Modelling Malaria Elimination in Dynamic Populations

Sai Thein Than Tun
Linacre College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Clinical Medicine
Michaelmas term 2024

Modelling Malaria Elimination in Dynamic Populations

Sai Thein Than Tun, Linacre College

Doctor of Philosophy in Clinical Medicine

Michaelmas term 2024

Abstract

Malaria is an ancient disease, yet it is still affecting the developing parts of the modern world with over 249 million malaria cases in 2022. Sustained financial commitment has been a challenge for malaria elimination. Mathematical models can explore how malaria interventions can be implemented effectively and efficiently for malaria elimination.

This DPhil explores two interventions for malaria elimination using two mathematical models of different complexity: (1) mass drug administration (MDA) with a Susceptible-Infected-Recovered-Treatment (SIRT) compartmental model, (2) reactive case detection and treatment (RACDT) with an individual-based, spatially explicit model with a realistic household structure. To enable the simulation of the second model, two data analyses were done: (i) movement pattern analysis of villagers in rural Thai-Myanmar border area yielding the demographic characteristics and mobility of forest goers, (ii) analysis of demographic surveys from four Southeast Asian countries to investigate age-distribution within household, and to develop an algorithm to generate a synthetic population with realistic household structure.

The thesis highlighted the significance of human mobility in the success of malaria interventions. It explained how we can leverage the meta-population version of herd effect resulting from human mobility to our advantage when implementing large-scale malaria interventions, such as MDA. The thesis found that in Thai-Myanmar rural area almost all forest goers were males (20-40 years), and they went into the forest on average twelve times/year. Their probability of staying in the forest each successive night was provided with a Kaplan-Meier estimate.

Incorporating a realistic household structure provided a more accurate predictions in the second model. When deploying RACDT, implementing it at the household level was predicted to be the most efficient configuration at the low transmission setting. Finally, the thesis discussed when and how to implement MDA and RACDT effectively in an integrated malaria elimination campaign involving other interventions.

Table of Contents

Acknowledgement

I would like to express my deepest gratitude to my supervisors, Professor Lisa White and Associate Professor Ricardo Aguas, whose expertise, understanding, and patience added significantly to my DPhil journey in mathematical modelling of malaria. I appreciate to learn from their vast knowledge and skill in many areas from conceptualization of research, to engaging the results of the research with different stakeholders. I am also thankful for their guidance in writing reports and presentations. Their invaluable insights and constructive feedback were instrumental in shaping this work.

In fact, my supervisors' support has gone beyond the scope of my DPhil. My life improved significantly since I joined Professor White's team in 2015 as a trainee in mathematical modelling. Through this opportunity, I was given a chance to engage with enthusiastic international scientists, to utilize state of the art equipment and tools for modelling, to secure a fellowship to study for a Masters degree in the UK for a year, to attend meetings and conferences in various countries to present my research. I will forever be grateful to the impact my supervisors have on my life.

My sincere thanks go to Associate Professor Daniel Parker for letting me collaborate on the human movement analysis project which forms a part of my thesis in Chapter 3. I extend my thanks to my research colleagues and friends, Dr. Shwe Sin Kyaw, and Dr. Sompob Saralamba, to name a few, for their support, encouragement, and for making my research journey an enjoyable experience. I would like to thank the administrative and supportive teams from the Mahidol-Oxford Research Unit, the Linacre College and the University of Oxford.

I suffered from a debilitating eye condition closer to the end of my DPhil. I was only able to work at most two hours a day with generous breaks, and subsequently my DPhil progress dragged on. But my wife, Dr. Yu Nandar Aung, supported me emotionally, and financially. Her encouragement kept me going until I completed my DPhil thesis by writing up one paragraph per day. I owe a great deal of gratitude to her for her unconditional love and understanding. I give my special thanks to my mom, Dr. Aye Aye Than, and my dad, U Win Maung, for their unwavering support during this challenging period and throughout my career.

Author's Contribution

The author was the primary contributor to all material presented in this thesis with support from DPhil supervisors. Additional contributions by others are outlined below.

Chapter 1: Primary author with review from DPhil supervisors

Chapter 2: Primary author with review and intellectual input from DPhil supervisors

Chapter 3: Primary author with review from DPhil supervisors. Intellectual input and access to data from Associate Professor Daniel Parker

Chapter 4: Primary author with review and intellectual input from DPhil supervisors

Chapter 5: Primary author with review from DPhil supervisors. The base model was developed by one of the DPhil supervisors (Dr Ricardo Aguas), and Dr. Bo Gao. The extension of the model and its simulation was conducted solely by the author

Chapter 6: Primary author with review from DPhil supervisors

Publications

Publications arising from DPhil work

Tun STT, Min MC, Aguas R, Fornace K, Htoo GN, White L, Parker, DM. Human movement patterns of farmers and forest workers from the Thailand-Myanmar border. Wellcome Open Research (2021)

Tun STT, Parker DM, Aguas R, White LJ. The assembly effect: the connectedness between populations is a double-edged sword for public health interventions. Malaria Journal (2021)

Tun STT, von Seidlein L, Pongvongsa T, Mayxay M, Saralamba S, Kyaw SS, Chanthavilay P, Celhay O, Nguyen TD, Tran TN, Parker DM, Boni MF, Dondorp AM, White LJ. Towards malaria elimination in Savannakhet, Lao PDR: mathematical modelling driven strategy design. Malaria Journal (2017)

Articles published during DPhil from other work

Aung YN, **Tun STT**, Vanisaveth V, Chindavongsa K, Kanya L. Cost-effectiveness analysis of G6PD diagnostic test for Plasmodium vivax radical cure in Lao PDR: An economic modelling study. PLoS ONE (2022)

Saita S, **Tun STT**, Parker DM. The Past, Present, and Future of Spatial Demography in Southeast Asia. Solicited article as part of "Looking Back, Looking Forward: Progress and Prospect for Spatial Demography." Spatial Demography (2021)

Chanthavilay P, Soukavong M, Aung YN, **Tun STT**, White LJ, Mayxay M. Comparing the potential effectiveness of interventions against coronavirus 2019 outbreak in the Lao PDR: a mathematical modeling approach. Lao Medical Journal (2020)

Alahmadi A, Belet S, Black A, Cromer D, Flegg JA, House T, Jayasundara P, Keith JM, McCaw JM, Moss R, Ross JV, Shearer FM, **Tun STT**, Walker CR, White LJ, Whyte JM, Yan AWC, Zarebski AE. Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges. Epidemics (2020)

Parker DM, **Tun STT**, White LJ, Kajeechiwa L, Thwin MM, Landier J, et al. Potential herd protection against plasmodium falciparum infections conferred by mass antimalarial drug administrations. eLife (2019)

List of figures

List of tables

List of abbreviations

ACT	Artemisinin-based combination therapy
API	Annual parasite incidence
BRB	Biased random bridge
BMGF	The Bill & Melinda Gates Foundation
COVID-19	Coronavirus Disease 2019 (an infectious disease caused by the SARS-CoV-2 virus)
CP	Constraint programming (algorithm in chapter 4)
DDT	Dichloro-diphenyl-trichloroethane
DHS	Demographic health surveys
EDAT	Early diagnosis and adequate treatment
G6PD	Glucose-6-phosphate dehydrogenase
GFATM	The Global Fund to fight AIDS, Tuberculosis and Malaria
GMEP	Global Malaria Eradication Programme
GMS	Greater-Mekong Subregion
GPS	Global positioning system
GTS	The Global Technical Strategy for Malaria 2016-2030
HS RDT	Highly sensitive rapid diagnostic test
HIC	High-income country
IBM	Individual-based model
ITFDE	International Task Force for Disease Eradication

IRS	Indoor residual spraying
ITN	Insecticide-treated nets
Lao PDR	Lao People's Democratic Republic
LLIN	Long-lasting insecticide-treated nets
malERA	The Malaria Eradication Research Agenda
MDA	Mass drug administration
MESA	Malaria Eradication Scientific Alliance
MSAT	Mass screening and treatment of imported cases
MICS	Multiple Indicator Cluster Survey
MVDA	Mass vaccination and drug administration
PCR	Polymerase chain reaction
PK/PD	Pharmacokinetic/pharmacodynamics
PRP	Percent reduction in prevalence (output metric of the model in Chapter 5)
RA	Random Assignment (algorithm in Chapter 4)
RACDT	Reactive active case detection and treatment
RBM	The Roll Back Malaria program
RDA	Reactive drug administration
RDT	Rapid diagnostic test
RMI	Reactive malaria interventions
SAE	Sum of absolute errors (comparison metric in Chapter 4)
SEA	Southeast Asia
SIR	Susceptible, Infected, Recovered (model compartments or disease states)

SIRT	Susceptible, Infected, Recovered, Treatment (model compartments or disease states)
uPCR	Ultrasensitive quantitative PCR (polymerase chain reaction)
UTM	Universal Transverse Mercator (coordinate grid system)
VHV	Village health volunteers
VMW	Village malaria worker
WGS 84	World Geodetic System 1984 (3D coordinate system that uses latitude, longitude, ellipsoidal height, and time to represent a location)
WHO	World Health Organization

1 Introduction

1.1 Summary

Malaria is a vector-borne parasitic disease predominantly caused by the parasites of *Plasmodium falciparum* and *P. vivax* species in humans. It is transmitted by the bite of the female *Anopheles* mosquitos. There were 249 million malaria cases, and over 600 thousand deaths from malaria worldwide in 2022. Global malaria incidence has been around 60 per 1,000 population at risk for the last four years which is not a significant progress compared to the two decades ago when the figure was 81 and when malaria elimination effort was re-initiated. Up to now, political will and financial backing for malaria elimination has been strong. But, for the malaria elimination to be sustainable, the interventions must be as efficient and effective as possible. Mathematical models can help evaluate malaria interventions of varying costs and effectiveness depending on the context and formulate optimal integrated strategies for malaria elimination. The thesis aims to explore the effective and efficient interventions for malaria elimination using mathematical models with emphasis on two important demographic properties: human mobility and household structure.

1.2 Malaria

1.2.1 Brief history of malaria

"Malaria" comes from ancient Italian words "mal", meaning bad and "aria", meaning air from the eighteenth century based on the belief that the disease was caused by foul air arising from the marshes. There were evidences that even prehistoric humans were

infected with malaria.[1,2] Presence of malaria in Europe and ancient Egypt has been confirmed by paleo-microbiological diagnoses.[2,3] Malaria has co-existed and co-evolved with humans for several centuries, acting as the evolutionary driving force for humankind in the development of sickle-cell disease, thalassemia, and glucose-6-phosphatase deficiency.[4,5]

Malaria parasites were first discovered as causal agents of malaria by Alphonse Laveran in 1880s. The role of *Anopheles* mosquito as a vector in malaria transmission was made in the 1890s by Ronald Ross. It took another 90 years to understand the complete life cycle of malaria parasites including the role of hypnozoites in *Plasmodium vivax*.[1]

The prevalence and impact of malaria before the development of accurate diagnostic tools could only be surmised through the descriptions of malaria-like symptoms (intermittent fever) in the historical records and the prevalence of co-evolved diseases such as β-thalassemia.[2,6,7] Before the mid-1900s, about 90% of the world population was at risk of malaria.[8] Around that time, the environmental and social changes such as the growth of urban centres, drainage of swamps for more agricultural land use, improved hosing and health care infrastructure, caused by the rapid economic growth in Europe and the United States rendered the malaria transmission in those areas to be unfavourable.[9] Malaria was thought to be endemic in Europe for over 2 millennia until it was eradicated there in the 1970s.[9]

1.2.2 Natural history

Malaria is a vector-borne parasitic disease caused by five protozoa species of *Plasmodium,* namely *Plasmodium falciparum, P. vivax, P. malariae, P. ovalae and P. knowlesi.* As per the nature of the parasitic pathogens, malaria parasites require living hosts to fulfil their

reproduction and survival. Female *Anopheles* mosquitos act as the primary hosts and transmission vectors while the humans are the intermediate hosts. The overview of parasites life cycle can be seen in the illustration (Figure 1) reproduced from Phillips 2001.[10]

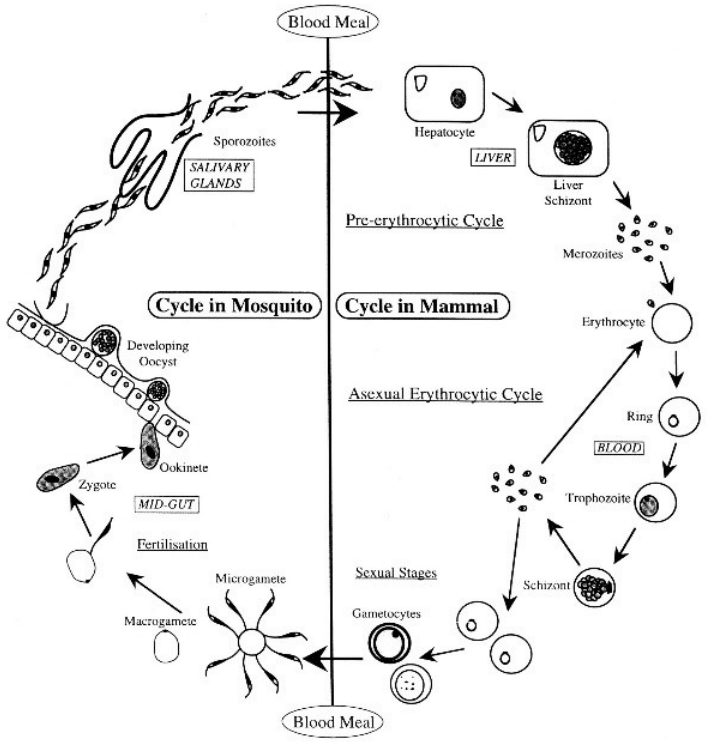


Figure 1: Life cycle of Plasmodium parasites across two different hosts.Reused with permission from the publisher. Source: Figure 1 from Phillips 2001, published in Clinical Microbiology Review. Available from [https://doi.org/10.1128/cmr.14.1.208-226.2001].

1.2.2.1 In human host

Lifecycle

When an infected female anopheline mosquito takes a blood meal from a human, it injects around 100 sporozoites into the person's skin where they stay for up to 6 hours before entering the lymphatics or the blood stream.[11] Sporozoites travel to the hepatocytes where

they mature into schizonts within 2 to 10 days, or lay dormant as hypnozoites. After merozoites are released from the ruptured schizonts, they enter the erythrocytic stage by invading red blood cells. Between 10,000 and 40,000 merozoites are produced from a single sporozoite.[12] In the erythrocytes, the merozoites undergo asexual reproduction, after maturing from the ring stage to mature trophozoite, and produce more schizonts (each erythrocytic schizont contains 6 to 36 merozoites) to repeat the cycle.[11] Erythrocytes are destroyed as part of this cycle, manifesting in rhythmic clinical symptoms such as chills and fever after the incubation period of 7 to 30 days when there are billions of parasites in the circulation. Some of the schizonts develop into male and female gametocytes that circulate in the peripheral circulation within 7 to 15 days after erythrocytes are invaded.[11] Gametocytes are picked up by a mosquito during a blood meal. Dormant hypnozoites could be reactivated months or years later to enter the erythrocytic stage and cause relapse.

Severe malaria

If left untreated, malaria infection could progress into severe malaria infection, which is a medical emergency, characterized by the presence of asexual parasitaemia with one or more of the following: impaired consciousness, prostration, multiple convulsions, acidosis, hypoglycaemia, severe malarial anaemia, renal impairment, jaundice, pulmonary oedema, significant bleeding, shock and hyperparasitaemia.[13] Untreated severe malaria, particularly cerebral malaria, has a mortality close to 100%.[13]

Chronic malaria

Humans could acquire immunity against the malaria parasites, especially in the malaria endemic regions. In such cases chronic malaria infections, the symptoms of which may not be severe enough to warrant a visit to the health care professional, are possible. Chen et al. 2016 argued that truly "asymptomatic" malaria does not exist since chronic malaria infection will result in anaemia, splenomegaly, and increase in the risk of systemic bacterial infections.[14] They presented the relationships between the parasite load, the risk of morbidity and mortality, the symptomology and the detection thresholds in a concise illustration (Figure 2).

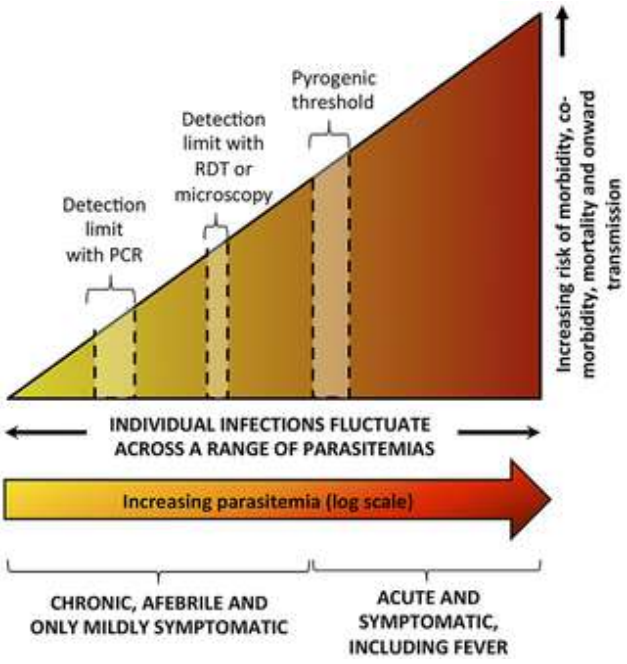


Figure 2: Spectrum of malaria infection. Source: Figure 1 from Chen et al. 2016, published in PLOS Medicine under CC BY 4.0. Available at [https://doi.org/10.1371/journal.pmed.1001942].

Immunity

Adaptive immunity against malaria can be naturally acquired after a single infection, but repeated infections are necessary to induce adequate protection against severe morbidity and mortality. Doolan et al. defined the types of acquired immunity against *Plasmodium* parasites as (1) antidisease immunity, which provides protection against clinical disease and reduces the risk of morbidity; (2) anitparasites immunity, which provides protection against parasitaemia thus reducing the density of parasites; and (3) premunition, which provides protection against new infections by having a low-grade, asymptomatic parasitaemia. Complete or sterilizing antiparasite and antidisease immunity could not be induced naturally.[11]

It takes longer to acquire protection against *P. falciparum* compared to those against *P. vivax* and *malariae,* and immunity is species specific and strain specific.[11] Even within a single strain of a species, the *Plasmodium* parasites could express unique sets of stage-specific genes and multiple antigens throughout its several developmental stages.[15] Additionally, the antigenic variation among clonal populations of the parasites further slows the development of acquired immunity.[11] Depending on the malaria endemicity, acquired immunity in the population will reduce the prevalence of severe malaria cases and the morbidity associated to malaria. A modelling study estimated that the antidisease immunity has a half-life of about 5 years and the antiparasite immunity has a half-life of over 20 years.[16] Many key aspects of the immunity against malaria are still not yet fully understood and much more research is needed before better immunological interventions could be developed.[17]

1.2.2.2 In mosquito vector

In the newly infected female *Anopheles* mosquito, the male and female gametocytes undergo sexual reproduction in the mosquito's stomach and produce zygotes (the fertilized eggs) which migrate into the midgut wall to become the oocysts. The oocysts mature and rupture to release sporozoites which move to the salivary glands to infect a human host.

1.2.3 Diagnosis

Before the discovery of the cause of malaria (i.e., the parasites) by Alphonse Laveran in 1880s, malaria was only recognized through its symptoms and signs such as fever and enlarged spleen, which are not disease specific.[1] With the advancement in medical technology and the better knowledge in the natural history of malaria, it can now be diagnosed with very high specificity and sensitivity.

In the 2021 WHO Guidelines for malaria[18], early diagnosis and prompt, effective treatment of malaria is recommended. Parasitological test confirmation with either light microscopy or immunochromatographic rapid diagnostic test (RDT) is recommended for all suspected malaria cases. But, in the absence of a confirmatory test, antimalarial treatment is still recommended for a suspected severe malaria case.

1.2.3.1 Light microscopy

To diagnose malaria with light microscopy, a lab technician needs to prepare two kinds of blood films from the peripheral blood on microscope slides. A thick film is used to detect the presence of malaria parasites and a thin film is used to confirm the species. Giemsa

stain is used to differentiate the parasites from the normal blood cells. The slides are then examined under the light microscope. The more detailed steps can be found in the WHO's Basic malaria microscopy guide[19]. Depending of the skills of the microscopist, the detection limit is about 100 parasites per microlitre[20]. It is still a gold standard against which other diagnostic methods are compared.

1.2.3.2 Rapid diagnostic test (RDT)

RDTs are lateral flow immune-chromatographic tests that detect parasite-specific antigens from the peripheral blood sample. They do not require specialized equipment nor electricity, and they are easy to perform and interpret. Some RDTs can detect multiple species. Their sensitivity, specificity and detection limit for a particular *Plasmodium* species can vary according to their manufacturer.[18]

The WHO has set up a product testing programme to evaluate the RDTs so that they must detect both a low parasite density (200 parasites/µL which is a minimal threshold that tests must detect to identify clinical malaria reliably in many different settings) and a higher parasite density (2000 parasites/µL).[21]

1.2.3.3 Molecular diagnostic tools

The diagnostic tools based on nucleic-acid amplifications such as PCR are very sensitive, specific, and costly. They may not be useful or cost effective in a clinical context, but they have been used in research and for some elimination strategies that will be discussed in the

later sections.

1.2.4 Treatment

Management of malaria depends on its severity, the causal species, and the national malaria treatment guidelines. For uncomplicated *Plasmodium falciparum* malaria, the WHO recommends artemisinin-based combination therapy (ACT) as the first-line treatment.[22] ACTs combine an artemisinin derivative, such as artemether or artesunate, with a partner drug like lumefantrine or amodiaquine. The standard regimen involves administering the drugs for three consecutive days to ensure parasite clearance and reduce the risk of drug resistance. In areas with high resistance to ACT partner drugs, the use of triple ACTs is still being evaluated.[22]

For uncomplicated *Plasmodium vivax* infections, treatment typically involves chloroquine in areas where resistance is not prevalent. In addition to clearing the blood-stage parasites, a radical cure using primaquine is recommended to eliminate liver-stage hypnozoites and prevent relapses. However, primaquine use requires prior screening for glucose-6-phosphate dehydrogenase (G6PD) deficiency to avoid haemolysis. In regions where chloroquine resistance is present, ACTs such as artemether-lumefantrine can be used for the blood-stage infection.[22]

Supportive care is essential in the treatment of uncomplicated malaria to address symptoms like fever and dehydration. Antipyretics, such as paracetamol, are commonly used to manage fever. Adequate hydration and monitoring for potential complications, such as anaemia, are also crucial during the recovery phase. The WHO emphasizes prompt

diagnosis and early treatment to minimize complications and reduce malaria transmission within communities. Adherence to national guidelines tailored to local resistance patterns is vital to optimize treatment outcomes.[22]

1.3 Malaria control and elimination

After the success of smallpox eradication in the 1980s, humans have strived to eradicate more preventable infectious diseases. Poliomyelitis, and dracunculiasis (Guinea worm disease) are on track to becoming eradicated.[23,24] Before achieving eradication of an infectious disease, where worldwide incidence is zero, elimination must be attained in all the geographical areas of the world. The distinction between eradication and elimination, and their definitions can be seen in the disease control spectrum illustrated in Figure 3.

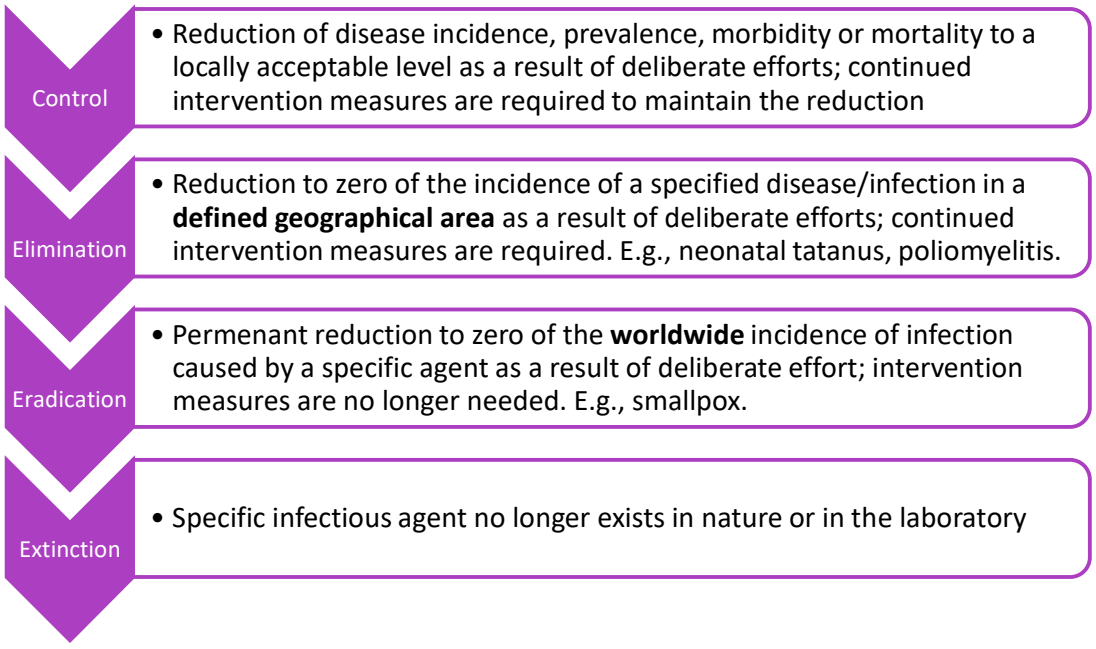


Figure 3: Definitions on disease control, elimination, eradication, and extinction [adapted from[25] Dowdle 1998, WHO]

1.3.1 Malaria burden

Even with today's state-of-the-art medical technology and advances, the malaria parasites have not been eradicated from our world. Socioeconomic disparity still plays a great role in the persistence of malaria in the developing countries.[26] In the World Malaria Report 2023[27], the WHO estimated 249 million malaria cases in 85 malaria endemic countries in 2022 which was an increase of 5 million from 2021. Globally, malaria case incidence was estimated to be 58 cases per 1000 population at risk. Out of all the malaria cases in 2022, 94% of cases (~233 million cases) were from the WHO African Region. The WHO South-East Asia Region only accounted for 2% of the global figure (~5 million cases). Deaths form malaria globally in 2022 is estimated to be 608,000. Global malaria mortality rate is 14.3 deaths per 100,000 population at risk.

1.3.2 Lessons learned from prior elimination attempt

Eradication of malaria has been attempted in 1955 when the WHO launched the Global Malaria Eradication Programme (GMEP). A residual insecticide, dichloro-diphenyl-trichloroethane (DDT) spraying was used extensively, in the GMEP, to kill indoor resting adult mosquitoes. Traditional methods of malaria control such as destruction of mosquito breeding marshes and prevention of mosquito bites were abandoned.[28] The GMEP eliminated malaria from several regions of the world, but failed to achieve global eradication. In the 1960s, some countries failed to progress, and some countries faced malaria resurgence after a long period of interruption. Nájera et. al. attributed the failure of GMEP to (1) waning international support for malaria control due to economic and

financial crises in the 1970s; (2) rising levels of drug and insecticide resistance; (3) having an inflexible and unsustainable vertical approach with limited community involvement regardless of the malaria levels; and (4) lacking integration with research bodies, health systems and surveillance systems.[28]

In the last few decades, there has been success stories in a few countries. The island nation of Sri Lanka was certified malaria-free in 2016, seventeen years after the end of 26-year long civil war.[29,30] In 2021, China was also certified malaria-free even though it still has to worry about the case importations from the neighbouring GMS countries.[31] With the exception of 2020, when the disruption to the malaria services during the COVID-19 pandemic caused the increase in malaria cases, the malaria burden is in a generally declining trend thanks to the elimination efforts explained in the following sections.[32]

1.3.3 Current malaria elimination efforts

1.3.3.1 Funding & political commitment

The International Task Force for Disease Eradication (ITFDE) was reactivated in 2000 by The Bill & Melinda Gates Foundation (BMGF).[33] In 2007 at the malaria forum, Melinda Gates and Margaret Chan, the then-secretary general of the WHO, call to action for malaria eradication.[34] The Global Malaria Action Plan (GMAP) was developed in 2008 by The Roll Back Malaria (RBM) partnership, when USD 3 billion was initially committed by donors and government agencies, to achieve a substantial reduction in the burden malaria in a sustainable framework in the short-term with a long-term goal of global malaria eradication.[35,36] With the lessons learned from GMEP, research was back on the agenda to support every step of the malaria eradication. The Malaria Eradication Research Agenda

(malERA) initiative was established to identify knowledge gaps and new generation of tools to interrupt malaria transmission until the Malaria Eradication Scientific Alliance (MESA) took its place in 2012.[37]

The total funding for malaria control and elimination in 2020 was estimated at USD 3.3 billion, 68% of which came from the international funding, according to the latest World Malaria Report.[32] The report states that out of the domestic funding of USD 1.1 billion, USD 0.3 billion was spent on malaria case management and USD 0.7 billion on other malaria control activities. It doesn't report the expenditure from the international funding however.

A notable funder is the Global Fund to fight AIDS, Tuberculosis and Malaria (GFATM) which is an international funding body committed to control and/or eliminate the diseases mentioned in its namesake. According to its 2021 report, the GFATM is the largest financier of malaria control programs, accounting for 56% of all international funds and supporting 107 countries. The GFATM had invested nearly US$ 15 billion for malaria since its inception in 2002.[38] So far, the global community has shown a sustained financial and political commitment to eliminate malaria and hopefully it will continue until the elimination is achieved.

1.3.3.2 Coordinating body & Guiding principles

Each nation/country or region could be anywhere on the malaria control spectrum (Figure 3) with some having the certified malaria-free status while others are still striving for malaria control. There is no one-size-fits-all approach to the malaria elimination, as a lesson-learned from the previous effort, each place has its targets and goals for the malaria control

and elimination while being coordinated at the cross-border areas. The WHO is serving as a guiding and coordinating body for the current concerted effort of global malaria elimination.

The WHO's "The Global Technical Strategy for Malaria 2016-2030" (GTS) includes three pillars of strategy: (Pillar 1) ensuring access to malaria prevention, diagnosis and treatment as part of universal health coverage; (Pillar 2) accelerating efforts towards elimination and attainment of malaria-free status; (Pillar 3) transforming malaria surveillance into a key intervention; with two supporting elements: (1) harnessing innovation and expanding research; (2) strengthening the enabling environment for more sustainable and equitable results. The Strategy is set with five-year milestones up to 2030, when malaria case incidence and mortality globally are targeted to be reduced by at least 90% compared with 2015, while individual countries would their own national targets.[39,40]

The GTS was updated in 2021 after reflecting upon the first five years of implementation. It reported that "the progress in reducing malaria mortality and morbidity slowed, stalled or even reversed in many moderate- and high-transmission countries".[40] The rapid population growth in sub-Saharan Africa, even though malaria incidence has decreased, was explained as a paradoxical effect for the progress being stalled. Many of the poorest and the most marginalized people are still unable to access health care and malaria services without the financial hardship. The updated GTS recommends reaching the unreached by stepping up significant expansion of existing interventions to all those in need, to make malaria a higher technical, financial and political priority, and to ensure that the development and use of interventions are maximized, guided by the stratification of data by risk to optimize impact and the cost-effectiveness of interventions. While the GTS serves as a high-level overarching strategy for malaria elimination globally, a more specific "A framework for

malaria elimination" was also developed by the WHO to help guide the malaria elimination effort in all malaria-endemic countries.[41]

1.3.4 Challenges

The WHO highlights the lack of predictable and sustained funding with sustained political commitment and regional collaborations as one of the challenges for malaria elimination.[40] It also acknowledges the biological challenges such as expansion and/or emergence of parasites' resistance to malaria drugs and mosquitoes' resistance to insecticides. Human mobility and the difficulty to track mobile and migrant population who are at high risk of malaria also played a significant role in the persistence of malaria. Meanwhile, total funding for malaria research and development (R&D) has been declining since 2019. In 2020, malaria R&D funding was estimated at USD 619 million—falling short by USD 232 million of the projected requirement of USD 851 million.[32] COVID-19 pandemic, and political instability such as the coup in Myanmar has been proven to disrupt the malaria services.[32,42,43] The war in Ukraine has already had financial repercussions globally, and if it leads to the global financial recession, malaria funding could be cut off.

1.3.5 Public health interventions on malaria

As long as there are parasites in the vector or the intermediate host, malaria transmission cycle will go on indefinitely. The public health measures designed to interrupt malaria transmission are referred to as "interventions" throughout the thesis. Interrupting the life cycle of the parasites, discussed in Section 1.2.2, is the key to prevention and treatment of malaria.

Malaria interventions and tools are categorized in the following section based on the type of preventions (Table 1). Primary prevention prevents the previously uninfected persons from contracting the infection whereas the secondary prevention aims for early diagnosis and prompt treatment. Tertiary prevention prevents the development of the complications arising from the infection. From the viewpoint of the human host, all interventions done to the mosquitos are primary preventions.

Primary prevention:
• Prevent humans from mosquito bites
o Insecticide-treated nets (ITNs), long-lasting insecticide-treated nets (LLINs)
o Mosquito repellents (skin ointments, mosquito coils and sprays)
o Housing improvements (mosquito proof doors and windows) so mosquitos will not enter the house
• Prevent humans from getting infected even when bitten by infectious mosquitos
o Vaccination
o Use of RTS,S vaccine was approved by the WHO in October 2021.[44] In children aged 5-17 months, four doses of RTS,S/AS01 vaccine has a

modest efficacy of 39% against clinical malaria and 29% against severe malaria.[45]
• Prevent mosquito breeding
○ Clear the stagnant water reservoirs in and around the house where the mosquitos breed
○ Cultivate lava eating fish in the water bodies
• Get rid of the mosquitoes
○ Mosquito coils/sprays/fumigations
○ Killing mosquitos with laser (not cost effective)
• Genetically modified mosquitos that are resistant to the *Plasmodium* parasites
Secondary prevention:
• Interrupt parasite lifecycle within humans
○ Early diagnosis and adequate treatment with multi-drug combinations such as Artemisinin-based combination therapy (ACT)

○ Primaquine along with ACTs to clear hypnozoites and prevent relapse of *P. vivax* and *P. ovale.* It is also used to kill gametocytes and prevent further transmission.
Tertiary prevention:
• Proper case management of severe malaria

Table 1: Examples of malaria intervention across the disease prevention continuum

Given the plethora of tools with varying costs and efficiencies depending on the context, putting these together into a cost-effective, integrated strategy could become a daunting task. Fortunately, with the advancement in mathematics and computational technologies, multiple combinations of available tools could be tested out *in silico* experiments and mathematical models for the myriad of scenarios. In fact, the WHO's GTS notes that "Mathematical models may help countries to understand the impact on malaria of the scenarios with different combinations of interventions".[40] The WHO's recommendation on the RTS,S/AS01 vaccine was partly based on the favourable cost-effectiveness estimates of mathematical models.[45] The next section explains the mathematical models briefly and the efforts that has been done in mathematical modelling of malaria and its elimination.

1.4 Mathematical models

1.4.1 What are mathematical models?

Models are the simplified representation of the real-world complex phenomena.[46] When real-world systems are too complex or develop too slowly to be analysed using experiments, models can be used instead. Just as an orrery model might use balls to represent the solar system, mathematical models use symbols and equations to represent the system under study.

Mathematics has been applied in the physical sciences for over three centuries. But the use of mathematical methods in biology was not seen until the 17th century. The first mathematical model of infectious disease appeared in a manuscript in 1760 by Daniel Bernoulli when he discussed the effectiveness of variolation in immunizing against smallpox.[47]

Mathematical models are being used in many disciplines of science, such as social science (political science, economics), and engineering (aerospace, energy and computer science).[48] Especially in the life sciences, they provide an opportunity to conduct experiments that may be impossible or impractical to conduct in real life because they are too complex, resource-intensive, or unethical.[49,50]

Developing mathematical models is multidisciplinary in nature – requiring mathematics, domain knowledge of the concept under investigation and computer programming to manipulate the model and experiment on. Bailey recognized the difficulties of practical scientists to understand advanced mathematics and professional mathematicians to understand the practical issues in greater depth, and the importance of effective

collaboration between them.[47] The process of mathematical model development is a balancing act between accuracy, transparency and flexibility.[50] It generally entails:[49]

- Identifying the problem
- Formulating the model: choosing appropriate type of modelling method, design its assumptions and algorithms
- Specifying model input parameters
- Setting up the model
- Validating the model
- Prediction and optimization

1.4.2 Classification of mathematical models

Mathematical models can be classified based on their important differentiating features. One feature is whether the model considers chance or probability in its processes. If a model doesn't incorporate probability, its output would be always the same on every model run (or pre-determined) provided the input parameters are kept the same. Such models are called **deterministic models**, and they describe what happens in the population on average. On the other hand, models that output different results every time because they incorporate probability are termed **stochastic models**.

Another way to classify mathematical models is based on how they conceptualize the population units. **Compartmental models** group individuals of the same characteristics and represent them as separate populations or compartments. Since the aspects of probability are negligible in a large enough population, most compartmental models tend to be deterministic. In the agent-based or **individual-based models (IBM),** each individual is

represented with its own attributes and choices in the model. Due to their explicit incorporation of variability/probability/uncertainty, IBM are almost always stochastic.

For this thesis, the classification of models into compartment model and IBM is more relevant and these two types of models are explained in the following brief introductions.

1.4.3 Introduction to compartmental models

A simple compartmental model for an acute infection is provided here as example. It could represent an acute infection that infect susceptible individuals, quickly develop an immunity, and then recovered such as influenza, chickenpox and rabies. The population is subdivided into three compartments – Susceptible, Infected, and Recovered in the model, thus it is named S-I-R model.[51] Once people in S compartment are infected, they will go to I compartment (Figure 4). From I, those recovered will go to R compartment. For simplicity, the immunity is assumed to be lifelong, and the population is constant (i.e., there is no births, deaths and migrations) and homogeneously mixed.

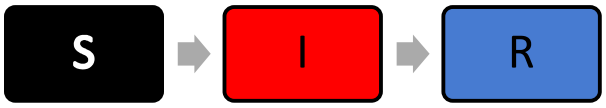


Figure 4: Flow diagram of a simple SIR model

The flow of people across these compartments can be modelled by the rate of change of each compartment which are represented mathematically by the following sets of ordinary differential equations:

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Equation 1 Rate of change of the respective compartments in a simple SIR model

In Equation 1, S, I and R denote the proportions of the total population in their respective compartments. The left side of the equations represents the change of each compartment (e.g., dS means the change in *S*) with respect to change in time (*dt*). For disease transmission (progression from S to I) to happen susceptible people in S must come into contact with infected cases in I first. Since not all contacts result in infection, the contact rate is scaled by the transmission probability. This effective contact rate is denoted by β which is the product of the contact rate and transmission probability. β is then multiplied by the prevalence of disease in the population (I) and the availability of the susceptible people (S). βSI becomes the rate at which people leave the S compartment and enter the I compartment. Infected cases eventually recover at the recovery rate γ, thereby leaving I compartment and entering R compartment. γ is calculated from the reciprocal of average infectious period. The dynamics of a typical SIR model can be seen in Figure 5.

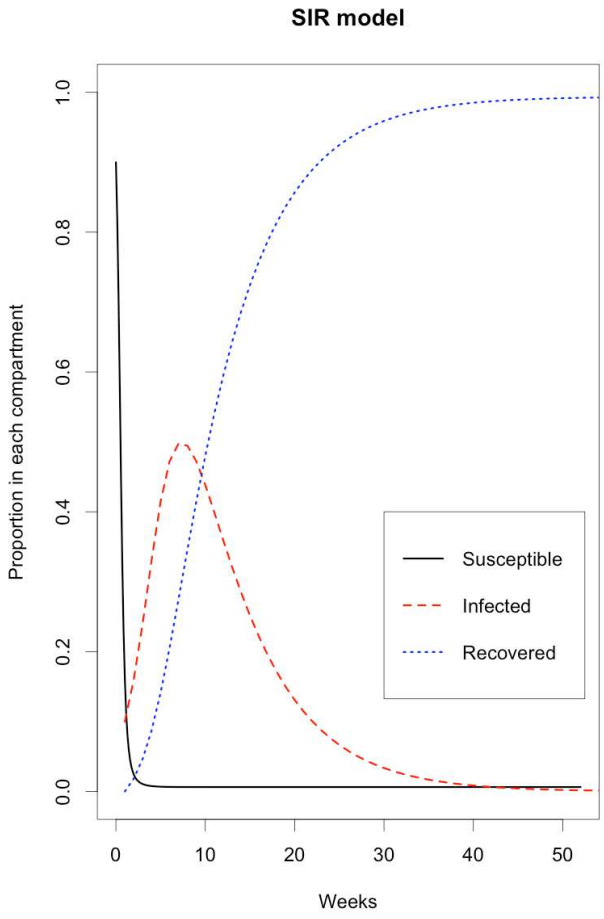


*Figure 5: Typical **SIR** compartmental model dynamics. Proportions in the respective compartments of **SIR** are in different line types and colour. The red line (Infected) is the epidemic curve. Beta=5/week, 1/Gamma=1week.*

By rearranging the equation of I, two important epidemiological quantities are achieved.

$$\frac{dI}{dt} = I(\beta S - \gamma)$$

In order to have more infected persons in the I compartment, $(\beta S - \gamma)$ must be positive. This warrants that S must be greater than the ratio $\frac{\gamma}{\beta}$ which is called the critical epidemic threshold or the herd immunity threshold or the relative removal rate. Once the

proportion of S gets below this threshold, the I compartment will shrink. If the epidemic has not started, it will not occur. If the epidemic has already started, and the proportion of the S is no longer below the critical epidemic threshold, the epidemic will start to dwindle.

The reciprocal of the critical epidemic threshold is the basic reproduction number (R_0) which is defined as "the average number of secondary cases arising from an average primary case in an entirely susceptible population". Therefore, if a primary case causes less than 1 secondary case (R_0 is less than 1), the epidemic cannot take hold. When the rate of change of all the compartments in the model does not change, then the model is at equilibrium. One obvious equilibrium is when there is no infection (disease-free equilibrium). Endemic equilibrium can occur when the disease persists in the population. Equilibrium analysis can be done by setting the rate of change of each compartment (such as those in Equation 1) to zero and solving them either analytically or numerically. Equilibrium analysis can provide the better understanding of the disease dynamics, evaluating control strategies, and likelihood of disease eradication.

In Chapter 2, a compartmental model is presented to explore the impacts of human mobility on targeted malaria interventions.

1.4.4 Introduction to individual-based models (IBM)

Unlike the compartmental models, individual-based models (IBM) conceive each individual as a separate entity in the model. Imagine an excel sheet with each row representing an individual and each column representing a property such as age, gender, disease state, location, resource, history, and so on. Given enough properties, an IBM can create a heterogeneous population with unique individuals who could act autonomously

based on their properties with one another and with their local environment. Interactions between individuals in the IBM are usually constrained to their neighbours in a geographic space or a social network.

Whereas the compartmental models usually have one level (the population level), IBM have at least two levels – the population level and the individual level. The effect of interactions in the individual level produces an effect on the population level, which is called the emergence property of IBM. The population level changes (such as an intervention for a disease) will impact the individuals as well.

IBMs run on a discrete time step that is sensible for the system under investigation. An IBM modelling acute infection would require a small time step such as a day or even smaller whereas a chronic disease IBM may need a larger time step like a month. During each time step, processes (e.g., growing old by a time step), and interactions (e.g., whether or not infection is acquired) take place for all eligible individuals in the model.

To build on the previous compartmental model example, a corresponding individual-based SIR model is presented here. Each individual has his/her own disease status S, I, or R. Each individual susceptibility could be different depending on the individual properties such as age, social connections, etc. For simplicity, this example assumes that all individuals are mixing together and equally susceptible to infection. For each time step, the following infection process happens. The population wide probability of getting infected could be calculated from the number of infected persons in the current time step. Every eligible person, a random number is drawn from a probability distribution. If the random number drawn is less than the probability of getting infected, the person changes his/her status into I. Otherwise, s/he remains with S status. The progression from I to R happens in the

similar process for each eligible individual on each time step. An output from a single run of a simple individual-based SIR model is presented in Figure 6.

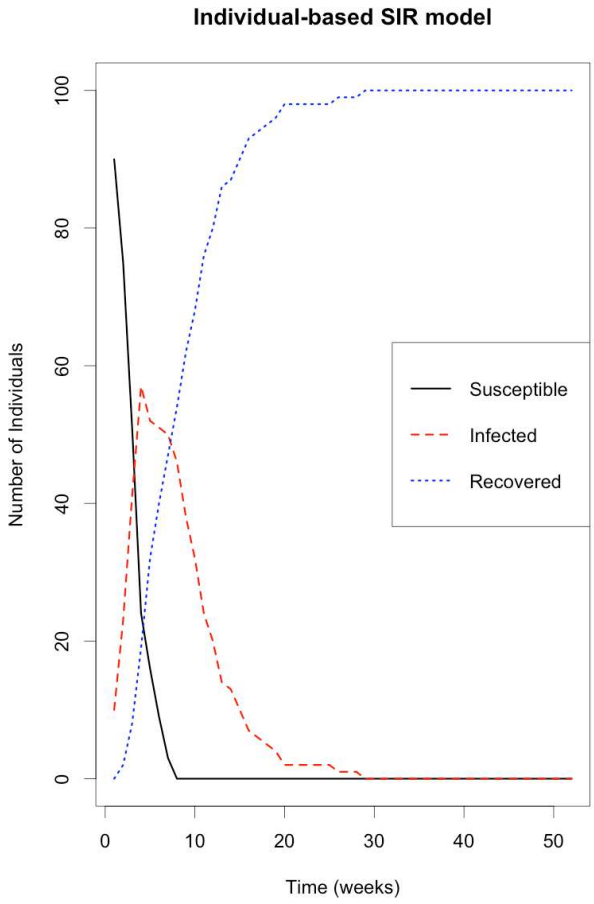


Figure 6: Output from a single run of the IBM version of SIR model.

Because IBMs are stochastic, output from a single run cannot be taken at face value. Results from multiple model runs are pooled together to get a distribution of outcome of interest which is then summarized with an average or median value with a 95% credible interval. Just from this simple example, it would become apparent that the IBMs are

computationally resource intensive. Tracking of hundreds or thousands of individuals with many of their properties over the period of the simulation, and repeating the model run multiple times require a lot of computation and memory. It doesn't help that the IBMs usually include processes that are known to be important but are too complex to include in simpler models. The rise of advanced IBM was only made possible by the increasing computational power of modern computers. Needless to say, formulating and developing an IBM is more complex than most other kinds of models. Dedication and a lot of software programming skills are required to implement models on computers and to observe, test, control and analyse the models.

In Chapter 5, an IBM is presented to explore a reactive malaria intervention.

1.4.5 Mathematical models for malaria

The first mathematical model for malaria appeared in 1908 in Ronald Ross's "Report on the prevention of malaria in Mauritus" where he suggested a threshold of mosquito populations which could be enough to eliminate malaria without the complete eradication of the vector.[52] The equations from Ross's 1910 book named "The prevention of malaria" and their modifications were used for over 40 years to control malaria.[47] Ross's model did not include seasonality nor the metrics for measuring the important components of transmission by mosquitoes.[47,53]

During 1950s, George Macdonald made substantial contributions and expanded the foundations on malaria models set by Ross. Most importantly, Macdonald provided a theoretical justification for using DDT for malaria eradication which became the core intervention in the failed GMEP.[53] Many other mathematicians and scientists also

contributed towards the refinement of the Ross-Macdonald type of models over 7 decades. The Ross-Macdonald model became a de facto basis for a broader theory of mosquito-borne disease transmission and control up to this day.[53]

Smith et al. 2012 comprehensively reviewed the development of Ross-Macdonald type models and noted that there is no canonical formula for "The" Ross- Macdonald model.[53] One of the several versions of Ross-Macdonald style model, published by Anderson and May can be seen in the following sets of differential equations:

$$\frac{dx}{dt} = mabz(1-x) - rx$$

$$\frac{dz}{dt} = acx(1-z) - gz$$

Where x is the prevalence of malaria, m is the ratio of mosquitoes to humans, a is the human blood feeding rate (the proportion of mosquitoes that feed on humans each day), b is the proportion of bites by infectious mosquitoes that infect a human, z is the fraction of infectious mosquitoes, r is the daily rate each human recovers from infection, c is the proportion of infected humans that are infectious, or alternatively, the probability a mosquito becomes infected after biting an infected human, and g is the death rate of mosquito.

Over a century has passed since Ross's first malaria model, and the field has improved greatly albeit the elimination of malaria has not achieved. Especially in the computer science, computers have become powerful enough to solve mathematical equations numerically which would have been impossible analytically. Powerful computers can now simulate far more details of the individual-based models. Many more models have followed to advise the WHO and the national malaria programs to plan for the malaria elimination.

The strategy to contain artemisinin-resistance in the Southeast Asia was advised by a modelling study.[54] The WHO's latest recommendation on RTS,S/AS01 vaccine was made through the favourable cost-effective modelling studies.[45]

1.5 Human mobility in the model

Human mobility is one of the most important factors for disease transmission across different geographical areas for infectious diseases. It is in fact the most attributed cause of malaria transmission.[55] Mobile and migrant populations are hard to track and they are the key challenge to successful malaria interventions in the Southeast Asia countries.[56,57] Inclusion of human mobility in the models depends on the type of model and the data availability. Model 1 in Chapter 2 includes an abstraction of human mobility between the two populations, while Model 2 in Chapter 5 has a much more sophisticated human mobility at the individual level.

In the absence of mobility data, modified gravity models have been used as a standard to represent human mobility between different population centres.[58,59] Gravity models, inspired by Newton's law of gravity, can estimate movement between populations based on their size and distance. However, they might not be good enough to model special cases such as the movement of forest goers in the rural areas. In fact, mobility of forest goers in the rural Southeast Asian countries has never been analysed in fine scale before. This is rectified in Chapter 3, and the resulting forest goers' demographics and mobility data was used to parameterize Model 2 in Chapter 5.

1.6 Composition of households

In IBMs, each individual has his or her own properties that translate to their risks against some infectious diseases. E.g., individuals who are forest goers will have higher risk of contracting malaria. But individuals rarely live alone. Rather, they live in social groups such as families and households which exert great influence on the individuals' wellbeing and reduce or accentuate their disease risks. For instance, a good housing of a well-to-do family will better protect all its members against mosquitos and subsequently against malaria while they are at home. A forest goer might live together with other forest goers in his household based on the occupational inheritance phenomenon. To include the nuances of such social structure, individuals must be grouped into households. The household age-composition problem is tackled in Chapter 4, and the resulting data is used to initialize the population for Model 2 in Chapter 5.

1.7 Thesis aims and outline

- To formulate and develop models of increasing complexity to explore the effective and efficient interventions for malaria elimination with the focus on *P. falciparum* malaria in the Greater Mekong Subregion (GMS) area
    - Model 1: Mass drug administration (MDA) is explored with a two-patch compartmental model that incorporates a simplified version of human mobility in Chapter 2
    - Model 2: Reactive Case Detection and Treatment (RACDT) in a spatially dynamic population that retains a realistic household age-distribution is explored with an IBM in Chapter 5

- To improve methodology and data for malaria models in the GMS
    - To highlight the impact of human mobility in success/failure of public health interventions particularly for malaria (Chapter 2)
    - To analyse the data on mobility of forest goers in the Thai-Myanmar border rural area so that they can be incorporated into the IBM (Chapter 3)
    - To analyse age composition in the households in the GMS (Chapter 4)
    - To develop algorithms to synthesize realistic household age-composition from readily available aggregated census data (Chapter 4)

2 Model 1: Exploring the impact of human mobility on malaria interventions

2.1 Summary

Many public health interventions lead to disruption or decrease of transmission, providing a beneficial effect for people in the population regardless of whether or not they individually participate in the intervention. This protective benefit has been referred to as a herd or community effect and is dependent on the participation of a sufficient proportion of the population. In practice, public health interventions are implemented at different spatial scales, such as at the village, district, or provincial level. Populations, however defined (i.e., neighbourhoods, villages, districts, countries) are connected to other populations through human mobility, and this connectedness can influence potential herd effects.

A two-patch compartmental model was developed with the connectedness between the two patches representing the human mobility between the two patches. The impact of a public health intervention (mass drug administration for malaria) was modelled, for different levels of connectedness between populations that have similar disease epidemiology (e.g., two nearby villages which have similar baseline malaria incidences and similar malaria intervention measures), or between populations of varying disease epidemiology (e.g., two nearby villages which have different baseline malaria incidences and/or malaria intervention measures).

The overall impact of the interventions deployed could be influenced either positively (adding value to the intervention) or negatively (reducing the impact of the intervention) by how much the intervention units are connected with each other (e.g., how frequent people

go to the other village or town) and how different the disease intensity between them are. This phenomenon is termed the “assembly effect”, and it is a meta-population version of the more commonly understood “herd effect”.

The connectedness of intervention units or populations is an important factor to be considered to achieve success in public health interventions that could provide herd effects. Appreciating the assembly effect can improve the cost-effectiveness of strategies for global disease elimination projects.

2.2 Background

2.2.1 Herd effect

Communicable diseases made up 44% and 31% of mortality in low and low-middle income countries as of 2017.[60] Public health interventions have been used for the control and prevention of diseases. Whenever a large enough proportion of the population take up an effective public health intervention for a communicable disease, the transmission of that disease will be reduced and there can be a community-level effect commonly referred to as the “herd effect”.[61] This herd effect provides a protective benefit to all members of a population, regardless of individual participation in the intervention. Conversely, when relatively few individuals in a population participate in an intervention there will be a negligible impact on transmission and, therefore, no herd effect.

2.2.2 Herd effect in different populations connected through human mobility

Herd effects have been documented for several interventions that reduce the transmission potential such as early detection and treatment of pulmonary tuberculosis, mass drug administration (MDA) against lymphatic filariasis [61], insecticide-treated nets (ITN) against malaria infections [62] and recently for MDA against *Plasmodium falciparum* malaria [63]. Herd effects depend on sufficient population adherence to an intervention in order to provide a protective benefit to all individuals in the population. This threshold of participation has been considered in the context of a single population, with little consideration of the existence of meta-populations (groups of spatially separated populations of the same species which interact at some level).[64] Here, how connectedness through human mobility between different populations from different areas influences the effectiveness of the public health interventions was explored, by using malaria elimination as a working example.

Types of intervention for effective malaria control depend on the level of malaria transmission.[39,41,65] In high burden malaria areas, malaria control measures, such as indoor residual spraying (IRS), insecticide-treated mosquito nets (ITNs/LLINs), and ensuring universal access to malaria prevention, diagnosis and treatment aim to reduce malaria prevalence. Population-wide parasite clearance by mass drug administration (MDA) could be used to accelerate the malaria elimination process. Investigation and treatment of residual cases should be done only when the malaria transmission intensity is low enough. Progression from malaria control to malaria elimination is a continuous process with different countries, subnational areas, and communities at different stages on the pathway towards malaria elimination.[41] To address the uneven landscape of malaria transmission in different areas, risk maps can be created through the combination of epidemiological data, geographical information system, and remote sensing of environmental features, followed by a stratification algorithm to allow for better targeting and improved efficiency of malaria

interventions.[41,66,67] Targeting high-risk areas would definitely have a high impact, but when the goal is the global elimination of malaria, the connectedness of the geographical areas through human and/or mosquito movement must also be taken into account. For example, a population movement survey done in the Thai-Myanmar border area found that 44% of participants in one malaria cluster crossed the international border at least once a month.[68] The two countries have different healthcare infrastructures and malaria transmission intensities [68,69] and such cross-border human movement could negatively impact the malaria elimination efforts on one side provided that no similar malaria elimination effort (e.g. mass drug administration, increased access to early diagnosis and treatment) is made across the border. Previous models have also suggested the importance of taking into consideration human movement for efficient deployment of malaria interventions.[70,71]

2.2.3 Model to explore herd effect between two connected populations

A theoretical framework with two interconnected populations, hereafter referred to as "patches", is presented here. Connectedness in the model is the abstraction of human mobility between patches causing humans to contribute to the infectious/ non-infectious pool of individuals in his/her non-native patches. As an example, when a person from patch 1 goes and spends some proportion of their time in patch 2, that person will partially contribute towards the force of infection of a disease in patch 2, either augmenting or diluting it, depending on their disease transmissibility status, and provided that there is a favourable condition for disease transmission.

The goal of this chapter was to evaluate the impact of human mobility on the success of public health interventions that could produce herd effects. In particular, the potential success of anti-malarial MDA deployment in two populations was explored for a range of

different malaria burdens, different intervention coverages, and different levels of connectedness between them through human mobility. This work has operational relevance for targeted anti-malarial campaigns, especially with regard to the spatial unit (household, village, district) that is being targeted. It also has relevance for other public health interventions, all of which have an inherent spatial unit that is being targeted.

2.3 Methods

2.3.1 Model

A two-patch compartmental model was developed using the R software version 3.6.0 [72] with the following packages: deSolve [73], Rcpp [74], and lattice [75]. Each patch had 8 compartments, representing the subgroups with different characteristics such as susceptibility and infectiousness of malaria (Figure 7, Table 2 explains the symbols in the figure). There were two types of susceptible compartments: Susceptible with active antimalarial drug (S_D), and those without drug (S). Likewise, there were two types of recovered compartments: R_D and R. Individuals in the compartments with active drugs were immune to infection until the drugs run out from the body. The infectious compartment was separated into three sub-compartments: I_C represented clinical cases, I_A represented the cases that were asymptomatic, but detectable through microscopy and rapid-diagnostic test (RDT), and I_U represented the cases that were asymptomatic and undetectable through microscopy and RDT.

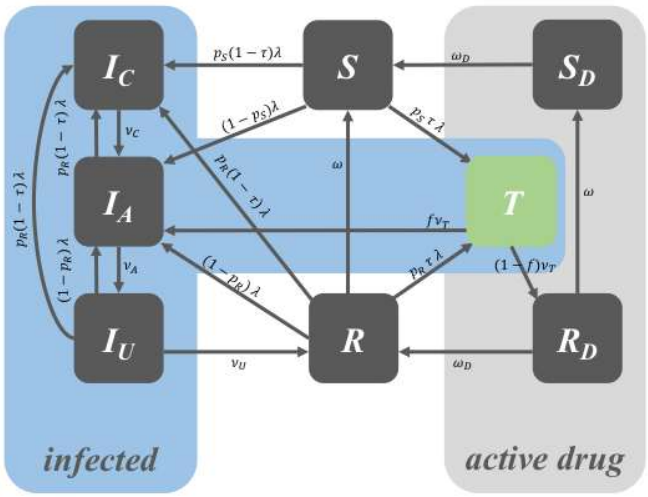


Figure 7: Model's compartments within a patch. Arrows represent the direction of flow between compartments at the corresponding rates shown.

The natural progression of malaria in the model was from S to I to R to S. Coverage and effect of early diagnosis and treatment is modelled through the parameter τ. When MDA was implemented, a proportion of the population under coverage received protection from the disease for some duration (i.e., Individuals from S & R were moved to their respective compartments with active drug, S_D & R_D, where they would remain until the prophylactic effect of the drug was lost. Individuals from I were moved to T) as shown in Figure 8, with the rate $m_i = -\frac{\ln(1-\kappa_{mi})}{3}$ where i is the ith round of MDA and κ is the coverage of that MDA. The descriptions of the parameters in the model can be found in Table 2.

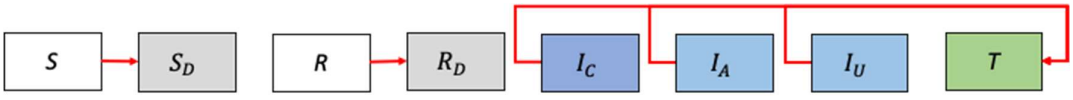


Figure 8: MDA deployment. Red arrows show the flow between compartments in a patch when MDA is deployed

For brevity in describing the two patch model, all sub-compartments of I were combined as an I in the subsequent equation and figure. There was a treatment compartment (T) to accommodate those from the infectious compartment who got treated.

The two patches are represented graphically as two intersecting circles (Figure 9). Human mobility between the two patches was modelled with the level of connectedness between them (C). The force of infection for patch i (λ_i) was defined as equation 1 so that when there was no human movement between the two patches ($C = 0$), λ had a different, independent value for each patch, and when C was 1, λ was identical for both patches.

$$\lambda_i = (1 - C)\,\beta_i \left(\frac{I_i}{P_i}\right) + \frac{C}{2}\frac{(\beta_1+\beta_2)\,(I_1+I_2)}{(P_1+P_2)}$$

Equation 2: Force of infection for patch i

where β is the contact rate between mosquito and human, adjusted by the effectiveness and coverage parameters of insecticide treated nets ($\zeta_{ITN}, \kappa_{ITN}$), I is the combination of I_A, I_C, and I_U compartments, P is the total population in the respective patches.

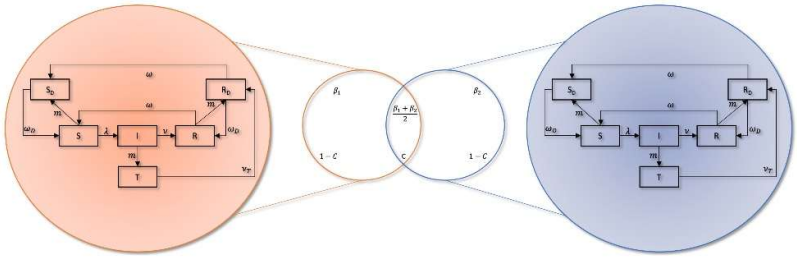


Figure 9: Overview of the two-patch compartmental model. Human mobility is represented by connectedness parameter C. When there is no human movement between the two patches (i.e., patches are isolated, C=0), they share no infections and each individual's risk of infection in a patch is completely independent of that in the other patch. At the other extreme of the connectedness spectrum (C=1), all individuals in the two patches are subject to the same force of infection (λ). β is the effective biting rate adjusted by vector interventions. Zoomed-in areas describe the simplified compartments within each patch- S: Susceptible. I: Infected and Infectious; subgroups of I to capture different detectability and infectiousness are explained in the methods section. R: Recovered. T: Treatment. Compartments with subscript D denote temporary protection by having drugs.

#	Name	Description	Value	Unit	Reference
1	ζ_{ITN}	Effectiveness of ITN (insecticide treated nets), proportion of new infections averted due to ownership of ITN	0.30	proportion	[3]
2	κ_{ITN}	Coverage of ITN	0.70	proportion	-
3	κ_{MSAT}	Coverage of MSAT to prevent case importation from other areas	0.90	proportion	-
4	μ	Birth/death rate	1/69	/year	[4]
5	μ_A	Rate of importation of asymptomatic patent cases from other areas	1	/year/1000	-
6	μ_C	Rate of importation of clinical cases from other areas	1	/year/1000	-
7	μ_{out}	Death rate + emigration rates for malaria cases	-	-	-

8	μ_U	Rate of importation of asymptomatic non-patent cases from other areas	1	/year/1000	-
9	ν_A	Rate of transition from asymptomatic patent state (IA) to asymptomatic non-patent state (IU)	365/60	/year	[5]
10	ν_C	Rate of relief from clinical symptoms in absence of treatment	365/3	/year	[6]
11	ν_U	Rate of transition from asymptomatic non-patent state (IU) to recovered state (R)	365/100	/year	[7]
12	ν_T	Recovery rate after treatment with anti-malarial drug	365/14	/year	[8]
13	ξ_A	Sensitivity of the detecting an asymptomatic, patent (microscopically detectable) case with MSAT	0.87	proportion	-
14	ξ_C	Sensitivity of the detecting a Clinical case with MSAT	0.99	proportion	-
15	ξ_U	Sensitivity of the detecting an asymptomatic, non-patent (microscopically undetectable) case with MSAT	0.44	proportion	[9]
16	ρ_A	Relative infectivity of super-microscopic asymptomatic infections compared with clinical infections	0.55	proportion	[10]

17	ρ_U	Relative infectivity of sub-microscopic asymptomatic infections compared with clinical infections	0.17	proportion	[10]
18	ω	Rate of immunity loss	½	/year	-
19	ω_D	Rate of loss of protection by anti-malarial drug	365/30	/year	-
20	p_R	Proportion of all immune new infections that are clinical	0.20	proportion	[11]
21	p_S	Proportion of all non-immune new infections that are clinical	0.90	proportion	[5]
22	f	Proportion of failed treatment	0.05-0.30	proportion	-

Table 2: Parameter descriptions and values of the two patch model

2.3.2 Simulations

The two-patch model was simulated for several scenarios where one parameter of interest was varied at a time. The outcome metric measured from each patch in each simulation was whether or not a malaria elimination threshold, defined as "less than 1 infection per 1000 population per year" [41], was achieved one-year after the completion of a three-month MDA campaign. Since there were two patches, four outcomes were possible: achieving malaria elimination (a) in none of the patches, (b) in patch 1 only, (c) in patch 2 only and (d) in both patches.

The results were plotted on a two-dimensional surface plot. On the X-axis, the connectedness parameter, C, was increased from 0% to 100% with 1% incremental steps. The MDA coverage in patch 2 was increased from 0% to 100% on the Y-axis, while the MDA coverage in patch 1 is fixed at a particular value for each surface plot. These permutations resulted in over 10,000 simulations, the outcomes of which were summarized in the surface plots (e.g., Figure 10).

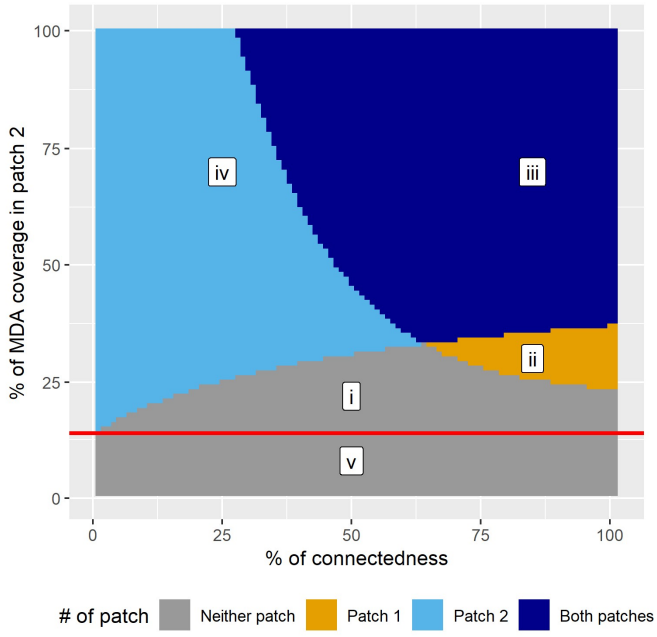


Figure 10: Illustrative example to guide in interpreting the surface plots in the result. The colour of each pixel represents whether the two patches achieve malaria elimination in the simulation, based on the corresponding values from X and Y axes. On the left edge of the plot, where connectedness is 0%, identify the point at which the colour changes from grey or orange to light blue or dark blue. A horizontal line drawn from this transition point (indicated as the red line in the figure) represents the "baseline" or "minimal" MDA threshold required for patch 2 to achieve elimination if it were completely isolated from other patches. As connectedness increases, MDA threshold for patch 2 increases. This deviation from the baseline MDA threshold reflects the assembly effect on patch 2. Lower-case Roman numerals mark different areas of the surface plot. The joint areas "i"+"ii" illustrate a negative assembly effect on patch 2—indicating that, as connectedness with another patch increases, a higher MDA coverage is needed in patch 2 to achieve elimination. Conversely, the joint areas "ii"+"iii" demonstrate a positive assembly effect on patch 1—showing that increasing connectedness with patch 2 (which has high-enough MDA coverage) supports elimination in patch 1. This figure is merely an illustration and it is not an actual model result.

These sets of simulations were repeated for the MDA coverage values in patch 1 from 0% to 90% with 10% increments and for the higher, identical, and lower pre-intervention disease intensities in patch 2.

2.4 Results

2.4.1 Interpreting the assembly effect from surface plots

Figure 10 serves as an example on how to interpret the surface plots. Different colours differentiate four possible outcomes: grey for not achieving malaria elimination in either of the patches (denoted by area "i" and "v" in the figure); orange for elimination in patch 1 only (area "ii"); light-blue for elimination in patch 2 only (area "iv"); and dark-blue for elimination in both patches (area "iii"). The required MDA coverage threshold for malaria elimination in patch 2 can be seen at the transition from the grey or orange area (area "i" or "ii") to the light-blue or dark-blue area (area "iii" or "iv"). For a given malaria incidence in an isolated patch, there exists a specific "baseline" or "minimal" threshold of MDA coverage above which elimination could be achieved. The "baseline" MDA threshold for patch 2 in Figure 10 is the MDA coverage that is required in patch 2 when the connectedness is 0%, indicated by the horizontal red line. Connectedness between the two patches is an indication of how much time humans from each patch spend in the other patch, with 100% connectedness indicating that the two patches are functionally the same patch and 0% connectedness indicating that there is no human movement between the patches (they are isolated). As connectedness increases through increasing human mobility between the two patches, the MDA coverage threshold deviates from the red line. Depending on the MDA coverage in patch 1, increasing connectedness will affect the patch 2 negatively (by requiring higher MDA coverage in patch 2) or positively (by requiring lesser MDA coverage in patch 2). **The beneficial or detrimental effect in a patch (in this case, the change in coverage threshold for successful intervention) due to its connectedness to another patch is hereafter referred to as an "assembly effect".** In Figure 10, "i"+"ii" is the assembly effect affecting negatively for patch 2, where increasing connectedness with patch 1 increases the MDA coverage threshold required for elimination in patch 2. From the

point of view of patch 1, a positive effect is seen in “ii”+“iii” – patch 1 does not achieve elimination when it is isolated, but it does after a certain level of connectedness.

2.4.2 Assembly effects in different scenarios

The simulation results in a collection of thirty plots (ten for each level of MDA coverage in patch 1, repeated for three relative pre-intervention disease intensities between the two patches). Only the last three MDA coverage levels (70%, 80%, and 90%) in patch 1 were focused on here, as the assembly effects in these scenarios are more pronounced for the demonstration purpose. The columns of sub-plots in Figure 11 represent the MDA coverage in the patch 1 (column 1 = 70%, column 2 = 80%, column 3 = 90%); and the rows represent the relative pre-intervention disease intensities in patch 2 compared to patch 1 (top = higher, middle = identical, and bottom = lower).

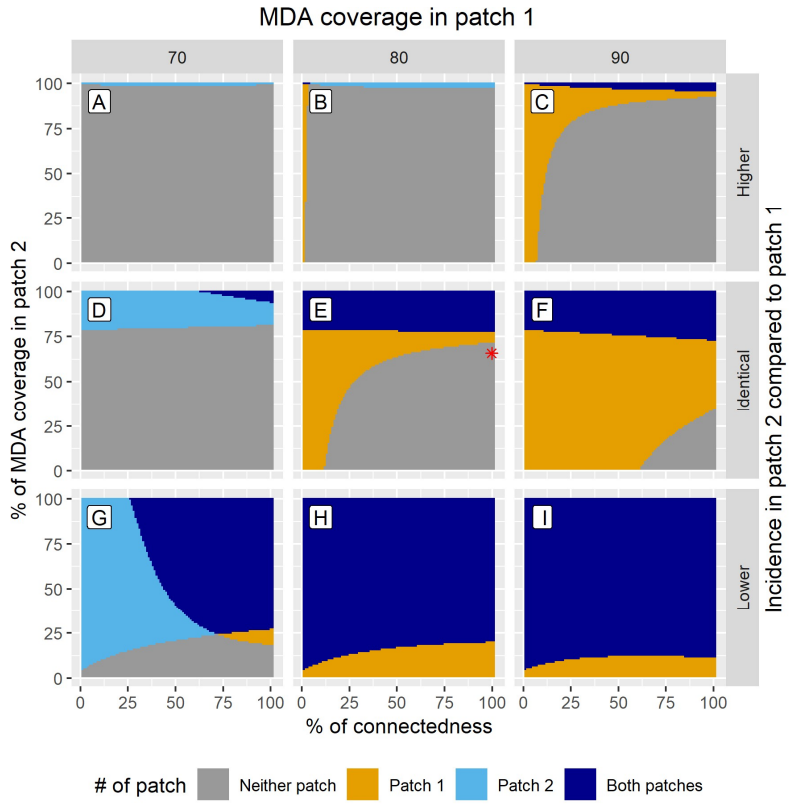


Figure 11: Achieving elimination in two connected patches by varying connectedness between the two populations (x-axis) and MDA coverage in the 2nd patch (y-axis) of each subplot. Columns represent different sets of MDA coverage in patch 1 (70%, 80%, and 90%, respectively). Each row represents the relative incidence level between the two patches. Panel A: No visually distinguishable assembly effect is found. Panel B: Patch 1 should achieve elimination on its own but did not achieve it because of its connectedness to patch 2 (negative effect from the viewpoint of patch 1). Panel C: Patch 1 should achieve elimination on its own, but it will not achieve it if its connectedness to patch 2 is high and if MDA coverage in patch 2 is not high enough. Panel D: Slight negative effect from the viewpoint of patch 2. Panel E: Slight positive assembly effect from the viewpoint of patch 2. The red asterisk represents the combination of parameter values matching the MDA trial implementation described in Parker et. al. Panel F: Increased positive assembly effect from the viewpoint of patch 2 compared to panel E. Panel G, H, I: MDA coverage in patch 1 is 70%, 80%, and 90% respectively. Patch 2 has lower pre-intervention incidence, thus its baseline MDA coverage threshold is low. From the viewpoint of patch 2, there is always a negative assembly effect but its magnitude diminishes as the MDA coverage in patch 1 is increased from 70 to 90%.

2.4.2.1 Between two patches with the same pre-intervention incidence

In the middle row of Figure 11, both patches have an identical pre-intervention incidence, requiring a baseline MDA threshold of 78% coverage to achieve elimination (when the patches are not connected). In Figure 11:panel D, there is a negative assembly effect for patch 2 (the grey area above the baseline MDA threshold) because of the increasing connectedness with patch 1, which has a relatively low MDA coverage (70%). However, the increasing connectedness is beneficial to patch 1 (a positive assembly effect). Despite patch 1 having 70% MDA coverage, and not being able to achieve elimination on its own, the increasing connectedness with patch 2 (when patch 2 has more than enough MDA coverage for itself e.g., 94% MDA coverage), makes elimination still attainable in patch 1 (dark blue triangle at the upper right corner).

An opposite effect is seen when patch 1 has higher MDA coverage (80% and 90%) than is necessary to achieve elimination on its own (Figure 11:panels E and F). In this scenario, patch 2 experiences the assembly effect positively, indicated by the extension of the dark blue areas below the baseline MDA coverage threshold of 78%. However, patch 1 experiences the assembly effect negatively; as connectedness increases, elimination in patch 1 is not predicted to occur for low MDA coverage in patch 2 (grey area in the lower right corners) because less-than-optimal coverage in patch 2 prevents patch 1 from achieving elimination at those levels of connectedness.

When the pre-intervention transmission intensities are the same in the two patches, the resulting assembly effects are purely due to differences in intervention coverage. To quantify the total assembly effect in patch 2 in each plot, the area between the "baseline" MDA threshold line (the red line in Figure 10) and the diverging MDA threshold for

increasing levels of connectedness (i.e. area “i”+“ii” in Figure 10) was integrated. The total effect is assigned positive if it is beneficial to patch 2, and it is assigned negative otherwise.

Figure 12 displays how the total assembly effect in a particular patch is modulated by its connectedness to the other patch for different relative incidence ratios. The total assembly effect in patch 2 increases with increasing intervention coverage in patch 1 (black dots in Figure 12). The switch from negative to positive total assembly effect occurs at the "baseline" coverage threshold for the particular disease intensity shared by both patches (78% coverage in this case).

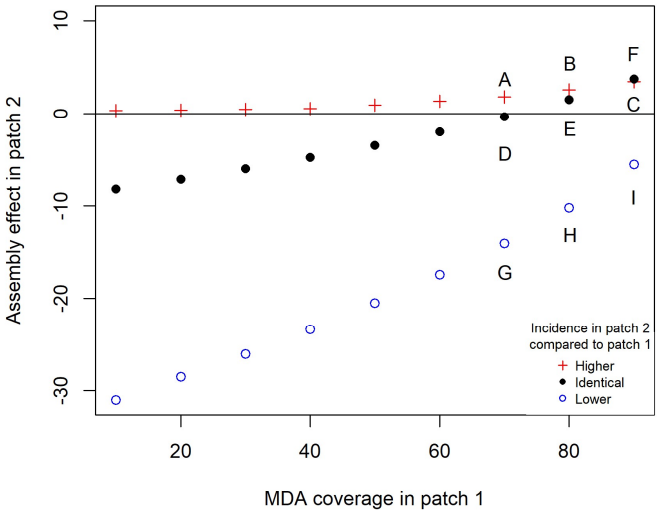


Figure 12: Total assembly effects in patch 2 where relative incidence is higher, identical, or lower compared to patch 1. The value of total assembly effects on the Y-axis was calculated by integrating the assembly effects in patch 2 over all levels of connectedness with patch 1. A, B, C, D, E, F, G, H, and I represent the assembly effects of respective panels in Fig. 3. Blue circles are analogous to the assembly effects in non-hotspots for different coverage in the hotspot. Red crosses represent the assembly effects when incidence in patch 2 is so high that MDA is not an effective intervention (i.e., Nearly 100% MDA coverage is required to achieve elimination in patch 2). Black dots represent the assembly effects when the two patches have identical incidence (i.e., assembly effects are the same from the point of view of both patches).

The model's prediction was compared against results from an MDA trial described in Parker et al. [63] where a village failed to achieve elimination presumably due to a cluster of non-participation in the MDA. This scenario was modelled as a set of two contiguous patches with 100% connectedness and with identical incidence. One patch received approximately 80% MDA coverage and the other 64%, with the latter representing the non-participation cluster. The model accurately predicted that neither patch would achieve elimination (the red asterisk in Figure 11:panel E).

2.4.2.2 Between two patches with different pre-intervention incidences (Hotspot vs non-hotspot)

In the bottom row of Figure 11, patch 2 has a 25% lower pre-intervention incidence compared to patch 1. This is analogous to a scenario where a low-incidence community (non-hotspot: patch 2) is connected to a high-incidence community (hotspot: patch 1). For this example, the following definition of malaria hotspots is used: "*geographical areas within a wider area of transmission in which the transmission intensity is significantly higher than the average level in the surrounding area of that setting and are widely observed in malaria-endemic regions*" [76]. When in isolation (no connectedness between patches), the MDA coverage threshold for elimination is very low at 5% for the non-hotspot, whereas it is 78% for the hotspot.

When MDA coverage in the hotspot is slightly below its required threshold for elimination (70% rather than the required 78%, Figure 11:panel G), the assembly effect impact negatively for the non-hotspot and positively for the hotspot (areas similar to negative

assembly effect for patch 2: “i”+“ii” and positive assembly effect for patch 1: “ii”+“iii” respectively in illustrative Figure 10).

This suggests that when MDA coverage in the non-hotspot is high, and when the connectedness between hotspot and non-hotspot is high, elimination could be achieved in both patches despite the hotspot having less-than-optimal MDA coverage. For instance, when there is 60% connectedness, MDA coverage over 30% in the non-hotspot is predicted to result in elimination in both patches.

In panels H and I of Figure 11, the hotspot has an adequate MDA coverage at 80% and 90% respectively. In those scenarios, the hotspot is predicted to always achieve elimination, regardless of the level of connectedness and the value of MDA coverage in the non-hotspot.

As seen in Figure 12, non-hotspots (blue circles) will always experience a negative total assembly effect. However, the magnitude of the negative total assembly effect decreases with increasing coverage in the connected hotspot. The opposite is true for the positive total assembly effect gained by the hotspot (i.e., it increases with increasing coverage in the connected non-hotspot as seen in the Appendix Figure 2). These trends suggest that the difference in transmission intensity is the main determinant of whether the assembly effects would be observed positively or negatively.

In Figure 11:panel I, the required intervention threshold for the non-hotspot plateaus between 40% and 80% of connectedness. Further increase in the connectedness decreases the required intervention threshold slightly.

2.4.2.3 When intervention is ineffective for one of the patches

An intervention may not be appropriate if the disease intensity is too high e.g., MDA may not work in a high-transmission setting unless a very high MDA coverage is achieved. This scenario was simulated in the first row of Figure 11 by setting patch 2 as a high-transmission setting. In isolation, patch 2 would require almost 100% of MDA coverage, while patch 1 would require more than 78% coverage of MDA for elimination to be attainable. As a consequence of being connected to patch 2, the prospects for elimination in patch 1 would be greatly diminished (large negative assembly effect for patch 1 represented by grey areas in Figure 11:panels B and C).

2.5 Discussion

In a single patch system, the success of an intervention depends on the pre-intervention disease intensity and the coverage of the intervention, provided the intervention is efficacious and its coverage is maintained for an adequate period. In the two-patch connected system, whilst those metrics are still relevant, the level of connectedness between the two patches through human mobility is another key determinant of the intervention's success. The results illustrate how connectedness can be advantageous to one patch, while potentially being disadvantageous to the other. This effect is designated as the *assembly effect*. An assembly effect can be seen when connectedness is as low as 1%. Its magnitude and direction of effect depend on transmission intensity and intervention coverage in the adjacent area.

When connected patches have identical pre-intervention disease intensity, but different intervention coverage, the required threshold for successful intervention in each patch will

equilibrate with increasing connectedness. In other words, the required intervention threshold in each connected patch approaches some average threshold values between them as their connectedness level is increased. The assembly effect would be disadvantage for a patch (by having to increase its required threshold) when it is connected to another patch that does not have enough intervention coverage to control its transmission intensity. At the same time, a positive assembly effect (decrement in the required threshold) may occur in the latter patch depending on how connected they are. Therefore, if one patch achieves a higher-than-optimal coverage of intervention, and its connected patch has a less-than-optimal coverage, it is still possible to attain a successful outcome in both patches, provided they are connected enough. This has implications for public health interventions in locations with low adherence. In settings where multiple communities or populations are highly connected, as long as a certain number of the populations achieve higher-than-optimal coverage, the remaining populations can have less-than-optimal coverage.

As countries move towards disease elimination and as disease transmission intensity distributions over space become extremely patchy [41], it becomes increasingly important to target disease hotspots with adequate intervention coverage. The results suggest that to achieve elimination, adjacent non-hotspot areas should not be left without interventions. Having some intervention coverage in the adjacent non-hotspots is also helpful when the optimal intervention coverage could not be achieved in the hotspots.[77] For highly connected patches, hotspots with sub-optimal intervention coverage are predicted to have a significant positive assembly effect because of the connectedness to the non-hotspot patches that have modestly increased MDA coverage above its required threshold (Figure 11:panel G and Appendix Figure 2).

Many public health interventions that reduce transmission and target populations that are not in complete isolation will likely also result in an assembly effect. By considering the

following: connectedness between populations, overall disease intensity, and adherence to the public health interventions being used, communicable diseases can more effectively be controlled and eliminated.

2.5.1 Implications for the focal malaria interventions

The WHO has recommended MDA as a potential tool to accelerate malaria elimination but recommended its deployment only with high coverage and only when core malaria interventions are already in place[41]. Even through improvement and maintenance of core malaria interventions, some patches with relatively higher incidence (hotspots) and relatively lower incidence (non-hotspots) could persist. In such a scenario, it would be tempting to just target malaria hotspots with MDA. The results from this study suggest that targeting only malaria hotspots may not be enough. It is often challenging and resource-intensive to achieve high coverage for MDA [78,79], and the imported asymptomatic infections from the connected non-hotspots could refuel transmission [80]. Therefore, when targeting hotspots in these scenarios, reinforcement of interventions in adjacent or connected non-hotspots would benefit the hotspots because of the positive assembly effect and improve the chance of a successful elimination campaign. We will discuss more about deploying malaria interventions in the Discussion chapter (Chapter 6) of the thesis.

2.5.2 Limitations

This model was developed as a theoretical framework to define the concept of the assembly effect in a general sense. There were many assumptions in the model structure and parameter values used. The way MDA was modelled in the compartmental system

may not be an accurate representation of a real-world MDA. The model has so far been validated on a single scenario. The time point for measuring the outcome was arbitrarily set as one year after the completion of MDA. Results will vary depending on where this time point is set.

2.6 Conclusions

For malaria elimination, improving and maintaining core malaria interventions is the first step, which could be followed by an acceleration to elimination. In implementing accelerating activities such as MDA, targeting malaria hotspots alone may not be optimal. Having positive assembly effects on the hotspots by additionally implementing MDA with some effective coverage on their connected non-hotspots will lower the required MDA coverage threshold in the hotspots and thus increase the feasibility of malaria elimination.

Assembly effect is a meta-population version of the herd effect and it occurs when populations of potentially different disease intensities and/or intervention coverages are connected through human mobility. The ultimate impact of an intervention in an area depends on how well it is connected with neighbouring areas. Information on the level of connectedness between populations will inform efficient control and elimination strategies. In the next chapter, we explore the human movement and space use of rural villagers in the Thai-Myanmar border area.

3 Movement patterns of farmers and forest goers along the Thailand-Myanmar border

3.1 Summary

Human mobility play an important role in infectious disease epidemiology and ecology. Movement into geographic spaces with high transmission can lead to increased risk of acquiring infections. As evident in the previous chapter, human movement and connectedness can impact the success of malaria interventions. Most fine scale studies of human travel patterns have been done in urban settings in wealthy nations. Research into human travel patterns in rural areas of low- and middle-income nations are useful for understanding the human components of epidemiological systems for malaria or other diseases of the rural poor. The goal of this chapter was to empirically measure human mobility in the remote and rural setting using GPS loggers, and to quantify differing mobility patterns by age, gender, and seasonality.

In this study we recruited 50 rural villagers from along the Myanmar-Thailand border to carry GPS loggers for the duration of a year. The GPS loggers were programmed to take a time-stamped reading every 30 minutes. We calculated daily movement ranges and multi-day trips by age and gender. We incorporated remote sensing data to assess patterns of days and nights spent in forested or farm areas, also by age and gender.

We found that older adults travelled farther distances than younger adults. Adult males of age between 20 and 40 were found to frequent farms and forest the most and spend more nights there compared to other age groups and gender. We calculated how often a forest goer goes into the forest and once in the forest, how the probability of staying overnight

decreased over the subsequent days. The results of this study are useful for informing the human movement in individual-based models of disease transmission.

3.2 Background

Human mobility is important with regard to infectious disease epidemiology and ecology.[81,82] Infectious diseases are heterogeneously distributed across landscapes. Individuals may be exposed to greater risk of acquiring infection if they move through transmission hotspots. Infectious individuals who travel may disperse pathogens across the landscape. Healthcare facilities are also heterogeneously distributed across landscapes, with ramifications for individual, household, and community access to diagnosis and treatment. Generally speaking, individuals who must travel long distances or through difficult terrain in order to seek diagnosis or treatment are less likely to receive adequate treatment.[83-85]

As early as the 1950s, human mobility was recognized as one of the most important factors for disease elimination and eradication.[86] We also saw the influence of human mobility on the success of intervention for the malaria intervention in the previous chapter. A growing number of research projects, some focused on health, are recording human movement patterns.[87-92] There have been attempts to map human mobility in the rural Thailand border areas to delineate and intervene the risks of malaria.[68,93,94] These projects can be broadly divided into those that are based on questionnaires/interviews and those that are based on empirical measurements (GPS devices, mobile phones, tweets, etc.) All approaches have strengths and weaknesses.[95] Interview/questionnaire-based approaches are prone to recollection bias and some movements may be unreported because of their nature (for example, if movements are made for illegal purposes or to places that participants don't want to discuss/report).

Mobile phone records provide a source of movement information across broad swaths of many populations.[96] However, the movement data are limited to the resolution of mobile tower density, and mobile phone towers are not evenly distributed across landscapes (they tend to be clustered in urban settings). There is bias in who owns and uses mobile phones as well [97] and mobile phone records will not allow for fine-scale mapping of the routes travelled in between locations.[98] Wearable GPS devices offer extremely detailed data, but are labour intensive and dependent on volunteer cohort members. However, as the devices have become more compact (increasing wearability) and have become more affordable, their use is increasingly common.[92,99-105]

The main goal of this chapter is to measure human movement patterns, including how they vary seasonally, among a cohort of participants. The results of this work have implications for further research in this region with regard to targeted public health interventions, normal travel patterns and related exposure to different environments, for individual risk of infection by various diseases (e.g. SARS-CoV-2, malaria, melioidosis), and with regard to human disease ecology. The resulting data can also be useful for calibrating human movement patterns of individuals in an individual based modelling system.

3.3 Methods

3.3.1 Context of the study area

The study area is on the Thailand-Myanmar border. Participants were recruited from clinics that serve rural, mostly underdeveloped, and low population density communities. Most participants were of the Karen ethnic group. Villages were made up of a few dozen of mostly multigeneration families living in stilt houses made of wood and thatched roof.

Villages didn't normally have schools, clinics, or sanatory toilets. The houses are normally located along the main dirt road of the village. The dirt roads then continued to connect to other villages and small towns through a hilly and rugged terrain with occasional watersheds and river basins which made traversing difficult, especially during the rainy season.

Villagers made their living mostly though agriculture, but they have to undertake various types of jobs throughout the year for their subsistence. They developed land in and around their villages into farms to cultivate rice and vegetables. They farm poultry and pigs under their stilt houses. Some villagers go into the forests for hunting, or for foraging wood, and to collect wild edibles. They would go to the farms and forests overnight occasionally, and sleep without much protection from mosquitos and insects.[106] We focused on farms and forests as places of interest in this study since apart from their homes, farms and forests might be the places the villagers spend significant amount of their time while being vulnerable to infectious diseases such as malaria.

3.3.2 Data

The study period began in March 2017 and ended in February 2018 and aimed to recruit 50 participants for a one-year duration of time (ClinicalTrials.gov Identifier: NCT03087214). The study size was purposive as this was an exploratory pilot study. Prior to the study beginning we held community engagement meetings with community elders in the Tak Province Community Ethics Advisory board (T-CAB) to explain the project. The study locations were selected because of community enthusiasm to participate and operational feasibility. Participants were recruited from 10 villages (the lowest level of administrative division in Myanmar) near two clinics on the Thailand-Myanmar border: Wang Pah and Maw Ker Tai Clinics (Figure 13). These clinics primarily serve migrant and cross border

populations and have connections to village health workers in nearby villages. We reached out to village health workers in the nearby villages to explain the project and to ask if they could help us recruit participants from their respective villages. Participants were recruited and interviewed at the respective clinics, and there were no house visits.

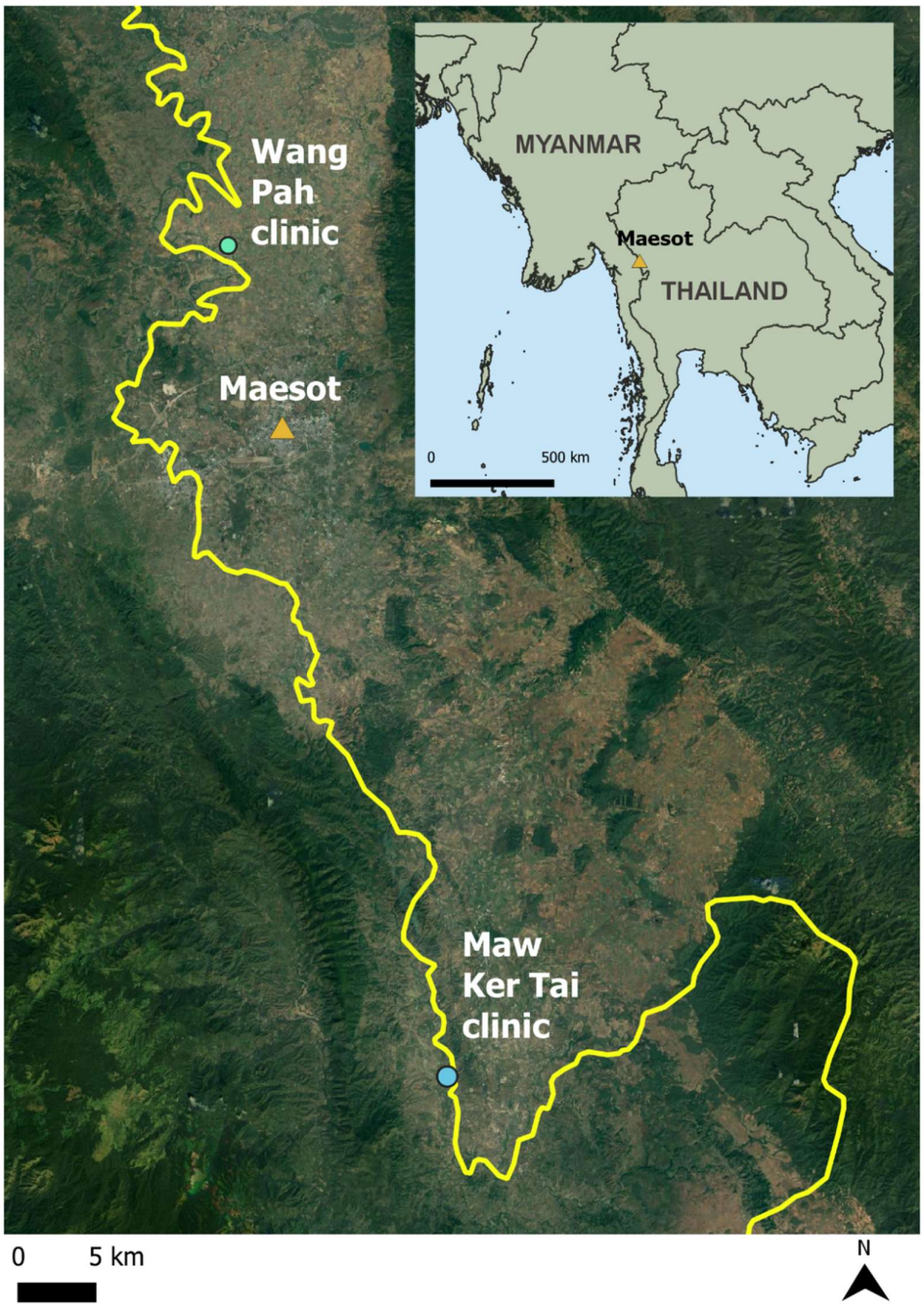


Figure 13: Study site and location of clinics that were used for recruitment

The study targeted individuals who were 18 years of age and above from the Karen or Burmese ethnic groups, who stated that they would be able to keep track of the GPS device,

who were capable of walking outside of village boundaries at recruitment, and who were willing to provide written consent to the study. As incentives, participants were provided with a waterproof handbag at the beginning of the study, a headlamp in the middle of the study, and a jacket at the end of the study. The total cost of incentives per person was less than 10 GBP.

Upon recruitment, the age and gender of each participant was recorded following receipt of written informed consent (in Karen language). Participants were asked to carry i-gotU GT-600 (46x41.5x14mm) mobile GPS devices for the study. They have a reported average location error of less than 10 meters [107]. They were programmed to take a reading every 30 minutes. The devices are equipped with motion sensors and were set to go into a dormant mode if they sat still for longer than one hour, and to resume taking GPS readings upon detection of movement. Devices were also set to take readings at one-minute intervals if the device was moving $\geq$ 15km/hour (travelling by vehicle rather than walking).

Field managers (one per clinic) met with study participants each month. During these meetings participants were questioned about their continued willingness to participate in the study by the field managers, their general movement patterns during the previous month and with regard to any illnesses. A newly charged GPS device was given to each participant (two GPS devices were devoted to each study participant, total of 100 devices used) during each of these monthly meetings and the GPS logger that had been carried during the previous month was collected for re-charging. The GPS device batteries last roughly 1 to 1 ½ months.

The data were transferred to a computer and stored in an encrypted folder with a unique code for each person to maintain security and anonymity. Separate data files were combined to obtain aggregated, longitudinal data for each participant. QGIS

(https://www.qgis.org/en/site/) version 3.4.9 was used to generate study location maps and to visually explore the raw data. R statistical software version 4.0.3 was used for the data processing and analysis [72,108], using the "sp", "rgdal", "raster", "proj4", "reshape" and "ggplot2" R packages [109-114]. GPS coordinates, which were originally recorded in 1984 World Geographical Coordinate System (WGS 84), were projected to UTM zone 47N to perform geographical calculations.

Land cover types (farms and forests) were classified manually (by hand) using satellite imagery from Google Earth (version 7.3.3.7786). Farms could be differentiated from forests by presence of human intervention on the vegetation cover e.g., vegetation cover in the farms were in more or less neatly arranged rows/columns. While formal ground truthing was not done after categorization, the locations of farms and forests do correspond to our experiences on the ground in these villages.

3.3.3 Analysis

Our analyses focused on quantifying daily movement ranges, multi-day trips, and time spent in farm or forest areas across population strata.

The last GPS point of the day between 6pm to 12 midnight was considered to be the location where an individual spent the night. The median center of these points was assumed to be the individual's home location. A buffer with 266 meters radius (which is the standard distance deviation of accuracy of the GPS device placed inside a bag inside a house; details in Appendix Figure 3) was created around each home to create a polygon (a GIS object with a series of x and y coordinate pairs that represents an enclosed area on a

map) for home area. Polygons for the farms and forests were manually classified using satellite imagery from Google Earth.

As a proxy for how far people move each day, we calculated the maximum daily Euclidian distance, which is the furthest Euclidian distance a person was away from the location he or she slept the previous night. Multiday trips away from home were identified when the minimal daily Euclidian distances were more than 266 meters from the individual's home location consecutively for two or more days.

The Wilcoxon rank-sum test was used to compare the distributions of maximum daily Euclidian distances. A negative-binomial generalized linear mixed-effects model was used to investigate potential associations between the total number of nights spent in the farms or forests (response variables) and other characteristics such as age group, gender, and season (exploratory variables). As there were multiple observations per individual (for each time step), a random intercept was used for individuals.

Utilization of places (home, farms or forests) for each person was estimated using two different approaches. The first method was by checking whether more than two temporally-consecutive GPS points of a person fall within a polygon designated for the person's home (for this particular measurement, the polygon is a circle of radius 266 meters around the person's home location), farms, or forests on each day. This is equivalent to checking if a person spent at least an hour within the same polygon. For each participant, the number of days spending in each category of place (home, farm, forest) was divided by the total number of days participated during the study period to obtain the proportion of being at the respective places.

The second method estimated the utilization of places by a biased random bridge (BRB) technique.[115,116] Unlike the prior method, BRB takes the activity time between successive

relocations into account and models space utilization as a time-ordered series of points to improve accuracy and biological relevance while adjusting for missing values. BRB estimate the probability of an individual being in a specific location during the study time period and can be used to estimate home range (the area where individuals spend a defined percentage of their time).

To parameterize BRB models for each individual, we considered points collected more than three hours apart to be uncorrelated. However, the two temporally-consecutive points that are deemed uncorrelated by the prior cutoff, may in fact be correlated (e.g., when individuals go to sleep for more than three hours in a single location). Without manually adding points between them, this method will underestimate the usage of homes. An individual is considered stationary when the distance between two consecutive points is less than 10 meters. The minimum standard deviation in relocation uncertainty is set at 30 meters. For each individual, estimation for the usage of different places was done for the whole study period (i.e. for the duration of his/her contribution) and for each season as described below.

In Central and Southern Myanmar, the monsoon rain starts in mid-May and ends in mid-October.[117,118] Therefore, we split the data on 15th May 2017 and 15th October 2017, and the period between the two dates was regarded as the “rainy season”. Mid-October to mid-March is the “cool and dry season”, mid-March to mid-May is the “hot and dry season”. Combinations of the two dry seasons had been used simply as the “dry season” in some of the analyses.

The forest goer's probability of sleeping in the forest for each consecutive day was estimated using a Kaplan-Meier's survival analysis.

3.3.4 Ethics statement

Approval for this research project was obtained from the Faculty of Tropical Medicine Ethics Committee, Mahidol University (TMEC 17-007); by the Oxford Tropical Research Ethics Committee (OxTREC reference: 503-17); and by the Tak Province Community Ethics Advisory Board (T-CAB reference: TCAB-04/REV/2016). All participants provided written informed consent in the Karen language.

3.4 Results

A total of 50 persons participated for at least two seasons during the one-year study period. The age and gender distribution of the participants can be found in Table 3. Female participation was low (n= 10). Efforts were made to increase female recruitment but many women declined, stating that they did not normally leave their homes or villages and therefore thought they would not be interesting for the study. Most participants (29 out of 50) were in the 20-40 age group. Individual duration of participation differs between participants (Figure 14). The mean percentage of days GPS points were actually observed for the participants over the study is 86.17% (median: 91.21, IQR: 79.54-96.93).

Age group	Less than 20	20-40	40 and above	Total
Male	7	24	9	40
Female	2	5	3	10

Total	9	29	12	50

Table 3: Age and gender distribution of the participants

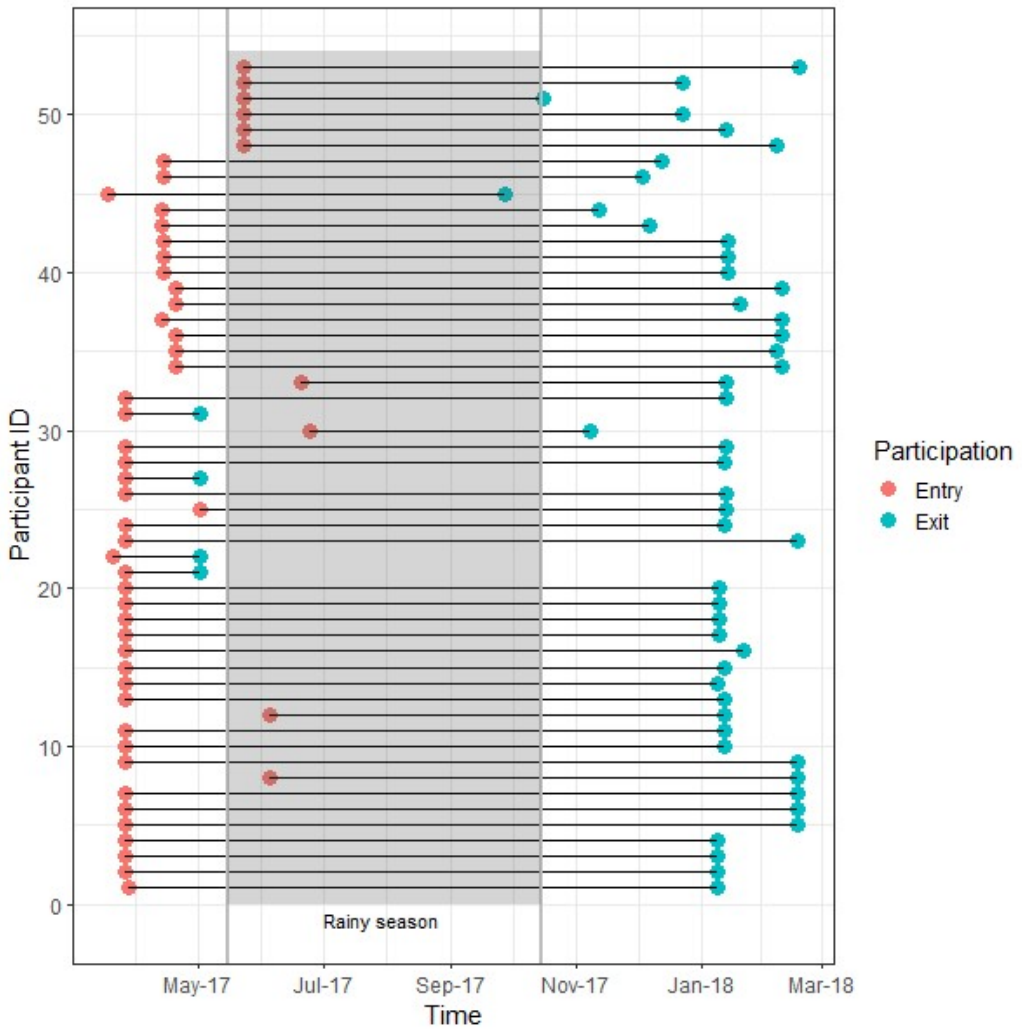


Figure 14: Duration of participation for each person, over the study period

3.4.1 Daily movement ranges

The violin plot of the maximum daily Euclidian distances travelled in kilometres in $\log_{10}$ scale (Figure 15) shows that there is a bimodal distribution for all three age groups. The

violin plot is a hybrid of kernel density plot and box-plot with the axes flipped that is particularly used to describe data with multimodal distribution. In the figure, the vertical axis is the distance value in kilometres with the smallest value at the bottom and the horizontal axis shows the density value. The heights and peaks in the following results refer to the width/broadness of the violins in the horizontal axis. The first peak was between 0.01 to 0.1 kilometres (10 to 100 meters) and the second peak was between 1 and 10 kilometres. The relative heights of the two peaks differ in different age groups. For under 20s, the first peak is over 20% higher (i.e. they have higher proportion of daily maximum distance close to where they were the previous night) compared to the second peak. The difference between the two peaks in the other two age groups is less than 10%.

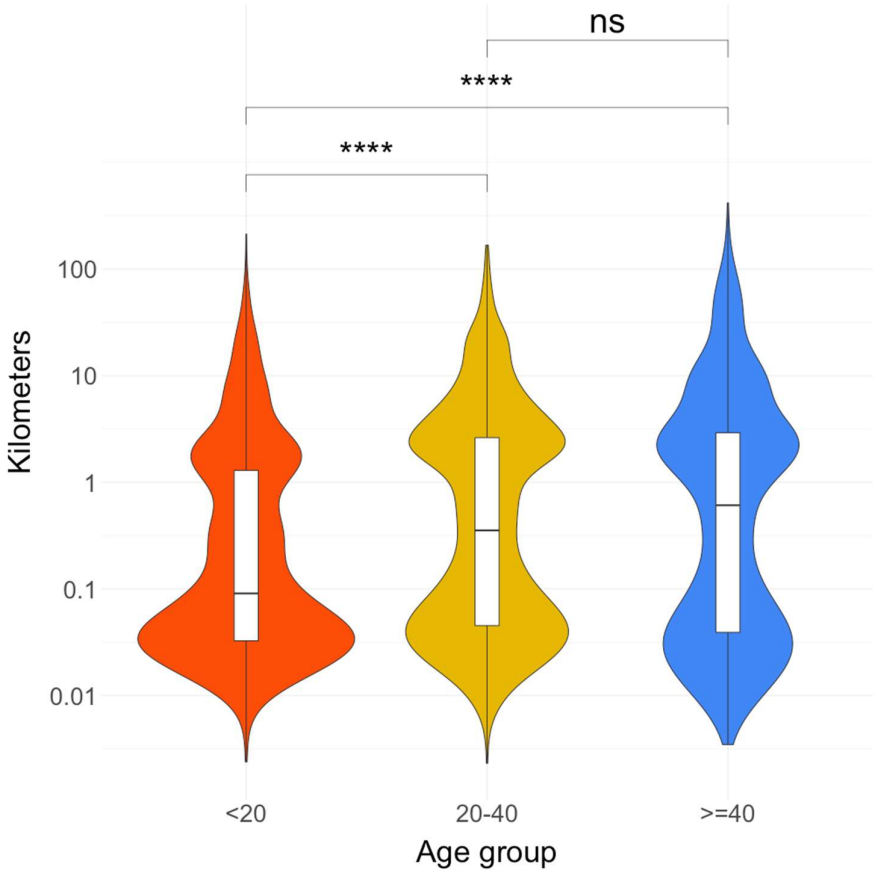


*Figure 15: Maximum daily Euclidian distances travelled by participants in kilometres. Distance was calculated from the location a person was at the end of the prior night (most often, this location is their home location). Wilcoxon rank-sum test results are shown on the top of the lines connecting the age groups chosen for the tests. "ns" represents a p-value of > 0.05. "**** represents a p-value of <= 0.0001.*

The Wilcoxon rank-sum tests provided evidence that 20-40 and over-40 age groups have greater maximum daily Euclidian distances away from home compared to under-20 age group on average. Further disaggregation of this data by gender, and age group can be found in the Appendix Figure 4.

3.4.2 Multiday trips

Participants may make trips that would last several days, either because their destination could not be reached within a single day or because they stayed at their destination for several days (e.g. staying at a farm hut). Using a buffer radius of 266 meters around their home GPS points as their home locations, we calculated the number of consecutive days they spent away from home. Aside from two participants (an over-40 male and an under-20 female), all other participants had at least one trip with more than two consecutive days away from home during their participation period. Trips of less than 10 consecutive days are the most frequent among the participants. There are male outliers of over 20-years old (n=6) who took shorter consecutive day trips (2-5 days) over 10 times. Making trips of over 10 consecutive days was relatively uncommon, but 21 participants still made at least one trip of over 20 consecutive days away from home. Details are available in Figure 16.

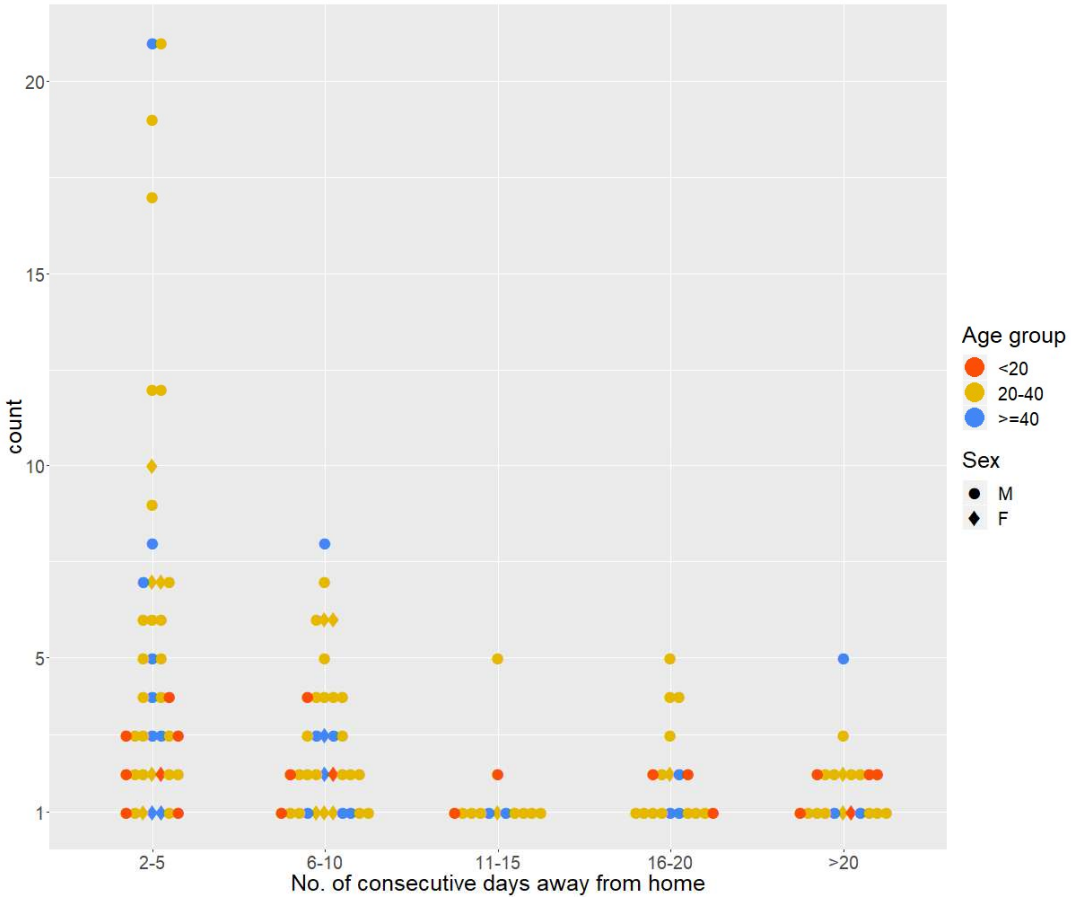


Figure 16: Multiday trips made by the participants

3.4.3 Days spent at the forest or farm

For each participant, we identified the number of days spent at farms, forests, or at one's home, and looked for an association between farm visits and forest visits. Here we assumed that having at least two GPS points in the polygon of a particular place constitutes using the respective place for that day, and that a person can be at various types of places in a single day. We found that if a person spent a higher proportion of days at the farms, she or he will likely spend a lower proportion of days at the forests, and vice versa, even though both being at the farms and being in the forests are possible on the same day.

Figure 17 shows the distribution of the proportion of the number of days being at the farms, forests or home for different age groups. All participants were found to be at their respective home for the majority of days. Compared to other age groups, the 20-40 age group had a higher proportion of time spent in the forests. The under-20 group had the highest proportion of time spent in the farms on average, followed by the 20-40 age group.

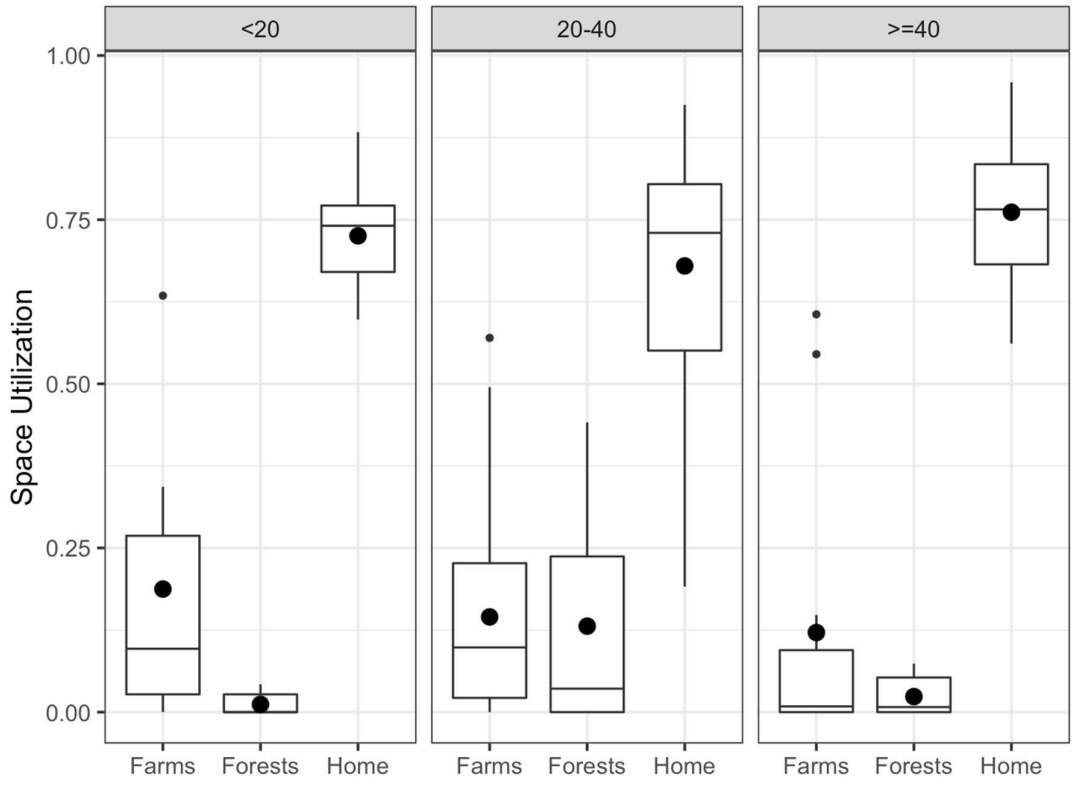


Figure 17: Utilization of the farm, forest, and home (calculated as a proportion of number of days being at the respective places) over the participation period for different age groups. The bigger dots represent the mean values, while the smaller dots represent the outliers.

3.4.4 Time spent at the forest or farm (using BRB algorithm)

We also combined the geographic information of farms and forests with the place utilization estimated from a biased-random bridge (BRB) algorithm, and calculated the utilization of each specific place over the study period Appendix Figure 5. An example of the place utilization of a person can be seen in Appendix Figure 6. On average, participants in the under-20 age group spent 20.0% and 2.2% of their time in farms and forests, respectively. For the participants from the 20-40 age group the percentages are 7.6% and 7.4%, and for those in the over-40 age group, the percentages are 7.2% and 3.8%, respectively.

3.4.5 Night spent at the forest or farm

Being in the farms and forests at night might impose increased risks of diseases such as malaria because of potential exposure to important mosquito vector species (i.e. *Anopheles dirus)*. As seen in Figure 18, we looked at the total number of nights participants spent in the farms or in the forests. Two female participants (20% of females) spent at least a night in the farm compared to 22 male participants (55% of males). As for spending at least a night in the forest, there were 21 males and only one female. Most participants in the 20-40 age group spent at least one night in the farm (18 out of 29, 62%) and in the forest (16 out of 29, 55%) whereas fewer than 35% of participants from under-20 and over-40 age groups spent a night in such places.

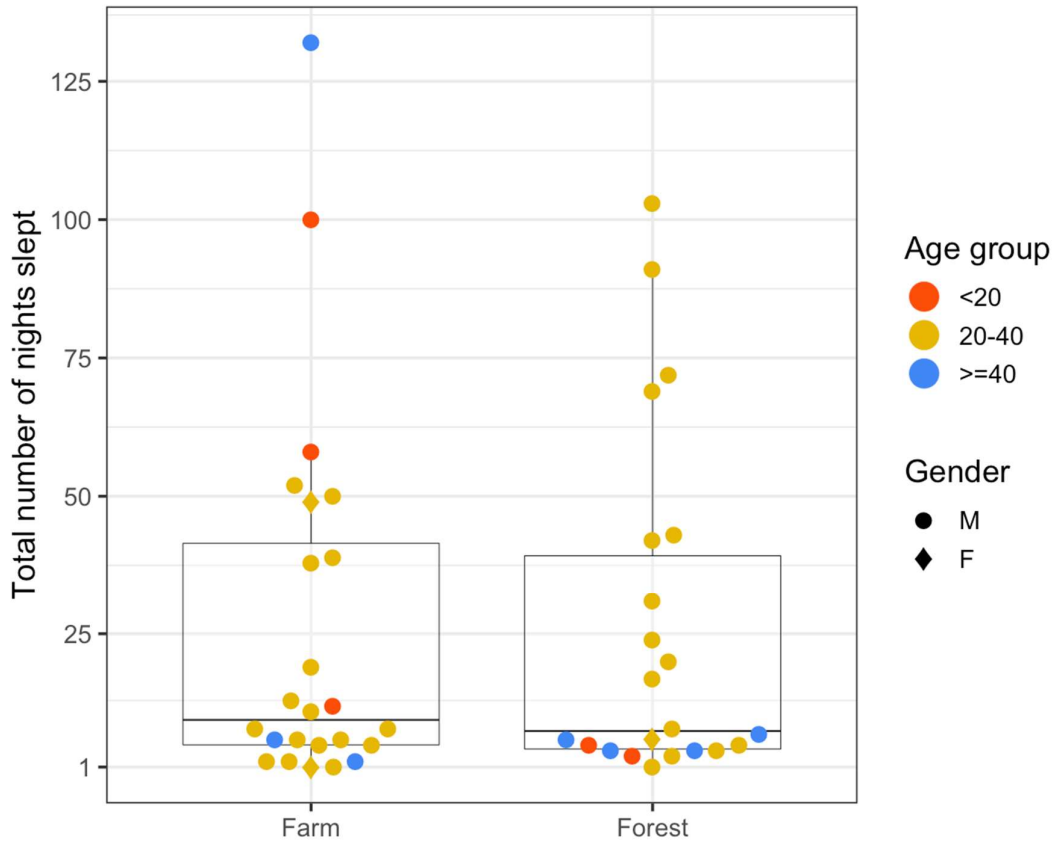


Figure 18: Total number of nights spent in the farms and the forests by each person over the participation period.

The negative binomial regression provided some evidence that males in this cohort were more likely to spend nights in farms (p=0.045) and in forests (p=0.01) compared to females, and that young adults (the 20-40 age group) were more likely to spend nights in the forest compared to the under-20 age group (p=0.043), after controlling for the remaining variables (Table 4).

	Nights slept in farms		*Nights slept in forests*	
Predictors	*IRR [95% CI]*	*p-value*	*IRR [95% CI]*	*p-value*

(Intercept)	0.06 [0.00 – 2.01]	0.116	0.00 [0.00 – 0.17]	**0.005**
Age [<20]: comparator				
Age [>=40]	0.15 [0.01 – 3.24]	0.228	2.03 [0.09 – 46.25]	0.658
Age 20-40	1.43 [0.11 – 18.52]	0.784	16.80 [1.09 – 259.75]	**0.043**
Gender [F]: comparator				
Gender [M]	14.80 [1.07 – 205.70]	**0.045**	46.34 [2.47 – 869.07]	**0.010**
Season [Dry]: comparator				
Season [Rainy]	1.20 [0.66 – 2.19]	0.554	0.74 [0.35 – 1.58]	0.435
Random effects				
σ^2	1.35		1.66	
τ_{00}	6.62_{pid}		5.66_{pid}	
ICC	0.83		0.77	
N	47_{pid}		47_{pid}	
Observations	89		89	
Marginal R^2 / Conditional R^2	0.210 / 0.866		0.341 / 0.850	

Table 4: Predictors of nights slept in farms or forests Incidence rate ratios (IRR), their 95% confidence intervals (95% CI) and p-values are reported.

Participants may spend consecutive nights in the farms or the forests without going back home. The number of consecutive nights spent in the farms, or the forests is the subset of the multiday trips mentioned in the previous section. Figure 19 quantifies this metric for different age groups and gender. Persons of all age groups and gender spent varying numbers of consecutive nights in the farms. An under-20 male spent the most consecutive nights (16-20 nights) in the farm. A female of 20-40 age-group and a male of over-40 age-group spent two episodes of 11-15 consecutive nights in the farm. In contrast, there was

little demographic heterogeneity among those who spent consecutive nights in the forests. A few males of the 20-40 age group not only spent long periods of consecutive nights (more than six consecutive nights), but also frequently spent many short periods of consecutive nights (two to five nights) in the forests.

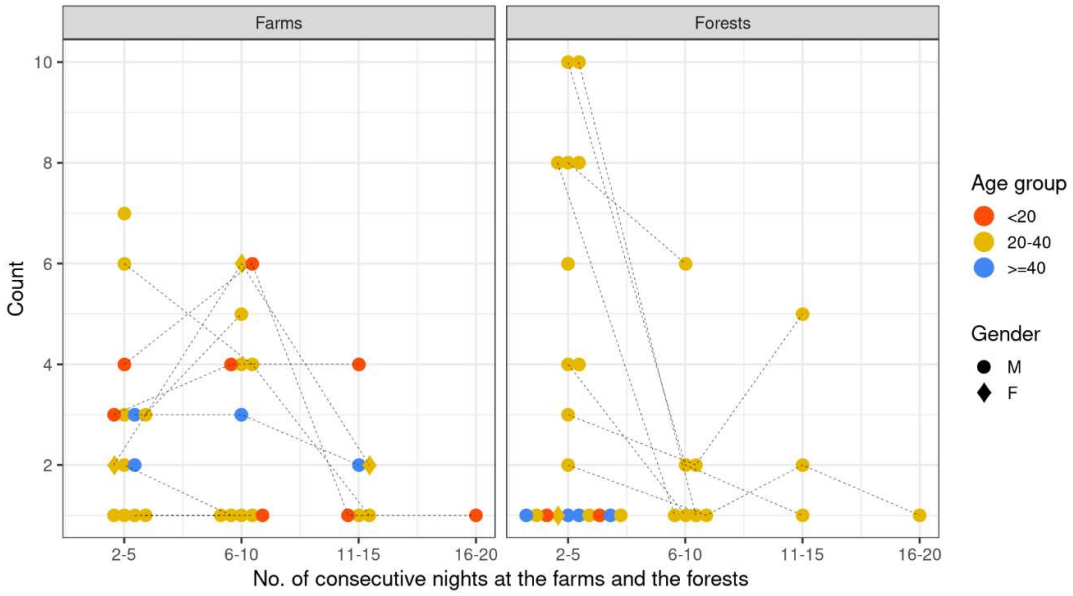


Figure 19: Number of consecutive nights spent in the farms and forests. In each panel, each of the points in each column represents a person of a specific age group and gender, defined in the legend. A single person may contribute a point in each of the columns (e.g., In the panel named Farms, a single person may contribute one point for each of the ranges of consecutive nights). Dotted lines connect the points contributed by the same person across different columns.

3.4.6 Forest going activities

The number of times going into the forest vary among forest goers in the study. After adjusting for the duration of participation, the forest goers in the study leave for the forest on average 12 times per year (**IQR**: 4, 18, see Appendix Figure 7). Once in the forest, the

probability of the forest goer to remain in the forest for each subsequent day was shown in Figure 20.

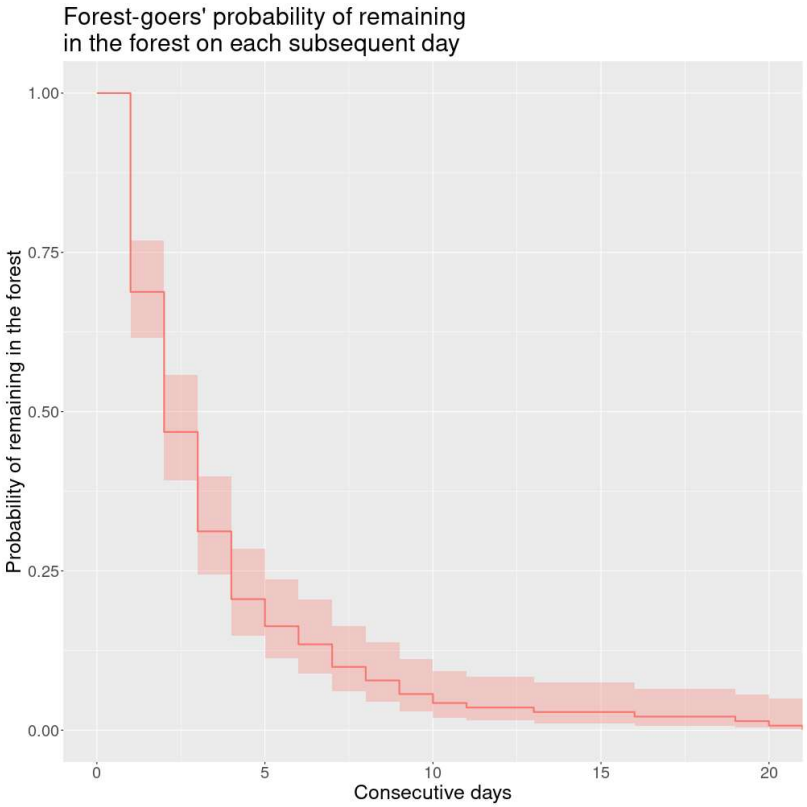


Figure 20: Forest goers' probability of remaining in the forest on each subsequent day. Shaded regions represent 95% confidence intervals.

3.5 Discussion

Many detailed human movement studies have been done, mainly in the regions of high socio-economic status. Our study presents an analysis of human movement in a remote rural area that has been under-studied with regard to human ecology (though there has been some work [119,120]). Compared to other studies where GPS loggers were used for a very

short period of time, there is a relatively long duration of participation in our study. This makes it possible to examine potential seasonal variation.

Our data suggest a bimodal pattern of movement away from participant homes, with one peak nearby (≤ 100m) and another one to three kilometres away from their homes (Figure 15). There were differences in these movement patterns by demography, with under-20s staying close to home on the majority of the days and both 20-40 and over-40 age groups tending to move farther away each day. We hypothesize that the reason for this difference is that over-20 age groups are more heavily involved in subsistence activities (e.g., farming and foraging which are conducted further away from home) than the under-20 age group.

Multiday trips of less than 10 days are common among the participants. The metrics of multiday trips do not signify anything unless they are associated with the activities done during the trip which vary from visits to friends/family, getting supplies at the nearby town, farming, foraging, and other economic or subsistence activities.

All age groups in this study visited farm areas and spent the night in the farms, with no statistically significant difference found between age groups. When they spent their nights in the farms, they did it consecutively and on several occasions during the study period. Farming is one of the major forms of subsistence for rural families and it must be regarded as relatively safe compared to subsistence activities in the forests that all age groups partake in it. There was no evidence of seasonal variation in the number of nights spent at the farms in these data.

In contrast, going to and sleeping in the forests, which may involve foraging, logging, mining etc., is found to be the task for males of the 20-40 age group. The median number of nights slept in the forest among those who ever spent the night in the forest was 7.5. Only males of the 20-40 age group spent a higher number of nights in the forest than the median value.

The same males (20-40 age group) were found to take frequent and successive overnight trips to the forests. We surmise that the males of 20-40, most likely being the breadwinners of the family, are subject to any possible subsistence activities and are regarded as the most suitable persons to venture into the forests overnight despite the danger from the wildlife and the harsh living conditions. No seasonal variation was found in the number of nights of sleeping in the forest.

Compared to home, sleeping places in the farms and forests may be more rudimentary, leaving people more vulnerable to medically important arthropods or other environmental risks (i.e. potentially more contact with venomous snakes, etc.) Spending several consecutive nights in the farms and forests may increase the chances of vector-borne diseases such as malaria since major malaria vectors in the area such as *Anopheles dirus*, and *Anopheles minimus* are found in the deep forests, forest edges, plantations and even in the rice fields.[121] Studies have found that the increased risk of malaria in forest-goers is contributed by inconsistent bed net usage, misconception that alcohol consumption or blankets provides protection against mosquito bites, non-participation in the malaria prevention activities held at the villages.[122]

Results from this study, particularly the mobility data of forest goers, would be useful in spatially explicit individual-based infectious disease model such as Gao 2020[123] which models the malaria elimination in the rural South East Asian region. Human mobility is a crucial part of many disease transmission dynamics, yet it has been ignored in many infectious disease models because of constraints on data and computational capacity. Compartmental models assume homogeneous mixing of individuals in their respective compartments. While they are quick to set up, they are not suitable for the disease elimination settings. Their homogeneous nature limits the modelers from exploring the impact of multiple interventions tailored towards different risk groups such as forest-goers

in malaria intervention. Individual-based models could have individual specific properties and their related movement patterns thus achieving a heterogeneous population. Calibrating on the space utilization data of this study, such models could become more realistic in terms of transmission dynamics. They could provide more accurate and precise estimations to tackle infectious diseases cost-effectively.

For the individual-based model in Chapter 5, the demographic characteristic of forest goers from this study was used to initialize forest goers in the model. The probability of going to the forest in the model was calibrated using the result from this study. An inverse of the probabilities in Figure 20 was used to model forest goers going back home after each night sleeping in the forest.

3.5.1 Limitations

The study has several limitations. It was a pilot study and had a limited sample size. Most participants were adult males. Potential female participants said they rarely go beyond village boundaries and thus were not eligible to be included in the study. It may have introduced a selection bias, but it points to the fact that the mobility preferences between the two genders were too different that they were essentially two different populations requiring separate analyses. The most commonly reported occupation was farming and most people in this study area, indeed, farm for at least part of the year. However, people in the study area usually perform different types of work according to the season and assigning a single occupation to a person may not be appropriate. Employment in this region is almost entirely informal, and most working-age men will work in agriculture for part of the year and in other types of labour during other parts of the year. Responses to

surveys about employment will therefore vary by the time of year, even within a single research participant.

While we believe that this cohort is representative of adult males in this setting, more studies that are demographically representative of rural villages in this setting could be useful for understanding differences in travel patterns by age and gender. Mobile GPS devices have their own limitations. During the study period, participants may have failed to carry the GPS device (intentionally if they engaged in activities that other people might think were illegal or sensitive in nature). Mechanical failures may also cause problems in data collection. Even though the utmost care was taken to preserve data integrity, there could be errors and bias from data collection (due to device inaccuracies) or data manipulation (described in the methods section under analysis and in Appendix Figure 3). Categorization of land types such as farms and forests were done manually using satellite imagery. While the categories do match our authors' understanding of the area, no validation was done on the ground after categorization for this analysis.

Finally, the estimation of land utilization regardless of the method used is imperfect. Having two temporally-consecutive GPS points to constitute usage of the land area may make the estimates less precise. While BRB method provide more precise estimates, it is not without its caveats. BRB assumed that consecutive points that were more than three hours apart were uncorrelated. Since the GPS logger went into sleep mode while stationary, the current land utilization estimation under-estimates the time spent motionless (e.g., sleeping) and hence resulting in lower usage of home in Appendix Figure 5 compared to that in Figure 17.

3.6 Conclusion

We found that younger age group spent more days around their home compared to older age groups. Older age groups spent almost equal amounts of time both around their home and at places one to three kilometres away from their home. Males of 20-40 age group spent more nights in the farms and forests. They spent the nights in the forest for longer consecutive days or more frequently compared to other groups. No seasonal variation was found in the mobility of the participants during the study period. The resulting human movement characteristics can be incorporated in infectious disease modelling studies in similar regions. In fact, we will be incorporating the data from this study in the model in the upcoming chapter (Chapter 5).

4 Generating a realistic household structure in a synthetic population

4.1 Summary

Age and gender are the most basic characteristics that provide individualism in the individual-based models (e.g., movement behaviour is dependent on age and gender). When detailed microdata for the population is not available either because they are not collected or because of the privacy reasons, synthetic populations are generated. Given only aggregated data, populating, and assigning individuals to social structures such as households or families while maintaining a realistic age distribution within each household is not a trivial matter.

To solve this problem, we developed a new approach based on a computational paradigm called constraint programming (CP) which imposes rules on the variables (e.g., age of each household members depending on the size of the household). The imposed rules must be satisfied to produce a solution (e.g., a synthetic population with realistic household composition). We devised two sets of CP algorithms, one of which has a more stringent set of rules than the other, to check their improvements in accuracy over the simplest algorithm. Rural household composition data from Cambodia, Lao PDR, Myanmar, and Vietnam were analysed to craft the rules for the algorithms, and to act as a gold-standard to compare the algorithms' performance. Overall, algorithms with more stringent rules performed better than the one with generic rules. Stringent rules were better for bigger household sizes, while generic rules were better for smaller household sizes. Mixing of rules with different flexibility for different households will produce a more performant algorithm as a whole. A synthetic population generated in this exercise was then used in an

individual-based model project for assessing the important aspects of reactive case detection and treatment of malaria discussed in the next chapter.

4.2 Background

In the previous chapter, we learned that human mobility such as going into the forest and staying overnight was greatly influenced by individual's age and gender. To properly simulate the human movement for an *in silico* experiment with individual-based models (IBM), each individual must have his/her respective age represented in the model. Depending on the level of detail required, generating a synthetic population could become a non-trivial task.

For example, when creating a population where each person's age needs to be initialized, the numbers could simply be drawn from an age distribution of the required population. However, individuals live in social groups such as families or households, in which resources are shared among its members. Households with high socio-economic status tend to have better housing that could protect all its members from mosquitos.[124,125] Health care decisions are usually made by a primary care giver of the households on behalf of its members.[126] Individual's treatment seeking behaviours such as treatment acceptance and affordability are dictated by which household he or she belongs to.[127,128] Occupation traits are also concentrated in households (e.g., A father who is going to the forest for subsistence activities might bring his adult son along), thus increasing the occupational risk in some households.[129] Inclusion of a household structure is particularly pertinent for the model in Chapter 5 where we will explore the effect of a reactive malaria intervention that is implemented at the household level. To generate households with a realistic age

composition, we need a better approach than drawing from an age distribution and assigning them randomly into households.

Several methods on generating synthetic population could be found in transportation literature[130-134]. Most of them used joint probability distributions of individuals and household properties. They have been tested with the household compositions in high income countries where the household sizes are small, and the variety of household types is limited compared to low-income countries. Their use becomes challenging with household sizes larger than 4.

Here, we present a novel method using constraint programming to generate a synthetic population that has realistic household compositions by age when the actual microdata is either not present or not accessible. The overview of the method, its requirements, and performances are described using the rural population data from four countries in South-East Asia (Cambodia, Lao PDR, Myanmar, and Vietnam).

4.3 Methods

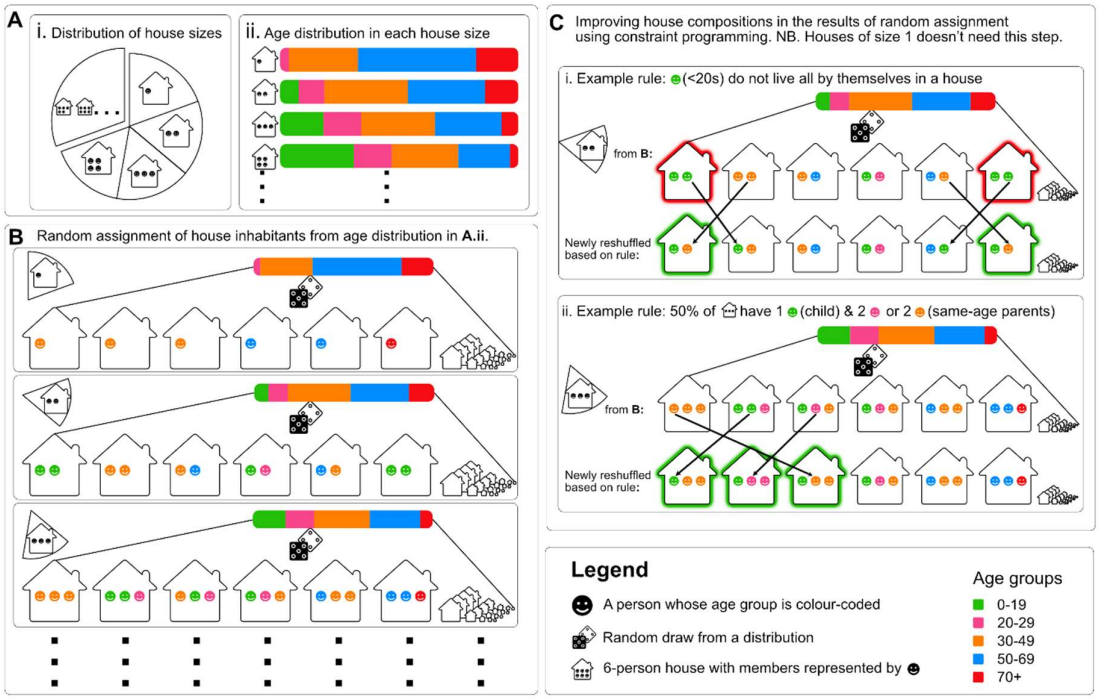


Figure 21. Schematic diagram of data and algorithms. ***A.*** *Demographic summary statistics of a hypothetical population. A.i. Distribution of the number of household members in houses in the population. The number of smiley faces represent the number of household member in the house. A.ii. Distribution of age groups within all houses of different capacity in the population.* ***B.*** *Based on the distributions from A., a synthetic population is generated randomly based on the distribution data in A (i.e., age groups of persons to be put in the houses are randomly drawn from the respective distributions. The results from the random draws depicted in the figure are for illustrative purpose only). This algorithm is named Random Assignment (RA) for further reference.* ***C.*** *Result from RA may not have optimal within-household composition (e.g. In the houses with red halo in C.i., have under-twenties living all by themselves in a house). Constraint programming could be used to set rules on household compositions to reshuffle people in the result from RA so that the houses could conform to the rules (houses with green halo). Only two examples are provided here, but each household size can have multiple rules. Generic rules such as children & teens don't live in a house by themselves (C.i.) can be applied to any household size, whereas specific rules such as a nuclear family of same age couple with one child (C.ii.) can be applied only to certain household size.*

To create a synthetic population with realistic household composition, summary statistics for the desired population is processed through an algorithm. The basic concepts of the algorithms are illustrated in Figure 21. The minimal amount of information required is the distribution of household sizes and the age distribution within those household sizes of the interested population (Figure 21. A.). The algorithms are implemented in R and C++ programming languages.[72,135]

4.3.1 Data

The required summary statistics and detailed household level data are extracted with permission from Demographic Health Surveys (DHS) for Cambodia, Myanmar, and Vietnam, from Multiple Indicator Cluster Survey (MICS) for Lao PDR, and from Census Ad Hoc Household Composition for the UK.[136-138] Inclusion of the UK data was to prove how different the household types were between the Southeast Asian (SEA) countries and a high-income western country. The following packages in R are used to process and visualize the data: foreign, magrittr, xlsx, GDAtools, tidyr, plyr, ggplot2.[139-145]

4.3.2 Household types

Assuming that all members of a household constitute a family, the within-household composition of different age groups can be represented by family types (e.g., a household with one child and two same age adults is a nuclear family of size three). For this reason, the terms "household" and "family" are used interchangeably throughout the thesis.

Five age groups with cut off values at 20, 30, 50, and 70 years of age are defined for this exercise. To differentiate between different family types (there are 15, 35, 70, 126, 210, 330, and 495 theoretically possible family types in household of sizes from two to eight respectively), a naming scheme is established such that the number of members in the house is encoded into five digits separated by space, with the first digit representing the number of person(s) with the first age group (0-19), the second digit representing the number of person(s) with the second age group (20-29) and so on. "1 0 2 0 0" encodes the three-person nuclear family in the example mentioned previously. A family of two elderly persons (70+ age) is represented by "0 0 0 0 2".

The detailed household level data is analysed to obtain the distribution of household types in each country. Household type distribution data is then used to craft the rules for the algorithms and to evaluate their performances.

4.3.3 Algorithm: Random Assignment (RA)

The distribution of household sizes is used to create a proportionate number of respective households for a synthetic population of approximately 40,000 people. Randomly drawing people's age from the age distribution associated with each household size, and placing them into houses is the simplest and easiest way to create the synthetic population (Figure 21. B.). This algorithm is denoted as "Random Assignment" or RA for future references. Unfortunately, RA does not take into account how each household is composed, resulting in

spurious age composition within the individual households. An example would be a four-person household inhabited by 4 persons of under 20 years of age.

4.3.4 Algorithm: Re-constructing households with constraint programming (CP)

Constraint programming (CP) is a computational paradigm to solve discrete optimization problems such as sudoku and optimal resource allocation. It imposes constraints or rules in order to reduce the set of values that each variable can take. Rules for CP are created using the gecode framework.[146] As mentioned previously, RA may not produce sensible family types in the synthetic population. CP can be applied on top of the results of RA to improve the household composition. The algorithm goes through all the households of particular size and shuffles the individuals so that the rules imposed are satisfied. Rather than optimizing an objective function, the algorithm stops when a satisfactory solution is reached. Since individuals are only shuffled within the houses that have the same household size, overall age distribution within that household size is not disturbed.

All the houses which have a particular household size are regarded as a series of houses within which shuffling of individuals will take place according to the rules. Rules can be applied to all or part of the series of houses. Some rules are generic as in Figure 21. (panel C.i.: Children and teens don't live by themselves in a household) which could be applied to all or many of the household sizes, and some are more specific as in Figure 21. (panel C.ii.: 50% of households have a nuclear family with two same-age couple and a child) which could only be applied to a particular household size.

Based on the major family types in the four Asian countries, two sets of constraint programming rules (denoted as CP1 and CP2) were devised manually. Each of them has several subsets of rules for each of the household sizes. The rules used in CP1 are more generic (e.g., Figure 21. C.i.) compared to those used in CP2 which are more specific (e.g., Figure 21. C.ii.). Once these sets of rules were established, both algorithms were applied to the same summary data used in RA. The details on rules for each algorithm are listed in the Appendix Table 1.

4.3.5 Performance of algorithms

The actual household type distribution data in each country serves as the target against which the performances of different algorithms were compared. The comparison is made for the proportions of different household types within each household size. The sum of absolute differences in proportion of major household types between the data and each algorithm (sum of absolute errors or SAE) for each household size is used to measure the accuracy.

4.4 Results

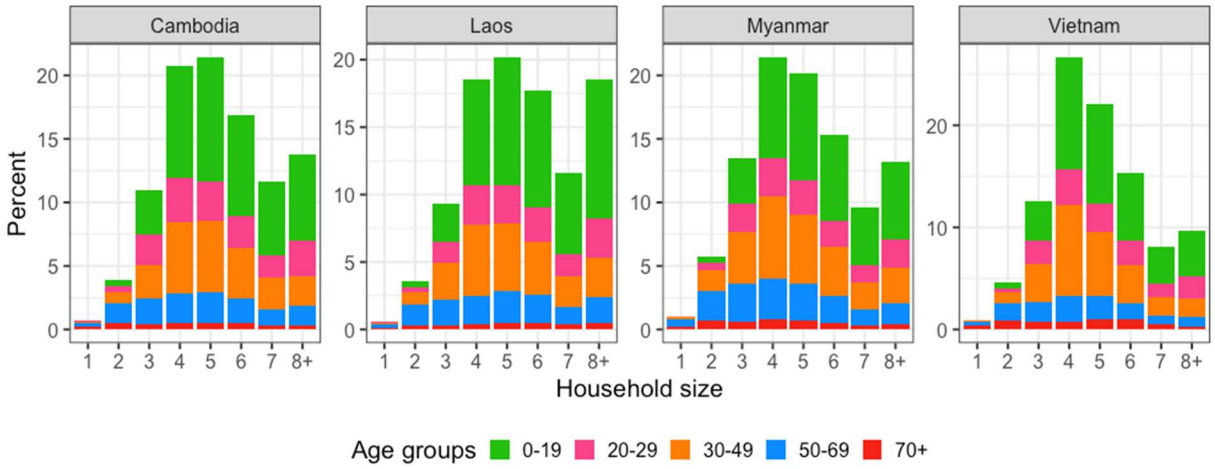


Figure 22. Distribution of household sizes and within-household age distribution in the rural areas of Cambodia, Laos, Myanmar and Vietnam. The height of each bar represents the percentage of the corresponding household size or age group. The sum of all the heights of the bars in each country is 100%. Data adapted from Demographic Health Surveys (DHS) and Multiple Indicator Cluster Survey (MICS), and filtered and categorized by the author.

4.4.1 Data

The distribution of household sizes and age-distribution within those household sizes for the SEA countries included in this exercise can be seen in Figure 22. The SEA countries have a similar household size distribution with a median household size of 4 to 5. They have a young population with significant portion of under-twenties and 30-49 age group occupying the household sizes above 3. Single-person households are relatively uncommon in these countries whereas in the UK they constitute 22% of the population. The UK has a median household size of 2, with relatively fewer number of household sizes that have 5 or more members. Its age distribution is dominated by older age groups above age 50. The age-distribution within different household sizes for the UK can be seen in Appendix Figure 8.

4.4.2 Household types

	No. houses in data	No. of unique family types in the respective household sizes							
		1	2	3	4	5	6	7	8
Theoretically possible	-	5	15	35	70	126	210	330	495
Cambodia	13086	5	15	33	59	74	85	85	83
Laos	14288	5	15	33	61	78	92	101	101
Myanmar	8895	5	15	33	61	89	98	101	86
Vietnam	4941	5	15	32	53	64	60	58	42
United Kingdom	~23 million	5	15	35	70	126	209	301	352

Table 5. Number of unique family types in each household size in each country based on five age groups. N.B. Data for Asian countries are extracted from surveys in which only a subset of population was sampled, whereas the data for the UK is from a census.

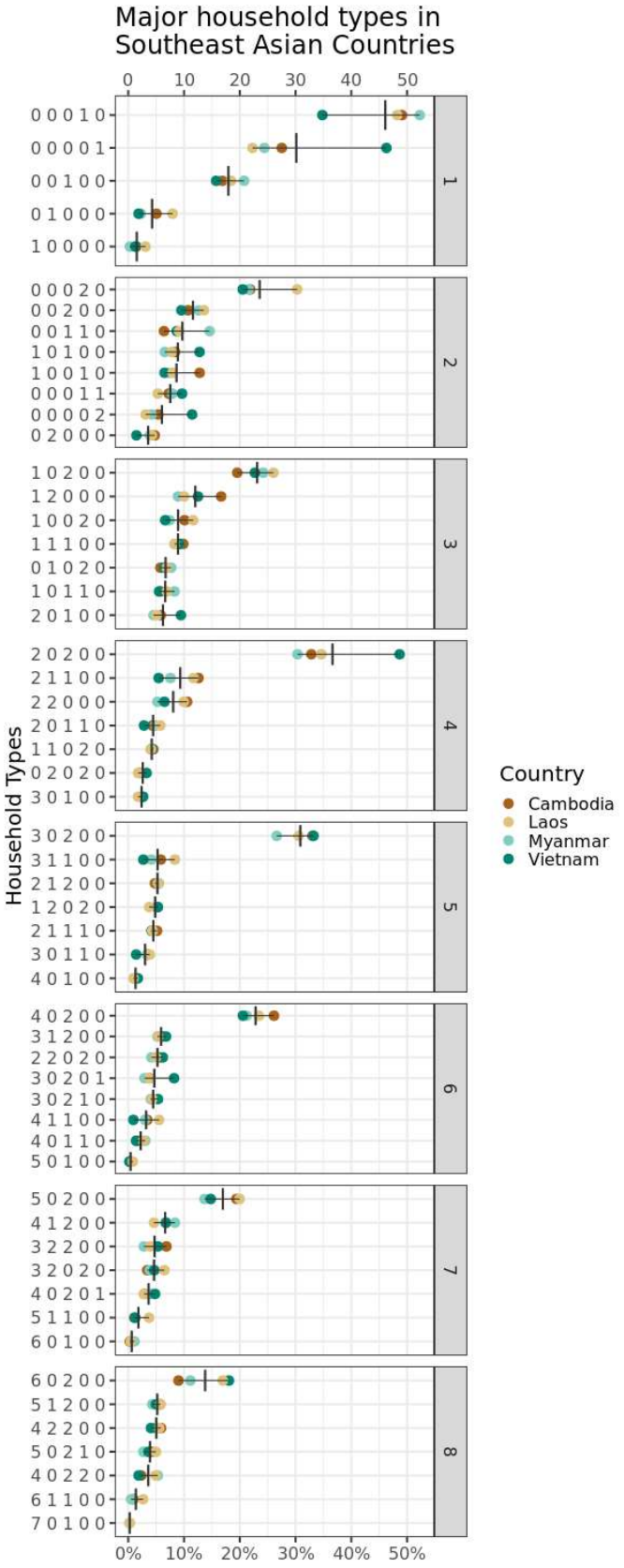


Figure 23. Percentages of major household types in four South-East Asian countries (coloured dots). Black horizontal and vertical bars show the range and the mean percentage among the countries. Each household size is represented by the respective horizontal panel. Household Types (y-axis) are encoded as explained in the Methods section. X-axis represents the percentage of the household types within the respective household size.

The number of unique household types in different household sizes can be found in Table 5. The larger the household size is, the more the number of theoretically possible types is. However, out of all the possible types, only a handful of types make up the majority in all household sizes (Figure 23 and for the major types including the UK, see Appendix Figure 9). Notably, the households that include the two persons with 30-49 age group and the remaining person(s) being in 0-19 age group are always the commonest type regardless of the household size. Even though there is variation in the percentages of household types across different countries, there is consistency in the order of the major types.

4.4.3 Performance of algorithms

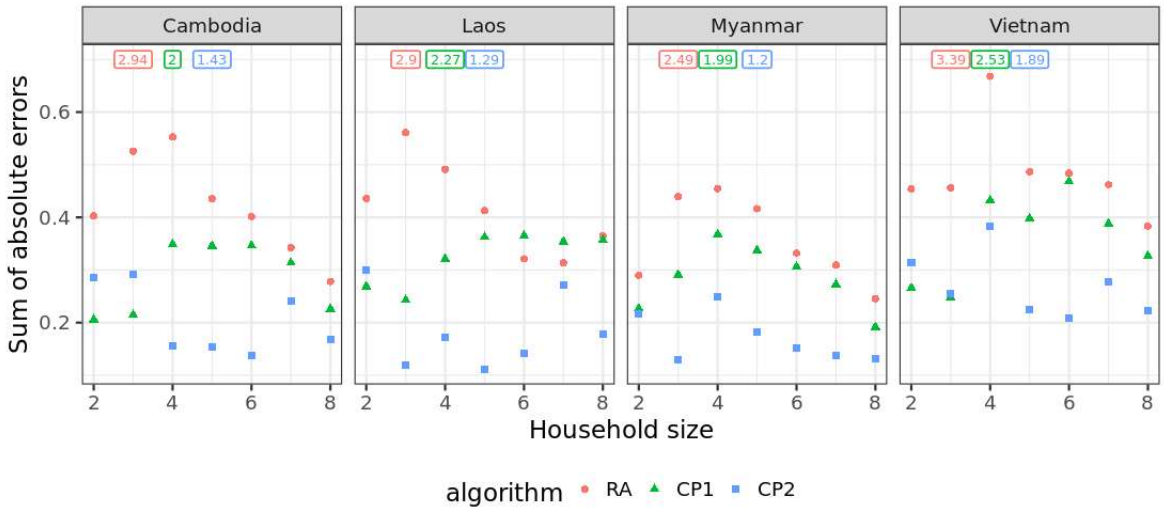


Figure 24. Performance of random assignment (RA) and constraint programming (CP) for different household sizes (x-axis) measured by sum of absolute errors (SAE, y-axis) deviating from the proportions of major family types present in the data. SAE value of 0 represents a perfect match to data. The higher the value of SAE, the worse off the algorithm is. Total SAE for

each algorithm for each panel is stated in the respective coloured boxes on the top left corner. Households of size one do not need to be processed by the algorithms.

After synthetic populations are generated from the algorithms, they are compared against the original data. The deviations from the real proportion of household types within a particular household size are summed up to calculate the comparison metrics (sum of absolute errors or SAE) shown in Figure 24 for different algorithms.

The SAE values for all the household sizes are aggregated to get the overall performance of each algorithm in each panel (inside the coloured boxes on top left). For the overall performance in all the comparisons, CP2 performs better than CP1 which in turn performs better than the RA. Performances of the algorithms for each household size follow the same pattern except in the household size of 2 where the performance of CP2 is worse than that of CP1 and in the household of size 6 where the performance of CP1 is worse than that of RA.

4.5 Discussion

4.5.1 Data

The choice of survey data for this exercise was based on the data availability and the convenience in accessing them, in addition to the fact that the first output of these algorithms will be used to initialize a population in the SEA region to simulate a malaria intervention in the next chapter. The resulting household age-composition data from the analysis has been essential in creating the algorithms and their rules.

Household composition is greatly different between the UK and the SEA countries. This would have risen from the differences in cultural, socio-economic status, distribution of equality, wealth, and health. The domination of under-twenties and 30-49 age groups in the SEA countries is most likely due to the parents-children pairs reflecting a growing population. Unlike the UK, living alone is rare and huge household sizes are common in these countries.

4.5.2 Household types

Even though the distribution of household sizes and age group distribution within households differ slightly between different SEA countries, there is consistency in the most prevalent family types. Most of the household sizes are dominated by a handful of family types. In all the countries, even the UK which is culturally and socioeconomically different from the Asian countries, the top family type is where two persons of same age group co-habit either by themselves or with one or more children/teens.

The bigger the size of the household, the greater the number of possible family types there could be. The prevalence of the rare family types in the data increases with the increasing household size and the increased number of samples in the data as in the UK data in Table 5. Likewise, the relative proportions of the most prevalent types are reduced with the increasing household size. Since the rules in constraint programming algorithms are constructed based on the distribution of these major family types, there will be an impact on the algorithms with increasing household size, as described in the next section.

4.5.3 Performance of algorithms

The simplest algorithm (RA) is the easiest to implement and the quickest to execute. However, the household composition it produces could not capture the popular types such as those with parents-children pairs. Among the CP algorithms, CP1 has rules that impose less strict constraints compared to those of CP2.

Based on the SAE metrics, the overall accuracy is the best with CP2 followed by CP1 and RA sequentially. This result is naturally expected as both CP1 and CP2 are based on the results of RA. But it should be reiterated here that the CP1 and CP2, which have been named to make the comparison simpler, are not the complete algorithms in and of themselves. They are made up of a combination of different sets of rules for each household size which could be modified individually.

In household size 2, CP2 is better than the RA, but worse than CP1. In household sizes 6 and 7 in Laos, CP1 somehow performs worse than RA. In such instances, revising the sets of rules specific for those household sizes may improve their accuracy. As they currently stand, CP2 performs better than both RA and CP1 in most of the household sizes in all SEA countries analysed. The algorithms were not tested for the UK, because a completely different sets of rules would be required.

Based on the experience of devising different sets of rules for CP in this exercise, more generic rules, such as those in CP1, are better for the household size 2. Specific rules, such as those in CP2, are better suited for household of sizes between 3 and 6. It should be noted that the algorithms with more stringent constraints may produce better results, but they run the risk of not finding any satisfactory solution, in which case the rules will have to be revised.

For the household sizes bigger than 7, CP algorithms have to struggle to perform better than RA. In fact, for the household size 8, the number of specific rules have to be reduced in CP2 to stay competitive with RA & CP1 since adding more specific rules make its performance worse (See notes on Appendix Table 1). Increasing household size increases the number of possible family types, which could be difficult for the CP rules to cover since the number of rules that could be implemented in CP is limited.

4.5.4 Limitations

The summary statistics used to recreate the synthetic population are a snapshot of a population at one point in time. The family composition will change over time with the dynamics of the demographic processes of the population.

It takes approximately an hour to process a synthetic population of size 40,000 on an Intel dual-core Core i7 computer with 16 GB of RAM. Currently, the algorithm doesn't scale well with increasing size of the population as it must go through the series of records multiple times to shuffle them to satisfy the rules. There may be a way to improve this process by reconfiguring the underlying Gecode framework that is used for the constraint programming.

Synthetic populations are representations of reality and there will never be a perfect synthetic population unless the actual microdata is used. By using constraint programming the inclusion of the rare family types is not guaranteed especially in the households with sizes greater than 7.

4.5.5 Future improvements

The results presented in this study is only an instance of the series of rules that has been specified. Individual rules may be refined to further improve the accuracy of the synthetic population. Even though only the age group composition is mentioned throughout the manuscript, other compositions such as gender, occupations could also be generated with the same methods.

A hybrid algorithm using a mixture of rules from CP1 and CP2 will further improve the accuracy. Additional steps, such as adding rare family types as some sort of noise to the synthetic population created by the algorithm may also be added. Comparisons could be made with other type of algorithms such as machine learning algorithms.

4.6 Conclusions

It is increasingly common to solve real-world problems by individual-based models. When these problems involve human population with the nuance of their social structure such as a family or household composition, the creation of synthetic populations with detailed and realistic characteristics becomes important. Despite its limitations, the constraint programming algorithm and the methods described here could be useful tools to quickly generate such synthetic populations using only the summary statistics of the distribution of household sizes and the age distribution within those household sizes of the interested population. A

combination of rules from the algorithms described here has been used to synthesize a population required for the modelling exercise in the next chapter.

5 Model 2: Simulating the effectiveness of reactive case detection and treatment interventions in dynamic populations with realistic household structures

5.1 Summary

Reactive malaria interventions (RMI) have been recommended by the WHO in the final phases of malaria and the prevention of re-establishment of malaria. However, one of the RMIs, the effectiveness and optimal implementation of reactive case detection and treatment (RACDT) is not well studied. We explored the effectiveness and efficiency of various RACDT configurations in two malaria transmission settings with an individual-based, spatially explicit model, that has a realistic household level age-composition. Two different implementation levels of RACDT were considered – the village level where reactive activities were done village wide and the household level where reactive activities were done only at the households where the index case was found. To properly assess the nuance of having households as intervention units, realistic household level demographic properties such as age composition were incorporated in the model. RACDT is predicted to work best when malaria transmission is low enough. Village level RACDT is more effective, but it must screen significantly more people. The predicted effect of household level RACDT is 15 times more efficient than the village RACDT in the low malaria transmission setting, achieving 9.94 percent reduction in prevalence per 1000 individuals screened over a five-year period. Provision of malaria kits to forest goers could further improve the household RACDT. Having large household sizes with

multi-generational members in the population is predicted to receive more effect from RACDT than having small household sizes with fewer multi-generation members. Efficient use of RACDT in the right malaria transmission setting could be an important part of a successful malaria elimination package.

5.2 Introduction

According to the WHO's estimation, there was an increase of 5 million malaria cases in 2022 (a total of 249 million malaria cases globally) compared with 2021.[27] The stalled progress on malaria elimination has been attributed to the disruption to the malaria services by the COVID-19 pandemic and other humanitarian and health emergencies in several malaria endemic countries over the past few years. Despite the setback, the WHO envisions a malaria free world by 2050. The WHO's Global Technical Strategy for Malaria 2016-2030 (GTS) has set an ambitious target of reducing malaria case incidence and mortality rates by at least 90% by 2030. The GTS also aims to eliminate malaria in at least 35 countries by 2030.[40]

Malaria cases have been found to cluster in households and neighbourhoods.[147-150] This clustering increases as the malaria prevalence decreases, and it gets closer to elimination phase.[151] In the Greater Mekong Subregion (GMS), 65% of households that had a patent malaria case also had another household member(s) with *Plasmodium* infection.[152] Aside from households and neighbourhoods, clustering of malaria cases could happen socially, for instance, when people shared the same type of occupation or travelled together to an endemic area.[153]

Identification and targeting of these malaria clusters was thus recommended through the reactive interventions in response to detection of confirmed malaria cases such as reactive drug administration (RDA) and reactive case detection and treatment (RACDT) in the latest WHO Guidelines for malaria for the final phase of malaria elimination and prevention of re-establishment of malaria transmission.[22] Unfortunately, because there's no single recommended approach on how to implement RACDTs, numerous variations have been used worldwide.[154] China used a "1-3-7" strategy which involved reporting malaria cases within 1 day, confirming and investigating them within 3 days, and implementing indoor residual spraying (IRS) or RACDT in the households to prevent further transmission within 7 days.[155] Detailed information on RACDT in other countries in the GMS has been scant at best.[154,156]

In addition, the RACDT's effectiveness and its optimal implementation in malaria elimination settings are not well understood.[22] In a systematic review on 72 empirical studies and 24 modelling studies for testing and treatment of malaria elimination, it was noted that because there were no intervention studies directly comparing RACDT to a control group without RACDT, the effectiveness of RACDT in reducing transmission was not possible to assess.[156] The majority of the modelling studies in the review did not explicitly incorporate the household level age composition, which might be an important concept to be included in a model simulating the targeting of the malaria clusters, given that these clusters are most likely to be at the household level.

In this chapter, we explore the different configurations of RACDT with an individual-based, spatially explicit model, that incorporates household level realistic age-composition. We modified a well-established model[123] to develop new functionalities such as incorporation of realistic household age structure, incorporation of forest goers' demographics and their

movement probabilities, and the RACDT intervention. The model was then simulated to assess the effect and efficiency of different RACDT configurations on two malaria transmission settings. Figure 25 describes how prior chapters of the thesis feed the necessary information and data required by RACDT into the model.

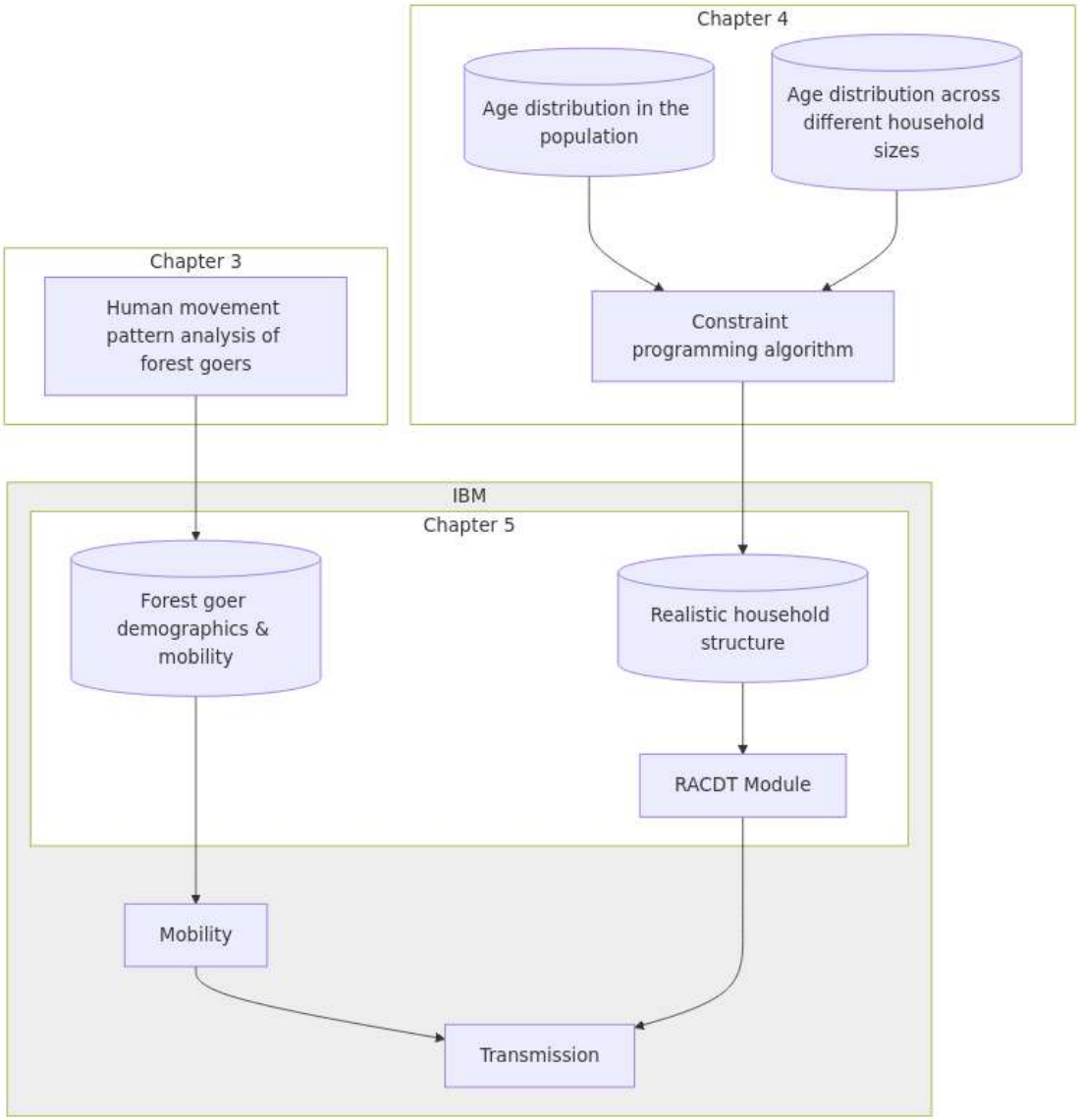


Figure 25: Data flow diagram

5.2.1 Definition of RACDT

RACDT is one of the reactive interventions (RMI) that is implemented in response to detection of confirmed malaria cases. According to the WHO guidelines, RACDT is the identification of malaria cases through active case detection in the vicinity of a confirmed malaria case, followed by treatment of the identified cases.

5.2.2 Objectives

- Explore and rank the important attributes of Reactive Case Detection and Treatment (RACDT) intervention
- Assess the potential impact of RACDT on *Plasmodium falciparum* malaria prevalence in low and moderate malaria transmission settings

5.3 Methodology

5.3.1 Overview

A well-established malaria model was adapted to incorporate new features: (1) realistic household age-composition, (2) initialization of forest goers' demographic properties from data, (3) forest goers' movement probabilities calibrated from data and (4) mechanisms for RACDT. Model simulations were done to investigate the impact of several different types of

RACDT on a hypothetical population of around 10,000 in the Greater Mekong Subregion. Our experiments focused on two malaria transmission settings as starting points of the simulations – a low malaria transmission setting represented by a starting prevalence of around 5% and a moderate malaria transmission setting represented by a starting prevalence of around 20% (the WHO categories of malaria transmission intensity is listed in Appendix 7.4 for reference). Each starting point of the simulation was at an equilibrium state (i.e., the prevalence would not change unless an intervention was introduced) which was achieved by running the model for 10 years in advance.

For a given malaria elimination scenario with a particular RACDT, a set of parameters was produced. Using each parameter set, the model was simulated for 100 times for a duration of 10 years – the first five years without the intervention and the last five years with the intervention. For each simulation, the model output daily prevalence and incidence data, which were then summarized for each scenario into aggregate metrics (see Metrics for outcome and efficiency section).

5.3.2 Model description

5.3.2.1 Original model

The original model is an individual based, discrete time, spatially explicit, stochastic model, with explicit mosquito population dynamics and human population movements, that provides realistic prediction for outcomes of malaria interventions at various levels.

The individuals in the model resided in villages with different mosquito densities. The villages were connected by a human flow network. The human mobility between villages was modelled using a modified gravity model which presumed that people who travelled below a certain critical distance threshold would return home without staying overnight somewhere else. The malaria transmission in each village assumed a pseudo-homogenous process in which each mosquito was equally likely to bite a given individual. For a specific transmission scenario, the desired malaria prevalence was calibrated by changing the mosquito biting rate parameter.

The model was run until an equilibrium was achieved which also acted as a control scenario. The integral of a particular metric such as incidence over a certain period was compared between the control scenario and a different scenario under investigation (usually an addition of a new intervention) to elucidate the difference between the scenarios and the effectiveness of the new intervention.

For one simulation, the model required a configuration file and an input file each to initialize the properties of villages and humans. When death occurred, an individual was replaced with a new infant agent whose age and immunity parameters were reset in order to keep the population size constant.

Very detailed logical processes for deploying interventions particularly MDA, IRS, village malaria worker (VMW) network, and bed net distribution were already implemented in the model. In a previously published manuscript, it was used to explore a malaria Mass Drug Administration (MDA) with a particular consideration for the logistics of MDA deployment. More specific details on human mobility, immunity, infectiousness, treatment, PK/PD, clinical outcome, mosquito dynamics and interventions can be found in the cited paper.[123]

Because the original model is modular, and customizable, we have been able to add new features to it while keeping the malaria transmission dynamics intact. On each step of the modification, the original model was tested against the modified model with the new feature turned on and off; with the expectation that only when the new features are off, the modified model should behave identically to the original model, provided that both uses the same parameter set.

5.3.2.2 Modifications

The major addition to the original model was the module on RACDT. During any period that RACDT policy is implemented, the module checks if there was any confirmed malaria case in the simulated population which acted as a trigger for the RACDT. Case detection and/or treatment was implemented either at the household level or at the village level depending on the implementation level parameter set for the intervention. For the household level RACDT, the trigger was always a confirmed malaria case within the household whereas for the village level RACDT, the trigger could be one or more confirmed malaria case(s) in the village that was set as an additional parameter. Once triggered, eligible individuals were screened with a diagnostic test with sensitivity of 90%, followed by the treatment of the malaria cases found during screening with Artemisinin Piperaquine combination therapy. The screened households were allowed a cooling-off period of 30 days before another RACDT screening could take place. A decision flowchart for RACDT is presented in Appendix Figure 10.

Treatment seeking behaviour was implemented as a binary property of each individual and based on two properties: whether or not an individual was willing to accept the

treatment/intervention and whether or not an individual was able to afford the treatment in cases where out of pocket payment was necessary. Since health care decisions were usually made by a primary care giver – usually the matriarch of the household, these properties were initialized to be the same within each household.

Assignment of who would be the forest goers was done depending on age and gender data from Chapter 3 [157], unlike in the original model where the assignment was done randomly. When the household structure permits, the forest goers were assigned to congregate together. The probabilities of forest goers going and staying in the forest overnight was calibrated using the data from the same research.

5.3.2.3 Metrics for outcome and efficiency

To compare the output of the various scenarios, prevalence reduction percentage or percent reduction in prevalence (PRP) was calculated by the difference between the two integrals of malaria prevalence over a five-year period over time – one where RACDT was implemented and one where no intervention was implemented. Unless mentioned otherwise, PRP was used throughout the manuscript.

For each type of RACDT implementation, the number of individuals screened and those with new malaria positive tests discovered through the reactive process were counted. The screening efficiency was then calculated as the new malaria cases found per 1,000 individuals screened. The intervention efficiency was calculated by the PRP per 1,000 individuals screened.

5.3.3 Sensitivity analyses

Multivariate sensitivity analysis was performed to assess the sensitivity of model parameters on the model output. The parameters explored and their values can be seen in Table 6.

Parameter	Explanation	Values	Default
Village RACD trigger threshold	The number of confirmed malaria case(s) that is needed to trigger village RACD. This only applies to village level RACD. The trigger threshold for family level RACD is fixed at 1.	1, 4, 8	1
Implementation level	Once RACD is triggered, the level at which screening and treatment is performed.	Village, Household	-
Reaction delay	The delay (days) between the day when RACD is triggered and the day when screening & treatment happens.	1, 7, 14	1
Days per RACDT	Number of consecutive days the treatment is provided for each RACD round	1, 3, 5	3

Forest treatment kit	Provision of treatment kit to forest goers so that they could be treated if they become malaria positive	Yes, No	No
Household type	Household level age structure which is based on either Southeast Asian (SEA) data or data from a western high income country (HIC)	SEA, HIC	SEA

Table 6: Parameters of RACDT

5.3.4 Tools and technologies used

Since the original model used C++ programming language[135], modifications had to be made in the same language. C++ libraries such as standard template library (STL), Boost, etc were used. R programming language was used for data analysis of the model output. The following R packages were used: tidyverse, sfsmisc, lemon, and rstatix.[158-161] Constraint programming[146] was used to synthesize household structures as described in Chapter 4. The human mobility information from previous analysis (described in Chapter 3) was fed into the model to inform age, gender and mobility of forest goers in the model.

5.4 Results

5.4.1 Overall effects of RACDT in low and moderate transmission settings

For a particular low transmission setting at 5% initial prevalence, the RACDT was predicted to achieve an average of 76 (95%CI: 61, 89) percent reduction in prevalence (PRP) over a 5-year period for the various RACDT configurations considered. For a moderate transmission setting at 20% initial prevalence, PRP for RACDT was 59 (95%CI: 50, 70) for the same period.

5.4.2 Different implementations of RACDT

Effects of village level and household level implementations of RACDT were predicted to differ significantly in the moderate malaria transmission setting whereas they were not in the low transmission setting (Figure 26). Village level RACDT had 64 PRP (95%CI: 57, 71) compared to 55 (95%CI: 49, 60) in household level RACDT in the moderate transmission setting. For the low transmission setting, both village and household level RACDTs, the effects were similar at 78 (95%CI: 62, 90) and 75 (95%CI: 61, 84) respectively. Differences in their effects for various starting prevalences are described later in the section.

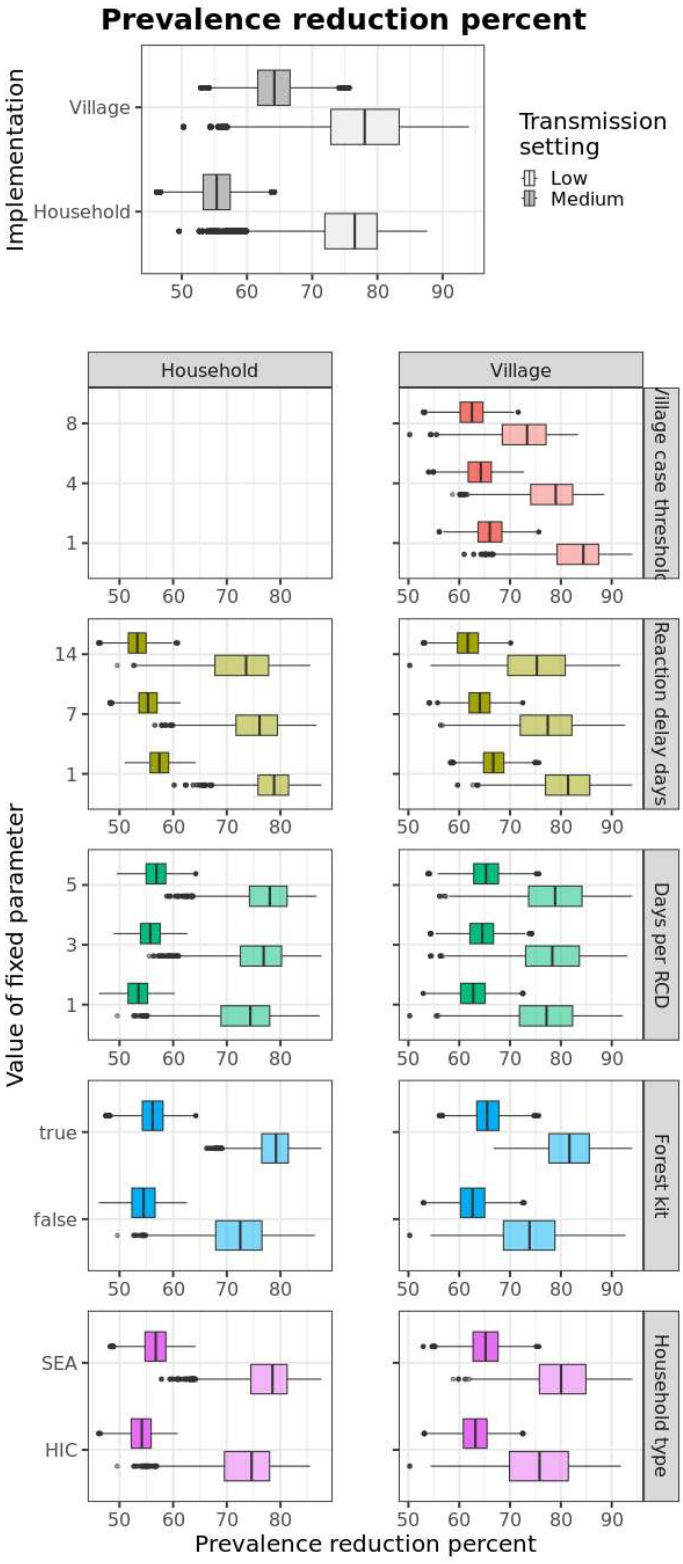


Figure 26: Sensitivity analysis of RACDT parameters on prevalence reduction percentage in low and moderate transmission settings

The village case threshold was the number of confirmed malaria cases in the village in a month's duration above which village level RACDT would be implemented. It was found that the lower the threshold, the higher the effect of RACDT was in both low and moderate transmission settings. Family level RACDT was always triggered at the detection of one confirmed malaria case and thus it was not explored in the sensitivity analysis.

Once the RACDT was triggered, there was a delay before testing and treatment of eligible cases were conducted. A delay of 1 day, 1 week and 2 weeks were explored. Our simulations found that the longer the delay, the less effective RACDT was. A week of delay could forego around 3 PRP in prevalence reduction in a low transmission setting and around 2 PRP in a moderate transmission setting.

For the treatment of the eligible cases, the higher the number of days the anti-malarial drug was given, the higher the effect of RACDT. The concurrent strategy of providing forest goers with treatment kits and treating them as necessary gave a boost to the effect of RACDT particularly in the low transmission setting.

To test whether the household structure matters on our simulation exercise, a counterfactual household structure from a western high-income country (HIC) was used. Having different household structure was found to have an impact on the effect of RACDT implementation, regardless of the transmission setting. Compared to a population with household structures from an HIC, RACDT in a population that has SEA household structures would yield 4-6 more PRP in the low transmission setting and 2 more PRP in the moderate transmission settings.

The order of importance of the RACDT's parameters for optimal effectiveness differed greatly depending on which transmission setting it was implemented. In the moderate transmission setting, the level of implementation had the most influence with nearly 9 PRP from changing village level to household level implementation. This was followed by reaction delay days, days per RACDT, household structure, forest kit, and village threshold.

In the low transmission setting, having the forest kit improved RACDT by 7 PRP on average. This was followed by reaction delay days, village threshold, household structure, days per RACDT, and implementation level.

In Figure 27, the difference between the two implementation levels was explored for the initial prevalences from 1% to 20%. The other parameters were fixed at their default values (see Table 6). Unlike in the previous results, the metrics used here was the absolute reduction in prevalence compared between the start and the end of the RACDT intervention. The effect of the village level implementation was higher than the household level implementation for the initial prevalence values of 6% and above. At or below 5% prevalence, the efficacy of village level RACDT is either worse or not differentiable from family level RACDT.

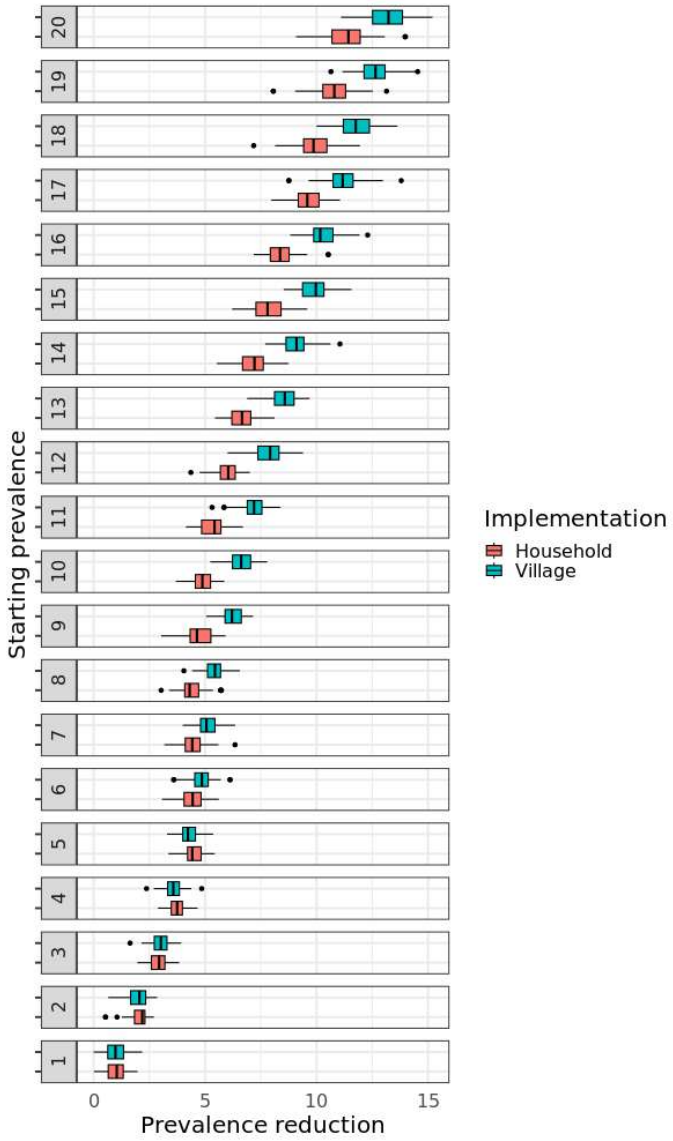


Figure 27: Prevalence reduction over different starting prevalence values. The prevalence reduction here is an absolute reduction calculated by the difference between prevalences at the start and the end of the intervention.

5.4.3 Efficiency

5.4.3.1 Screening efficiency

On average, the household level implementation was predicted to find 227 new cases per 1000 individuals screened (95%CI: 207, 247) in the moderate transmission setting and 167 new cases per 1000 individuals screened (95%CI: 132, 180) in the low transmission setting (Figure 28). The village level implementations found 53 (95%CI: 44, 63) new cases and 12 (95%CI: 9, 15) new cases for the moderate and low transmission settings respectively. Thus, the household level implementation on average was expected to find 174 more cases per 1000 individuals screened in moderate transmission setting and 155 more cases per 1000 individuals screened in low transmission setting compared to the village level implementation. Changing the values of the remaining RACDT parameters had very little impact on the screening efficiency.

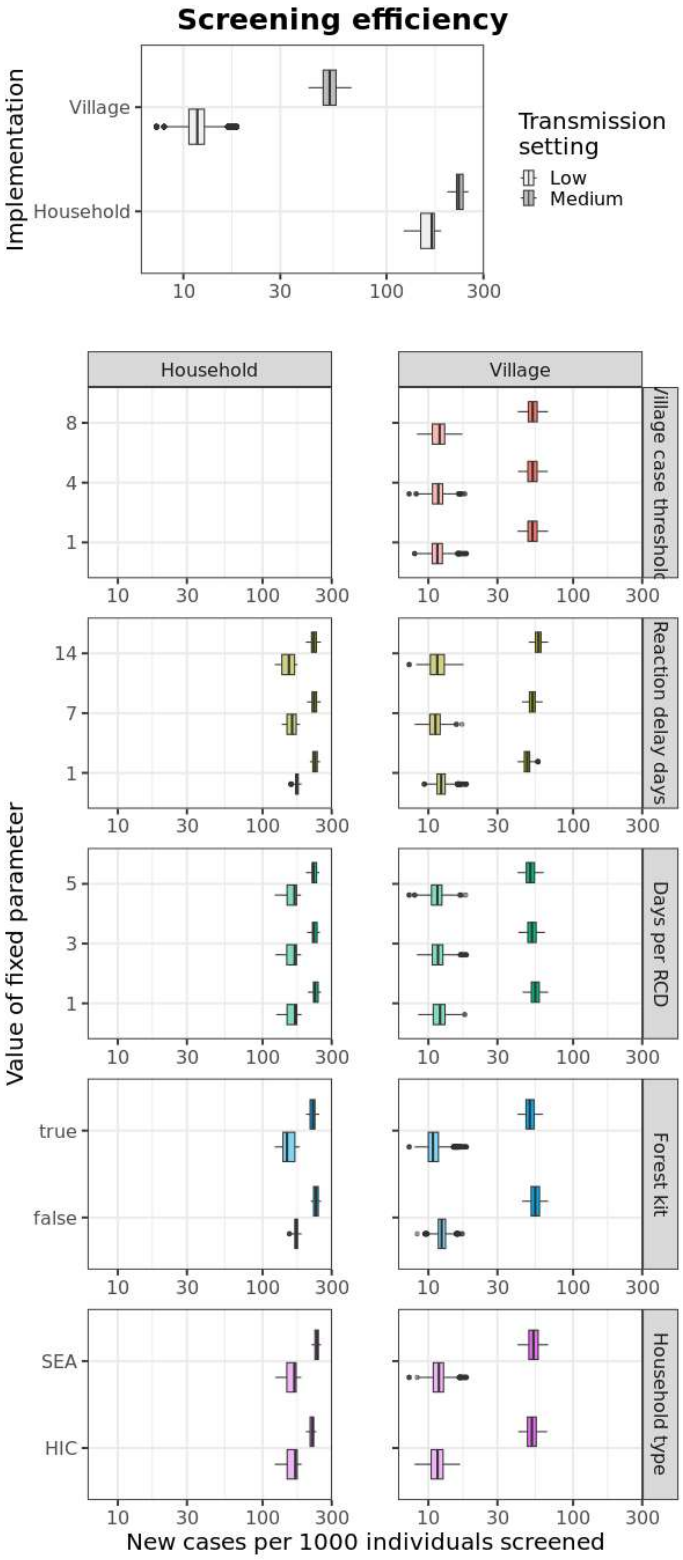


Figure 28: Screening efficiency of RACDT. X-axis was scaled logarithmically at base 10, but the value labels were in their unscaled values.

5.4.3.2 Efficiency of RACDT intervention

In the moderate transmission setting, the efficiency of RACDT on prevalence reduction was predicted to be low. The household level implementation which achieved a PRP per 1000 individuals screened of 1.14 (95%CI: 0.99, 1.29) was slightly better than the village level implementation for which PRP was 0.22 (95%CI: 0.19, 0.26) (Figure 29).

RACDT worked the most efficiently in the low transmission setting with household level implementation – achieving 9.94 PRP per 1000 individuals screened (95%CI: 4.89, 17.2). Village level implementation in the low transmission setting however was much less efficient at only 0.67 PRP per 1000 individuals screened (95%CI: 0.37, 1.36). On average, the household level implementation was nearly 15 times more efficient in prevalence reduction than the village level implementation in the low transmission setting. The second most influential parameter in the low transmission setting was the forest kit, with which a PRP per 1000 individuals screened of 1.64 was achieved on average.

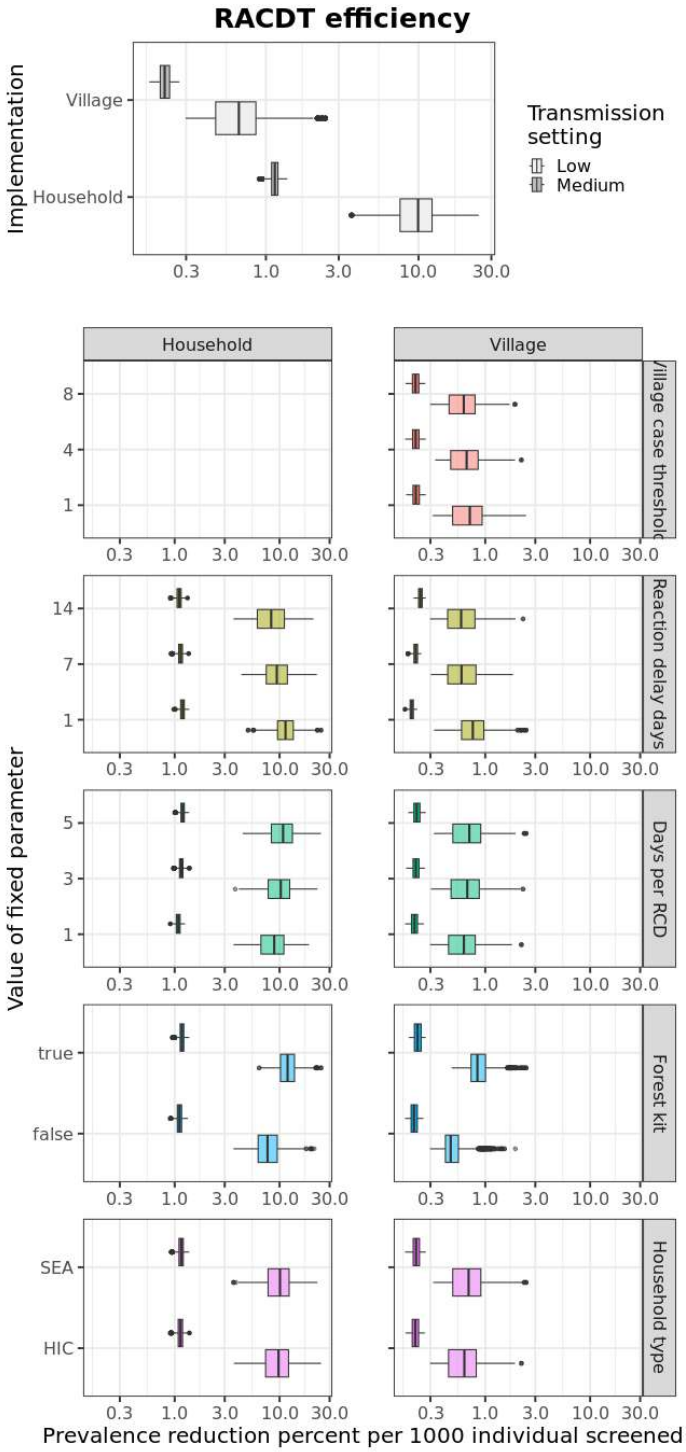


Figure 29: RACDT intervention efficiency. X-axis was scaled logarithmically at base 10, but the value labels were in their unscaled values.

5.4.4 RACDT in an integrated malaria elimination campaign

A scenario analysis was performed for both low and moderate malaria transmission intensities in which RACDT was implemented with a few other malaria interventions. The total simulation time was 10 years. In the presence of malaria posts and vector control, MDA with a coverage of around 53% was performed near the end of the 5th year of the simulation. Three rounds of MDA were then followed by household level RACDT with default parameters for the remainder of the simulation (see default values in Table 6). For a moderate transmission setting, the predicted prevalence rebounded after the MDA rounds were stopped even when MDA was followed by the RACDT, though the rebound was not as high as MDA alone (Figure 30). Elimination was not predicted to be achieved with any of the interventions compared in the moderate transmission setting.

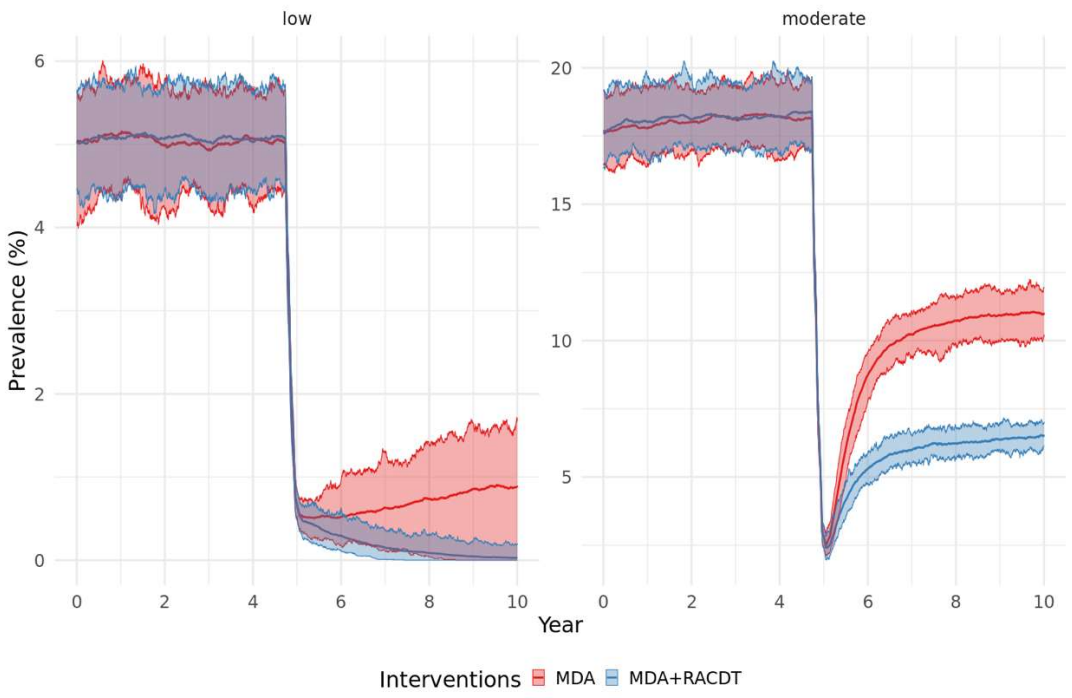


Figure 30: Parasite prevalence after MDA alone vs. MDA followed by RACDT in low and moderate transmission settings. Malaria posts and vector control were implemented as core interventions in both scenarios. The shaded regions indicate the 95% confidence interval around the mean.

For a low transmission setting, when MDA was not followed by RACDT, elimination was predicted in 7% of the simulations over the 5-year period, but on average the prevalence rebounded toward the 1% prevalence level (Figure 30). When MDA was followed by RACDT, the chance of elimination increased over ten-fold. A few of the simulations started to have complete elimination as early as 2.5 years after the RACDT was started. Formulating an integrated malaria elimination campaign involving RACDT is further discussed in the Chapter 6.

5.5 Discussion

Exploring the nuances of reactive interventions in small operational units such as the household level RACDT would be unrealistic and challenging with compartmental models that simulate at the population scale. It requires advanced models that can simulate at the individual scale. We have thus used an established malaria IBM to further develop and integrate the RACDT as a new module.

An essential operational unit of RACDT – a household level grouping of individuals – was not present in the original model, and grouping them randomly would create households which have unrealistic age and gender distribution within each household. We overcame this by using constraint programming techniques explained in detail in Chapter 4 while still maintaining the overall age and gender distribution of the whole population. Especially in the low transmission setting, the importance of household structure for the RACDT in the sensitivity analysis was observed in the model. We found a higher effect of RACDT on a population with household structure from SEA compared to a counterfactual population with different household structure, where the majority of households were smaller in size and with fewer multi-generational household members.

In addition to integrating RACDT as a new intervention, we incorporated the demographic and movement data of forest goers that was analysed in Chapter 3. Compared to the original model where forest goers were assigned randomly and thus could be of any age and gender, the assignment in the modified model was based on the actual data of the forest goers. The forest goers' probability of going into and out of the forest was also calibrated based on the human mobility data tracked by the GPS devices.

From the outset, it was predicted that RACDT alone won't achieve elimination even in a low transmission setting. Many other studies corroborated this prediction.[156,162,163] A combination of interventions with a sustained effort would be required for malaria elimination.[164] And RACDT could be a component of a successful malaria elimination. We explored the various configurations of RACDT so that it could be implemented efficiently and effectively in the right setting at the right time.

5.5.1 Effectiveness

In the two transmission settings we have explored, the predicted effect of RACDT was on average 17% lower in a moderate transmission setting. From a pure effectiveness standpoint, RACDT may not be suitable in a moderate transmission setting, which will become more apparent when we discuss about the efficiency in the following section. Other studies have also shown that RACDT would work well only in the low transmission setting.[154,165]

Choosing the level at which RACDT would operate made the most difference on its effectiveness in the moderate transmission setting. The effectiveness of village and household level implementation converges as the transmission decreases, with both having similar effects at around 5% prevalence. Therefore, the implementation level doesn't matter as long as the transmission is low enough.

At the low transmission setting, the most important factor is the provision of forest kit to forest goers, so that they could be treated when in the forest they become malaria cases. In contrast, the effect of forest kit is predicted to be low in the moderate transmission setting. This suggests

that forest kit provision is not worth it while there are more infections elsewhere in the population.

Remaining factors that are explored in the sensitivity analysis are found to have a small influence on the effect of RACDT on both transmission settings. Logically, the lower number of reaction delay days, the higher number of days per RACDT, and the lower village threshold for RACDT would yield a slightly better result compared to their respective counterparts.

Screening of cases around the index case is a resource intensive part of the RACDT. Efficiency may not matter when the aim is elimination or whenever there is enough funding. In such cases, more aggressive village level RACDT may be preferable because of its slightly higher effectiveness. It may even find more asymptomatic cases, which might have been missed by the limited screening scope of household level RACDT. But when there are resource constraints, their efficiencies must be taken into account.

5.5.2 Efficiency

The predicted effect of village level RACDT is always higher than its household counterpart until the 5% prevalence is reached. This is mainly because more new cases are found at the cost of having to screen more people. For the same reason, screening in village level RACDT based on the index case is always less efficient than the household level RACDT. Implementation level was the single most important parameter that influenced the screening efficiency.

Compared to the village level implementation, the predicted screening efficiency for the household level implementation is 4.3 times and 14.3 times higher in the moderate and low transmission settings respectively (Figure 28). Within the same implementation level, the higher the initial prevalence, the more efficient the screening was. Even though the remaining parameters have minimal influence, configurations that result in higher transmission such as having longer delays in reaction time, having no treatment to the forest goers, etc. increased the screening efficiency by a small margin. The way the screening efficiency translates into the efficiency of RACDT intervention is not straightforward for these parameters.

In the effectiveness section, we stated that the choice of the implementation level doesn't matter if the transmission is low enough. But when factoring in the efficiency, achieving a similar effect for much lower costs from screening less individuals, the household level RACDT would be a better choice. In fact, in the low transmission setting, the household level RACDT is predicted to be 15 times more efficient than the village level RACDT. Having forest kit provision together with the household level RACDT can further improve its efficiency.

For a Southeast Asian population in a low transmission setting, the most efficient configuration that adheres to current malaria treatment of 3 days is the household level RACDT with at most 1 day of reaction delay, and provision of forest kits which is predicted to reduce 81 PRP (95%CI: 75, 91) over a 5-year period. Such efficient RACDT would greatly improve the chance of malaria elimination in combination with other malaria interventions such as MDA, malaria posts, indoor-residual spraying, and provision of long-lasting insecticide treated nets.

5.5.3 Limitations and future work

This study inherits the limitations from the original model. In addition, household structure generated from constraint programming, even though it is better than random family assignment, is far from perfect. The RACDT interventions assume perfect detection of clinical cases, thus the results presented may have been overestimated. In addition, the majority of the malaria clusters in SEA countries are made up of asymptomatic individuals who have low parasite density, escaping the detection by microscopy, RDTs and even by standard dried blood spot PCR.[152] Depending on the sensitivity of screening tests used, the effectiveness and efficiency will vary greatly.

Uncertainty in the model can arise from two sources. Uncertainty due to the stochastic nature of the model was managed by running the model hundreds of times to average out random noise, providing a stable estimate of the expected outcome. Parameter uncertainty, which stems from incomplete knowledge about key biological, behavioral, or intervention-related inputs, was assessed through sensitivity analysis. This study focused specifically on the sensitivity analysis of RACDT-related parameters.

Inclusion of one day of provision of anti-malarial drug in the sensitivity analysis was just to check if the treatment portion of the model worked as intended. Providing only one dose of drug especially an artemisinin based one, would increase the drug resistance and jeopardize the elimination effort. Although the drug resistance was a pertinent issue, its development was not explicitly modelled in the model.

RACDT may have usefulness beyond malaria elimination. It could serve as one of the sentinels of malaria importation and reaction to the importation. Because of the newly added

features into the model, more elaborate investigations for malaria elimination could be conducted in the future.

5.6 Conclusion

Inclusion of a realistic household structure in the model is required to properly assess the impact of reactive malaria interventions particularly when households are considered as an intervention unit. The lower the transmission level, the more effective RACDT is. Without resource limitation, village level RACDT is the most effective type of RACDT. When efficiency is considered, the household level RACDT is predicted to be nearly 15 times more efficient with 9.24 percent reduction in prevalence (PRP) per 1000 individuals screened over a five-year period compared to 0.67 PRP of the village level RACDT in the low transmission setting. Provision of forest kit in the low transmission setting gives a boost to the RACDT by a further 1.84 PRP per 1000 individuals screened. Implementing the household level RACDT with at most 1 day of reaction delay, at least 3 days of treatment per RACDT round while forest kits are provided to the forest goers, is predicted to yield 83 PRP (95%CI: 76, 92) over a 5-year period on average in a low transmission setting in a Southeast Asian country. When used efficiently, RACDT could be an impactful component of a successful malaria elimination strategy.

6 Discussion

6.1 A tale of two models

Throughout the thesis, we have seen two mathematical models of varying complexity and purpose. Presented in Chapter 2, Model 1 is a deterministic compartmental model with SIRT compartments. It focuses on how human mobility could affect the success of an MDA, an intervention suitable for moderate to high malaria transmission settings. Model 2, presented in Chapter 5, is a stochastic spatially-explicit individual-based model which focuses on how to effectively and efficiently implement RACDT, an intervention suitable for low to very low malaria transmission settings.

In a realistic-abstract continuum, Model 1 is on the abstract side of the continuum whereas Model 2 is on the realistic side. Model 1 represents malaria status at the population level while in Model 2, each individual has his/her own malaria status. The process of malaria transmission and recovery is a population level event in Model 1 whereas it is an individual event in Model 2. Model 1 simulates the human mobility as a crude abstract concept. Model 2 incorporates a more refined human mobility at the individual level with special cases such as the forest goers who would go into and out of the forest based on their individual demographic properties.

Because of the difference in structural and conceptual foundations between Model 1 and 2, their assumptions are different as well. Individuals in the subpopulations in Model 1 are assumed to be mixing homogenously ignoring any spatial or social network structure with the exception of the connectedness between the two patches. Individuals in Model 2 live together with their family members in households which in turn are located in their respective villages.

As a person without a computer science background, the biggest hurdle for me during the development of the models especially for Model 2 was the knowledge and skills of the programming language. With that being said, developing Model 1 was extremely easy compared to developing Model 2. The highly abstract nature of compartmental models made Model 1 fit into a handful of mathematical equations. In comparison, even though Model 2 was based on an established malaria model, several thousand lines of new code has to be added to the original model code.

Model 2 required a lot more development resources aside from several months of development time. It has more aspects of reality than Model 1 and subsequently requires a lot more data and parameters. In fact, Model 2 has nearly 20 times more configurable parameters compared to Model 1.

Demographic and mobility information on forest goers in the SEA region was not available at that time. And neither was the household level demographic composition required for household level intervention in Model 2. These additional data requirements of Model 2 commanded two data analyses (Chapter 3 & 4). A mobility data analysis was conducted to calibrate forest goers' mobility. A demographic data analysis and algorithm development were carried out to establish realistic household age composition.

After the models are developed, they are ready to be initialized and run the simulations. In practice, initializing and running the model simulations are vastly different between Model 1 and Model 2. Model 1 can be run directly from an R script in which all the parameters and initial population sizes are defined. For a set of parameters, there is only one output for Model 1. On the other hand, in Model 2, all the C++ code has to be compiled first into an executable

file. Then, one configuration file and at least five data files have to be read in before running. Because of the stochastic nature of Model 2, it has to run multiple times for a single set of parameters, resulting in thousands of output files for a few scenarios. The output files have to be organized, aggregated, and analysed. If there is a need to modify a code or a parameter value, the mentioned steps have to be done all over again. The additional steps involved in running Model 2 on top of the computational overhead of simulating all individuals for the movement and transmission processes for each timestep mandate serious computer resources (at least 16 GB of memory, quad core CPU, 100 GB of free hard drive space, etc).

A list of discerning features between the two models has been tabulated in Table 7.

Aspect	Model 1	Model 2
Type	Deterministic compartmental model	Individual-based model
Disease states	Susceptible 3 types of Infected cases: Clinical, Asymptomatic and Undetected Treated Recovered	Susceptible 2 types of Infected: Clinical and Asymptomatic Treated Recovered (susceptible with added level of immunity) The states are much more nuanced than Model 1 with the possibility of multiple infections in the same individual.

Development process	Once the model structure was conceived, it was quick and easy to develop and test the model. Ready-made packages such as deSolve were available to solve differential equations of the model numerically. The packages were easy to use since just a few lines of code have to be modified between models of similar types.	Developing an IBM for malaria from the ground up would take several years. I based my model on an established model and added new features and capabilities. Even then, learning how the code of the original model work took me several months. Modifying the original code for new features was also a tedious process: requiring re-testing (comparing outputs with the original model for the same input) on every step of the way so as not to break it. Understanding of object-oriented data structures was required to develop the steps of RACDT intervention.
Required programming skills	R programming language was easy to learn within 1 to 2 months.	R was not efficient enough to run IBMs and a low-level programming language like C++ had to be used. Learning C++ took me more than a year.

Initializing the model	The only thing that is needed to initialize was the number of people in each compartment of the model.	Every individual's properties must be initialized before the model runs. This includes the demographic characteristics such as age, gender, occupation (whether a forest goer or not), treatment seeking behaviour (weather a treatment acceptor/afforder or not), infection status, immunity, etc. All of these information must come from real world data for which you are modelling, or some assumptions have to be made.
No. of parameters	23	409
Assumptions and limitations	Homogeneous mixing in each model compartments. Fixed rates for many transitions e.g., recovery rates, rate of immunity loss, etc. Infections in mosquitos were not explicitly modelled.	Individuals live together with other members in households which are located in villages. Their movement between villages is modelled with modified gravity model. Forest goers' movement is calibrated with data from Chapter 3. Transition rates are different for each individual (based on their properties).

		Mosquito infection is modelled at the village level, and each mosquito is equally likely to bite a given individual.
Time taken to run the model	A single model run in a low-end machine for 3650 timesteps (10 years) took at most a few seconds.	A single model run in a high-end machine for 3650 timesteps (10 years) with ~10,000 individuals took over a few minutes.
Processing the output	A set of parameters will output the same result no matter how many times the model is run. Therefore, interpreting the output is straight forward.	Inherent nature of the IBM makes the model stochastic. The output of a single model run cannot be taken at face value. Multiple runs are required and the result of a set of parameter is usually the summary statistics of the output with some credible interval. Additional data analysis steps are required to interpret the output of the model.

Table 7: Comparison of Model 1 and Model 2

The tedious development process and the amount of resources (computation and time) required to run Model 2 would beg the question why we need to use it instead of Model 1. The short answer is because each type of models has its own advantages over the other, and one is more suitable for some specific use cases than the other.

The compartmental models like Model 1 are suitable for quick prototyping and exploring emerging epidemics in a large population. They are suitable for broad-scale interventions like MDA, and for high-level analysis of malaria elimination strategies. They make an excellent starting point or complement to more complex individual-based models. But they cannot accommodate heterogeneity within the population and the accompanying nuances. Interventions like RACDT requires the data resolution at the household composition level to tease out their effect more accurately. IBMs like Model 2 could accommodate such data-driven approach. They are also more adept at capturing the very low malaria transmission level, thus more suitable for modelling the scenarios closer to the malaria elimination.

6.2 Human mobility in the models: from abstraction to data driven approach

Human mobility dictates the spread of a majority of infectious diseases. It is also the main reason for the spread of malaria geographically given that suitable vectors (E.g., *Anopheles minimus*, and *A. dirus* for *P. falciparum* malaria in Thai-Myanmar border area[121]) are available in the area. Thus, malaria models should invariably incorporate human mobility if the model structure and resources permit.

Model 1 has a very crude representation of human mobility. While the assumption that humans mix randomly in the corresponding compartments of the model remains intact, the human mobility between two populations is modelled as a connectedness parameter. Despite its abstract nature, Model 1 is able to explain the impact human mobility has on malaria elimination.

Model 2 incorporates human mobility at the individual level, but it comes with a challenge of making sure the mobility of individuals in the model approximates reality. In the absence of any mobility data, gravity model has been useful in modelling human mobility between population centres of different sizes.[58,59]

In the remote and rural parts of SEA countries, individuals have to go into the forests for foraging, logging, mining, etc for their subsistence. While they are in the forests, they are unprotected for insects such as mosquitos.[106] They might stay overnight in the forest which increases the chances of malaria infection. The mobility of these forest goers has been understudied. There was no fine-scale data available to calibrate their mobility in the malaria models in the region, even though it has been found to be the crucial part of a sustained malaria transmission. To have the demographic characteristics and mobility of forest goers in the model based on data, a movement pattern analysis for the farmers and forest goers in Thai-Myanmar border area was done in Chapter 3.

The analysis found that adults tended to spend more time away from their home compared to children and teens. Among all the forest goers who ever spent overnight in the forest, the median number of overnight stay was 7.5 nights. The majority of forest going activities were done by males with age between 20 and 40 years. They took the most frequent and successive overnight trips to the forests. Through this exercise, we obtained the real-world data on forest goers such as their age and gender, the number of times they go into the

forest per year, and the Kaplan-Meier probability of staying each consecutive night in the forest. These data were used to calibrate the demographic and movement properties of forest goers in the Model 2.

6.3 Household matters

Individual-based models like Model 2 explicitly have individuals' properties and interactions between them to represent reality in a finer scale.[49] As we have seen in Chapter 3, human mobility is greatly influenced by individual's age and gender. However, individuals do not exist by themselves in reality. They live and share resources with other individuals together in households which are a basic unit of organization in society. Members of a household are often linked by family ties or other social relationships. Health care decisions and treatment seeking behaviour are made at the household level. As a consequence, an individual's interactions with the environment (malaria transmission in this case) are further influenced by which household he or she belongs to.

Moreover, when the important aspects of effective and efficient RACDT implementation were explored with Model 2, one of the aspects was the RACDT implementation at the household level. To investigate the nuance of household level RACDT, realistic household structure must be present in the model i.e., the age and gender composition of members within the household must be realistic.

In order to include realistic household structure into Model 2, an analysis of demographic surveys from four SEA countries (Cambodia, Lao PDR, Myanmar, and Vietnam) was performed in Chapter 4. The analysis revealed that the major household types were

common among these SEA countries. These popular household types accounted for the majority of the types in each household size.

Based on these major household types, the rules for constraint programming (CP) algorithm were crafted. The developed algorithm was then used to process the two commonly available aggregated data: the distribution of different household sizes, and the age distribution of household members in them, to generate a synthetic population where major household types were retained. Chapter 4 documents the data analysis and the development of the algorithm in detail.

CP algorithms performed much better than the random assignment and grouping of individuals into households. The demographic properties of the population generated by the best performing CP algorithm was used in the Model 2. In the sensitivity analysis of Model 2, the effect of RACDT was found to be higher in a population that has a SEA household structure compared to a counterfactual population which didn't have a SEA household structure, showing that realistic household structure should be included in the malaria IBMs to improve the accuracy of the predictions particularly when the intervention modelled is implemented at the household level.

6.4 Malaria interventions explored in the thesis

Two malaria interventions were explored in the thesis – MDA which is a type of mass malaria intervention and RACDT which is a type of reactive malaria intervention. Despite their common aim of malaria elimination, the approach they take to achieve it is different.

6.4.1 Mass malaria interventions

Mass drug administration (MDA) involves providing anti-malarial drugs to a large number of people. The major challenge with MDA concerns with achieving high coverage.[78] Fear of adverse events, misinformation and rumours, lack of trust in health authorities, cultural beliefs and perceived low risk of malaria could make people refuse to participate in the MDA.[166-168] Without sufficient coverage, malaria infectious reservoirs may remain unaddressed, undermining the effectiveness and purpose of MDA. It's also important to note that humans, who are the recipient of MDA, are involved in only half of the malaria lifecycle.

With Model 1, we predicted that human mobility between intervention units could impact the success of a malaria MDA and that we could use the connectedness arising from human mobility in our favour when planning for elimination. A meta-population version of herd effect that we termed "assembly effect" was observed by both of the interconnected population when MDA was implemented. The pre-intervention malaria intensity, intervention coverage and the connectedness between the populations were the deciding factors for having a beneficial or detrimental effect from the human mobility.

Model 1 could explain why in Thai-Myanmar boarder area, villages on Thailand side couldn't eliminate malaria no matter how high the malaria intervention coverage they have. Human mobility from Myanmar side which had low malaria intervention coverage diluted the effect of interventions on the Thailand side.

Targeted mass malaria interventions were believed to be cost-effective because resources were directed at the population and places that require the most.[169] They would have high impact because they focus on transmission hotspots thus significantly reduce malaria prevalence and break transmission cycles. But Model 1 pointed out the pitfall of targeted

MDA only in hotspots. Because of the assembly effect, when targeting malaria hotspots, the connected non-hotspot areas should not be neglected. Providing some coverage in the connected non-hotspot could lower the intervention threshold required to achieve elimination in the hotspot. Since achieving high coverage of malaria interventions is already difficult and resource intensive especially for an intervention like MDA, a way to boost the effect of the intervention is necessary to improve the probability of malaria elimination. An example scenario planning for malaria elimination campaign is described in Table 8.

6.4.2 Reactive malaria interventions

Reactive case detection and treatment (RACDT) is one of the reactive malaria interventions. As the name suggests, RACDT is triggered as a reaction to finding a malaria case. Once triggered, it involves screening for malaria in people close to the index case and providing treatment to those who are malaria positive. Unlike MDA, it is applied to a small subset of the population.

After incorporating the data from Chapter 3 (forest goer demographics and mobility) and Chapter 4 (realistic household structure), and building the RACDT module on top of an established malaria IBM, different configurations of RACDT and their impact were explored with Model 2 in Chapter 5.

Since malaria cases tend to cluster in households and neighbourhoods as the malaria prevalence gets lower, RACDT gets more effective and more efficient when the malaria transmission is close to elimination. At the low malaria transmission intensity, the household level RACDT was predicted to be 15-fold more efficient than the village level

RACDT. Implementation at various sizes of radii around the index case was not explicitly modelled but their effect would fall between the household level implementation (smallest radius) and the village level implementation (largest radius). The model also predicted that providing malaria treatment kits to forest goers boosted the impact of the RACDT further.

There are many other types of malaria interventions and their variants available currently, and it is out of the scope of this thesis to explore them all. However, an example guideline for an integrated malaria elimination strategy involving MDA and RACDT is presented in the next section.

6.5 Towards malaria elimination

Each intervention explored in the thesis should not be deployed alone to achieve malaria elimination. MDA alone might achieve temporary elimination, but it would not be sustainable unless other types of interventions are also in place. Several interventions introduced at appropriate time or transmission level are required to eliminate malaria and sustain the elimination status. Based on the lessons we learned throughout the thesis, we can formulate an integrated malaria elimination campaign as described in Table 8.

Steps	**Details**
Adequate core malaria interventions and identification of malaria hotspots	First, we must ensure the quality coverage of core malaria interventions such as early diagnosis and treatment, and long-lasting insecticide-treated nets (LLINs) in all villages within

	the province. We then need to identify the hotspot villages based on prevalence surveys or incidence reports.
Information on connectedness	Depending on the budget and the available timeframe, connectedness between villages can be inferred in several ways. Remote sensing and GIS analysis may be used to infer connectedness through metrics such as distance, estimated population size, and estimated travel time. Human mobility surveys may be conducted to inform connectedness. GPS logger studies may be more expensive and labour intensive but could produce more detailed measures of connectedness. A multi-patch or individual-based model may be used to fit historical data of a similar area to yield an estimate of the connectedness.
Optimisation of intervention coverage across hotspots and non-hotspots	Armed with some information on the connectedness between villages and the location of hotspots, we can strategize to ensure the efficiency and effectiveness of the focal MDA is optimized. All malaria hotspots should aim to reach an MDA coverage over the minimal threshold (i.e., 80% in most contexts). Non-hotspot villages that are connected to the hotspots should get an MDA coverage of at least 30%.

Household level RACDT implementation following the MDA rounds	After MDA has significantly reduced the transmission, there will be fewer cases which could be used as index cases for RACDT. Since the transmission is now lower, implementing RACDT has become more effective and efficient. Core interventions and RACDT must continue until elimination is achieved.
Prevent re-establishment of malaria	RACDT could still be used as a sentinel approach. In island nations without porous borders, stricter measures such as screening of cases at the immigration points could be used to prevent importation of malaria.

Table 8: Example guidelines for a malaria elimination scenario

The results of a modelling exercise with Model 2 conducted for a low transmission setting based on the example guideline (Table 8) is shown in Figure 31. Core interventions brought down malaria transmission intensity lower over time, but they were not predicted to achieve elimination (red lines). When MDA was added to the strategy around year-5, the prevalence was reduced immediately and significantly (blue lines). The prevalence after the MDA was predicted to vary widely – a few simulations even achieved elimination, but on average, the prevalence was predicted to rebound up to the level of simulations with only the core interventions. When MDA was followed by RACDT, the prevalence kept declining and most of the simulations were predicted to achieve elimination (green lines).

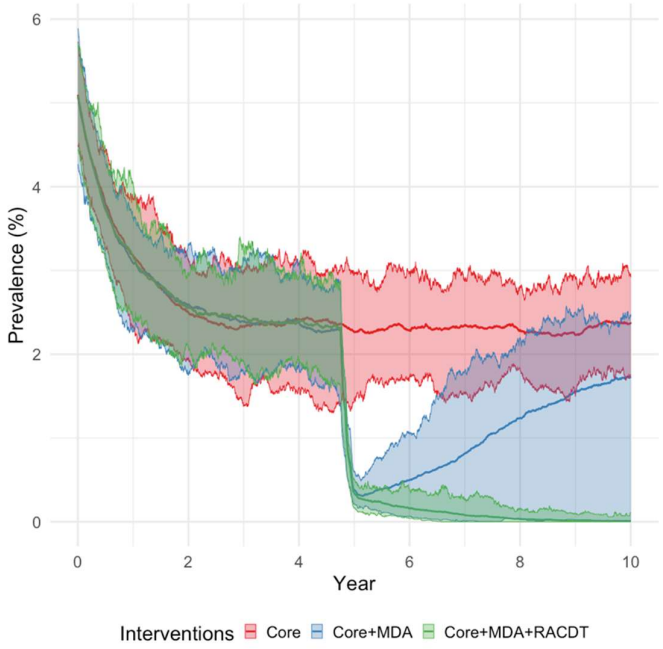


Figure 31: Parasite prevalence where malaria interventions were deployed sequentially in a low transmission setting. The prevalence values for each time point from 100 simulations were summarized into median (thick lines) and 95% credible interval (coloured area bounded by thin lines). Core interventions include early diagnosis and adequate treatment of malaria through village malaria workers or malaria posts, and vector control. MDA was implemented around year 5, which was followed by RACDT until the end of the simulation (year 10).

6.6 Conclusion

This DPhil explored two malaria interventions in the context of malaria elimination using two of mathematical models of varying complexity. The experiences of developing and simulating these two models can inform future modellers on what to expect with each choice of model type and guide them throughout their modelling journey. The thesis established the importance of human mobility for the success of malaria interventions and explained how we could harness its meta-population version of herd effect to our advantage when implementing mass malaria interventions like mass drug administration. The thesis also included the analysis of fine-scale human mobility of rural villagers in Thai-Myanmar

border area. It generated demographic and movement patterns particularly of forest goers who were crucial individuals for malaria transmission. It analysed demographic surveys in 4 Southeast Asia countries to identify the most frequent household types (household level age-distribution for different household sizes) based on which a novel algorithm for generating a synthetic population with realistic household structure was formulated. The resulting mobility data and household structure were parameterized in an IBM to explore the impactful aspects of RACDT implementation. Effectiveness and efficiency of RACDT was predicted to improve with decreasing malaria transmission. At low transmission setting, RACDT was predicted to be the most efficient with household level implementation. Having multi-generational households in the population, and providing forest treatment kits to forest goers were found to improve the effectiveness of household level RACDT meaningfully. Each intervention explored in the thesis may not bring malaria elimination sustainably on their own. But they could be effective and efficient when they are used in a suitable malaria transmission setting. Strategically combining multiple malaria interventions and implementing them at the appropriate time will bring the world towards malaria elimination.

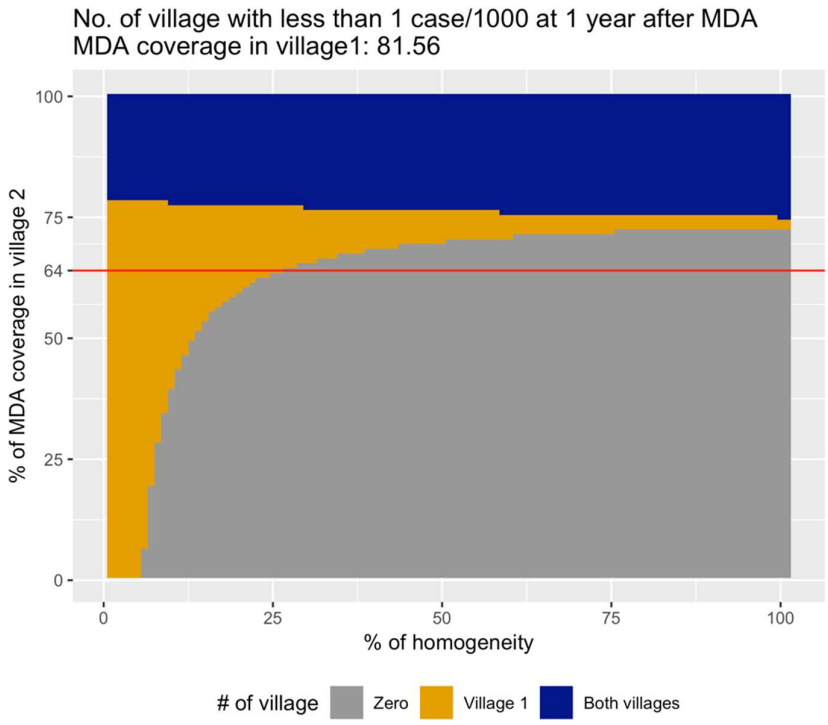
7 Appendix

7.1 Additional materials for Chapter 2

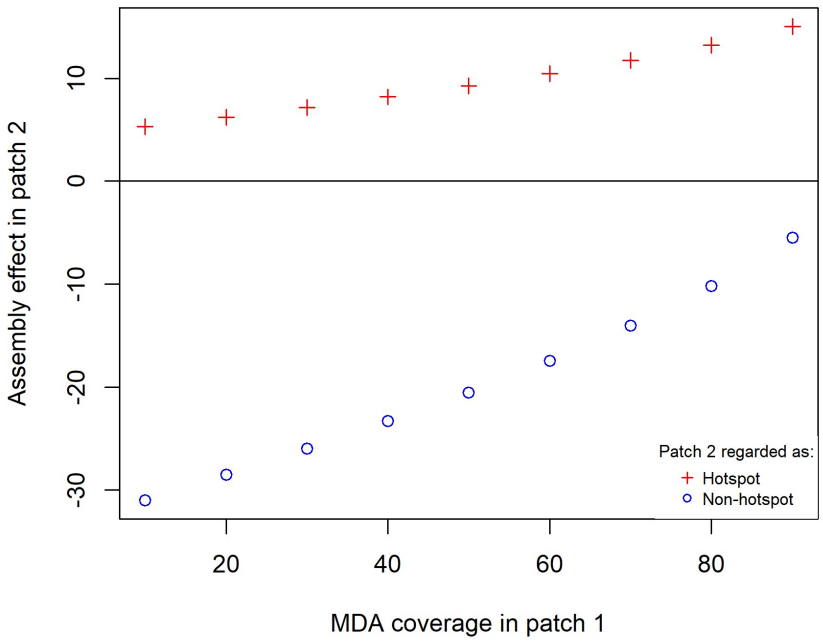
Validation of Model 1

The parameters for the model are calibrated based on the data from Parker et al.[63] In order to do so, we simulated a scenario with the MDA coverages corresponding to the two parts of the village. Human biting rate and the size of the patches were assumed to be the same. In the result of this scenario analysis (Appendix Figure 1), the red line indicates that no elimination threshold would be achieved in the part of the village with lower MDA coverage (64%), at all level of connectedness. But because the two patches were from a single village, the connectedness would be close to 100%. At this connectedness level, both parts of the village will not achieve the elimination threshold. Therefore, the model can explain why there were clinical cases throughout the village 12 months after the completion of the MDA. Table 2 lists of all other model parameter values.

Additional figures

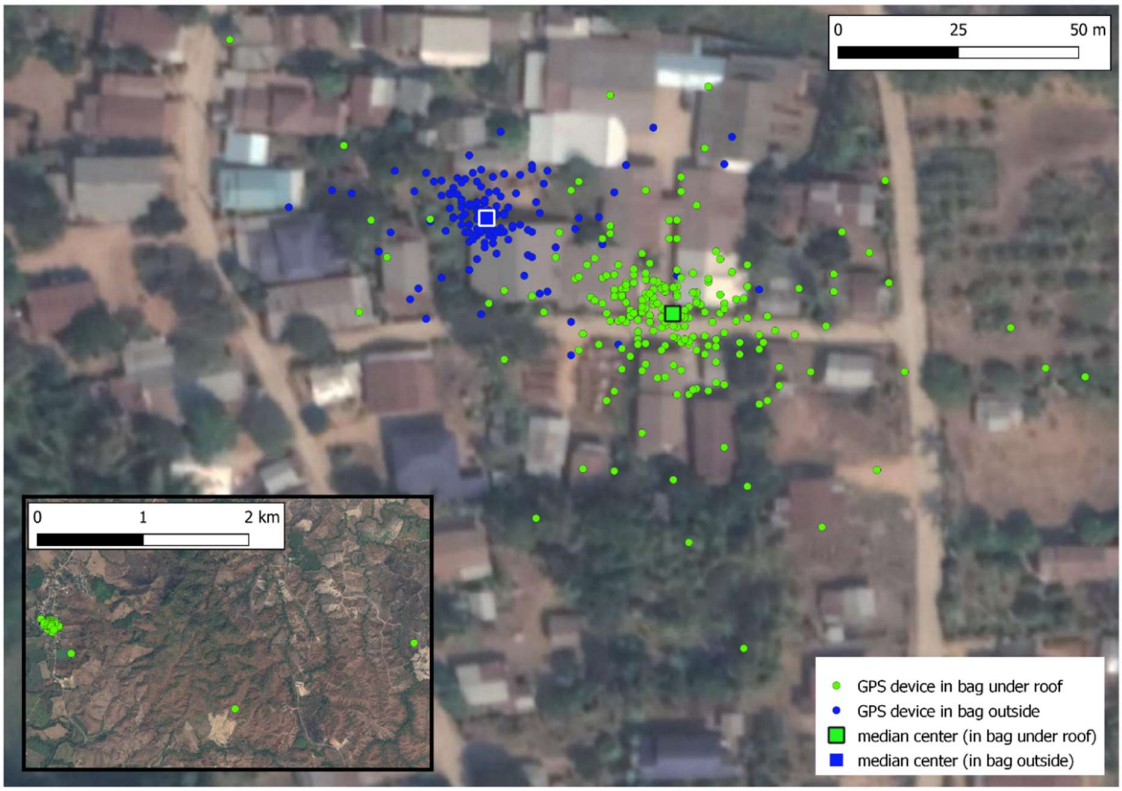


Appendix Figure 1: Result of the model calibration based on Parker et al. 2019

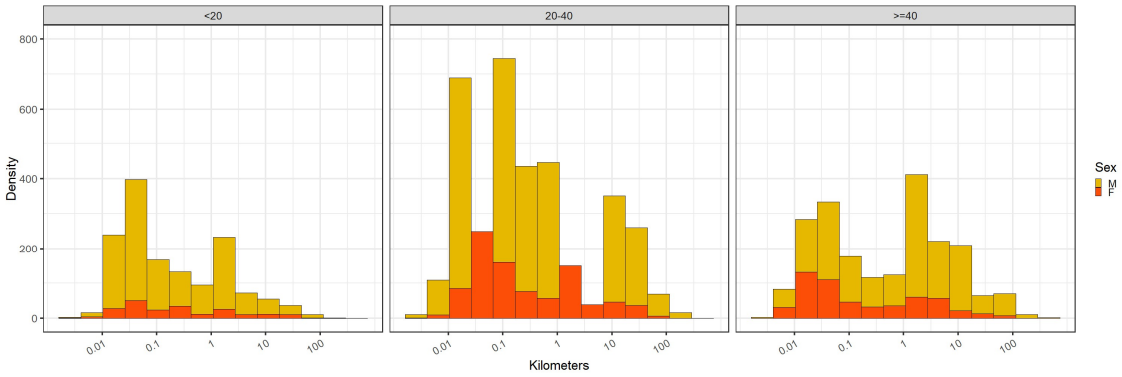


Appendix Figure 2: Assembly effects between a hotspot and a non-hotspot. The non-hotspot has 25% lower pre-intervention incidence compared to the hotspot.

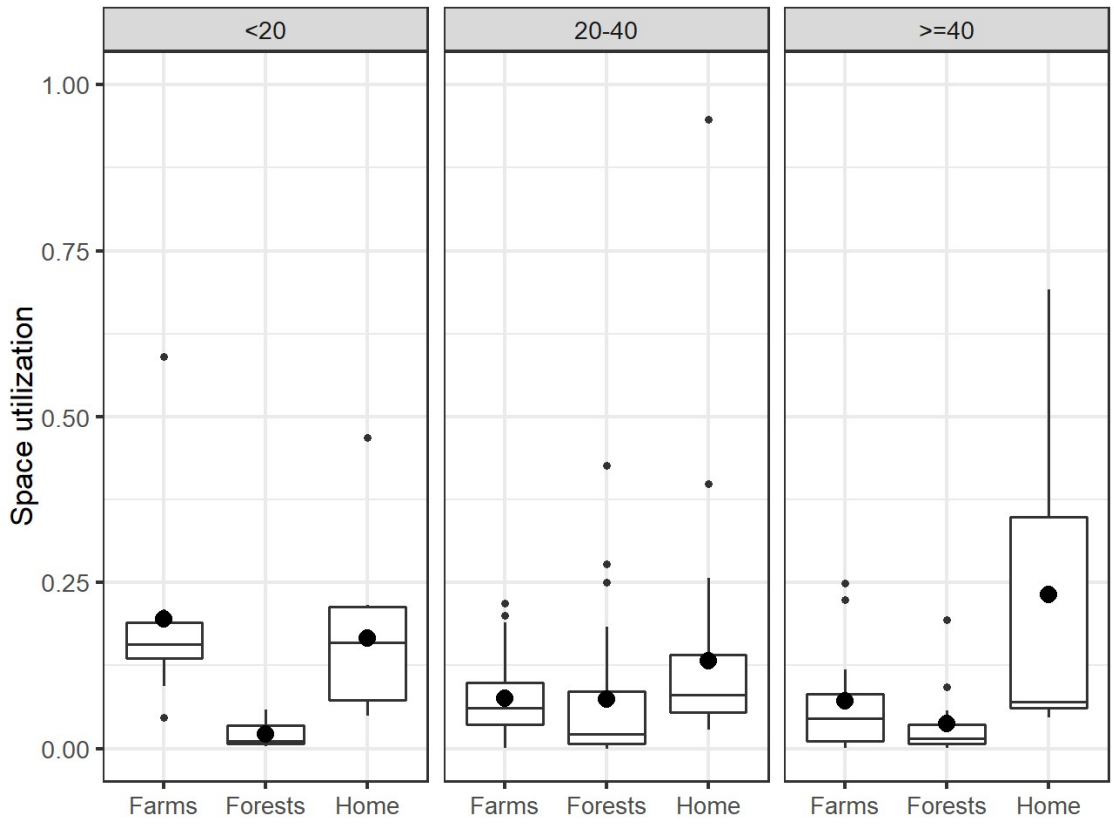
7.2 Additional materials for Chapter 3



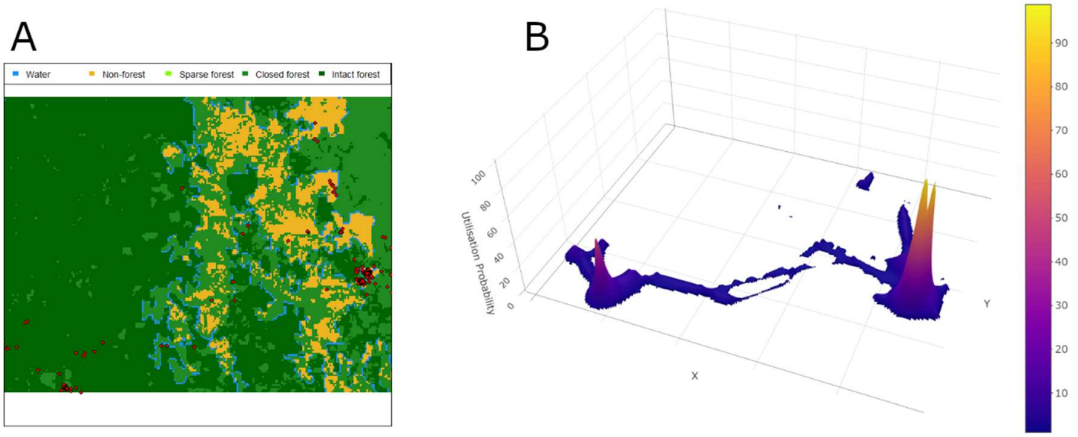
Appendix Figure 3: GPS reading errors in stationary devices. Field tests of GPS device error under stationary conditions were conducted. These tests consisted of placing GPS loggers in stationary locations (tied to a bamboo pole, on a shelf in a house), plotting the points from the device over a period of one week, and measuring the geographic distribution of those points from their geographic centre. Devices were placed inside bags, as this would also be likely for carriage/storage by participants. The mean locational error recordings was larger for the in-doors device. Most erroneous points were within 50 meters of the house. However, a few points were far outside of this range (inset map on bottom left). The maximum distance away from the centre for any of the erroneous points was over 3km away. Only one reading was recorded at this distance and the next reading (30 minutes later) was back within the 50m radius around the house. A standard deviation from the median centre for the worst performing test (in a bag, inside a house) was 266m radius and the value was used as a basis for judging whether or not a participant's movements were likely real or the result of measurement error. This is a conservative estimate.



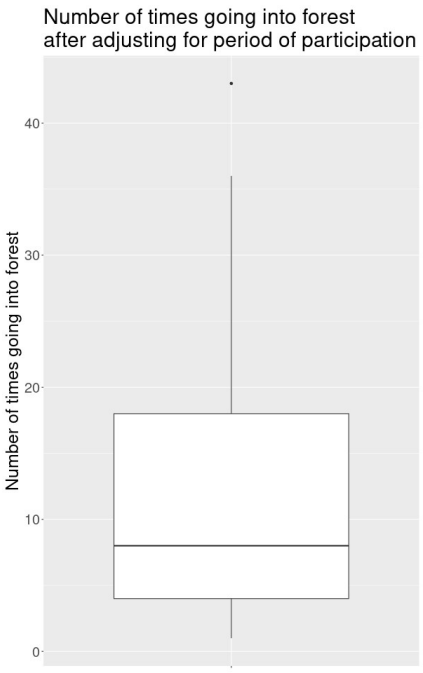
Appendix Figure 4: Frequency histogram of maximum Euclidian distance travelled by the participants in kilometres



*Appendix Figure 5: **BRB** estimation of the utilization of the farm, forest, and home over the participation period for different age groups. The bigger dots represent the mean values, while the smaller dots represent the outliers. Usage of Home was underestimated because of the limitation explained in the Methods section.*

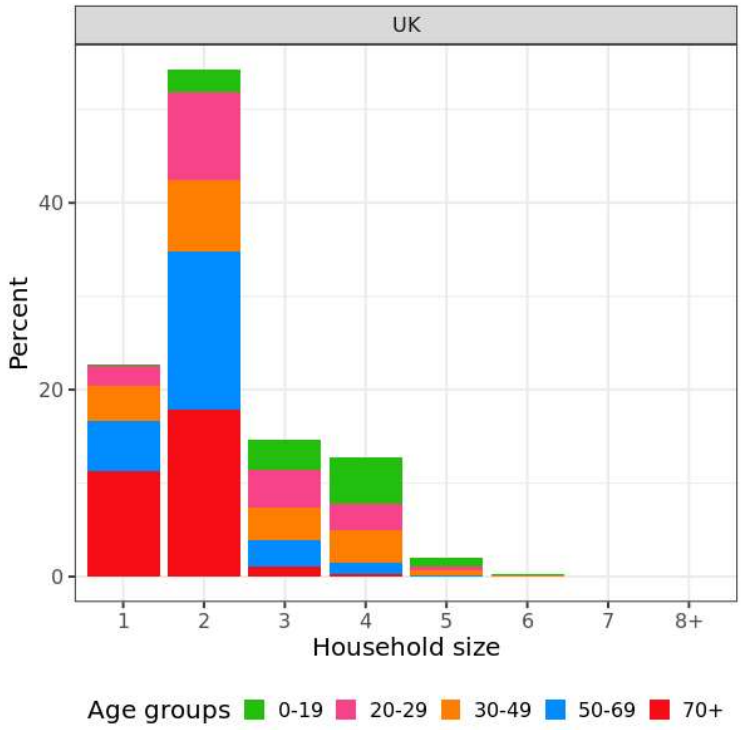


Appendix Figure 6: From GPS data to utilization probability. A. Example of GPS points (red points) recorded for a person. B. Corresponding utilization probability calculated from the GPS points. Its 3D version can be found here: https://hum-mov-patt-fig3-3d.bitbucket.io/

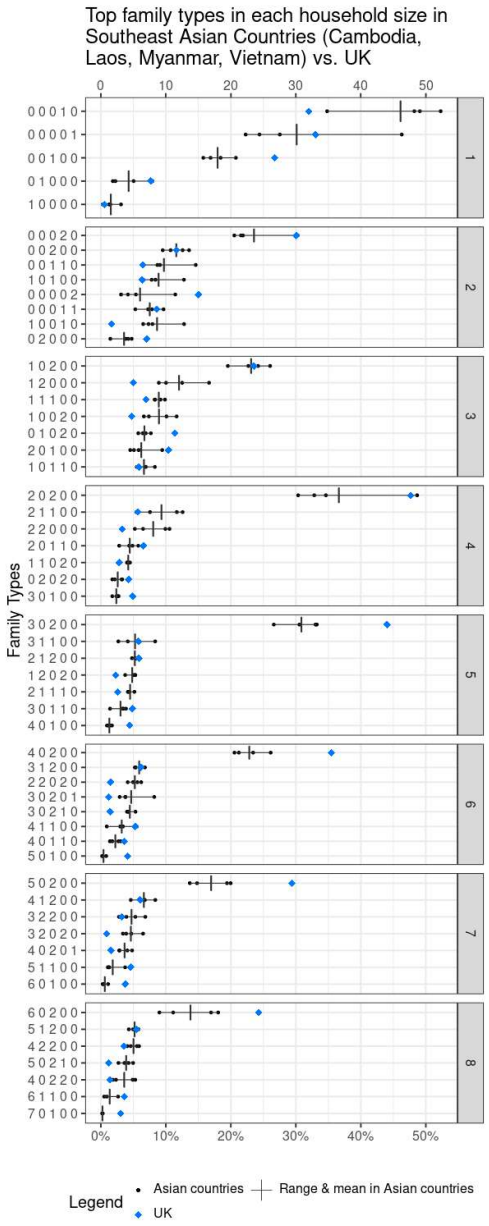


Appendix Figure 7: Frequency of going to the forest by forest goers in a year after adjusting for their participation period

7.3 Additional materials for Chapter 4



Appendix Figure 8: Distribution of household sizes and within-household age distribution in the UK



Appendix Figure 9: Top family types and their percentages within each household size (represented by each horizontal panel) in four South-East Asian countries (black dots) and the UK (blue diamond). Family Types (y-axis) are encoded as explained in the Methods section. X-axis represents the percentage of the family types within the respective household size.

Household Size	CP1	CP2
2	1.2.1. In at least 50% of households, the two members have identical age groups (AG1, AG2, AG3 or AG4)	2.2.1. In at least 18% of households, both members have AG3
	1.2.2. In at least 40% of households, the two members have different age groups	2.2.2. In at least 11% of households, both members have AG2
	1.2.3. No household has two members of AG0 by themselves	2.2.3. In at least 40% of households, the two members have different age groups
3	1.3.1. In at least 40%, there are two parents of same age group (AG 1 or AG2) and one child/teenager (AG0)	2.3.1. In at least 21%, there are two parents of same age group (AG2) and one child/teenager (AG0)

	1.3.2. In at least 20%, there are 2 parents of different age groups (AG1 or AG2) and 1 child/teenager (AG0)	2.3.2. In at least 18%, there are two parents of same age group (AG1 or AG2) and one child/teenager (AG0)
	1.3.3. None of the household have all 3 members being AG0	2.3.3. In at least 14%, all three members have different age groups
4	1.4.1. In 45% of households, there are two parents of same age group (AG 1 or AG2) and two child/teenager (AG0)	2.4.1. In at least 30%, there are two parents of same age group (AG2) and two child/teenager (AG0)
	1.4.2. In 10% of households, there are two parents of different age groups (AG 1 or AG2) and two child/teenager (AG0)	2.4.2. In at least 15%, members could be either (1xAG1 + 1xAG2 + 2xAG0) or (2xAG1 + 2xAG0)
	1.4.3. In the remaining households, number of child/teenage (AG0) is either none, one or three.	2.4.3. In at least 6%, members could be (1xAG2 + 1xAG3 + 2xAG0) or (2xAG3 + 2xAG0)

5	1.5.1. In at least 33%, there are two same-age group parents (AG1, AG2 or AG3) and three children/teenagers (AG0)	2.5.1. In at least 27%, members are 3xAG0 + 2xAG2
	1.5.2. In at least 7%, there are two parents with different age groups (AG1, AG2 or AG3) and three children/teenagers (AG0)	2.5.2. In at least 4%, members are 1xAG0 + 2xAG1 + 2xAG3
	1.5.3. In at least 20%, there are two same-age group parents (AG1, AG2 or AG3), two children/teenagers (AG0) and an aunt/uncle/grandparent (AG2, AG3 or AG4)	2.5.3. In at least 10%, there are three children/teenagers (AG0) with two parents of either identical or different age group (AG1, AG2, or AG3)
	1.5.4. In at least 13%, there are two pairs of same-age adults (each pair can have AG1, AG2, AG3 or AG4) and 1 child/teenager (AG0)	2.5.4. In at least 22%, there are two children/teenagers (AG0) with two parents and an aunt/uncle/grandparent
6	1.6.1. In at least 25%, there are two parents of same age group (AG1, AG2 or AG3) and four children/teenagers (AG0)	2.6.1. In at least 19%, there are four children/teenagers (AG0) and two parents (AG2)

	1.6.2. In at least 6%, there are two parents of different age group (AG1, AG2 or AG3) and four children/teenagers (AG0).	2.6.2. In at least 4%, there are two children/teenagers (AG0), two AG1 and two AG3.
	1.6.3. In at least 20%, there are two parents of same age group (AG1, AG2 or AG3), three children/teenagers (AG0) and one aunt/uncle/grandparent (AG1, AG2, AG3 or AG4)	2.6.3. In at least 13%, there are three children/teenagers (AG0), two parents (AG2) and one aunt/uncle/grandparent whose age is different from the parents (AG1, AG3 or AG4)
	1.6.4. In at least 30%, total number of members with same age group is less than or equal to 2	2.6.4. In at least 10%, there are two children/teenagers (AG0), two parents (AG2) and either one or two of aunt/uncle/grandparent (AG1, AG3 or AG4)
7	1.7.1. In at least 15%, there are two parents of the same age group (AG1, AG2 or AG3), and five children/teenagers (AG0)	2.7.1. In at least 10%, there are four children/teenagers (AG0), two parents (AG2), and one aunt/uncle/grandparent (AG1, AG3 or AG4)
	1.7.2. In at least 19%, there are two parents of the same age group (AG1, AG2 or AG3), four children/teenagers (AG0) and one aunt/uncle/grandparent (AG1, AG2, AG3 or AG4)	2.7.2. In at least 12%, there are three children/teenagers (AG0), two parents (AG2) and a pair of aunt/uncle/grandparents (AG1, AG3 or AG4)

	1.7.3. In at least 25%, there are two parents of the same age group (AG1, AG2 or AG3), three children/teenagers (AG0) and two of aunt, uncle or grandparent (AG1, AG2, AG3 or AG4)	2.7.3. In at least 13%, there are five children/teenagers (AG0) and two parents (AG2)
	1.7.4. In at least 30%, total number of members with same age group is less than or equal to 2 except for AG2 and AG3 for which the maximal possible is 4	2.7.4. In at least 50%, the maximal possible number of members for AG0, AG1, AG2, AG3, and AG4 are 4, 3, 2, 2, and 1 respectively.
		2.7.5. In at least 13%, the maximal possible number of members for AG0, AG1, AG2, AG3, and AG4 are 5, 4, 2, 2, and 1 respectively.
8	1.8.1. In at least 33% of households, there are at five children/teenagers (AG0) or more and at least two adults (AG1, AG2, AG3 or AG4)	2.8.1. In at least 11%, there are six children/teenagers (AG0), and two parents (AG2)
	1.8.2. In at least 21%, there are four children/teenagers (AG0) and the maximum number of members in other age groups is less than 3.	* Adding more rules to household size 8 deteriorates the algorithm rather than improves it. The number of possible household type compositions has greatly

		increased at this point and having many specific constraint rules is no longer useful.
	1.8.3. In at least 20%, the maximal possible number of members for AG0, AG1, AG2, AG3, and AG4 are 3, 3, 3, 2, and 2 respectively.	
	1.8.4. In at least 15%, the maximal possible number of members for AG0, AG1, AG2, AG3, and AG4 are 2, 4, 4, 2, and 2 respectively.	

Appendix Table 1: Rules for algorithms: CP1 and CP2. AG represents age groups. AG0: [0,20), AG1: [20,30), AG2: [30,50), AG3: [50,70), AG4: 70+

7.4 Additional materials for Chapter 5

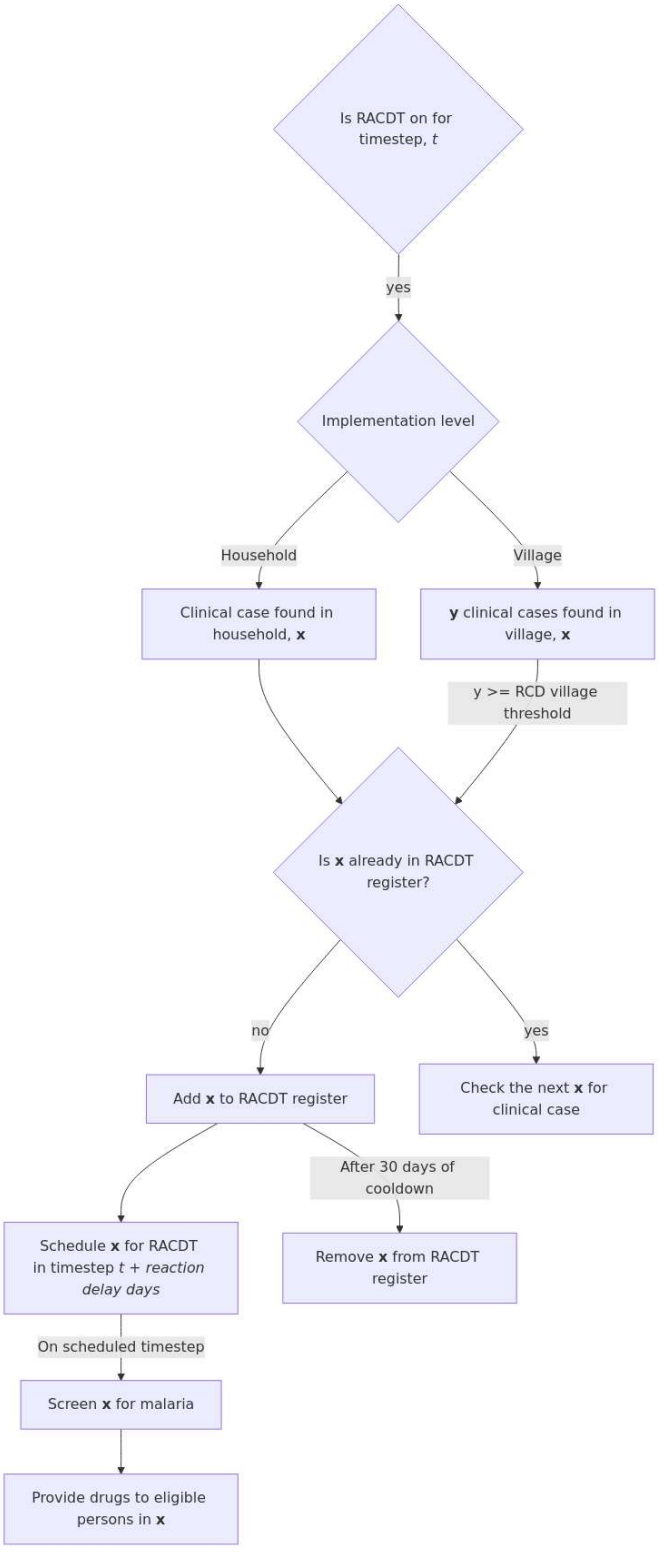
WHO categories of malaria transmission intensity[41]

High transmission: annual parasite incidence of 450 or more cases per 1000 population and a *P. falciparum* prevalence rate of ≥35%.

Moderate transmission: annual parasite incidence of 250–450 cases per 1000 population and a *P. falciparum*/*P. vivax* prevalence rate of 10–35%.

Low transmission: annual parasite incidence of 100–250 cases per 1000 population and a *P. falciparum*/*P. vivax* prevalence rate of 1–10%.

Very low transmission: annual parasite incidence of <100 cases per 1000 population and a *P. falciparum*/*P. vivax* prevalence rate of >0 but <1%.



Appendix Figure 10: Decision flowchart for RACDT intervention

8 References

1 Cox, F. E. G. History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors* **3**, 5 (2010). https://doi.org/10.1186/1756-3305-3-5

2 Boualam, M. A., Pradines, B., Drancourt, M. & Barbieri, R. Malaria in Europe: A Historical Perspective. *Front Med (Lausanne)* **8**, 691095-691095 (2021). https://doi.org/10.3389/fmed.2021.691095

3 Miller, R. L. *et al.* Diagnosis of Plasmodium falciparum infections in mummies using the rapid manual ParaSight™-F test. *Trans. R. Soc. Trop. Med. Hyg.* **88**, 31-32 (1994). https://doi.org/10.1016/0035-9203(94)90484-7

4 Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am J Hum Genet* **77**, 171-192 (2005). https://doi.org/10.1086/432519

5 Gelabert, P., Olalde, I., de-Dios, T., Civit, S. & Lalueza-Fox, C. Malaria was a weak selective force in ancient Europeans. *Sci Rep* **7**, 1377 (2017). https://doi.org/10.1038/s41598-017-01534-5

6 Carter, R. Speculations on the origins of Plasmodium vivax malaria. *Trends Parasitol* **19**, 214-219 (2003). https://doi.org/10.1016/s1471-4922(03)00070-9

7 De Sanctis, V. *et al.* β-Thalassemia Distribution in the Old World: an Ancient Disease Seen from a Historical Standpoint. *Mediterr J Hematol Infect Dis* **9**, e2017018 (2017). https://doi.org/10.4084/mjhid.2017.018

8 Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M. & Snow, R. W. The global distribution and population at risk of malaria: past, present, and future. *The Lancet Infectious Diseases* **4**, 327-336 (2004). https://doi.org/https://doi.org/10.1016/S1473-3099(04)01043-6

9 Piperaki, E. T. & Daikos, G. L. Malaria in Europe: emerging threat or minor nuisance? *Clinical Microbiology and Infection* **22**, 487-493 (2016). https://doi.org/https://doi.org/10.1016/j.cmi.2016.04.023

10 Phillips, R. S. Current Status of Malaria and Potential for Control. *Clinical Microbiology Reviews* **14**, 208-226 (2001). https://doi.org/10.1128/CMR.14.1.208-226.2001

11 Doolan, D. L., Dobaño, C. & Baird, J. K. Acquired Immunity to Malaria. *Clinical Microbiology Reviews* **22**, 13-36 (2009). https://doi.org/10.1128/CMR.00025-08

12 Hoffman, S. L., Vekemans, J., Richie, T. L. & Duffy, P. E. The march toward malaria vaccines. *Vaccine* **33**, D13-D23 (2015). https://doi.org/https://doi.org/10.1016/j.vaccine.2015.07.091

13 Geneva: World Health Organization. *Guidelines for the Treatment of Malaria (3rd edition)*. (World Health Organization, 2015).

14 Chen, I. *et al.* "Asymptomatic" Malaria: A Chronic and Debilitating Infection That Should Be Treated. *PLoS Med.* **13**, e1001942 (2016). https://doi.org/10.1371/journal.pmed.1001942

15 Long, C. A. & Zavala, F. Immune Responses in Malaria. *Cold Spring Harbor perspectives in medicine* **7**, a025577 (2017). https://doi.org/10.1101/cshperspect.a025577

16 Filipe, J. A., Riley, E. M., Drakeley, C. J., Sutherland, C. J. & Ghani, A. C. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. *PLoS Comput Biol* **3**, e255 (2007). https://doi.org/10.1371/journal.pcbi.0030255

17 Langhorne, J., Ndungu, F. M., Sponaas, A.-M. & Marsh, K. Immunity to malaria: more questions than answers. *Nature Immunology* **9**, 725-732 (2008). https://doi.org/10.1038/ni.f.205

18 Geneva: World Health Organization. *WHO Guidelines for malaria*. (2021 (WHO/UCN/ GMP/2021.01)).

19 Geneva: World Health Organization. *Basic Malaria Microscopy: Part I. Learner's guide*. Second edition edn, (World Health Organization, 2010).

20 Geneva: World Health Organization. Malaria diagnosis: memorandum from a WHO meeting. *Bulletin of the World Health Organization* **66** (1988).

21 Geneva: World Health Organization. Malaria rapid diagnostic test performance. (Geneva, 2016).

22 Geneva: World Health Organization. *WHO guidelines for malaria - 16 October 2023*. (2023).

23 Molyneux, D. H. *et al.* Certifying Guinea worm eradication: current challenges. *The Lancet* 396, 1857-1860 (2020). https://doi.org/10.1016/S0140-6736(20)32553-8

24 Chumakov, K., Ehrenfeld, E., Agol, V. I. & Wimmer, E. Polio eradication at the crossroads. *The Lancet Global Health* 9, e1172-e1175 (2021). https://doi.org/10.1016/S2214-109X(21)00205-9

25 Dowdle, W. R. The principles of disease elimination and eradication. *Bulletin of the World Health Organization* 76 Suppl 2, 22-25 (1998).

26 Geneva: World Health Organization. State of inequality: HIV, tuberculosis and malaria. (2021).

27 Geneva: World Health Organization. *World malaria report 2023*. (Geneva, WHO, 2023).

28 Nájera, J. A., González-Silva, M. & Alonso, P. L. Some Lessons for the Future from the Global Malaria Eradication Programme (1955–1969). *PLoS Med.* 8, e1000412 (2011). https://doi.org/10.1371/journal.pmed.1000412

29 Geneva: World Health Organization. *WHO certifies Sri Lanka malaria-free*, <https://www.who.int/southeastasia/news/detail/05-09-2016-who-certifies-sri-lanka-malaria-free> (2016).

30 Ahmed, A. *et al.* Eliminating malaria in conflict zones: public health strategies developed in the Sri Lanka Civil War. *BMJ Global Health* **6**, e007453 (2021).

31 Cao, J. *et al.* Achieving malaria elimination in China. *The Lancet Public Health* **6**, e871-e872 (2021). https://doi.org/10.1016/S2468-2667(21)00201-2

32 Geneva: World Health Organization. World malaria report 2021. (2021).

33 *Bill & Melinda Gates Foundation Supports International Task Force For Disease Eradication*, <https://www.gatesfoundation.org/ideas/media-center/press-releases/2000/11/international-task-force-for-disease-eradication> (2000).

34 Gates, M. *Melinda French Gates - Malaria Forum - Bill & Melinda Gates Foundation*, <https://www.gatesfoundation.org/ideas/speeches/2007/10/melinda-french-gates-malaria-forum> (2007).

35 Roll Back Malaria Partnership. *THE GLOBAL MALARIA ACTION PLAN: For a malaria free world*. (UNHCR, 2008).

36 UN News. *Leaders at UN launch campaign to virtually eliminate malaria deaths by 2015*, <https://news.un.org/en/story/2008/09/274952-leaders-un-launch-campaign-virtually-eliminate-malaria-deaths-2015> (2008).

37 *Malaria Eradication Research Agenda (malERA) Initiative*, <https://www.isglobal.org/en/-/malaria-eradication-research-agenda-malera-initiative> (2018).

38 The Global Fund. Results Report 2021. (The Global Fund to fight AIDS, Tuberculosis and Malaria, Geneva, 2021).

39 Geneva: World Health Organization. *Global technical strategy for malaria, 2016-2030*. (2015).

40 Geneva: World Health Organization. *Global technical strategy for malaria 2016-2030, 2021 update*. (2021).

41 Geneva: World Health Organization. *A framework for malaria elimination*. (2017).

42 Aung, Y. N. The post-coup health crisis in Myanmar is not a local issue, it is a ticking time-bomb for the region. <https://blogs.lse.ac.uk/globalhealth/2021/06/23/the-post-coup-health-crisis-in-myanmar-is-not-a-local-issue-it-is-a-ticking-time-bomb-for-the-region> (2021).

43 Moos, B., Roberts, R. & Aye, M. The Myanmar military coup: Propelling the 2030 milestones for neglected tropical diseases further out of reach. *PLoS Negl Trop Dis* **15**, e0009532 (2021). https://doi.org/10.1371/journal.pntd.0009532

44 Geneva: World Health Organization. *WHO recommends groundbreaking malaria vaccine for children at risk*, <https://www.who.int/news/item/06-10-2021-who-recommends-groundbreaking-malaria-vaccine-for-children-at-risk> (2021).

45 Geneva: World Health Organization. Full Evidence Report on the RTS,S/AS01 Malaria Vaccine. (Geneva, 2021).

46 *Oxford English Dictionary*. (Oxford University Press, 2020).

47 Bailey, N. *The biomathematics of malaria*. First edition edn, (Charles Griffin & Company Ltd, 1986).

48 *Mathematical model - Wikipedia*, <https://en.wikipedia.org/wiki/Mathematical_model> (2022).

49 Vynnycky, E. & White, R. *An introduction to infectious disease modelling*. (OUP oxford, 2010).

50 Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals*. (Princeton university press, 2011).

51 Kermack, W. O. & McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **115**, 700-721 (1927).

52 Ross, R. *Report on the Prevention of Malaria in Mauritius*. (J. & A. Churchill, 1908).

53 Smith, D. L. *et al.* Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. *PLoS Pathogens* **8** (2012). https://doi.org/10.1371/journal.ppat.1002588

54 Maude, R. J. *et al.* The last man standing is the most resistant: Eliminating artemisinin-resistant malaria in Cambodia. *Malaria Journal* **8** (2009). https://doi.org/10.1186/1475-2875-8-31

55 Cohen, J. M. *et al.* Malaria resurgence: a systematic review and assessment of its causes. *Malaria Journal* **11**, 122 (2012). https://doi.org/10.1186/1475-2875-11-122

56 Sturrock, H. J. W., Roberts, K. W., Wegbreit, J., Ohrt, C. & Gosling, R. D. Tackling Imported Malaria: An Elimination Endgame. *Am J Trop Med Hyg* **93**, 139-144 (2015). https://doi.org/10.4269/ajtmh.14-0256

57 Lyttleton, C. Deviance and resistance: Malaria elimination in the greater Mekong subregion. *Social Science & Medicine* **150**, 144-152 (2016). https://doi.org/https://doi.org/10.1016/j.socscimed.2015.12.033

58 Erlander, S. & Stewart, N. F. *The gravity model in transportation analysis: theory and extensions*. Vol. 3 (Vsp, 1990).

59 Kwon, O.-H., Hong, I., Jung, W.-S. & Jo, H.-H. Multiple gravity laws for human mobility within cities. *EPJ Data Science* **12**, 57 (2023). https://doi.org/10.1140/epjds/s13688-023-00438-x

60 Network., G. B. o. D. C. Global Burden of Disease Study 2017 (GBD 2017) Results., (Institute for Health Metrics and Evaluation (IHME), Seattle, United States, 2019).

61 John, T. J. & Samuel, R. Herd immunity and herd effect: New insights and definitions. *Eur. J. Epidemiol.* **16**, 601-606 (2000). https://doi.org/10.1023/A:1007626510002

62 Hawley, W. A. *et al.* Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. *American Journal of Tropical Medicine and Hygiene* **68**, 121-127 (2003). https://doi.org/10.4269/ajtmh.2003.68.121

63 Parker, D. M. *et al.* Potential herd protection against Plasmodium falciparum infections conferred by mass antimalarial drug administrations. *eLife* **8**, e41023 (2019). https://doi.org/10.7554/eLife.41023

64 Wells, J. V. & Richmond, M. E. Populations, metapopulations, and species populations: What are they and who should care? *Wildlife Society Bulletin* **23**, 458-462 (1995).

65 Eradication, W. S. A. G. o. M. Malaria eradication: benefits, future scenarios and feasibility: executive summary. (Geneva: World Health Organization, 2019).

66 Carter, R., Mendis, K. N. & Roberts, D. Spatial targeting of interventions against malaria. *Bulletin of the World Health Organization* **78**, 1401-1411 (2000).

67 Bousema, T. *et al.* Hitting Hotspots: Spatial Targeting of Malaria for Control and Elimination. *PLoS Med* **9**, e1001165 (2012). https://doi.org/10.1371/journal.pmed.1001165

68 Saita, S. *et al.* Human population movement and behavioural patterns in malaria hotspots on the Thai–Myanmar border: implications for malaria elimination. *Malaria Journal* **18** (2019). https://doi.org/10.1186/s12936-019-2704-3

69 Zaw, M. T. *et al.* Asymptomatic and sub-microscopic malaria infection in Kayah State, eastern Myanmar. *Malaria Journal* **16**, 138 (2017). https://doi.org/10.1186/s12936-017-1789-9

70 Cosner, C. *et al.* The effects of human movement on the persistence of vector-borne diseases. *J. Theor. Biol.* **258**, 550-560 (2009). https://doi.org/10.1016/j.jtbi.2009.02.016

71 Prosper, O., Ruktanonchai, N. & Martcheva, M. Assessing the role of spatial heterogeneity and human movement in malaria dynamics and control. *J. Theor. Biol.* **303**, 1-14 (2012). https://doi.org/10.1016/j.jtbi.2012.02.010

72 R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2019).

73 Soetaert, K., Petzoldt, T. & Setzer, R. W. Solving differential equations in R: package deSolve. *Journal of Statistical Software* **33**, 1-25 (2010). https://doi.org/10.18637/jss.v033.i09

74 Eddelbuettel, D. & François, R. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40**, 1-18 (2011). https://doi.org/10.18637/jss.v040.i08

75 Sarkar, D. *Lattice: Multivariate Data Visualization with R*. (Springer, 2008).

76 Kangoye, D. T. *et al.* Malaria hotspots defined by clinical malaria, asymptomatic carriage, PCR and vector numbers in a low transmission area on the Kenyan Coast. *Malaria Journal* **15** (2016). https://doi.org/10.1186/s12936-016-1260-3

77 Kajeechiwa, L. *et al.* Community engagement for the rapid elimination of malaria: the case of Kayin State, Myanmar. *Wellcome Open Research* **2** (2017). https://doi.org/10.12688/wellcomeopenres.12051.1

78 Newby, G. *et al.* Review of Mass Drug Administration for Malaria and Its Operational Challenges. *Am J Trop Med Hyg* **93**, 125-134 (2015). https://doi.org/10.4269/ajtmh.14-0254

79 Pell, C. L. *et al.* Community engagement, social context and coverage of mass anti-malarial administration: Comparative findings from multi-site research in the Greater Mekong sub-Region. *PLOS ONE* **14**, e0214280 (2019). https://doi.org/10.1371/journal.pone.0214280

80 von Seidlein, L. *et al.* The impact of targeted malaria elimination with mass drug administrations on falciparum malaria in Southeast Asia: A cluster randomised trial. *PLoS Med.* **16** (2019). https://doi.org/10.1371/journal.pmed.1002745

81 Stoddard, S. T. *et al.* The role of human movement in the transmission of vector-borne pathogens. *PLoS Negl Trop Dis* **3**, e481 (2009).

82 Reiner Jr, R. C., Stoddard, S. T. & Scott, T. W. Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal. *Epidemics* **6**, 30-36 (2014).

83 Stock, R. Distance and the utilization of health facilities in rural Nigeria. *Social science & medicine* **17**, 563-570 (1983).

84 Jordan, H., Roderick, P., Martin, D. & Barnett, S. Distance, rurality and the need for care: access to health services in South West England. *Int J Health Geogr* **3**, 1-9 (2004).

85 Blanford, J. I., Kumar, S., Luo, W. & MacEachren, A. M. It'sa long, long walk: accessibility to hospitals, maternity and integrated health centers in Niger. *Int J Health Geogr* **11**, 1-15 (2012).

86 Prothero, R. M. Disease and mobility: a neglected factor in epidemiology. *Int J Epidemiol* **6**, 259-267 (1977). https://doi.org/10.1093/ije/6.3.259

87 Tatem, A. J. *et al.* The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. *Malaria journal* **8**, 1-12 (2009).

88 Bharti, N. *et al.* Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. *Science* **334**, 1424-1427 (2011).

89 Wesolowski, A. *et al.* Quantifying the impact of human mobility on malaria. *Science* **338**, 267-270 (2012).

90 Vazquez-Prokopec, G. M. *et al.* Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment. *PloS one* **8**, e58802 (2013).

91 Blanford, J. I., Huang, Z., Savelyev, A. & MacEachren, A. M. Geo-located tweets. Enhancing mobility maps and capturing cross-border movement. *PloS one* **10**, e0129202 (2015).

92 Searle, K. M. *et al.* Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia. *Royal Society open science* **4**, 170046 (2017).

93 Prothero, R. M. Malaria, Forests and People in Southeast Asia. *Singapore Journal of Tropical Geography* **20**, 76-85 (1999). https://doi.org/https://doi.org/10.1111/1467-9493.00044

94 Singhanetra-Renard, A. Malaria and mobility in Thailand. *Soc Sci Med* **37**, 1147-1154 (1993). https://doi.org/10.1016/0277-9536(93)90254-2

95 Pindolia, D. K. *et al.* Human movement data for malaria control and elimination strategic planning. *Malaria journal* **11**, 1-16 (2012).

96 Wesolowski, A. *et al.* Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones. *Scientific reports* **4**, 1-7 (2014).

97 Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W. & Buckee, C. O. The impact of biases in mobile phone ownership on estimates of human mobility. *Journal of the Royal Society Interface* **10**, 20120986 (2013).

98 Williams, N. E., Thomas, T. A., Dunbar, M., Eagle, N. & Dobra, A. Measures of human mobility using mobile phone records enhanced with GIS data. *PloS one* **10**, e0133630 (2015).

99 Elgethun, K., Yost, M. G., Fitzpatrick, C. T., Nyerges, T. L. & Fenske, R. A. Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time–location patterns. *Journal of exposure science & environmental epidemiology* **17**, 196-206 (2007).

100 Elgethun, K., Fenske, R. A., Yost, M. G. & Palcisko, G. J. Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument. *Environ Health Perspect* **111**, 115-122 (2003).

101 James, P. *et al.* Comparing GPS, log, survey, and accelerometry to measure physical activity. *American journal of health behavior* **40**, 123-131 (2016).

102 Duncan, D. T. *et al.* Feasibility and acceptability of global positioning system (GPS) methods to study the spatial contexts of substance use and sexual risk behaviors among young men who have sex with men in New York City: A p18 cohort sub-study. *PloS one* **11**, e0147520 (2016).

103 Stothard, J. R., Sousa-Figueiredo, J. C., Betson, M., Seto, E. Y. & Kabatereine, N. B. Investigating the spatial micro-epidemiology of diseases within a point-prevalence sample: a field applicable method for rapid mapping of households using low-cost GPS-dataloggers. *Trans. R. Soc. Trop. Med. Hyg.* **105**, 500-506 (2011).

104 Carrasco-Escobar, G. *et al.* Open-source 3D printable GPS tracker to characterize the role of human population movement on malaria epidemiology in river networks: A proof-of-concept study in the Peruvian Amazon. *Frontiers in public health* **8**, 563 (2020).

105 Fornace, K. M. *et al.* Local human movement patterns and land use impact exposure to zoonotic malaria in Malaysian Borneo. *Elife* **8**, e47602 (2019).

106 Bannister-Tyrrell, M. *et al.* Forest Goers and Multidrug-Resistant Malaria in Cambodia: An Ethnographic Study. *Am J Trop Med Hyg* **100**, 1170-1178 (2019). https://doi.org/10.4269/ajtmh.18-0662

107 Morris, G. & Conner, L. M. Assessment of accuracy, fix success rate, and use of estimated horizontal position error (EHPE) to filter inaccurate data collected by a common commercially available GPS logger. *PLOS ONE* **12**, e0189020 (2017). https://doi.org/10.1371/journal.pone.0189020

108 QGIS Geographic Information System (Open Source Geospatial Foundation Project, 2019).

109 ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York, 2009).

110 Urbanek, S. *proj4: A simple interface to the PROJ.4 cartographic projections library*. (2012).

111 Pebesma, E. & Bivand, R. *sp: Classes and Methods for Spatial Data*. (2018).

112 Wickham, H. *reshape: Flexibly Reshape Data*. (2018).

113 Bivand, R., Keitt, T. & Rowlingson, B. *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. (2019).

114 Hijmans, R. J. *raster: Geographic Data Analysis and Modeling*. (2019).

115 Benhamou, S. Dynamic Approach to Space and Habitat Use Based on Biased Random Bridges. *PLOS ONE* **6**, e14592 (2011). https://doi.org/10.1371/journal.pone.0014592

116 Calenge, C. & Fortmann-Roe, c. f. S. *adehabitatHR: Home Range Estimation*. (2019).

117 Lwin, T. The prevailing synoptic situations in Myanmar. (2000). A report submitted at the Workshop Agricultural Meteorology. Department of Meteorology and Hydrology, Yangon, Myanmar

118 Htway, O. & Matsumoto, J. Climatological onset dates of summer monsoon over Myanmar. *International Journal of Climatology* **31**, 382-393 (2011). https://doi.org/10.1002/joc.2076

119 Edwards, H. M. *et al.* Transmission risk beyond the village: entomological and human factors contributing to residual malaria transmission in an area approaching malaria elimination on the Thailand–Myanmar border. *Malaria journal* **18**, 1-20 (2019).

120 Haddawy, P. *et al.* Effects of COVID-19 government travel restrictions on mobility in a rural border area of Northern Thailand: A mobile phone tracking study. *Plos one* **16**, e0245842 (2021).

121 Parker, D. M., Carrara, V. I., Pukrittayakamee, S., McGready, R. & Nosten, F. H. Malaria ecology along the Thailand–Myanmar border. *Malar J* **14** (2015). https://doi.org/10.1186/s12936-015-0921-y

122 Nofal, S. D. *et al.* How can interventions that target forest-goers be tailored to accelerate malaria elimination in the Greater Mekong Subregion? A systematic

review of the qualitative literature. *Malar J* **18**, 32 (2019). https://doi.org/10.1186/s12936-019-2666-5

123 Gao, B. *et al.* Determinants of MDA impact and designing MDAs towards malaria elimination. *eLife* **9**, e51773 (2020). https://doi.org/10.7554/eLife.51773

124 Tusting, L. S. *et al.* Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis. *Lancet (London, England)* **382**, 963-972 (2013). https://doi.org/10.1016/S0140-6736(13)60851-X

125 Degarege, A., Fennie, K., Degarege, D., Chennupati, S. & Madhivanan, P. Improving socioeconomic status may reduce the burden of malaria in sub Saharan Africa: A systematic review and meta-analysis. *PLoS One* **14**, e0211205 (2019). https://doi.org/10.1371/journal.pone.0211205

126 Pierce, H., Gibby, A. L. & Forste, R. Caregiver Decision-Making: Household Response to Child Illness in sub-Saharan Africa. *Popul Res Policy Rev* **35**, 581-597 (2016). https://doi.org/10.1007/s11113-016-9396-y

127 Sulaiman, S. K. *et al.* A systematic review and meta-analysis of the prevalence of caregiver acceptance of malaria vaccine for under-five children in low-income and middle-income countries (LMICs). *PLoS One* **17**, e0278224 (2022). https://doi.org/10.1371/journal.pone.0278224

128 Tangcharoensathien, V. *et al.* Children and Women's Health in South East Asia: Gap Analysis and Solutions. *Int J Environ Res Public Health* **17** (2020). https://doi.org/10.3390/ijerph17103366

129 Laband, D. N. & Lentz, B. F. Occupational inheritance in agriculture. *American Journal of Agricultural Economics* **65**, 311-314 (1983).

130 Moeckel, R., Spiekermann, K. & Wegener, M. in *Proceedings of the 8th international conference on computers in urban planning and urban management (CUPUM).* 1-18.

131 Ye, X., Konduri, K., Pendyala, R. M., Sana, B. & Waddell, P. in *88th Annual Meeting of the transportation research Board, Washington, DC.*

132 Barthelemy, J. & Toint, P. L. Synthetic Population Generation Without a Sample. (2012). https://doi.org/trsc.1120.0408

133 Lenormand, M. & Deffuant, G. Generating a synthetic population of individuals in households: Sample-free vs sample-based methods. (2012). https://doi.org/10.18564/jasss.2319

134 Fournier, N., Christofa, E., Akkinepally, A. P. & Azevedo, C. L. Integrated population synthesis and workplace assignment using an efficient optimization-based person-household matching method. *Springer US* (2021). https://doi.org/https://hdl.handle.net/1721.1/130430

135 ISO. *ISO\IEC 14882:2014 Information technology — Programming languages — C++*. Fourth edn, (International Organization for Standardization (ISO), 2014).

136 ICF. *Demographic and Health Surveys (various) [Datasets]* (Rockville, Maryland: ICF [Distributor], 2005-2016).

137 UNICEF. *Multiple Indicator Cluster Surveys (MICS6) - Lao PDR [Dataset]* (ed UNICEF) (2017).

138 Office for National Statistics. *2011 Census Ad Hoc Household Composition (Age Groups) Safeguarded Tables (Lower Layer Super Output Area): England and Wales* (UK Data Service, 2021).

139 foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ... (2020).

140 magrittr: A Forward-Pipe Operator for R (2020).

141 xlsx: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files (2020).

142 GDAtools: A Toolbox for Geometric Data Analysis and More (2021).

143 tidyr: Tidy Messy Data (2021).

144 Wickham, H. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software* **40**, 1--29 (2011).

145 Wickham, H. *et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. (2019).

146 Gecode: Generic Constraint Development Environment (2019).

147 Sluydts, V. *et al.* Spatial clustering and risk factors of malaria infections in Ratanakiri Province, Cambodia. *Malar J* **13**, 387 (2014). https://doi.org/10.1186/1475-2875-13-387

148 Pongvongsa, T. *et al.* Household clustering of asymptomatic malaria infections in Xepon district, Savannakhet province, Lao PDR. *Malar J* **15**, 508 (2016). https://doi.org/10.1186/s12936-016-1552-7

149 Gaudart, J. *et al.* Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC Public Health* **6**, 286 (2006). https://doi.org/10.1186/1471-2458-6-286

150 Smith, J. L. *et al.* Spatial clustering of patent and sub-patent malaria infections in northern Namibia: Implications for surveillance and response strategies for elimination. *PLoS One* **12**, e0180845 (2017). https://doi.org/10.1371/journal.pone.0180845

151 Stresman, G., Whittaker, C., Slater, H. C., Bousema, T. & Cook, J. 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 Med* **17**, e1003370 (2020). https://doi.org/10.1371/journal.pmed.1003370

152 Mukaka, M. *et al.* Clustering of malaria in households in the Greater Mekong Subregion: operational implications for reactive case detection. *Malar J* **20**, 351 (2021). https://doi.org/10.1186/s12936-021-03879-9

153 Sandfort, M. *et al.* 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. *Malar J* **19**, 413 (2020). https://doi.org/10.1186/s12936-020-03482-4

154 Reiker, T., Chitnis, N. & Smith, T. Modelling reactive case detection strategies for interrupting transmission of Plasmodium falciparum malaria. *Malar J* **18**, 259 (2019). https://doi.org/10.1186/s12936-019-2893-9

155 Zhou, S. S. *et al.* China's 1-3-7 surveillance and response strategy for malaria elimination: Is case reporting, investigation and foci response happening according to plan? *Infect Dis Poverty* **4**, 55 (2015). https://doi.org/10.1186/s40249-015-0089-2

156 Newby, G. *et al.* Testing and treatment for malaria elimination: a systematic review. *Malar J* **22**, 254 (2023). https://doi.org/10.1186/s12936-023-04670-8

157 Tun, S. T. T. M., Myo Chit; Aguas, Ricardo; Fornace, Kimberly; Htoo, Gay Nay; White, Lisa; Parker, Daniel M. Human movement patterns of farmers and forest workers from the Thailand-Myanmar border. *Wellcome Open Research* (2021).

158 Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686 (2019). https://doi.org/10.21105/joss.01686

159 Maechler, M. sfsmisc: Utilities from 'Seminar fuer Statistik' ETH Zurich. (2023).

160 Edwards, S. M. lemon: Freshing Up your 'ggplot2' Plots. (2023).

161 Kassambara, A. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. (2023).

162 Gerardin, J. *et al.* Effectiveness of reactive case detection for malaria elimination in three archetypical transmission settings: a modelling study. *Malar J* **16**, 248 (2017). https://doi.org/10.1186/s12936-017-1903-z

163 Perera, R., Caldera, A. & Wickremasinghe, A. R. Reactive Case Detection (RACD) and foci investigation strategies in malaria control and elimination: a review. *Malar J* **19**, 401 (2020). https://doi.org/10.1186/s12936-020-03478-0

164 Tun, S. T. T. *et al.* Towards malaria elimination in Savannakhet, Lao PDR: Mathematical modelling driven strategy design. *Malaria Journal* **16** (2017). https://doi.org/10.1186/s12936-017-2130-3

165 Steinhardt, L. C. *et al.* Reactive Case Detection and Treatment and Reactive Drug Administration for Reducing Malaria Transmission: A Systematic Review and Meta-Analysis. *Am J Trop Med Hyg* **110**, 82-93 (2024). https://doi.org/10.4269/ajtmh.22-0720

166 Schneider, Z. D. *et al.* Mass Drug Administration: Contextual Factor Considerations. *Am J Trop Med Hyg* **110**, 30-37 (2024). https://doi.org/10.4269/ajtmh.22-0767

167 Legendre, E. *et al.* Prospective acceptability of mass drug administration for malaria in Kedougou region in Senegal: a mixed method study. *Malar J* **23**, 279 (2024). https://doi.org/10.1186/s12936-024-05078-8

168 Galatas, B. *et al.* Community acceptability to antimalarial mass drug administrations in Magude district, Southern Mozambique: A mixed methods study. *PLoS One* **16**, e0249080 (2021). https://doi.org/10.1371/journal.pone.0249080

169 Tusell, M. *et al.* Targeted Drug Administration to Reduce Malaria Transmission: A Systematic Review and Meta-Analysis. *Am J Trop Med Hyg* **110**, 65-72 (2024). https://doi.org/10.4269/ajtmh.22-0754