routinely collected health facility data in urban Dar es Salaam, Tanzania. *Geospatial health*, 12 (1), p.494.

Mockenhaupt, F. P. et al. (2004). Manifestation and outcome of severe malaria in children in northern Ghana. *The American journal of tropical medicine and hygiene*, 71 (2), pp.167–172.

Mogeni, P. et al. (2017). Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. *BMC medicine*, 15 (1), p.121.

Molnar, C. (2020). *Interpretable Machine Learning*. Lulu.com.

Monnahan, C. C., Thorson, J. T. and Branch, T. A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in ecology and evolution*, 8 (3), pp.339–348.

Mordecai, E. A. et al. (2013). Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology letters*, 16 (1), pp.22–30.

Mordecai, E. A. et al. (2019). Thermal biology of mosquito-borne disease. *Ecology letters*, 22 (10), pp.1690–1708.

Mosha, J. F. et al. (2022). Effectiveness and cost-effectiveness against malaria of three types of dual-active-ingredient long-lasting insecticidal nets (LLINs) compared with pyrethroid-only LLINs in Tanzania: a four-arm, cluster-randomised trial. *Lancet*, 399 (10331), pp.1227–1241.

Mourou, J.-R. et al. (2012). Malaria transmission in Libreville: results of a one year survey. *Malaria journal*, 11, p.40.

Mousa, A. et al. (2020). The impact of delayed treatment of uncomplicated P. falciparum malaria on progression to severe malaria: A systematic review and a pooled multicentre individual-patient meta-analysis. *PLoS medicine*, 17 (10), p.e1003359.

Moyes, C. L. et al. (2014). Defining the geographical range of the Plasmodium knowlesi reservoir. *PLoS neglected tropical diseases*, 8 (3), p.e2780.

Moyes, C. L. et al. (2020). Evaluating insecticide resistance across African districts to aid malaria control decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (36), pp.22042–22050.

Mueller, I. et al. (2009). High sensitivity detection of Plasmodium species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea. *Malaria journal*, 8 (1), p.41.

Müller, G. C. et al. (2010). Field experiments of Anopheles gambiae attraction to local fruits/seedpods and flowering plants in Mali to optimize strategies for malaria vector control in Africa using attractive toxic sugar bait methods. *Malaria journal*, 9 (1), p.262.

Murray, C. J. L. et al. (2012). Global malaria mortality between 1980 and 2010: a systematic analysis. *The Lancet*, 379 (9814), pp.413–431.

Murray, L. et al. (2017). Multiplication rate variation in the human malaria parasite Plasmodium falciparum. *Scientific reports*, 7 (1), p.6436.

Musiime, A. K. et al. (2019). Impact of vector control interventions on malaria transmission intensity, outdoor vector biting rates and Anopheles mosquito species composition in Tororo, Uganda. *Malaria journal*, 18 (1), p.445.

Mwesigwa, J. et al. (2019). Mass Drug Administration With Dihydroartemisinin-piperaquine and Malaria Transmission Dynamics in The Gambia: A Prospective Cohort Study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 69 (2), pp.278–286.

Nagpal, B. N. et al. (2003). Spiracular indices in Anopheles stephensi: a taxonomic tool to identify ecological variants. *Journal of medical entomology*, 40 (6), pp.747–749.

Nájera, J. A., González-Silva, M. and Alonso, P. L. (2011). Some lessons for the future from the Global Malaria Eradication Programme (1955-1969). *PLoS medicine*, 8 (1), p.e1000412.

Nanda, N. et al. (2000). Studies on Anopheles fluviatilis and Anopheles culicifacies sibling species in relation to malaria in forested hilly and deforested riverine ecosystems in northern Orissa, India. *Journal of the American Mosquito Control Association*, 16 (3), pp.199–205.

Nanda, N. et al. (2012). Prevalence and incrimination of Anopheles fluviatilis species S (Diptera: Culicidae) in a malaria endemic forest area of Chhattisgarh state, central India. *Parasites & vectors*, 5, p.215.

Neal, R. (2011). MCMC using Hamiltonian dynamics. In: *Chapman & Hall/CRC Handbooks of Modern Statistical Methods*. Chapman and Hall/CRC.

Nguyen, M. et al. (2020). Mapping malaria seasonality in Madagascar using health facility data. *BMC medicine*, 18 (1), p.26.

Njuguna, P. et al. (2019). Observational study: 27 years of severe malaria surveillance in Kilifi, Kenya. *BMC medicine*, 17 (1), p.124.

Obsomer, V., Defourny, P. and Coosemans, M. (2007). The Anopheles dirus complex: spatial distribution and environmental drivers. *Malaria journal*, 6, p.26.

Odongo-Aginya, E. et al. (2005). Relationship between malaria infection intensity and rainfall pattern in Entebbe peninsula, Uganda. *African health sciences*, 5 (3), pp.238–245.

Okell, L. C. et al. (2009). Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis. *The Journal of infectious diseases*, 200 (10), pp.1509–1517.

Okell, L. C. et al. (2012). Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nature communications*, 3, p.1237.

Okell, L. C., Griffin, J. T. and Roper, C. (2017). Mapping sulphadoxine-pyrimethamine-resistant Plasmodium falciparum malaria in infected humans and in parasite populations in Africa. *Scientific reports*, 7 (1). [Online]. Available at: doi:10.1038/s41598-017-06708-9.

Okello, P. E. et al. (2006). Variation in malaria transmission intensity in seven sites throughout Uganda. *The American journal of tropical medicine and hygiene*, 75 (2), pp.219–225.

Olotu, A. et al. (2016). Seven-year efficacy of RTS,S/AS01 malaria vaccine among young African children. *The New England journal of medicine*, 374 (26), pp.2519–2529.

O'Meara, W. P. et al. (2008). Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet*, 372 (9649), pp.1555–1562.

Ouédraogo, A. L. et al. (2009). Substantial contribution of submicroscopical Plasmodium falciparum gametocyte carriage to the infectious reservoir in an area of seasonal transmission. *PloS one*, 4 (12), p.e8410.

Parker, D. M. et al. (2017). Scale up of a Plasmodium falciparum elimination program and surveillance system in Kayin State, Myanmar. *Wellcome open research*, 2 (98), p.98. [Accessed 2 May 2022].

Pascual, M. et al. (2008). Shifting patterns: malaria dynamics and rainfall variability in an African highland. *Proceedings. Biological sciences / The Royal Society*, 275 (1631), pp.123–132.

Paton, R. S. et al. (2021). Malaria infection and severe disease risks in Africa. *Science*, 373 (6557), pp.926–931.

Paul, R. E. et al. (2000). Sex determination in malaria parasites. *Science*, 287 (5450), pp.128–131.

Phillips, M. A. et al. (2017). Malaria. *Nature reviews. Disease primers*, 3, p.17050.

Pluess, B. et al. (2010). Indoor residual spraying for preventing malaria. *Cochrane database of systematic reviews* , (4), p.CD006657.

Plummer, M. and Others. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*. 124. 2003. Vienna, Austria. pp.1–10.

Pollard, E. J. M. et al. (2020). Protecting the peri-domestic environment: the challenge for eliminating residual malaria. *Scientific reports*, 10 (1), p.7018.

Prakash, A. et al. (1997). Seasonal prevalence of Anopheles dirus and malaria transmission in a forest fringed village of Assam, India. *Indian journal of malariology*, 34 (3), pp.117–125.

Prakash, A. et al. (2001). Estimation of vectorial capacity of Anopheles dirus (Diptera: Culicidae) in a forest-fringed village of Assam (India). *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 1 (3), pp.231–237.

Price, R. N. et al. (2020). Plasmodium vivax in the Era of the Shrinking P. falciparum Map. *Trends in parasitology*, 36 (6), pp.560–570.

Protopopoff, N. et al. (2018). Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal net and indoor residual spray interventions, separately and together, against malaria transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two factorial design trial. *Lancet*, 391 (10130), pp.1577–1588.

Raman, J. et al. (2020). High levels of imported asymptomatic malaria but limited local transmission in KwaZulu-Natal, a South African malaria-endemic province nearing malaria elimination. *Malaria journal*, 19 (1), p.152.

Ramos Júnior, W. M. et al. (2010). Clinical aspects of hemolysis in patients with P. vivax malaria treated with primaquine, in the Brazilian Amazon. *The Brazilian journal of infectious diseases: an official publication of the Brazilian Society of Infectious Diseases*, 14 (4), pp.410–412.

Ranson, H. and Lissenden, N. (2016). Insecticide resistance in African anopheles mosquitoes: A worsening situation that needs urgent action to maintain malaria control. *Trends in parasitology*, 32 (3), pp.187–196.

Rasmussen, C. E. (2004). Gaussian processes in machine learning. In: *Advanced Lectures on Machine Learning*. Lecture notes in computer science. Berlin, Heidelberg: Springer Berlin Heidelberg. pp.63–71.

Reyburn, H. et al. (2005). Association of transmission intensity and age with clinical manifestations and case fatality of severe Plasmodium falciparum malaria. *JAMA: the journal of the American Medical Association*, 293 (12), pp.1461–1470.

Robert, V. et al. (2003). Malaria transmission in urban sub-Saharan Africa. *The American journal of tropical medicine and hygiene*, 68 (2), pp.169–176.

Roberts, L. and Enserink, M. (2007). Malaria. Did they really say ... eradication? *Science*, 318 (5856), pp.1544–1545.

Robinson, L. J. et al. (2015). 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 medicine*, 12 (10), p.e1001891.

Rodriguez-Barraquer, I. et al. (2016). Quantifying Heterogeneous Malaria Exposure and Clinical Protection in a Cohort of Ugandan Children. *The Journal of infectious diseases*, 214 (7), pp.1072–1080.

Rodriguez-Barraquer, I. et al. (2018). Quantification of anti-parasite and anti-disease immunity to malaria as a function of age and exposure. *eLife*, 7. [Online]. Available at: doi:10.7554/eLife.35832.

Romeo-Aznar, V. et al. (2018). Mosquito-borne transmission in urban landscapes: the missing link between vector abundance and human density. *Proceedings. Biological sciences / The Royal Society*, 285 (1884). [Online]. Available at: doi:10.1098/rspb.2018.0826.

Roser, M. and Ritchie, H. (2019). Malaria. *Our World in Data*. [Online]. Available at: https://ourworldindata.org/malaria [Accessed 2 May 2022].

Ross, A. et al. (2011). Determinants of the cost-effectiveness of intermittent preventive treatment for malaria in infants and children. *PloS one*, 6 (4), p.e18391.

Rossi, G. et al. (2018a). Adapting Reactive Case Detection Strategies for falciparum Malaria in a Low-Transmission Area in Cambodia. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 66 (2), pp.296–298.

Rossi, G. et al. (2018b). Closing in on the Reservoir: Proactive Case Detection in High-Risk Groups as a Strategy to Detect Plasmodium falciparum Asymptomatic Carriers in Cambodia. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 66 (10), pp.1610–1617.

Roth, G. A. et al. (2018). Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392 (10159), pp.1736–1788.

RTS,S Clinical Trials Partnership et al. (2011). First results of phase 3 trial of RTS,S/AS01 malaria vaccine in African children. *The New England journal of medicine*, 365 (20), pp.1863–1875.

RTS,S Clinical Trials Partnership. (2015). Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. *Lancet*, 386 (9988), pp.31–45.

Russell, T. L. et al. (2011). Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malaria journal*, 10 (1), p.80.

Ruwende, C. and Hill, A. (1998). Glucose-6-phosphate dehydrogenase deficiency and malaria. *Journal of molecular medicine*, 76 (8), pp.581–588.

Ryan, S. J. et al. (2015). Mapping physiological suitability limits for malaria in Africa under climate change. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 15 (12), pp.718–725.

Saavedra, M. P. et al. (2019). Higher risk of malaria transmission outdoors than indoors by Nyssorhynchus darlingi in riverine communities in the Peruvian Amazon. *Parasites & vectors*, 12 (1), p.374.

Safi, N. H. Z. et al. (2019). Status of insecticide resistance and its biochemical and molecular mechanisms in Anopheles stephensi (Diptera: Culicidae) from Afghanistan. *Malaria journal*, 18 (1), p.249.

Sahu, S. S. et al. (2017). Bionomics of Anopheles fluviatilis and Anopheles culicifacies (Diptera: Culicidae) in Relation to Malaria Transmission in East-Central India. *Journal of medical entomology*, 54 (4), pp.821–830.

Sandfort, M. et al. (2020). Forest malaria in Cambodia: the occupational and spatial clustering of Plasmodium vivax and Plasmodium falciparum infection risk in a cross-sectional survey in Mondulkiri province, Cambodia. *Malaria journal*, 19 (1), p.413.

Sang, R. et al. (2016). Effects of irrigation and rainfall on the population dynamics of rift valley fever and other arbovirus mosquito vectors in the epidemic-prone Tana river county, Kenya. *Journal of medical entomology*, p.tjw206.

Schneider, P. et al. (2007). Submicroscopic Plasmodium falciparum gametocyte densities frequently result in mosquito infection. *The American journal of tropical medicine and hygiene*, 76 (3), pp.470–474.

von Seidlein, L. et al. (2019a). Novel approaches to control malaria in forested areas of southeast Asia. *Trends in parasitology*, 35 (6), pp.388–398.

von Seidlein, L. et al. (2019b). The impact of targeted malaria elimination with mass drug administrations on falciparum malaria in Southeast Asia: A cluster randomised trial. *PLoS medicine*, 16 (2), p.e1002745.

Selvaraj, P., Wenger, E. A. and Gerardin, J. (2018). Seasonality and heterogeneity of malaria transmission determine success of interventions in high-endemic settings: a modeling study. *BMC infectious diseases*, 18 (1), p.413.

Seydel, K. B. et al. (2015). Brain swelling and death in children with cerebral malaria. *The New England journal of medicine*, 372 (12), pp.1126–1137.

Shapiro, L. L. M., Whitehead, S. A. and Thomas, M. B. (2017). Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria. *PLoS biology*, 15 (10), p.e2003489.

Sharp, S. J. and Thompson, S. G. (2000). Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in medicine*, 19 (23), pp.3251–3274.

Sherrard-Smith, E. et al. (2018). Systematic review of indoor residual spray efficacy and effectiveness against Plasmodium falciparum in Africa. *Nature communications*, 9 (1), p.4982.

Sherrard-Smith, E. et al. (2019). Mosquito feeding behavior and how it influences residual malaria transmission across Africa. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (30), pp.15086–15095.

Sherrard-Smith, E. et al. (2020). The potential public health consequences of COVID-19 on malaria in Africa. *Nature medicine*, 26 (9), pp.1411–1416.

Singh, B. et al. (2004). A large focus of naturally acquired Plasmodium knowlesi infections in human beings. *Lancet*, 363 (9414), pp.1017–1024.

Singh, N. et al. (1999). Population dynamics of Anopheles culicifacies and malaria in the tribal area of central India. *Journal of the American Mosquito Control Association*, 15 (3), pp.283–290.

Singh, R. K. et al. (2013). Bionomics and vectorial capacity of Anopheles annularis with special reference to India: a review. *The Journal of communicable diseases*, 45 (1–2), pp.1–16.

Singh, R. K. et al. (2014). Bionomics and vector potential of Anopheles subpictus as a malaria vector in India: An overview. *Int J Mosq Res*, 1 (1), pp.29–37.

Sinka, M. E. et al. (2010a). The dominant Anopheles vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasites & vectors*, 3, p.117.

Sinka, M. E. et al. (2010b). The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasites & vectors*, 3, p.72.

Sinka, M. E. et al. (2011a). Erratum to: The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasites & vectors*, 4 (1), p.210.

Sinka, M. E. et al. (2011b). The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasites & vectors*, 4, p.89.

Sinka, M. E. et al. (2012). A global map of dominant malaria vectors. *Parasites & vectors*, 5, p.69.

Sinka, M. E. (2013). Global Distribution of the Dominant Vector Species of Malaria. In: Manguin, S. (Ed). *Anopheles mosquitoes*. Rijeka: IntechOpen.

Sinka, M. E. et al. (2020). A new malaria vector in Africa: Predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk. *Proceedings of the National Academy of Sciences of the United States of America*, 117 (40), pp.24900–24908.

Slater, H. C. et al. (2019). The temporal dynamics and infectiousness of subpatent Plasmodium falciparum infections in relation to parasite density. *Nature communications*, 10 (1), p.1433.

Slater, H. C. et al. (2020). Ivermectin as a novel complementary malaria control tool to reduce incidence and prevalence: a modelling study. *The Lancet infectious diseases*, 20 (4), pp.498–508.

Smit, M. R. et al. (2018). Safety and mosquitocidal efficacy of high-dose ivermectin when co-administered with dihydroartemisinin-piperaquine in Kenyan adults with uncomplicated malaria (IVERMAL): a randomised, double-blind, placebo-controlled trial. *The Lancet infectious diseases*, 18 (6), pp.615–626.

Smit, M. R. et al. (2019). Human direct skin feeding versus membrane feeding to assess the mosquitocidal efficacy of high-dose ivermectin (IVERMAL trial). *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 69 (7), pp.1112–1119.

Smith, D. L. et al. (2011). Infectious disease. Solving the Sisyphean problem of malaria in Zanzibar. *Science (New York, N.Y.)*, 332 (6036), pp.1384–1385.

Smith, D. L. and McKenzie, F. E. (2004). Statics and dynamics of malaria infection in Anopheles mosquitoes. *Malaria journal*, 3, p.13.

Snounou, G. et al. (1993). The importance of sensitive detection of malaria parasites in the human and insect hosts in epidemiological studies, as shown by the analysis of field samples from Guinea Bissau. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87 (6), pp.649–653.

Sougoufara, S., Ottih, E. C. and Tripet, F. (2020). The need for new vector control approaches targeting outdoor biting Anopheline malaria vector communities. *Parasites & vectors*, 13 (1), p.295.

Spiegelhalter, D. J. et al. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 76 (3), pp.485–493.

Stan Development Team. (2022). *RStan: the R interface to Stan*. [Online]. Available at: https://mc-stan.org/.

Steenkeste, N. et al. (2010). Sub-microscopic malaria cases and mixed malaria infection in a remote area of high malaria endemicity in Rattanakiri province, Cambodia: implication for malaria elimination. *Malaria journal*, 9, p.108.

Stone, W. et al. (2022). Pyronaridine-artesunate or dihydroartemisinin-piperaquine combined with single low-dose primaquine to prevent Plasmodium falciparum malaria transmission in Ouélessébougou, Mali: a four-arm, single-blind, phase 2/3, randomised trial. *The Lancet. Microbe*, 3 (1), pp.e41–e51.

van de Straat, B. et al. (2021). A global assessment of surveillance methods for dominant malaria vectors. *Scientific reports*, 11 (1), p.15337.

Stresman, G. et al. (2020). Quantifying Plasmodium falciparum infections clustering within households to inform household-based intervention strategies for malaria control programs: An observational study and meta-analysis from 41 malaria-endemic countries. *PLoS medicine*, 17 (10), p.e1003370.

Stresman, G., Bousema, T. and Cook, J. (2019). Malaria Hotspots: Is There Epidemiological Evidence for Fine-Scale Spatial Targeting of Interventions? *Trends in parasitology*, 35 (10), pp.822–834.

Stresman, G. H. et al. (2018). Do hotspots fuel malaria transmission: a village-scale spatio-temporal analysis of a 2-year cohort study in The Gambia. *BMC medicine*, 16 (1), p.160.

Sturrock, H. J. W. et al. (2013a). Reactive case detection for malaria elimination: real-life experience from an ongoing program in Swaziland. *PloS one*, 8 (5), p.e63830.

Sturrock, H. J. W. et al. (2013b). Targeting asymptomatic malaria infections: active surveillance in control and elimination. *PLoS medicine*, 10 (6), p.e1001467.

Subbarao, S. K. et al. (1987). Egg-float ridge number in Anopheles stephensi: ecological variation and genetic analysis. *Medical and veterinary entomology*, 1 (3), pp.265–271.

Surendran, S. N. and Ramasamy, R. (2005). Some characteristics of the larval breeding sites of Anopheles culicifacies species B and E in Sri Lanka. *Journal of vector borne diseases*, 42 (2), pp.39–44.

Tadesse, F. G. et al. (2017). The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. *Malaria journal*, 16 (1). [Online]. Available at: doi:10.1186/s12936-017-1749-4.

Tadesse, F. G. et al. (2018). The relative contribution of symptomatic and asymptomatic Plasmodium vivax and Plasmodium falciparum infections to the infectious reservoir in a low-endemic setting in Ethiopia. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 66 (12), pp.1883–1891.

Tadesse, F. G. et al. (2021). Anopheles stephensi Mosquitoes as Vectors of Plasmodium vivax and falciparum, Horn of Africa, 2019. *Emerging infectious diseases*, 27 (2), pp.603–607.

Tangena, J.-A. A. et al. (2020). Indoor residual spraying for malaria control in sub-Saharan Africa 1997 to 2017: an adjusted retrospective analysis. *Malaria journal*, 19 (1), p.150.

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

Tessema, S. et al. (2019). Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. *eLife*, 8. [Online]. Available at: doi:10.7554/eLife.43510.

Thang, N. D. et al. (2009). Long-Lasting Insecticidal Hammocks for controlling forest malaria: a community-based trial in a rural area of central Vietnam. *PloS one*, 4 (10), p.e7369.

The Lancet - Editorial. (2007). Is malaria eradication possible? *The Lancet*, 370 (9597), p.1459.

Thomas, S. et al. (2016). Overhead tank is the potential breeding habitat of Anopheles stephensi in an urban transmission setting of Chennai, India. *Malaria journal*, 15 (1), p.274.

Timmann, C. et al. (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature*, 489 (7416), pp.443–446.

Tiwari, S. et al. (2010). Reduced susceptibility to selected synthetic pyrethroids in urban malaria vector Anopheles stephensi: a case study in Mangalore city, South India. *Malaria journal*, 9, p.179.

Traore, M. M. et al. (2020). Large-scale field trial of attractive toxic sugar baits (ATSB) for the control of malaria vector mosquitoes in Mali, West Africa. *Malaria journal*, 19 (1), p.72.

Trape, J. F. and Zoulani, A. (1987). Malaria and urbanization in central Africa: the example of Brazzaville. Part III: Relationships between urbanization and the intensity of malaria transmission. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81 Suppl 2, pp.19–25.

Tripathy, A. et al. (2010). Distribution of sibling species of Anopheles culicifacies sl and Anopheles fluviatilis sl and their vectorial capacity in eight different malaria endemic districts of Orissa, India. *Memorias do Instituto Oswaldo Cruz*, 105 (8), pp.981–987.

Tripura, R. et al. (2018). A Controlled Trial of Mass Drug Administration to Interrupt Transmission of Multidrug-Resistant Falciparum Malaria in Cambodian Villages. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 67 (6), pp.817–826.

Tukei, B. B., Beke, A. and Lamadrid-Figueroa, H. (2017). Assessing the effect of indoor residual spraying (IRS) on malaria morbidity in Northern Uganda: a before and after study. *Malaria journal*, 16 (1), p.4.

United Nations. (2018). Revision of world urbanization prospects. *United Nations: New York, NY, USA*, 799.

Unwin, H. J. T. et al. (2022). Quantifying the direct and indirect protection provided by insecticide treated bed nets against malaria. *bioRxiv*. [Online]. Available at: doi:10.1101/2022.01.21.22269650.

Uwimana, A. et al. (2020). Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. *Nature medicine*, 26 (10), pp.1602–1608.

Valletta, J. J. et al. (2022). Individual-level variations in malaria susceptibility and acquisition of clinical protection. *Wellcome open research*, 6, p.22.

Vatandoost, H. and Hanafi-Bojd, A. A. (2012). Indication of pyrethroid resistance in the main malaria vector, Anopheles stephensi from Iran. *Asian Pacific journal of tropical medicine*, 5 (9), pp.722–726.

Viana, M. et al. (2016). Delayed mortality effects cut the malaria transmission potential of insecticide-resistant mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (32), pp.8975–8980.

Walker, P. G. T. et al. (2016). Estimating the most efficient allocation of interventions to achieve reductions in Plasmodium falciparum malaria burden and transmission in Africa: a modelling study. *The Lancet. Global health*, 4 (7), pp.e474-84.

Walker, T. et al. (2021). Stable high-density and maternally inherited Wolbachia infections in Anopheles moucheti and Anopheles demeilloni mosquitoes. *Current biology: CB*, 31 (11), pp.2310-2320.e5.

Wang, G.-H. et al. (2021). Combating mosquito-borne diseases using genetic control technologies. *Nature communications*, 12 (1), p.4388.

Wang, S.-J. et al. (2006). Rapid Urban Malaria Appraisal (RUMA) IV: epidemiology of urban malaria in Cotonou (Benin). *Malaria journal*, 5, p.45.

Wangdi, K. et al. (2016). Malaria elimination in India and regional implications. *The Lancet infectious diseases*, 16 (10), pp.e214–e224.

Warrell, D. A. and Gilles, H. M. (2017). *Essential Malariology, 4Ed*. CRC Press.

Watson, O. J. et al. (2019a). False-negative malaria rapid diagnostic test results and their impact on community-based malaria surveys in sub-Saharan Africa. *BMJ global health*, 4 (4), p.e001582.

Watson, O. J. et al. (2019b). Impact of seasonal variations in Plasmodium falciparum malaria transmission on the surveillance of pfhrp2 gene deletions. *eLife*, 8. [Online]. Available at: doi:10.7554/eLife.40339.

Watson, O. J., FitzJohn, R. and Eaton, J. W. (2019). rdhs: an R package to interact with The Demographic and Health Surveys (DHS) Program datasets. *Wellcome open research*, 4, p.103.

Weiss, D. J. et al. (2014). An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS journal of photogrammetry and remote sensing: official publication of the International Society for Photogrammetry and Remote Sensing* , 98, pp.106–118.

Weiss, D. J. et al. (2020). Global maps of travel time to healthcare facilities. *Nature medicine*, 26 (12), pp.1835–1838.

Weiss, D. J. et al. (2021). Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis. *The Lancet infectious diseases*, 21 (1), pp.59–69.

Wellems, T. E. and Plowe, C. V. (2001). Chloroquine-resistant malaria. *The Journal of infectious diseases*, 184 (6), pp.770–776.

Wesolowski, A. et al. (2012). Quantifying the impact of human mobility on malaria. *Science (New York, N.Y.)*, 338 (6104), pp.267–270.

White, M. T. et al. (2011). Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. *Parasites & vectors*, 4, p.153.

White, M. T. et al. (2014). Modelling the contribution of the hypnozoite reservoir to Plasmodium vivax transmission. *eLife*, 3. [Online]. Available at: doi:10.7554/eLife.04692.

White, M. T. et al. (2016). Variation in relapse frequency and the transmission potential of Plasmodium vivax malaria. *Proceedings. Biological sciences*, 283 (1827), p.20160048.

White, M. T. et al. (2018). Mathematical modelling of the impact of expanding levels of malaria control interventions on Plasmodium vivax. *Nature communications*, 9 (1), p.3300.

White, N. J. (2011). Determinants of relapse periodicity in Plasmodium vivax malaria. *Malaria journal*, 10, p.297.

Whittaker, C. et al. (2021). The ecological structure of mosquito population seasonal dynamics. *bioRxiv*, medRxiv. [Online]. Available at: doi:10.1101/2021.01.09.21249456.

WHO. (2014). *World malaria report 2014*. World Health Organisation Geneva.

Wilson, A. L. and IPTc Taskforce. (2011). A systematic review and meta-analysis of the efficacy and safety of intermittent preventive treatment of malaria in children (IPTc). *PloS one*, 6 (2), p.e16976.

Winskill, P. et al. (2017). The US President's Malaria Initiative, Plasmodium falciparum transmission and mortality: A modelling study. *PLoS medicine*, 14 (11), p.e1002448.

Wood, O. et al. (2010). Cuticle thickening associated with pyrethroid resistance in the major malaria vector Anopheles funestus. *Parasites & vectors*, 3, p.67.

World Health Organization. (2014). *WHO policy brief for the implementation of intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP)*.

World Health Organization. (2015). *Global Technical Strategy for Malaria 2016-2030*. World Health Organization.

World Health Organization. (2018a). *Artemisinin resistance and artemisinin-based combination therapy efficacy: status report*. World Health Organization. [Online]. Available at: https://apps.who.int/iris/bitstream/handle/10665/274362/WHO-CDS-GMP-2018.18-eng.pdf.

World Health Organization. (2018b). *High burden to high impact: a targeted malaria response*. World Health Organization. [Online]. Available at: https://apps.who.int/iris/bitstream/handle/10665/275868/WHO-CDS-GMP-2018.25-eng.pdf.

World Health Organization. (2018c). *Malaria surveillance, monitoring and evaluation: a reference manual*. World Health Organization.

World Health Organization. (2020). *World malaria report 2019 (World Health Organization)*.

World Health Organization. (2021a). Global technical strategy for malaria 2016-2030. 2021 update. *World Health Organization: Geneva*. [Online]. Available at: https://africa.northernrailextension.com/files-https-apps.who.int/iris/handle/10665/342995.

World Health Organization. (2021b). *World Malaria Report 2021*.

World Health Organization. (2021c). *Zeroing in on malaria elimination: final report of the E-2020 initiative*. [Online]. Available at: https://apps.who.int/iris/handle/10665/340881?locale-attribute=ar&order=desc&scope=&query=Zeroing%20in%20on%20malaria%20elimination&sort_by=score&rpp=10&search-result=true.

Wright, M. N. and Ziegler, A. (2015). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *arXiv [stat.ML]. arXiv* [Online]. Available at: http://arxiv.org/abs/1508.04409.

WWARN Artemisinin based Combination Therapy (ACT) Africa Baseline Study Group. (2015). Clinical determinants of early parasitological response to ACTs in African patients with uncomplicated falciparum malaria: a literature review and meta-analysis of individual patient data. *BMC medicine*, 13 (1). [Online]. Available at: doi:10.1186/s12916-015-0445-x.

Yared, S. et al. (2020). Insecticide resistance in Anopheles stephensi in Somali Region, eastern Ethiopia. *Malaria journal*, 19 (1), p.180.

Zhou, Z. et al. (2016). Assessment of submicroscopic infections and gametocyte carriage of Plasmodium falciparum during peak malaria transmission season in a community-based cross-sectional survey in western Kenya, 2012. *Malaria journal*, 15 (1), p.421.

Zomer, R. J. et al. (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, ecosystems & environment*, 126 (1), pp.67–80.