Australian National University

NOVEL EPIDEMIOLOGICAL TOOLS TO INFORM MALARIA CONTROL AND ELIMINATION IN MELANESIA

Eimear Cleary

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

DECLARATION BY THE AUTHOR

I declare that this thesis presents my original work and does not contain, in part or in full, material previously published or written by another person except where it is otherwise acknowledged in the text. I have clearly indicated the contribution by others to all jointly-authored works that I have included in my thesis. The work presented in this thesis is an accurate account of research undertaken during a PhD candidature in the Research School of Population Health at the Australian National University, and has not been previously submitted for any other degree or diploma at any university or institution.

_____

Eimear Cleary

LIST OF PAPERS AND STATEMENT OF AUTHORS CONTRIBUTIONS TO JOINTLY AUTHORED ARTICLES

This thesis by compilation is based on the following papers:

Paper I: Cleary E, Hetzel M, Clements ACA. A review of malaria epidemiology and control and elimination programmes in Papua New Guinea. (Manuscript in preparation for submission to Malaria Journal)

Paper II: Cleary E, Hetzel M, Siba P, Lau C, Clements ACA. Spatial Ecology and Predicted Risk of Malaria in Papua New Guinea, 2010/2011: A comparison of models from which to generate spatial risk maps. (Manuscript in preparation for submission to Malaria Journal)

Paper III: Eimear Cleary, Abby Harrison, Stuart Lee, Livingstone Tavul, Manuel Hetzel, Ivo Mueller, Melanie Bahlo, Alyssa Barry, Archie Clements. Landscape genomics of *Plasmodium falciparum* in Papua New Guinea reveals major population subdivisions associated with ecological niches and a routes of malaria transmission. . (Manuscript in preparation for submission to Nature Communications)

Paper IV: Cleary E, Barry A, McCaw J, Clements ACA. Examining the impact of human movement on malaria resurgence using a Ross-Macdonald meta-population model and the example of Solomon Islands. (Manuscript in preparation for submission to Epidemics)

For each publication in the thesis, I was first author. My contribution to each of the thesis chapters is as follows:

1. ACAC and I conceived the theme of this literature review. MH recommended substantial proportion of published the research literature reviewed for this paper. I conducted a review of the published work and wrote the manuscript. ACAC provided comments and I undertook revisions accordingly.

2. ACAC and I developed the concept of the analysis. MH and PS collected the observational data used in the analysis. I obtained and processed ecological covariate data used for the analyses. I reviewed the published literature, conducted the analysis, produced output, tables and figures, and drafted the manuscript. ACAC and I interpreted the results. CL provided intellectual support in developing statistical models. ACAC provided comments and I undertook revisions on the manuscript as recommended.

3. ACAC, AB and I developed the concept of the analysis. AH, SL, TL, MH, IM, MB and AB collected and genotyped the data used in the analyses. I carried out statistical analysis, reviewed published literature, drafted the manuscript and produced output, tables and figures. AB and AH produced supplementary material figures and tables for the chapter appendix. ACAC and AB provided comments on the drafted manuscript and I coordinated input and undertook revisions based on co-author comments.

ACKNOWLEDGEMENTS

There are several people I'd like to acknowledge and thank for their support and encouragement over the last four years.

Firstly, I'd like to sincerely thank my primary PhD supervisor, Archie Clements for the generous giving of his time, help and advice during the course of this PhD. I'd also like to acknowledge and thank the other members of my supervision panel, Manuel Hetzel for his generous provision of data, Aparna Lal for her regular supervision meetings, Darren Gray, and Susana Nery for her intellectual input in the planning aspects of parts of this thesis. I'd also like to thank Colleen Lau and Helen Mayfield for their research and analytical support.

I'd like to extend a thank you to James McCaw and the mathematical and computational biology group at the University of Melbourne for accommodating me as a visiting researcher and for offering me such a warm welcome. In particular I'd like to thank Alex Zarebski, Ada Yan, Michael Lydeamore and Jackson Kwon for their analytical and modelling support.

I would especially like to express my gratitude to Alyssa Barry at the Walter and Eliza Hall of Medical Research for her immense contribution to parts of this research thesis, not least for the generously sharing data with me which formed a central component of the thesis research.

I'd also like to thank Charles Spillane and the Plant and AgriBiosciences Research Centre group at NUI Galway for providing me with a research space and support while I was finishing my thesis write up and who also have been incredibly hospitable and welcoming. Also thank you to Farouq and Rose in the Lane Café for providing me with endless hours of desk and write up space.

I'd like to say a very special thank you to the people in the Global Health Department, and to the people in the Research School of Population Health at the ANU for being so welcoming

when I moved to Canberra, and to the ANU IT staff, in particular Omar Ibrahim, who provided me with a huge amount of remote technical support.

I'd like to thank my fellow PhD students who I've gone through the PhD process with over the last number of years, and for their incredible. Particularly Maura Tilbury for always being an incredible support and friend, Patrizio Mancuso for his exceptionally generous hospitality and for going through the write up process with me.

Thank you to my fantastic group of friends at home, many of whom have proof read parts of this thesis. I'm very lucky to have all of you.

A very special thank you to all the Black Gate crew for their moral support and fun times!

I especially would like to thank my wonderful sisters and brother Iobhan, Deirdre, Mairead and Eoin for all their help, support, counselling and long phone calls before, during, and after I moved to Australia.

Finally, a most important thank you to my mother Peig for her unwavering support, kindness, generosity and motivation always. You've always told me that I could do whatever I wanted as long as I put my mind to it. I couldn't ask for a better mother. Thank you so much.

I'd like to dedicate this thesis to the memory of my father, PJ Cleary.

This research was generously funded by a Postgraduate Award from The Australian National University.

ABSTRACT

Background:

Malaria is a vector-borne parasitic disease that in 2017 was responsible for an estimated 219 million clinical cases of infection and an estimated 435000 deaths globally, over 60% of which were among children under five years of age. Estimating the spatial distribution of malaria within endemic countries, and risk factors for transmission, is essential to the effective planning and allocation of malaria prevention interventions such as long lasting insecticide treated nets (LLINs), indoor residual spraying (IRS), enhanced case detection using rapid diagnostic tests (RDTs), malaria chemoprevention and antimalarial drugs.

The aims of this PhD thesis were to: 1) describe the epidemiology of malaria in Papua New Guinea (PNG) and summarise previous control strategies and outcomes in PNG over the last century; 2) compare the accuracy of multilevel generalised linear regression models (GLMs) with Bayesian decision network (BDN) models in the spatial prediction of prevalence of malaria in PNG based on associations of national parasite surveillance data with ecological and demographic covariates; 3) predict the geographic niches of eight genotypes of *Plasmodium falciparum* in PNG to ascertain patterns of connectivity in the human population in terms of malaria transmission; and 4) examine the impact of human movement between high and low transmission intensity locations on malaria transmission using a mathematical model based on the example of two islands of Solomon Islands.

Methods:

Data for this research was obtained from a number of different sources: published literature was obtained from online archives of science literature PubMed, Google Scholar, and the Australian National University library online resources; a national malaria indicator survey (MIS) which was conducted in five villages randomly sampled from a geo-referenced village

database in 17 of the 20 provinces of PNG by the Papua New Guinea Institute of Medical Research in 2010 and 2011, and; household-based national malaria indicator survey data collected in PNG and Solomon Islands between October 2008 and August 2009, where samples were collected and genotyped using the highly polymorphic *Pfmsp2* marker. Climate data at 1km resolution was obtained from the WorldClim online climate data repository and environmental remote sensing image data were obtained from Earthdata, the NASA hosted remote satellite imagery online database, at 250m resolution. Modelling approaches included: a comparison of GLMs with BDN models using point prevalence and ecological data to predict the spatial distribution of *P. falciparum* and *P. vivax* malaria in PNG; a Dirichlet regression model examining associations of *P. falciparum* genotype predominance with ecological covariates for the prediction of geographic niches of distinct parasite genotypes in PNG; and a Ross-Macdonald mathematical model using estimates of *P. falciparum* prevalence in two island of Solomon Islands (representing a high and low transmission location) for the estimation of the impact of varying rates of human migration on malaria transmission with relaxed or sustained use of vector control interventions. Geographic information system software ArcGIS version 10.3 (ESRI, Redlands, California) was used for processing and collating data and statistical analyses were carried out using the R open-source statistical software package version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria) and Stata version 14 (StataCorp, College Station, Texas).

Results:

In terms of *P. falciparum* and *P. vivax* spatial distribution in Papua New Guinea, BDN models were found to have improved accuracy in spatial predictions when compared with generalised linear models. Statistically significant associations were found between *P. falciparum* prevalence and maximum average temperature during the three hottest months of the year, enhanced vegetation index (EVI) during one of the hottest months of the year, distance to the

coastline, and precipitation during the driest three months of the year. *P. vivax* prevalence had statistically significant associations with EVI during one of the coldest and wettest months of the year, distance to the coastline, elevation, average precipitation during the three wettest months of the year and gradient, or slope, of the landscape. The predicted spatial distribution of *P. falciparum* and *P. vivax* based on BDN models followed a similar pattern to survey data with higher predicted prevalence on the islands to the East of PNG and northern coastline of the mainland, and lower predicted prevalence in the highlands and south coast. The results of the Dirichlet regression model identified geographic niches of eight distinct *P. falciparum* genotypes in PNG based on associations with population density, elevation, distance to the coastline, latitude and longitude, and their quadratic terms. The results of the mathematical model predicted that in the absence of sustained vector control post-elimination, resurgence of malaria may occur relatively quickly in low-transmission intensity locations where connectivity with high-transmission intensity locations exists due to human migration, such as in the islands of Solomon Islands.

Conclusions:

This PhD research provides a comprehensive review of literature on the control strategies for and challenges to, achieving goal of global malaria elimination, and a review of the current epidemiology of malaria, and major periods of malaria control in Papua New Guinea, drawing conclusions on reasons for failure of control programmes in the past and future directions for current control efforts. This thesis also identifies novel epidemiological methods for improved prediction accuracy in the spatial distribution of malaria based on environmental and climate predictors, a method for inferring human connectivity in terms of malaria transmission in PNG using parasite genotype data and the application of a mathematical model to in examining the transmission dynamics of malaria transmission in two islands in Solomon Islands under

varying rates human migration between both locations, vector control, and biological assumptions underlying malaria transmission dynamics.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 6

LIST OF FIGURES

Chapter 4

Chapter 5

Figure 2. Map of Papua New Guinea showing sampling locations and mean ancestry coefficients for each of 27 geographic areas.

Figure 3 Spatial predictions of eight distinct plasmoduim genotype clusters based on the results of a Dirichlet regression

Figure S1 Mean log posterior probability for each K in the STRUCTURE analysis

Figure S2 DeltaK estimates for each K in the STRUCTURE analysis

Chapter 6	
Figure 1	Geographic location of Solomon Islands, Nggela (patch 1) and Guadalcanal (patch 2).
Figure 2	Model simulation showing time to malaria resurgence in patch 1 with migration rate between patch 1 and 2 = 0.1 and migration rate from patch 2 to 1 = 0.1. Human biting rate = 0.1; infection prevalence in patch 2 = 0.096; relative mosquito to human population abundance = 10.
Figure 3	Model simulation showing time to malaria resurgence in patch 1 with migration rate from patch 1 and 2 = 0.1, migration rate from patch 2 to 1 = 0.1 and 50% increase in vector abundance in patch 1 and 2. Human biting rate = 0.1; infection prevalence in patch 2 = 0.096; relative vector abundance = 15.
Figure 4	Model simulation showing time to malaria resurgence in patch 1 with migration rate from patch 1 and 2 = 0.1, migration rate from patch 2 to 1 = 0.1 and a 50% increase in human biting rate. Human biting rate = 0.15; infection prevalence in patch 2 = 0.096; relative vector abundance = 10.
Figure 5	Model simulation showing time to malaria resurgence in patch 1 with ten-fold increase in migration between patch 1 and 2. Migration rate from patch 1 to 2 = 0.5; migration rate from patch 2 to 1 = 0.5; human biting rate = 0.1; infection prevalence in patch 2 = 0.096; relative vector abundance = 10.
Figure 6	Model simulation showing time to malaria resurgence in patch 1 with migration rate between patch 1 and 2 increased by a factor of 10 and increase in infection prevalence in patch 2. Migration rate from patch 1 to 2 = 1.0; migration rate from patch 2 to 1 = 1.0; human biting rate=0.1; infection prevalence in patch 2 = 0.2; relative vector abundance = 10.
Figure 7	Results of sensitivity analysis showing increasing infection prevalence in patch 1 at increasing biting rate (a) and increasing migration rate (k_{12}; k_{21}).

LIST OF ACRONYMS AND ABBREVIATIONS

ACT	Artemisinin-based combination therapy
AIC	Akaike information criterion
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area under the curve
BDN	Bayesian decision Network
DDT	Dichlorodiphenyltrichloroethane
DHS	Demographic and Health Surveys
EVI	Enhanced vegetation index
GIS	Geographic information system
GDEM	Global digital elevation model
GLM	Generalised linear model
GPS	Global positioning system
GTS	Global technical malaria strategy
HBHI	High burden to high impact
IRS	Indoor residual spraying
LLIN	Long lasting insecticide treated nets
MDA	Mass drug administration
MIS	Malaria Indicator Surveys
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	The National Aeronautics and Space Administration
PNG	Papua New Guinea
RBM	Roll Back Malaria
RCD	Reactive case detection
RDT	Rapid diagnostic tests
ROC	Receiver operating characteristic
SDG	Sustainable development goals
SDSS	Spatial decision support system
WHO	World Health Organisation
WHOPES	World Health Organisation Pesticide Evaluation Scheme
WWARN	World-Wide Antimalarial Resistance Network

CHAPTER 1

Introduction

CHAPTER 1. INTRODUCTION

1.1 Context

In 2017, the global number of malaria cases was an estimated 219 million, a decrease in 20 million cases compared with 2010[1]. This is primarily the result of a renewed commitment to global malaria elimination since 2007, through national and regional programmes of malaria control and increasing access to vector control interventions and improved treatment for infection[2-4]. This current commitment to malaria elimination is the first global attempt at controlling and eliminating malaria since the malaria eradication campaign of the 1950s[5], proposed by the World Health Organisation (WHO) in 1955[6], following the success of malaria elimination in the United States of America (USA) in 1951.

The 1950s campaign aimed to achieve a systematic process of elimination from the north to south of the northern hemisphere, and south to north of the southern hemisphere, through vector control by residual household spraying with dichlorodiphenyltrichloroethane (DDT)[5-7]. Although malaria elimination was achieved in 37 countries in Europe, North America, the Caribbean and parts of Asia and South-America during this campaign[8], the emergence of vector resistance to DDT and antimalarial drug resistance in *Plasmodium* parasites led to the programme being largely perceived as a failure[5,9,10]. The aim of malaria eradication was subsequently abandoned until 2007 when the current goal of global elimination was proposed by Bill and Melinda Gates, amongst others.

The current strategy for malaria elimination, outlined by the Global Technical Strategy for Malaria 2016-2030, and adopted by the World Health Assembly in 2015 in line with the Sustainable Development Goals (SDG) 2030, aims to reduce malaria incidence and mortality by 90%, eliminate malaria in 35 new countries, and prevent resurgence of malaria in countries

that were malaria free in 2015[4]. This involves building strong partnerships and government commitment, generating evidence-based policy for the elimination of malaria in diverse transmission zones, and investing in effective control interventions and surveillance tools[11]. Underpinning the strategy, just as with the global eradication campaign of the 1950s, is the importance of understanding the epidemiology of malaria and using novel epidemiological tools to direct resources and target interventions.

1.2 Background

Global malaria

In 2016, 91 countries in the world reported indigenous malaria cases, with 90% of the global malaria burden in sub-Saharan Africa[12]. The majority of global infection is caused by the *P. falciparum* parasite, accounting for 99% of cases in sub-Saharan Africa in 2016, whereas *P. vivax* is the dominant parasite in large parts of Asia and the Americas[12]. Malaria burden in the Asia-Pacific region is the highest outside of sub-Saharan Africa[13] but a substantial reduction in malaria burden has been achieved in this region over the past 15 years[14].

Prevention of malaria

The primary interventions used in the reduction of global malaria incidence are (i) vector control through distribution of long lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS), (ii) preventative therapies through seasonal malaria chemoprevention programmes, (iii) enhanced case finding using diagnostic approaches such as rapid diagnostic tests (RDT) and microscopy, and (iv) rapid, effective treatment with anti-malarial combination drugs. Effective surveillance is essential for providing regular estimates of malaria incidence, and identifying areas and populations where malaria intervention resources can be targeted for maximum impact[12].

Study area

PNG is a Pacific Island nation bordering Australia to the south west and consisting of the eastern half of New Guinea and a collection of several large and several hundred small islands[15]. In PNG, the epidemiology of malaria varies considerably across the country and small-area spatial variations in malaria prevalence exist, attributed to vector density, village bed net use, the availability of antimalarials and access to health care facilities[16,17]. The national malaria prevalence in 2008/2009 ranged from 0% to 49.7% by community (?) with a weighted national average of 12.1%[18]. In 2010/2011 infection prevalence was 6.7% in villages below 1600m, composed of 3.4% *P. falciparum,* 2.1% *P. vivax* and 0.06% *P. malariae* and mixed infections[19]. There are considerable urban-rural and regional disparities associated with access to quality healthcare and health infrastructure varies considerably between different regions. Poverty rates are high with people of lower income being at a marked disadvantage in terms of health care access[20,21]. Use of LLINs among people who have access to them is high (estimated to be approximately 78.7%)[22,23], but access can be restricted due to poor infrastructure and remoteness of villages[24]. Additional barriers to LLIN use include perception of low malaria risk, indifference, and reluctance to use LLINs in the heat[23,25].

Solomon Islands is an archipelago in the South Pacific, bordering PNG to the west and with Vanuatu as its closest neighbour to the south east[26]. In 2011, the national coverage of LLINs was 91%, the ability of health centres to diagnose malaria with confirmatory testing was 97%[27], and in 2014, malaria prevalence had decreased from over 170 cases per 1000 population in 2005 to just over 25 cases per 1000 population. By 2016, however, this number had increased to just under 100 confirmed cases per 100 population[28]. Health care is provided largely by provincial public health programmes with health care centres which range in size and complexity from nurse aid posts, located in remote areas to larger area health centres and

provincial hospitals[26]. Access to LLINs is high in Solomon Islands with 86% of households surveyed in 2015 in possession of at least one LLIN, but with reported usage of only 57%[29,30].

PNG and Solomon Islands are both aiming for malaria elimination by 2030[12,14]. However, in the context of stage of elimination, PNG is controlling malaria whereas Solomon Islands is technically nearing elimination. It is therefore likely that different epidemiological tools, and methodologies for examining drivers of transmission in these different endemic settings, will be required.

The aim of this thesis is to review the current objectives and strategies for achieving global malaria elimination, review the epidemiology and current and historical elimination programmes in PNG, and apply three different epidemiological approaches to support malaria interventions: 1) predicting the spatial distribution of malaria prevalence in a control setting using different spatial statistical methods, and comparing those methods in terms of prediction accuracy; 2) predicting the spatial distribution of parasite genotype clusters to infer how parasite populations in distinct geographic areas, and by extension their human hosts, are connected; and 3) simulating the impact of human migration between a high-transmission intensity location and a low-transmission intensity location with an absence of local transmission, on resurgence of malaria in the latter, under conditions of sustained or ceased vector control. These tools are applied in the context of malaria control, elimination, and prevention of reintroduction in Melanesia, specifically PNG and Solomon Islands

1.3 Research goals and objectives

Objective 1:

To summarise existing epidemiological tools for informing malaria control programmes and targeting interventions and how they can be applied to the control, elimination and prevention of reintroduction settings. This is detailed in chapter 2.

Objective 2:

To describe the current epidemiology of malaria in PNG and summarise the historical periods of malaria control in the country. This review is contained in chapter 3.

Objective 3:

To compare the predictive accuracy of different multivariable models in the prediction of the spatial distribution of *P. falciparum* and *P. vivax* infections using environmental and climatic variables as predictors of malaria risk in PNG. This manuscript forms chapter 4.

Objective 4:

To infer patterns of connectivity and human migration in PNG using a Dirichlet regression model of associations between *P. falciparum* genotype data and environmental covariates. This work is described in chapter 5.

Objective 5:

To use a mathematical model to examine the impact of human migration from a high-transmission intensity location on resurgence of malaria in a low-transmission location where local transmission of malaria has ceased, using the example of Nggela, one of the Solomon Islands. This paper forms chapter 6.

1.4 Approach and methods

Data for this research includes *P. falciparum* and *P. vivax* prevalence data, *P. falciparum* genotype cluster data, measures of climate and other aspects of the biophysical environment, migration and socio-demographics which were obtained from numerous different sources, and published research obtained from online archives of published scientific literature. Published research comprising both literature review chapters were obtained from a comprehensive search and review of literature published in various science journals and online books. *P. falciparum* and *P. vivax* prevalence data, and data on vector control interventions, and socio-

demographics were collected as part of the national malaria indicator survey (MIS), conducted in five villages randomly sampled from a geo-referenced village database in each of 17 of the 20 provinces of PNG by the Papua New Guinea Institute of Medical Research in 2010 and 2011. *P. falciparum* genotype data were collected as part of a household-based national malaria indicator survey conducted in PNG and Solomon Islands between October 2008 and August 2009, following which positive samples were genotyped using the highly polymorphic *Pfmsp2* marker. Climate data on precipitation and temperature, at 1km resolution aggregated over 50 years, were obtained from the WorldClim online climate data repository[31]. Environmental remote sensing image data for enhanced vegetation index (EVI) and digital elevation were obtained from Earthdata[32], the NASA hosted remote satellite imagery online database, at 250m resolution. Slope data were derived from the digital elevation model, and distance to the coastline for each village point from which parasite prevalence data was collected was calculated.

Modelling approaches included a comparison of the predictive accuracy of generalized linear models (GLM) with Bayesian decision network (BDN) models using point prevalence *P. falciparum* and *P. vivax* data, and ecological data to predict the spatial distribution of malaria in PNG. A Dirichlet regression model was used to examine associations of *P. falciparum* genotype predominance with ecological covariates, latitude and longitude, and population density, to predict geographic niches of distinct parasite genotypes in PNG. A Ross-Macdonald model using estimates of *P. falciparum* prevalence in Solomon Islands was used to simulate the impact of varying rates of human migration on malaria transmission between two islands in the absence of, or with sustained use of, vector control interventions. Geographic information system (GIS) software ArcGIS version 10.3 was used for processing and collating data and statistical analyses were carried out using R statistical software and Stata version 14.

1.5 Contribution of this thesis

This thesis comprises a literature review summarising an overview of the strategies used to prevent and treat malaria, the types of surveillance used in different endemic settings, and the current challenges facing national and global elimination objectives. It summarises general statistical modelling techniques for malaria prediction and examines ecological and demographic drivers of transmission. This thesis also presents an up-to-date review of malaria epidemiology and national malaria control efforts in PNG, including a description of historical elimination programmes, dating back to the start of the 1900's. Major contributions of this thesis include an examination and comparison of epidemiological methods for predicting malaria risk and estimating the impact of connectivity of distinct geographic areas and human mobility on malaria transmission, which can be utilised in different endemic settings.

BDN models are being used in infectious disease risk prediction with increasing frequency, and this thesis compares the ability of these models to predict spatial risk of malaria with conventional regression models. In doing so, this research produces spatial predicted risk maps of *P. falciparum* and *P. vivax* in PNG, and quantifies the predictive accuracy of these maps and the spatial distribution of uncertainty in spatial risk predictions, which can be used by the national malaria control programme in PNG for allocation of interventions and malaria control resources. This work also gives insight into the environmental drivers of malaria transmission in PNG, and using BDN models, presents these drivers in an intuitive manner to help communicate environmental risk factors in a way that is accessible to at-risk populations and policy makers.

This thesis also contributes to the growing field of research that aims to examine how malaria transmission is facilitated by the connectivity of distinct geographic areas and the movement of people, and advances the use of genomic data as an epidemiological tool for understanding

malaria transmission dynamics at a population level. Developing maps showing the predicted distribution of eight *P. falciparum* genotypes in PNG allows us to explore how different communities in PNG might be connected in terms of malaria transmission. This is, to our knowledge, the first study of this type using malaria genotype data and spatial statistical prediction methods. Understanding how malaria transmission dynamics are influenced by connectivity of distinct populations can help to inform surveillance and control operations, particularly if malaria is eliminated from provinces before elimination is achieved in endemic provinces to which they may be connected by human mobility.

This theme is expanded upon in the following study, which uses a mathematical model to estimate the impact of human movement on malaria transmission dynamics. The model simulates human mobility between two distinct island regions assuming local transmission has ceased in one location but is ongoing in the area to which it is connected by population movement. This work examines model parameters associated with resurgence and estimates the time it would take for transmission to be re-established after vector control interventions cease in the eliminating area. It also explores the effect of sustained vector control interventions through reducing the man biting rate (simulating LLIN use) and relative vector abundance (simulating IRS). In doing so we consider important questions for control programmes about continued surveillance and vector control interventions post elimination in places that remain connected to an endemic area by human mobility.

1.6 Research and thesis structure

This thesis consists of seven chapters; an introduction, background, four research chapters (one narrative review and three analytic research chapters) structured for submission as journal publications, and a discussion chapter summarising the findings and general conclusions of the thesis as well as future research directions. The three analytic research chapters examine the

use of epidemiological tools for informing malaria control and elimination programmes in different endemic settings: controlling malaria; and eliminating malaria/preventing the reintroduction of malaria. The details of each chapter are listed below.

Chapter one is the thesis introduction, providing on overview of the current global malaria situation, context of the thesis and background on malaria control strategies and the main geographic focus of the research presented in this thesis, that being PNG and Solomon Islands. It also outlines the main thesis objective, research approach and methods, and thesis contribution and structure.

Chapter two is a literature review of the types of surveillance used in control, elimination and prevention of reintroduction settings and current challenges faced by malaria control and elimination programmes. It also describes epidemiological tools for predicting malaria risk in control and elimination settings and for simulating the dynamics of malaria, both as a result of different interventions and of human population movement. These tools are the focus of, and primary research methods used, in the subsequent thesis chapters.

Chapter three is a narrative review of current and historical malaria control programmes in PNG. It summarises the current epidemiology of *P. falciparum* and *P. vivax* epidemiology in PNG and the three major historical periods of malaria control in PNG and discusses current control strategies, challenges to elimination and future directions of control programmes.

Chapter four is an analytic research chapter comparing two statistical methodologies, BDN models and GLMs, in terms of their predictive accuracy for malaria spatial risk prediction. Models are developed examining associations between *P. falciparum* and *P. vivax* with ecological drivers of transmission in PNG, a country classified as controlling malaria. Predictive accuracy was compared using model cross-validation.

Chapter 5 explores the use of genomic data in predicting the spatial predominance of *P. falciparum* genotype clusters in PNG. The application of this methodology is an attempt to examine how distinct populations are connected in terms of malaria transmission, in order to direct surveillance and control resources and prevent transmission between endemic and eliminating areas. Although PNG is classified as controlling malaria only, the genetic parasite population structure in PNG, due to relative isolation of distinct human populations, is indicative of a country with low transmission of malaria. Therefore this methodology can also be applied in the context of pre-elimination or elimination.

Chapter 6 is an analytic research chapter that focuses on simulation of the impact of human mobility on resurgent malaria transmission in a location where local transmission has ceased, but ongoing transmission may be sustained by population connectivity with a neighbouring endemic location. This chapter uses the example of two of the Solomon Islands for illustrative purposes.

Chapter 7 presents a general discussion of the main findings of the thesis, assesses the limitations of the research, makes recommendations for future investigations and presents the main conclusions. It summarises the key research findings of each chapter, interprets potential reasons for results obtained and discusses applications of the research carried out here in broader contexts. It also compares the research findings of this thesis with published research from other malaria endemic and eliminating areas.

References

1 World Health Organization. World Malaria Report 2018. (Geneva, 2018).
2 Roberts, L. & Enserink, M. Did they really say... eradication? *Science* **318**, 1544-1545 (2007).
3 Roll Back Malaria. *About RBM*, <http://www.rollbackmalaria.org/about/about-rbm/rbm-mandate> (2016).
4 World Health Organization. *Global technical strategy for malaria 2016-2030*. (World Health Organization, 2015).
5 Black RH. Malaria. *Papua New Guinea Medical Journal* **17**, 1-3 (1974).
6 Centers for Disease Control and Prevention. *The History of Malaria, an Ancient Disease*, <http://www.cdc.gov/malaria/about/history/> (2016).
7 McMahon, J. A note on changing concepts of malaria control. *Papua New Guinea Medical Journal* **16**, 78-79 (1973).
8 Carter, R. & Mendis, K. N. Evolutionary and Historical Aspects of the Burden of Malaria. *Clinical microbiology reviews* **16**, 173 (2003).
9 World Health Organization. Global malaria control and elimination: report of a technical review. (2008).
10 Greenwood, B. Can malaria be eliminated? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **103**, S2-S5 (2009).
11 Bill & Melinda Gates Foundation. *Malaria Strategy Overview* <http://www.gatesfoundation.org/What-We-Do/Global-Health/Malaria> (2015).
12 World Health Organization. World Malaria Report 2017. (Geneva, 2017).
13 United Nations. Overview: Malaria in the Asia-Pacific. (Asia Pacific Leaders Malaria Alliance, 2014).
14 APMEN. *Elimination 2030: Working together for a Malaria-free Asia Pacific*, <https://www.apmen.org/web/1/elimination/> (2019).
15 Campos-Outcalt, D. Health services in Papua New Guinea. *Public Health* **103**, 161-169 (1989).
16 Cattani, J. *et al.* Small-area variations in the epidemiology of malaria in Madang Province. *Papua New Guinea Medical Journal* **29**, 11-17 (1986).
17 Davy, C. P. *et al.* Seeking treatment for symptomatic malaria in Papua New Guinea. *Malaria journal* **9**, 268 (2010).
18 Hetzel, M. W. *et al.* Prevalence of malaria across Papua New Guinea after initial roll-out of insecticide-treated mosquito nets. *Tropical Medicine & International Health* **20**, 1745-1755 (2015).
19 Hetzel, M. W. *et al.* Country-wide household survey 2010/11: malaria control intervention coverage and prevalence of parasitaemia. *Goroka: Papua New Guinea Institute of Medical Research* (2012).
20 Bauze, A. E. *et al.* Equity and geography: the case of child mortality in Papua New Guinea. *PLoS One* **7**, e37861 (2012).
21 Serageldin, I., Shluger, E. & Martin-Brown, J. Papua New Guinea-Poverty and access to public services.
22 Hetzel, M. W. *et al.* Progress in mosquito net coverage in Papua New Guinea. *Malaria journal* **13**, 242.241-242.214 (2014).
23 Pulford, J. *et al.* Indifferent to disease: a qualitative investigation of the reasons why some Papua New Guineans who own mosquito nets choose not to use them. *Social Science & Medicine* **75**, 2283-2290 (2012).
24 Hetzel, M. W. *et al.* Ownership and usage of mosquito nets after four years of large-scale free distribution in Papua New Guinea. *Malaria journal* **11**, 1 (2012).
25 Hetzel, M. W. *et al.* Progress in mosquito net coverage in Papua New Guinea. *Malaria journal* **13**, 1 (2014).
26 World Health Organization. *Solomon Islands health system review*. (Manila: WHO Regional Office for the Western Pacific, 2015).
27 World Health Organization. (Manila: WHO Regional Office for the Western Pacific, 2011).

28 World Health Organisation. *Malaria Country Profiles* <http://www.who.int/malaria/publications/country-profiles/profile_slb_en.pdf> (2017).
29 Quah, Y. W. *et al.* Molecular epidemiology of residual Plasmodium vivax transmission in a paediatric cohort in Solomon Islands. *Malaria journal* **18**, 106 (2019).
30 National Vector-borne Disease Control Program, S. I. Annual Malaria Report 2015., (2016).
31 WorldClim. *Global climate data.* , <http://www.worldclim.org/> (2019).
32 The National Aeronautics and Space Administration. *Earthdata,* <https://earthdata.nasa.gov/> (2018).

CHAPTER 2

Background:

An overview of malaria prevention strategies and challenges facing national and global elimination objectives

CHAPTER 2.

BACKGROUND: AN OVERVIEW OF MALARIA PREVENTION STRATEGIES AND CHALLENGES FACING NATIONAL AND GLOBAL ELIMINATION OBJECTIVES

1. Introduction

1.1 Context

Considerable progress has been made toward global malaria elimination since a renewed commitment to this goal was made in 2007. Between the years 2000 and 2015 the global incidence of malaria was reduced by 41%, malaria attributed mortality rates reduced by 62%[1], and 19 previously endemic countries achieved zero indigenous cases for three consecutive years[2]. In 2016 the World Health Organisation (WHO) identified 21 countries with the potential to eliminate malaria by 2020, and in 2017 44 countries reported fewer than 10,000 malaria cases, up from 37 countries in 2010 (notably, 26 countries reported fewer than 100 cases)[3].

The strategy outlined for the global elimination of malaria involves intensified control strategies in endemic areas to achieve low transmission and a reduction in mortality[4] and shrinking the malaria map by progressively eliminating malaria from the endemic margins inward[4,5]. Elements that are intrinsic to the success of elimination programmes are interruption of transmission through control and reduction of the mosquito vector population, stemming the flow of imported infections from endemic areas[6] and collaborative regional partnerships to support and strengthen national elimination programmes[7]. Robust surveillance systems for estimating incidence of malaria infection, and epidemiological tools for identifying at risk populations, ecologically suitable habitats for vectors, and heterogeneities in infection prevalence are essential in guiding interventions and aiding planning for control programmes.

In 2016 the number of countries classified as endemic for malaria had reduced to 91 from 108[2] and in 2018 the number of countries reporting fewer than 100,000 cases was 49, up from 46 in 2017[8]. In 2019, the number of countries with fewer than 100 indigenous cases had increased to 25 from 17 countries in 2010[8]. In the eleven high burden to high impact (HBHI) countries however, there were an estimated 155 million malaria cases in 2018, ten countries of which were in sub-Saharan Africa[8,9]. The path toward malaria elimination is a continuous process. Depending on environmental, biological and financial determinants, variation in national and subnational endemicity may exist as progress is made toward a malaria free status[10]. As elimination efforts continue, heterogeneity in malaria transmission arises in response to control interventions and elimination programmes. Surveillance strategies should be tailored in response to this heterogeneity in infection risk[10]. The aim of this Chapter is to summarise key strategies of control and methods of surveillance, review some of the current challenges to achieving global malaria elimination and describe epidemiological tools used for informing interventions and surveillance in control, elimination and prevention of reintroduction settings.

2. Strategies for control and prevention of malaria

Integral to the success of malaria programmes is the scaling up and maintaining of vector control interventions such as long lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS), rapid infection diagnosis facilitated by use of rapid diagnostic tests (RDTs), and effective malaria drug treatment, most often achieved with artemisinin-based combination therapy (ACT)[6]. Since publication of the first and second editions of the WHOs guidelines on treatment of malaria in 2006 and 2010, all countries in which *P. falciparum* is endemic have updated their treatment policy to recommend ACTs as the recommended front-line therapy for malaria infection[14]. Between 2015 and 2017 national malaria programmes delivered 624 million LLINs, 85% of which were distributed as part of free distribution programmes, and 206

million ACTs[3]. However intense selection pressure due to high levels of drug administration, the unregulated availability of poor-quality medications and exposure of vectors to insecticides over a long period of time can result in emergence of resistance in the parasite and vector[15,16]. The most effective way of preventing malaria from progressing to severe clinical infection is in the prompt diagnosis of infection, in particular by use of RDTs, 276 million of which were distributed in 2017[3]. However income and socio-economic disparities, living in rural areas as opposed to urban areas and distance to health centres still present barriers to effective diagnosis and access to treatment. Occupational risks are also important, particularly those associated with forested areas (for example, logging and gem mining), and these occupational groups are often missed by LLIN distribution campaigns which focus on residential homes and villages[17]. This highlights the importance of surveillance programmes for detection of malaria cases and estimation of heterogeneity in national prevalence, to ensure access to diagnosis, treatment and vector control interventions, and targeting of resources for maximum impact[3].

3. Surveillance

An ideal surveillance system requires capability for early detection of all malaria infections and prompt, complete reporting of detected cases to a centralised database[3,18], with timely analysis and reporting of results back to local jurisdictions to inform an effective response[19] and robust data collection on surveillance activities themselves[20]. Surveillance can involve monitoring of infection incidence as well as entomological indicators, which can help identify populations at higher malaria risk[3]. Malaria surveillance falls mainly under two categories, passive and active, which receive greater or lesser relative emphasis depending on the different stages of elimination.

3.1 Passive Surveillance

Passive surveillance involves collecting malaria case data among a cohort of patients with clinical infection attending a health clinic and involves no active search for cases[21]. This type of surveillance records clinical burden of malaria[22] and is useful for estimating the population incidence of febrile malaria cases without the logistical and financial burdens inherent in population-wide surveys. For this method of surveillance to be effective, data should be regularly reported by local health jurisdictions to a central health authority database[21], which in most countries is based on a malaria-specific information system, for collation, estimation and analysis of malaria incidence[23,24,25].

In settings with a large proportion of sub-clinical infection however, cases are likely to be underreported and as a result prevalence estimates determined by passive surveillance may suffer from health-seeking biases, often characterised by underestimation of actual burden[25,26]. This can be improved by encouraging health seeking behaviour and increasing the availability of sensitive diagnostic tests[6]. One effective approach for improving access to malaria diagnosis and treatment is extending health access through community health care workers to deliver health care services to vulnerable and hard-to-reach populations with limited access to health care. This strategy, known as integrated community case management (iCCM), can also involve the systematic gathering, aggregating, analysing and reporting of data to identify gaps in treatment and intervention coverage for more equitable delivery of health care services[27].

Passive surveillance is generally characteristic of countries where the programmatic focus is the control of malaria as a public health problem[6]. However, passive surveillance may be inadequate for effective spatial targeting of resources as countries progress towards elimination. In such settings, systematically collected data may be preferential for determining accurate spatial distribution of malaria risk and reservoirs of infection, including asymptomatic infection[28].

3.2 Active Case Detection

As countries move toward elimination, control strategies require a shift from early case detection and treatment of symptomatic cases to more active surveillance of malaria infection[3,29]. Active case detection involves the detection of malaria infection and residual parasite carriers through mass screening of high-risk populations[30,31]. Active case detection and treatment among high risk migrants or mobile populations (section 4.3) is crucial for interrupting transmission and preventing reintroduction[4,32]. However, mass population screening may be logistically unfeasible and financially prohibitive for a national malaria control programme, and mass population screening of this type is usually conducted as part of large scale research projects. Malaria control programme resources may be best directed at more targeted active screening, which can be informed by risk assessment methods including spatial prediction. Georeferenced point data collected during active surveillance or baseline surveys can be used for spatial prediction of the distribution of infection in non-sampled locations (predictive risk maps) using a variety of statistical modelling approaches (section 5.2). To generate such predictive risk maps (section 5.1), models can incorporate associations with environmental and climate data, which can be visualised, processed and integrated with malaria survey data using geographic information software (GIS).

3.2.1 Reactive Case Detection

Reactive case detection (RCD) is active surveillance carried out among people and households situated within close proximity of a detected malaria infection, supported by georeferencing and mapping of confirmed malaria cases within a defined region or 'buffer zone' of the index case[20,33]. A reactive case detection surveillance programme is typically employed in elimination or prevention of reintroduction settings where surveillance includes collection of data on origin of infection and recent migration[34]. When malaria transmission occurs in

spatially proximal clusters[6,35,36], RCD can identify undiagnosed and asymptomatic carriers of infection for treatment and assist targeted interventions for prevention of sustained transmission. Evidence from simulation studies suggests that RCD is more beneficial in low-transmission settings with a recent decrease in transmission[37]. In RCD surveillance systems where all individuals within a defined catchment zone of confirmed cases are screened, the likelihood of detecting asymptomatic cases can be substantially higher than that of asymptomatic cases detected by passive case detection[36,38].

As a method of surveillance, however, RCD can be logistically and financially demanding[36] and may be limited in detecting asymptomatic cases if presence of fever is a prerequisite for reactive screening[36]. There is also evidence to suggest that RCD practices such as guidelines for the defined radius where screening around the index case takes place varies between control programmes, and is often decided arbitrarily and by what is operationally feasible[6]. Screening of all household members as opposed to symptomatic individuals only, also varies to a large extent between the control programmes of different countries[52,37-39], and a standard, evidence based protocol for implementation of reactive case detection would be beneficial to control programmes of countries wishing to achieve elimination[20].

3.2.2 Proactive case detection

Proactive case detection (PACD) involves parasitaemia screening and treatment among high risk population groups in the absence of a passively detected index case[40]. This method of active case detection can be useful for identifying clusters or hotspots of asymptomatic infection among high risk groups which can act as reservoirs of infection, maintaining year round transmission[41], but can be challenging to encourage population participation in areas with low perceived transmission risk[40]. PACD can be a useful surveillance strategy in low to moderate transmission settings with defined spatial or temporal risk such as elevated risk

among hotpops or higher seasonal transmission[40]. PACD through repeated household screening can substantially reduce parasite prevalence when compared with households screened at only one time point[42]. Repeated active household screening is resource intensive however and may not be logistically feasible for control programmes in many countries.*3.3 Cross-sectional surveys*

Georeferenced cross-sectional surveys such as Demographic and Health surveys (DHS) or Malaria Indicator Surveys (MIS)[43] are surveys administered to a randomly selected subset of the population which collect key malaria indicator data such as usage and ownership of LLINs, socio-demographic information, IRS coverage and population level malaria prevalence[44]. Cross-sectional surveys of this type can be used to assess heterogeneities in, and drivers of, malaria risk[7,45], and as a way of estimating the spatial distribution of baseline disease prevalence before implementation of control interventions.

Surveys such as MIS are useful in moderate transmission settings as an impact evaluation tool or as a means of assessing change in prevalence but become less useful as transmission decreases[40,46]. In low transmission settings diagnostic tools such as microscopy and RDTs may not be sensitive enough to detect low population levels of infection and as these surveys capture prevalence at only one time point, evaluation of progress toward elimination becomes difficult in low transmission settings[46]. Cross-sectional surveys such as these therefore become less cost-effective for assessing community level infection using household surveys and targeted surveillance of high risk populations should instead be prioritised[47]. In low transmission or elimination settings, case-control analysis of routine surveillance data to identify risk factors for malaria can provide a more suitable epidemiological tool where cross-sectional surveys are inappropriate[47,48].

4. Challenges to malaria elimination

Challenges to achieving malaria elimination are often multi-factorial, and are often outside the control of national malaria programmes, such as a lack of adequate funding, internal conflict and mass displacement of populations[2]. This section focusses on some of the challenges to elimination that can be approached from an epidemiological perspective.

4.1 Heterogeneity in transmission

Drivers of malaria transmission are complex and varied and include environmental factors, vector population dynamics, and anthropogenic and socio-demographic factors[49]. These influencing factors can differ by country, as well as exhibit small area variation, which gives rise to heterogeneity in efficacy of malaria control programmes[4,50]. A contributing factor to this heterogeneity arises from the changing epidemiological profile of at-risk populations, as progress is made toward elimination. In low transmission, and decreasing prevalence areas, a higher proportion of adults and males[7], hard to reach populations[10], people involved in high-risk occupational activities[47,51], migrants and rural communities with poor access to health services are often at disproportionate risk of infection[10]. As transmission decreases, surveillance and epidemiological tools must be tailored for examining heterogeneity in malaria transmission and detecting residual foci of transmission[12,52]. Decisions made regarding malaria control need to take local heterogeneity in transmission and vector and parasite resistance into account[53,54] and focus on targeted control operations rather than a universal approach to intervention coverage and treatment[12,55].

4.2 Asymptomatic malaria and low density infections

As malaria prevalence decreases under intensified control[41,56-58], and heterogeneity in malaria transmission arises[58,59], malaria transmission may recede to small pockets of residual transmission, where asymptomatic infections can still occur despite the lower intensity of

transmission and waning population immunity. Asymptomatic or undiagnosed infections are invisible to the health system[10], but contribute to sustained transmission, and detection of these infections is a priority for national control, surveillance and intervention programmes[17].

In endemic settings with higher levels of population level immunity, low density infections pose a challenge to screening programmes as they are difficult to detect through active screening using diagnostics with lower sensitivity/specificity such as RDTs, and go undetected by passive surveillance systems as low-density infections are often asymptomatic[60]. In some low transmission settings, *P. vivax* parasites have been found to contribute disproportionately to the asymptomatic and low density reservoirs of infection[61-63]. *P. vivax* infections pose additional challenges to elimination programmes due to infection recrudescence attributed to reactivation of dormant liver-stage hypnozoites and the transmission reservoir that is maintained as a consequence[64].

Residual malaria hotspots may occur as a result of differences in environmental and ecological suitability for anopheles breeding habitats[22,65], differences in intensity of vector control intervention coverage, or may be driven by elements of the climate, environment, ecology or socio-demographic characteristics of the human population[35]. As a result, the same control strategies may not be appropriate for all endemic settings within a country or at all time-points[10]. A shift in intervention strategies, tailored to the local endemic context of individual countries[20] is therefore required and foci of infection and high risk areas should be targeted as a country nears elimination [4,6,50]. Additionally, "hotpops", or high-risk sub-groups in the population should be identified and targeted for reactive case detection, particularly for detection of asymptomatic infection[57].

4.3 Imported malaria

As malaria control programmes progress toward elimination, particular attention needs to be paid to mitigating the potential for malaria resurgence facilitated by human mobility and cross-border migration[66]. Malaria prevalence among mobile populations, at international borders and in forested areas may often go undetected and provide reservoirs of introduced parasite infections in low transmission locations or areas where malaria elimination has already been achieved [17,67]. Migratory workers in particular may contribute to a large proportion of the infection reservoir[68].High-risk groups within this population such as forest dwellers[7] and miners[18] are often missed by malaria screening operations and LLIN distribution programmes, which typically target resources at village and household level[68]. Migratory workers may also have barriers to health access due to their remote location and mobility[69].

Active surveillance and screening of migratory populations which records travel history and origin of infection, and regional collaboration initiatives will be integral to preventing re-establishment of local transmission post elimination[6,20,12,34]. Receptivity to reintroduction, assessed through examination of the interactions between entomological, ecological, and epidemiological factors favourable to transmission[70,71], will aid in assessing the risk to resurgence in an area where elimination has been achieved but is connected by human mobility to an area where transmission remains endemic. Detection of imported cases, strengthening of control programmes in countries whose neighbours are aiming to achieve elimination, and regional collaborations for malaria control will be mutually beneficial for achieving regional elimination across connected areas[37].

4.4 Drug and insecticide resistance

Artemisinin combination therapies (ACTs) are compounds derived from the artemether plant, combined with drugs from a different class in the one tablet. Artemisinin is paired with a partner drug to ensure simultaneous administration of both drugs and to clear remaining

parasites after three days of ACT treatment administration, as artemisinin has a short half-life[72]. Artemisinin derivatives include dihydroartemisinin, artesunate and artemether[73]. Partner drugs include mefloquine, sulfadoxine/pyrimethamine, lumefantrine, amodiaquine and chlorproguanil/dapsone[73].

ACTs were introduced as malaria treatment in South East Asia in the 1990s following emergence of antimalarial resistance to chloroquine and sulfadoxine-pyrimethamine, and became the front-line treatment for malaria following WHO recommendations in 2005[74]. Malaria parasites show 'a remarkable ability to develop resistance'[75], and ACT resistant parasites are increasingly widespread in parts of South East Asia[76,77]. Current dihydroartemisinin-piperaquine and mefloquine-artesunate failure rates are over 30% in northern Cambodia and the Myanmar-Thailand border areas, which severely threatens malaria control and elimination efforts[78,79].

Cost effective and user friendly molecular methods and surveillance tools for detecting antimalarial resistance are required to detect the emergence of resistant parasites[80]. The Worldwide Antimalarial Resistance Network (WWARN) ACT Partner Drugs Molecular Surveyor, for example, is an online tool and freely accessible map interface for a database of over 86,000 samples from over 76 countries on antimalarial resistance biomarkers obtained from published data[81]. The aim of this tool is to provide publically accessible data and maps to aid countries and organisations in planning their response to drug resistance[53].

Resistance to the four insecticide classes has emerged in malaria vector populations globally, but of particular concern is emergence to pyrethroids, the group of insecticides cleared for use in all LLINs, and for IRS in many countries, by the WHO Pesticide Evaluation Scheme (WHOPES) [82]. Of the 79 malaria endemic countries from which data on insecticide resistance of *Anopheles* vectors was collected by the WHO in 2016, only 10 were found to have no evidence of resistance to any of the four insecticide classes[2,83]. Resistance of malaria vectors

to insecticide can be phenotypic resistance, a vectors ability to resist and survive the effects of an insecticide, or emerge through resistance mechanisms, the underlying genes that confer inherited traits of resistance. Resistance may be considered as impacting on intervention effectiveness when physiological resistance is established as the cause for increasing malaria transmission[83]. Widespread emergence of malaria vectors to insecticides, can result in reduction in LLIN efficacy and control programme failure, and monitoring and surveillance of vector resistance is essential to inform selection of appropriate vector control interventions[83,84].

5. Epidemiological tools for informing control and surveillance strategies

5.1 Maps of Malaria risk and infection

As countries approach elimination, high-resolution maps of malaria, including spatial clusters of cases or spatial predictions of risk, are useful in detecting the remaining foci of transmission[33], informing national strategic plans for malaria control and elimination, and targeting interventions where they may have the most impact[85,86]. The type of maps that are most useful may be dependent on whether a programme is aiming for control or elimination, or the geographic scale of the area of focus. Visualisation of the spatial distribution of malaria cases on a map enables an assessment of changing transmission patterns and heterogeneity in disease burden [87 88], including in response to interventions[89-91]. Maps of disease data may be better at conveying information in an accessible way to a wider audience than more traditional approaches such as tables, although preferences may differ depending on target populations[79]. Such maps can be produced for the purpose of visualisation using GIS, which are software systems for collating and analysing georeferenced data.

5.2 Spatial analysis

Georeferenced disease survey data can be spatially joined with remotely sensed climate and environment data in a GIS to produce a dataset that can be used to develop spatial statistical models, which can be used to identify and explain spatial patterns in the data and to predict the spatial risk of infection[22,92]. Spatial prediction is a form of interpolation, whereby a model is used to estimate a variable of interest (e.g. malaria risk) at non-sampled locations within a geographical area. There are multiple different approaches for spatial prediction, including generalized linear models (GLM) and other regression approaches, polynomial trend surfaces, geostatistical models, multiple criteria decision analysis, and Bayesian Decision Networks (BDN).

In recent years, an approach that integrates geostatistical concepts into a GLM framework using Bayesian methods for parameter estimation, called model-based geostatistics (MBG), has become increasingly popular. In model-based geostatistics, Bayesian inference is used to estimate model parameters and the GLM framework allows for the incorporation of covariate effects and non-Gaussian distributed outcomes (which are challenging within a traditional geostatistical approach), whereas the geostatistical component allows the explicit inclusion of spatial structure [87,93]. This approach has enabled spatial risk prediction of malaria endemicity and mapping of vector habitats[94,95,22], and is the underpinning method used by the globally influential Malaria Atlas Project[96].

Model based geostatistics and other Bayesian modelling approaches have the advantage of being able to provide a robust quantitative estimate of parameter uncertainty, including in spatial predictions[97,98,22]. This uncertainty may arise from specification of the model and from uncertainty inherent in parasite survey and remotely sensed environmental data due to observational measurement error and natural variation among populations being surveyed[92,99,100]. Mapping the uncertainty of predictions allows visualisation of how reliable

risk estimations are, and provides information on the potential consequences that could arise from relying on the spatial predictions for intervention planning[101].

5.3 Genetic epidemiological tools

Genetic diversity of *Plasmodium* parasite populations may be reflective of malaria transmission intensity, and the extent of parasite mixing between distinct geographic locations influenced by isolation or connectivity of parasites and their hosts from other populations[102]. In high transmission areas, genetic populations tend to be more diverse with little structure in the parasite population due to lots of genetic crossover and transmission between distinct geographic locations[103]. In low transmission areas a parasite population structure, low genetic diversity and high genetic differentiation arises in the parasite population as a result of intensive vector control interventions causing malaria transmission to recede to geographic pockets of transmission[50,104]. High genetic parasite diversity has also been observed in low-transmission settings where a high proportion of malaria infection is attributed to importation of malaria infection[105]. Genetically distinct parasite strains found in abundance in different geographic areas may imply a certain amount of connectivity between those locations[106,50,107]. Therefore, examination of predominant genotypes and associated gene flow may give insight into parasite and human migration or the geographic source of new infections[108,109,110].

Genomic profiling, or 'barcoding' of *Plasmodium* parasites using haplotypes specific to different geographic locations can help identify these geographically distinct clusters, and molecular surveillance can help identify sources of new infection[110]. Combining genetic tools with spatial analytical methods could allow for spatial visualisation and prediction of parasite genetic structure. The benefit of such an approach could be to assist in identifying spatial dominance and geographical niches of parasite genotypes, and, consequently, help infer parasite migration patterns and how this movement has impacted on malaria transmission

between different geographical areas[111]. However, there may be financial and logistic barriers to introducing genetic surveillance in some elimination countries. Therefore, more work needs to be done to evaluate the usefulness of these tools and to establish how they can be cost-effectively and sustainably implemented in elimination settings.

5.4 Mathematical Models

Mathematical models are useful tools for understanding malaria transmission in different endemic settings and predicting patterns of diseas[37,112,113]and for informing ways to interrupt transmission by simulating the effects of different interventions[114,115,116]. Within a modelling framework, parasite transmission is governed by vector capacity, density, longevity and biting rates[66]. The basic reproduction number in mathematical models, a measure of potential transmission intensity, is defined by the number of secondary infections arising from the introduction of an infectious individual to a totally susceptible population[115,117]. The first mathematical model for malaria, developed by Ronald Ross, was a basic representation of malaria transmission incorporating susceptible, infectious and recovered (SIR) population compartments with transmission governed by parameters on vector biting rate and proportion of bites that produce infection in humans. The model was tested by fitting functions to epidemic curves[118,119] and provided evidence for malaria prevention by vector control interventions[119,120]. MacDonald updated this model by including the extrinsic within-vector development period of the parasite[66,119]. Since then, mathematical models have been used to examine and explore the effect of combined interventions for elimination of artemisinin-resistant parasites, examine the impact of ceasing interventions too early on resurgence of anti-malarial drug resistant parasites, amongst many other applications to malaria[121]. Models have also been used for simulation of malaria transmission under varying conditions of human migration and vulnerability to reintroduction once malaria has been eliminated[52], an application that will be further explored in this thesis.

6. Conclusions

Effective epidemiological tools are essential for informing intervention efforts, developing a better understanding of malaria transmission heterogeneity and understanding the impact of human (and parasite) movement on malaria transmission dynamics[122]. Novel tools for informing targeted interventions are needed to address challenges that arise as the global prevalence of malaria decreases and the goal of global malaria elimination advances. In particular, epidemiological tools need to be appropriately tailored to control, elimination and prevention of reintroduction settings.

References

1 World Health Organization. World Malaria Report 2016. (Geneva, 2016).
2 World Health Organization. World Malaria Report 2017. (Geneva, 2017).
3 World Health Organization. World Malaria Report 2018. (Geneva, 2018).
4 Feachem, R. G. *et al.* Shrinking the malaria map: progress and prospects. *The Lancet* **376**, 1566-1578 (2010).
5 Kho, W. G. *et al.* A multiplex polymerase chain reaction for a differential diagnosis of Plasmodium falciparum and Plasmodium vivax. *Parasitol Int* **52**, doi:10.1016/s1383-5769(03)00028-x (2003).
6 Moonen, B. *et al.* Operational strategies to achieve and maintain malaria elimination. *The Lancet* **376**, 1592-1603 (2010).
7 Cotter, C. *et al.* The changing epidemiology of malaria elimination: new strategies for new challenges. *The Lancet* **382**, 900-911 (2013).
8 World Health Organization. World malaria report 2019. (2019).
9 World Health Organization. High burden to high impact: a targeted malaria response. (World Health Organization, 2018).
10 World Health Organization. *Global technical strategy for malaria 2016-2030.* (World Health Organization, 2015).
11 World Health Organization. Global malaria control and elimination: report of a technical review. (2008).
12 World Health Organization. Malaria elimination: a field manual for low and moderate endemic countries. *Malaria elimination: a field manual for low and moderate endemic countries.* (2007).
13 World Health Organization. *A framework for malaria elimination.* (World Health Organization, 2017).
14 World Health Organization. *Guidelines for the treatment of malaria.* (World Health Organization, 2015).
15 Nayyar, G. M., Breman, J. G., Newton, P. N. & Herrington, J. Poor-quality antimalarial drugs in southeast Asia and sub-Saharan Africa. *The Lancet infectious diseases* **12**, 488-496 (2012).
16 Hemingway, J., Field, L. & Vontas, J. An overview of insecticide resistance. *Science* **298**, 96-97 (2002).
17 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. *Malaria journal* **18**, 32 (2019).
18 Eer, E. D., Bretas, G. & Hiwat, H. Decreased endemic malaria in Suriname: moving towards elimination. *Malaria journal* **17**, 56 (2018).
19 Hofmann, N. E. *et al.* The complex relationship of exposure to new Plasmodium infections and incidence of clinical malaria in Papua New Guinea. *Elife* **6**, e23708 (2017).
20 Gueye, C. S. *et al.* Active case detection for malaria elimination: a survey among Asia Pacific countries. *Malaria Journal* **12**, 358 (2013).
21 World Health Organisation. *National Passive Surveillance*, <https://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/passive/en/> (2019).
22 Clements, A. C., Reid, H. L., Kelly, G. C. & Hay, S. I. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? *The Lancet infectious diseases* **13**, 709-718 (2013).
23 Heng, S. *et al.* Passive case detection of malaria in Ratanakiri Province (Cambodia) to detect villages at higher risk for malaria. *Malaria journal* **16**, 104 (2017).
24 Ohrt, C. *et al.* Information systems to support surveillance for malaria elimination. *The American journal of tropical medicine and hygiene* **93**, 145-152 (2015).
25 Sturrock, H. J. *et al.* Mapping malaria risk in low transmission settings: challenges and opportunities. *Trends in parasitology* **32**, 635-645 (2016).
26 Byass, P. *et al.* The long road to elimination: malaria mortality in a South African population cohort over 21 years. *Global health, epidemiology and genomics* **2** (2017).

27 World Health Organisation / UNICEF. Integrated Community Case Management (iCCM); An equity-focused strategy to improve access to essential treatment services for children., (2012).
28 Mathanga, D. P. *et al.* Malaria control in Malawi: current status and directions for the future. *Acta tropica* **121**, 212-217 (2012).
29 Malaria Eradication Science Alliance. *Enhanced responses with various approaches to Active Case Detection: Active case detection and treatment*, 2016).
30 Wickremasinghe, R., Fernando, S. D., Thillekaratne, J., Wijeyaratne, P. M. & Wickremasinghe, A. R. Importance of active case detection in a malaria elimination programme. *Malaria journal* **13**, 186 (2014).
31 World Health Organization. Disease surveillance for malaria elimination: an operational manual. 2012. *Geneva: World Health Organization Google Scholar*.
32 World Health Organisation. *Overview of malaria elimination*, <http://www.who.int/malaria/areas/elimination/overview/en/> (2015).
33 Wangdi, K. *et al.* Development and evaluation of a spatial decision support system for malaria elimination in Bhutan. *Malaria journal* **15**, 180 (2016).
34 Velarde-Rodríguez, M. *et al.* Origin of malaria cases: a 7-year audit of global trends in indigenous and imported cases in relation to malaria elimination. *Global health action* **8**, 29133 (2015).
35 Wangdi, K. *et al.* Analysis of clinical malaria disease patterns and trends in Vietnam 2009–2015. *Malaria journal* **17**, 332 (2018).
36 Sturrock, H. J. *et al.* Reactive case detection for malaria elimination: real-life experience from an ongoing program in Swaziland. *PloS one* **8**, e63830 (2013).
37 Gerardin, J. *et al.* Effectiveness of reactive case detection for malaria elimination in three archetypical transmission settings: a modelling study. *Malaria journal* **16**, 248 (2017).
38 Aidoo, E. K. *et al.* Reactive case detection of Plasmodium falciparum in western Kenya highlands: effective in identifying additional cases, yet limited effect on transmission. *Malaria journal* **17**, 111 (2018).
39 Eijk, A. M. *et al.* What is the value of reactive case detection in malaria control? A case-study in India and a systematic review. *Malaria journal* **15**, 67 (2016).
40 Sturrock, H. J. *et al.* Targeting asymptomatic malaria infections: active surveillance in control and elimination. *PLoS medicine* **10** (2013).
41 Ogutu, B. *et al.* Treatment of asymptomatic carriers with artemether-lumefantrine: an opportunity to reduce the burden of malaria? *Malaria journal* **9**, 30 (2010).
42 Sutcliffe, C. G. *et al.* Reduced risk of malaria parasitemia following household screening and treatment: a cross-sectional and longitudinal cohort study. *PloS one* **7** (2012).
43 Roca-Feltrer, A., Lalloo, D. G., Phiri, K. & Terlouw, D. J. Rolling Malaria Indicator Surveys (rMIS): a potential district-level malaria monitoring and evaluation (M&E) tool for program managers. *The American journal of tropical medicine and hygiene* **86**, 96-98 (2012).
44 Ministry of Health National Malaria Control Programme. Malawi Malaria Indicator Survey 2017 Key Indicators. (The DHS Program, Rockville, Maryland, USA, 2017).
45 Baidjoe, A. Y. *et al.* Factors associated with high heterogeneity of malaria at fine spatial scale in the Western Kenyan highlands. *Malaria journal* **15**, 307 (2016).
46 Hsiang, M. S. *et al.* Surveillance for malaria elimination in Swaziland: a national cross-sectional study using pooled PCR and serology. *PLoS one* **7** (2012).
47 Jacobson, J. O. *et al.* Surveillance and response for high-risk populations: what can malaria elimination programmes learn from the experience of HIV? *Malaria journal* **16**, 33 (2017).
48 Smith, J. L. *et al.* Malaria risk in young male travellers but local transmission persists: a case–control study in low transmission Namibia. *Malaria journal* **16**, 70 (2017).
49 Alonso, P. L. *et al.* A research agenda to underpin malaria eradication. *PLoS medicine* **8**, e1000406 (2011).
50 Nabet, C. *et al.* Genetic diversity of Plasmodium falciparum in human malaria cases in Mali. *Malaria journal* **15**, 353 (2016).
51 Thanh, P. V. *et al.* Epidemiology of forest malaria in Central Vietnam: the hidden parasite reservoir. *Malaria journal* **14**, 86 (2015).

52 malERA Refresh Consultative Panel on Combination Interventions Modelling. malERA: An updated research agenda for combination interventions and modelling in malaria elimination and eradication. *PLoS medicine* **14**, e1002453 (2017).
53 Sibley, C. H., Barnes, K. I. & Plowe, C. V. (BioMed Central, 2007).
54 Tinto, H., Valea, I., Ouedraogo, J.-B. & Guiguemdé, T. Lessons learnt from 20 years surveillance of malaria drug resistance prior to the policy change in Burkina Faso. *Annals of parasitology* **62** (2016).
55 Tatem, A. J. *et al.* Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria journal* **13**, 52 (2014).
56 Mosha, J. F. *et al.* Hot spot or not: a comparison of spatial statistical methods to predict prospective malaria infections. *Malaria journal* **13**, 53 (2014).
57 Bousema, T. *et al.* Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS medicine* **9**, e1001165 (2012).
58 Mogeni, P. *et al.* Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. *BMC medicine* **15**, 121 (2017).
59 Bejon, P. *et al.* Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS medicine* **7**, e1000304 (2010).
60 Björkman, A. B. Asymptomatic low-density malaria infections: a parasite survival strategy? *The Lancet Infectious Diseases* **18**, 485-486 (2018).
61 Motshoge, T. *et al.* Molecular evidence of high rates of asymptomatic P. vivax infection and very low P. falciparum malaria in Botswana. *BMC infectious diseases* **16**, 520 (2016).
62 van Eijk, A. M. *et al.* The burden of submicroscopic and asymptomatic malaria in India revealed from epidemiology studies at three varied transmission sites in India. *Scientific reports* **9**, 1-11 (2019).
63 Sattabongkot, J. *et al.* Prevalence of asymptomatic Plasmodium infections with sub-microscopic parasite densities in the northwestern border of Thailand: a potential threat to malaria elimination. *Malaria journal* **17**, 329 (2018).
64 Adekunle, A. I. *et al.* Modeling the dynamics of Plasmodium vivax infection and hypnozoite reactivation in vivo. *PLoS neglected tropical diseases* **9** (2015).
65 Coulibaly, D. *et al.* Spatio-temporal analysis of malaria within a transmission season in Bandiagara, Mali. *Malaria journal* **12**, 82 (2013).
66 Macdonald, G. Theory of the eradication of malaria. *Bulletin of the World Health Organization* **15**, 369 (1956).
67 Wangdi, K., Gatton, M. L., Kelly, G. C. & Clements, A. C. in *Advances in parasitology* Vol. 89 79-107 (Elsevier, 2015).
68 Kounnavong, S., Gopinath, D., Hongvanthong, B., Khamkong, C. & Sichanthongthip, O. Malaria elimination in Lao PDR: the challenges associated with population mobility. *Infectious diseases of poverty* **6**, 81 (2017).
69 Canavati, S. E. *et al.* Risk factor assessment for clinical malaria among forest-goers in a pre-elimination setting in Phu Yen Province, Vietnam. *Malaria Journal* **18**, 435 (2019).
70 Ejov, M., Davidyants, V. & Zvantsov, A. Regional framework for prevention of malaria reintroduction and certification of malaria elimination 2014–2020. (2014).
71 Franke, J., Gebreslasie, M., Bauwens, I., Deleu, J. & Siegert, F. Earth observation in support of malaria control and epidemiology: MALAREO monitoring approaches. *Geospatial health*, 117-131 (2015).
72 Nsanzabana, C. Resistance to Artemisinin Combination Therapies (ACTs): Do Not Forget the Partner Drug! *Tropical medicine and infectious disease* **4**, 26 (2019).
73 Malaria Consortium. *Artemisinin-based combination therapy*, <https://www.malariaconsortium.org/pages/112.htm> (2019).
74 Ashley, E. A. *et al.* Spread of artemisinin resistance in Plasmodium falciparum malaria. *New England Journal of Medicine* **371**, 411-423 (2014).
75 White, N. Qinghaosu in combinations. *Medecine tropicale: revue du Corps de sante colonial* **58**, 85-88 (1998).
76 Noedl, H. *et al.* Evidence of artemisinin-resistant malaria in western Cambodia. *New England Journal of Medicine* **359**, 2619-2620 (2008).

77 Phyo, A. P. *et al.* Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *The Lancet* **379**, 1960-1966 (2012).
78 WorldWide Antimalarial Resistance Network. *Tracking resistance to artemisinin collaboration II*, <https://www.wwarn.org/working-together/partner-projects/tracking-resistance-artemisinin-collaboration-ii> (2019).
79 Henry, N. L. Knowledge management: a new concern for public administration. *Public Administration Review*, 189-196 (1974).
80 Nsanzabana, C., Djalle, D., Guérin, P. J., Ménard, D. & González, I. J. Tools for surveillance of anti-malarial drug resistance: an assessment of the current landscape. *Malaria journal* **17**, 75 (2018).
81 Otienoburu, S. D. *et al.* An online mapping database of molecular markers of drug resistance in Plasmodium falciparum: the ACT Partner Drug Molecular Surveyor. *Malaria journal* **18**, 12 (2019).
82 Zaim, M., Aitio, A. & Nakashima, N. Safety of pyrethroid-treated mosquito nets. *Medical and veterinary entomology* **14**, 1-5 (2000).
83 World Health Organization. Global report on insecticide resistance in malaria vectors: 2010–2016. (2018).
84 Ranson, H. *et al.* Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends in parasitology* **27**, 91-98 (2011).
85 Omumbo, J. A., Noor, A. M., Fall, I. S. & Snow, R. W. How well are malaria maps used to design and finance malaria control in Africa? *PLoS One* **8**, e53198 (2013).
86 Carter, R., Mendis, K. N. & Roberts, D. Spatial targeting of interventions against malaria. *Bulletin of the World Health Organization* **78**, 1401-1411 (2000).
87 Brooker, S. & Clements, A. C. Spatial heterogeneity of parasite co-infection: Determinants and geostatistical prediction at regional scales. *International journal for parasitology* **39**, 591-597 (2009).
88 Xia, J. *et al.* Spatial, temporal, and spatiotemporal analysis of malaria in Hubei Province, China from 2004–2011. *Malaria journal* **14**, 145 (2015).
89 Sturrock, H. J. *et al.* Mapping Malaria Risk in Low Transmission Settings: Challenges and Opportunities. *Trends in parasitology* (2016).
90 Onyiri, N. Estimating malaria burden in Nigeria: a geostatistical modelling approach. *Geospatial health* **10** (2015).
91 Magalhães, R. J. S. *et al.* Geographical distribution of human Schistosoma japonicum infection in the Philippines: tools to support disease control and further elimination. *International journal for parasitology* **44**, 977-984 (2014).
92 Dalrymple, U., Mappin, B. & Gething, P. W. Malaria mapping: understanding the global endemicity of falciparum and vivax malaria. *BMC medicine* **13**, 140 (2015).
93 Diggle, P. J., Tawn, J. A. & Moyeed, R. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47**, 299-350 (1998).
94 Gebreslasie, M. T. A review of spatial technologies with applications for malaria transmission modelling and control in Africa. *Geospatial health* **10** (2015).
95 Shankar, S. & Agrawal, D. K. Role of geospatial technology in identifying natural habitat of malarial vectors in South Andaman, India. *Journal of vector borne diseases* **53**, 54 (2016).
96 Millar, J. *et al.* Detecting local risk factors for residual malaria in northern Ghana using Bayesian model averaging. *Malaria journal* **17**, 343 (2018).
97 Amoah, B., Giorgi, E., Heyes, D. J., Burren, S. & Diggle, P. J. Geostatistical modelling of the association between malaria and child growth in Africa. *International journal of health geographics* **17**, 7 (2018).
98 Diggle, P. & Lophaven, S. Bayesian geostatistical design. *Scandinavian Journal of Statistics* **33**, 53-64 (2006).
99 Hamm, N. A., Magalhães, R. J. S. & Clements, A. C. Earth observation, spatial data quality, and neglected tropical diseases. *PLoS neglected tropical diseases* **9**, e0004164 (2015).
100 Basáñez, M. a.-G., Marshall, C., Carabin, H., Gyorkos, T. & Joseph, L. Bayesian statistics for parasitologists. *Trends in parasitology* **20**, 85-91 (2004).

101 Diggle, P. J. & Ribeiro Jr, P. J. Bayesian inference in Gaussian model-based geostatistics. *Geographical and Environmental Modelling* **6**, 129-146 (2002).
102 Schultz, L. *et al.* Multilocus haplotypes reveal variable levels of diversity and population structure of Plasmodium falciparum in Papua New Guinea, a region of intense perennial transmission. *Malaria journal* **9**, 336 (2010).
103 Razak, M. R. M. A. *et al.* Genetic diversity of Plasmodium falciparum populations in malaria declining areas of Sabah, East Malaysia. *PloS one* **11**, e0152415 (2016).
104 Khaireh, B. A. *et al.* Population genetics analysis during the elimination process of Plasmodium falciparum in Djibouti. *Malaria journal* **12**, 201 (2013).
105 Roh, M. E. *et al.* High Genetic Diversity of Plasmodium falciparum in the Low-Transmission Setting of the Kingdom of Eswatini. *The Journal of infectious diseases* **220**, 1346-1354 (2019).
106 Daniels, R. F., Rice, B. L., Daniels, N. M., Volkman, S. K. & Hartl, D. L. The utility of genomic data for Plasmodium vivax population surveillance. *Pathogens and global health* **109**, 153-161 (2015).
107 Hastings, I. & Wedgwood-Oppenheim, B. Sex, strains and virulence. *Parasitology Today* **13**, 375-383 (1997).
108 Preston, M. D. *et al.* A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. *Nature communications* **5** (2014).
109 Delgado-Ratto, C. *et al.* Population genetics of Plasmodium vivax in the Peruvian Amazon. *PLoS neglected tropical diseases* **10**, e0004376 (2016).
110 Wong, V. K. *et al.* An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. *Nature communications* **7**, 12827 (2016).
111 Chang, H.-H. *et al.* The geography of malaria elimination in Bangladesh: combining data layers to estimate the spatial spread of parasites. *bioRxiv*, 421578, doi:10.1101/421578 (2018).
112 Stuckey, E. M., Smith, T. A. & Chitnis, N. Estimating malaria transmission through mathematical models. *Trends in parasitology* **29**, 477-482 (2013).
113 Anderson, R. M., May, R. M. & Anderson, B. *Infectious diseases of humans: dynamics and control*. Vol. 28 (Wiley Online Library, 1992).
114 Tanner, M. *et al.* Malaria eradication and elimination: views on how to translate a vision into reality. *BMC medicine* **13**, 167 (2015).
115 Smith, D. L. *et al.* A sticky situation: the unexpected stability of malaria elimination. *Phil. Trans. R. Soc. B* **368**, 20120145 (2013).
116 Acevedo, M. A. *et al.* Spatial Heterogeneity, Host Movement and Mosquito-Borne Disease Transmission. *PloS one* **10**, e0127552 (2015).
117 Vynnycky, E. & White, R. *An introduction to infectious disease modelling*. (Oxford University Press, 2010).
118 Ross, R. Some a priori pathometric equations. *British medical journal* **1**, 546 (1915).
119 Mandal, S., Sarkar, R. R. & Sinha, S. Mathematical models of malaria-a review. *Malaria journal* **10**, 202 (2011).
120 Ross, R. Inaugural lecture on the possibility of extirpating malaria from certain localities by a new method. *British medical journal* **2**, 1 (1899).
121 Maude, R. J. *et al.* The last man standing is the most resistant: eliminating artemisinin-resistant malaria in Cambodia. *Malaria journal* **8**, 31 (2009).
122 Tanner, M. & Savigny, D. d. Malaria eradication back on the table. *Bulletin of the World Health Organization* **86**, 82-82 (2008).

CHAPTER 3

A review of malaria epidemiology and control and elimination programmes in Papua New Guinea

CHAPTER 3. A REVIEW OF MALARIA EPIDEMIOLOGY AND CONTROL AND ELIMINATION PROGRAMMES IN PAPUA NEW GUINEA

CONTEXT

Chapter 3 summarises the current epidemiology of malaria in Papua New Guinea, including environmental drivers of transmission and vector habitats and species distribution. In this chapter we also review the three major historical time periods of malaria control in PNG, the current control programme in operation in the country, and future directions for control and elimination.

The research presented here is a result of a comprehensive review of literature on malaria in PNG published over the last century, beginning with expeditions to PNG in the early 1900s and leading up to published research on current control and elimination efforts. By reviewing and summarising the three main periods of malaria control in PNG, which we hope to identify common themes regarding strategies and challenges which have contributed to relative successes of previous control programmes, and draw conclusions about factors which led to their eventual breakdown. In doing so we extrapolate that integrated control interventions across all periods of control had success in reducing malaria prevalence however, logistical and financial pressures of such concerted control efforts, poor planning and administration, and lack of political will eventually contributed to the programmes being abandoned. Previous control programmes and interventions have also altered the epidemiology of malaria in PNG, which lends support for evaluating current interventions through national surveillance of malaria prevalence, incidence, transmission and species distribution as the control programme progresses.

This paper is formatted as a manuscript for submission to Malaria Journal.

CHAPTER 3. A REVIEW OF MALARIA EPIDEMIOLOGY AND CONTROL AND ELIMINATION PROGRAMMES IN PAPUA NEW GUINEA

Cleary E.[1], Hetzel M[2,3], Clements, A.C.A[4].

1. Research School of Population Health, Australian National University, Canberra, Australia.

2. Swiss Tropical and Public Health Institute, Basel, Switzerland

3. University of Basel, Basel, Switzerland

4. Curtin University, Perth, Australia

Abstract

The research and control of malaria has a long history in Papua New Guinea, sometimes resulting in substantial changes to the distribution of infection and transmission dynamics in the country. There have been three major periods of malaria control in PNG, and a current control programme commenced in 2004. Each previous control programme had success in reducing malaria burden in the country, but multiple factors led to programme failures and the eventual breakdown of control. A comprehensive review of literature dating from 1900 to 2018 was undertaken to summarise the current control, epidemiology, vector ecology and environmental drivers of malaria transmission in PNG. Furthermore, control strategies employed in the past and reasons underpinning the ultimate failure of these previous control and elimination programmes are discussed.

1. Introduction

The population of Papua New Guinea (PNG) consists of over 8 million people spread across 22,000 villages. Many of these villages are situated in rugged landscapes with difficult terrain

resulting in many highly isolated communities and a population that is one of the most culturally and ethnically diverse in the world[1,2]. The human settlement patterns in PNG have historically been shaped by malaria transmission, with the population mostly living above 1300m, where temperatures are too low for sustained malaria transmission[3,4], and below 600m where the local population has developed acquired immunity to infection[5-7]. The main species of human malaria in PNG are *Plasmodium falciparum (P. falciparum), Plasmodium v*ivax (*P. v*ivax) and, to a lesser extent, *Plasmodium malariae (P. malariae)*. *P. falciparum* is the predominant parasite in PNG, although a decrease in *P. falciparum* infection, observed since 2010, has resulted in an increase in relative abundance of *P. vivax*[3,4,8]. *P. falciparum* and *P. vivax* have a spatial distribution that covers the entire country with no geographic preference of the two parasite species observed in surveyed locations[9,10].

A renewed commitment to malaria control in 2004, and free nation-wide LLIN distribution campaigns carried out since 2005 has resulted in a substantial decrease in national malaria prevalence[2,9,11]. Despite this however, prevalence remains one of the highest outside Sub-Saharan Africa and over 90% of the population are considered to be at risk[12]. The aim of this review is to examine the current epidemiology, vector species and drivers of malaria transmission in PNG. We also aim to summarise progress that has been made by the current national malaria control programme, and learnings from the challenges that have led to the breakdown of control programmes in the past.

2. Epidemiological drivers of malaria transmission in PNG

Malaria transmission in PNG is mostly perennial, with year round transmission in the islands and coastal lowlands, and seasonal transmission occurring only on the south coast where transmission halts during the dry season[3]. Marked heterogeneity in transmission and complex

epidemiology exists, driven by diversity in vectors, as well as environmental and climate drivers, which can vary significantly between geographic locations. Small area variation also exists with marked heterogeneity observed between villages, and even households within the same village[9,13-15,16]. In addition to the environmental factors that drive these entomological and parasitological phenomena, human genetic factors such as red blood cell polymorphisms, which cluster in familial groups[17,18], differences in sporozoite and inoculation rates between villages, and response to control interventions also contribute to this heterogeneity in transmission[9,16].

Temperature, rainfall and altitude are the most important environmental drivers of transmission[3], and temporal variations in transmission associated with temperature and rainfall are observed in all areas of the country[19,20]. Temperature is inversely related to altitude and mosquito infectivity[21] while surface gradient and incline is associated with vector ecology and formation of larval habitat. The extent of these associations differ between different provinces, and during the wet and dry seasons[22]. In the mountainous interior and highland provinces of PNG for example, steep slopes are less suitable for the formation of standing pools of water as rainfall runs off into small streams. In the dry season, when the water level of small streams is low, this run off may add abundance to small streams, helping formation of suitable breeding habitats. In the wet season however, when stream water levels are higher, the extra abundance of water and heavy flow of the stream water may flush out breeding sites[22]. The heterogeneity in malaria transmission in PNG can in part be attributed to the variation in environmental drivers of transmission associated with vector abundance and habitat across the country, with malaria vectors in PNG exhibiting wide variation in preferred larval and breeding habitats.

3. Vector species, spatial distribution and habitats

The three main malaria vector species of human importance in PNG are *Anopheles farauti (An. farauti), An. koliensis* and *An. punctulatus*[23-25]. Each species has a wide spatial distribution[10], coexisting to a certain extent, but with distinct ecology and habitats[22]. The abundance of these three species is associated predominantly with rainfall and vegetation cover, however each of the dominant vector species in PNG have distinct ecological preferences[10,22]. The habitat of *An. farauti* is predominantly coastal[10,26], being found in streams, brackish water and coastal villages[3,22]. As a result of its widespread distribution along the PNG coastline[27], this species has the widest distribution among all malaria vectors in the country. Although *An. farauti* mainly have a coastal habitat, *An. farauti 6* is also found at high altitudes, being the only species with a preference for high altitudes[10]. *An. farauti* is most abundant at the end of the dry season[22].

The preferred habitat of *An. koliensis* is fresh water pools in low altitude, inland areas proximal to vegetation, temporary pools in grasslands, and around the edge of forests[3,22]. The abundance of this species may increase over 250% in the wet season[16,28]. *An. punctulatus* is found most commonly in the hills, although their habitats are diverse and extend from lowland coasts and valleys up to elevations of 2000 meters, with preferred breeding sites in shallow, sunlit pools of water created by ruts and drains[3,27]. The larvae of *An. puctulatus* develop rapidly under optimal conditions and have the ability to quickly colonise previously uninhabited areas[22]. The dispersal and breeding habitats of *An. punctulatus* have been impacted and expanded by the activities and movements of humans such as logging and mineral exploration, changes in settlement patterns and agricultural cultivation[29,27].

The three main vectors in PNG have different host and biting preferences with *An. koliensis* and *An. punctulatus* being more anthropophilic than *An. farauti*[30]. *An. koliensis* and *An. punctulatus* have higher inoculation rates than *An. farauti,* mainly because of higher abundance

and biting rates of these species[16]. *An. koliensis* bite at night-time, both indoors and outdoors, and generally rest close to their breeding site after feeding[31], rather than close to their next blood meal indoors[22]. Adult *An. punctulatus* are night-time biters, mainly feeding between midnight and the early morning hours.[3,22] They may rest indoors for prolonged periods of time[31], or close to their breeding site[22]. *An. farauti* are mostly night time and early evening biters, although they do occasionally bite during the day[33]. Biting and resting habits of vectors may influence susceptibility of vectors to IRS interventions[22]. Indoor spraying pilot projects conducted during the 1950s demonstrated that dichloro-diphenyl-trichloroethane (DDT) was effective against *An. punctulatus* and *An. koliensis* but not against *An. farauti*, which subsequently affected vector species composition[32,33].

4. History of malaria control in PNG

4.1 1900 Robert Koch's malaria expedition

Since the early 1900s, malaria in PNG has been the subject of ongoing control programmes, interventions and research[34,35]. Some of the earliest research into the chemotherapeutic application of quinine in the treatment of malaria and interruption of vector breeding was conducted by Robert Koch in the Madang district of PNG in 1900[34-37]. Koch also led early vector control efforts of mosquito elimination through drainage of breeding sites, elimination of larvae by use of the larvicidal fish, *Gambusia affinis*, indoor residual spraying with DDT, infrastructure development around settled areas to prevent vector breeding and routine surveillance among at-risk populations[38,3,5].

Control programmes implemented during this time had some success in halting epidemics, primarily in highlands provinces, and achieved a substantial reduction in prevalence in the rest of the country. However, problems with administration, training and discontent with the

programme, and reluctance of the general population to adhere to a disagreeable regimen of malaria prophylaxis with quinine, eventually contributed to a breakdown of the control programme [3,5]. Programme operations were also less effective on the islands due to poor coordination of activities during peaks and troughs in transmission and the change in vector biting behaviour in response to spray interventions[5].

4.2 During and post-World War II

During the Pacific War of World War II, control efforts in PNG were again intensified, but focussed predominantly on protecting the health of troops, members of the local population involved in national defence[5,39], and labourers working on coastal plantations and in highland regions[36]. Malaria transmission in the highlands manifested as seasonal epidemics and outbreaks, often in tea and coffee plantations where people from the endemic lowlands travelled to work[40,41], and were often marked with substantial morbidity and mortality[42,23].

Beginning in 1942, the malaria control campaign focused on vector control through use of mosquito nets, protective clothing and dimethyl-phthalate repellant, larviciding and draining of breeding grounds near campsites[5,39,43,44]. Infectious disease control interventions around Port Morseby between 1946 and 1948 included the burning off of vegetation in the surrounding areas to expose habitats suitable for *Aedes aegypti* breeding (a major arbovirus vector). However, this burning off of vegetation facilitated instead the formation of breeding grounds suitable for *Anopheles* species[43]. At this point mepacrine, an antimalarial drug related to chloroquine and mefloquine, was widely used as a malaria preventative, predominantly among allied forces during the second world war[45], but subsequent emergence of wide-scale resistance resulted in discontinuation of its use[43].

At the time, DDT was considered the gold standard of insecticides,[43] however the residual properties of DDT as an effective indoor insecticide had not yet been fully realised and its use

at the time was in aerial spraying interventions. An integrated malaria control strategy of 'bonification', a programme of village improvement or community development was instead employed for malaria control. This involved the distribution of antimalarials, drainage of breeding grounds and larval control with *Gambusia.* Interventions were facilitated by aid post orderlies, with one person from each village being trained in control methods and supplied with antimalarials from the Department of Public Health[43]. From about 1950 however, as the burden of disease was decreasing, efforts to sustain the bonification control programme were beginning to wane, as was investment in the training of control specialists and malaria education. The intensified integrated intervention approach of aerial spraying with DDT, and provision of protective clothing, mosquito nets, screens for houses, antimalarials and insect repellent to all at-risk populations, which had been employed in the protection of troops during war time, were subsequently considered to be too expensive to be sustainable in 'peace time administration' and were subsequently abandoned[43].

4.3 1950 to 1980

The first pilot project on residual spraying with DDT as a means of malaria control and eradication was undertaken in Maprik, Sepik province in 1957[5,46]. At the time, DDT was used as a malaria control intervention in well over half the population of PNG as it was considered safe, effective and economically viable[5], and the persistent nature of the insecticide on indoor surfaces meant that contact time with resting vectors was increased[43]. By the early 1970s, the malaria control programme predominantly involved spraying with DDT, supplemented with mass drug administration (MDA) during malaria outbreaks, and covered 14 of the 19 administrative districts of PNG[5]. In the highlands, the spraying regimen operated once per year, and case detection was carried out by passive surveillance. Active monthly surveillance was continued among a population of 20,000 people in the highlands for a period of 18 months but was discontinued following detection of only three positive slides over the surveillance period.

The low rates of infection detected during this surveillance time period impacted the morale of surveillance teams and led to the conclusion that surveillance of this type was too resource intensive for an area of such low parasite rates[5].[47] Prior to commencement of indoor spraying in the highlands, parasite prevalence was 5 - 10%[47], In the islands and lowlands, where prevalence was higher, spraying was carried out twice a year and case detection was carried out by active surveillance[5].

The success of this period of control operations, however, suffered from inconsistencies between different administrative areas in the running of the control programme, caused by a lack of inter-district coordination and communication[5]. The programme was also hampered by inadequate water transport facilities, financial constraints, and too small a number of technical, administration and training staff. Death of pets and destruction of household building materials associated with DDT spraying also led to a general refusal of IRS and a reluctance to allow strangers carrying out these spray operations into households[5,25]. In 1970, a four-year-long integrated MDA and residual spraying campaign in Karamui came to an end and an evaluation survey in 1971 deemed it to have been a failure due to administrative, operational and technical reasons. No regular evaluation of vector susceptibility to DDT was carried out during the campaign and concerns were starting to emerge regarding insecticide resistance[48]. Resistance resulted in a resurgence in local infection rates in some areas to a level higher than had existed pre-control programme commencement[5].

By 1974, after ten years of integrated interventions, and in excess of 30 rounds of spraying in some parts of PNG, the control programme was yielding poor results[25,36] and was operational in only part of the country[49]. It was suggested that the training received by malariologists during the early 1970s placed too much of an emphasis on vector control through IRS with DDT and as a result of broader approaches employed by earlier programmes being neglected[49]. Confidence in an eradication programme was abating in favour of a more realistic goal of

control[49,50]. While the programme had achieved some success in reduction of prevalence and interrupting transmission in the highlands[5], an eradication programme was only believed to have been feasible had the following been achieved: nationwide coverage of the interventions, adequate planning, administration, operation and assessment, sustained financial backing, visible economic benefits, full government support, integration of the programme with good national and local health services, and adequate health education[25,49].

5. Global Fund to Fight Aids, Tuberculosis and Malaria: 2004 to present

Between 2005 and 2009, the National Department of Health, in partnership with Rotarians Against Malaria, led a free nation-wide distribution campaign of 2.4 million LLINs, resulting in 65% national coverage, with the aid of a US$16 million grant awarded by the Global Fund to fight Aids Tuberculosis and Malaria[9,11]. A further grant of US$102 Million awarded in 2009, facilitated distribution of an additional 2.5 million LLINs between 2009 and 2011 to all households in all provinces, resulting in an increase in country level coverage to 81.8% of households owning at least one LLIN[2,51]. In the highly malaria-endemic islands of PNG, coverage increased from 29.3% to 98.3%[51,52] and a significant reduction in prevalence was reported one year after LLIN distribution, with malaria prevalence in the lowlands decreasing from 11.1% in 2008-2009 to 5.1% in 2010-2011 to 0.9% in 2013-2014[4]. Prevalence of *P. falciparum* and *P. vivax* along the north coast of PNG has decreased 12 and 6 fold, respectively following 8 years of malaria control interventions, and surveys carried out in 2013-2014 have recorded a historic low in national transmission[8].

National distribution of LLINs has also resulted in a decrease in malaria prevalence in the highlands since 2010, both directly and indirectly as a consequence of lower rates of infection and importation from the lowlands[4]. Prior to commencement of the current national malaria control programme, little malaria control had been undertaken in the highlands provinces since

the 1980s, and by the early 2000s malaria prevalence in the highlands had rebounded to pre-control levels[40]. Infections at this time were attributed predominantly due to *P. falciparum* and locations and timings of epidemics were largely similar to epidemics of the 1960s and 70s, associated with substantial morbidity and high prevalence of clinical infection[53]. Malaria in the PNG highlands continues to be unstable and is characterised by seasonal epidemics, with endemic transmission occurring in the mountain valleys[7,40].

As well as the free distribution of LLINs, the national malaria control programme expanded to include RDTs or microscopy for diagnosis of febrile illness and ACT for malaria treatment. Until 2002, diagnosis of malaria and treatment was administered on a presumptive basis[2,3,29] and fever cases were often treated as malaria even with a negative diagnosis by RDT[54,55]. New malaria treatment guidelines introduced in 2009 advocate for parasitological diagnosis of fever cases with RDT or light microscopy and treatment only in the case of a positive diagnosis[40,56] and this protocol is now applied in most health centres[8]. Prior to 2010, treatment of the hypnozoite stage of infection with primaquine was not a formal part of the PNG treatment guidelines and 87% of children had a relapse of *P. vivax* infection within 6 weeks of treatment[57]. Since the distribution of these antimalarials and diagnostics to health centres in 2008—2009 and 2010—2011[8,58] there has been an increase in the percentage of health centres stocking RDTs from 17.5% in 2010 to 90.2% in 2012[59,60,61], and in the use of RDTs for diagnosis of febrile illness[61]. In 2014, 85% of health centres surveyed were able to provide first-line treatment for uncomplicated malaria, and 42% of health facilities had first-line treatment available for severe malaria[59].

6. Challenges to malaria control in PNG

While recent increases in funding and mass distribution campaigns in PNG have substantially increased access to LLINs, malaria diagnostics and treatment, use of these interventions may still need to be increased, and barriers to access remain. Firstly, success of implementation of these distribution campaigns depends on the management, logistics and operation of the distribution programme[44,62]. The primary barriers to LLIN ownership and use are undersupply and limited accessibility [52,63]. The mountainous terrain, coupled with the high proportion of the population living in remote villages with poor infrastructure and access to basic services, makes distribution of interventions challenging[4,21,51]. Often villages are only reachable by boat, air, or walking on foot for several days[2]. Additional barriers to use of LLINs include perceptions of low malaria risk, indifference, sparing LLINs for later use and reluctance to use a LLIN in the heat[52,51]. Visitors to households are also at an increased risk of infection as they do not travel with an LLIN[52] and do not usually have access to one in the households they are visiting and spending the night[51].

Despite the scale up in availability of RDTs and ACTs, recent evidence suggests that less than half of confirmed or suspected malaria cases were prescribed treatment with ACTs at health centres[4]. Research suggests that RDT diagnostics are largely being used in place of complementary diagnostics with the result that levels of assessment are largely similar to what they were prior to the scale up of RDT diagnostics[61]. In addition to vector control and treatment interventions, good access to well-functioning, staffed and resourced health centres is important in achieving malaria elimination[64,65,49]. However distance to the nearest health facility has been reported as a factor in whether or not formal treatment is sought for suspected malaria[65] and considerable urban-rural and within-region disparities exist in access to quality health care[66].

7. Lessons learned and future of malaria control and elimination in PNG

Critical evaluations following the breakdown of previous control programmes uncover common themes in breakdown of control operations. Poor planning, a lack of cohesion in administration between different provincial districts, as well as difficulty in sustaining investment and control efforts once a visible decrease in malaria burden was observed, were inadequacies evident in all three major historical control periods. Poor communication with populations where interventions were being targeted, as well as general dissatisfaction with drug treatment and insecticide spraying regimens were also cited as creating challenges to control efforts. Malaria control programmes in the past have significantly altered the transmission dynamics of malaria resulting in emergence of resistance and resurgence of malaria at higher rates than had existed prior to commencement of control interventions[5,29]. As noted following the breakdown of control programmes in the 1970s, success of malaria control programmes in PNG are contingent on a broader approach to control through nationwide coverage of interventions, directed by adequate planning and administration[25,49] and this, as well as obstacles faced by previous control programmes, should be taken into consideration as current control efforts progress.

Heterogeneity and small area variation in transmission should be examined when planning distribution of control interventions, and change of transmission dynamics in response to control interventions evaluated by sustained surveillance[26]. Surveillance is currently carried out as cross-sectional surveys or passive case detection. Novel statistical methodologies for interpolating risk of infection outside of surveyed areas may improve allocation of resources and directing additional surveillance operations to where they will have most impact. [13,67]. Novel statistical methods and epidemiological tools for interpolating infection risk and heterogeneity in spatial distribution of *P. falciparum and P. vivax* transmission outside of

surveyed areas will be useful in guiding and planning allocation of control interventions to areas where they will have most impact.

Migration between the highlands and coastal lowlands continues to be associated with epidemic outbreaks of imported *P. falciparum*, posing a major challenge for malaria control in the highlands[68], and a major focus of control interventions in PNG should be the reduction of risk of importation to the highlands via human migration[53]. Continued surveillance to identify foci of residual transmission as transmission decreases and high risk areas for post elimination resurgence, is necessary to achieve and sustain elimination[40]. Improved methods for examining the impact of human mobility and connectivity of distinct geographic locations on malaria transmission in PNG will be beneficial to control programmes for guiding surveillance operations. Such epidemiological methods may include examining the geographic niches of distinct Plasmodium parasite genotypes to understand how connectivity of distinct populations influences malaria transmission dynamics. Methods for assessing risk of resurgence in parts of the country where elimination has been achieved, but remain connected to endemic areas via human mobility will be useful for targeting vector control interventions and estimating when control interventions should be sustained or can be discontinued.

8. Conclusions

Although it has been argued that perhaps the most realistic goal for PNG in terms of malaria elimination is stable low level transmission and perhaps elimination from the highlands, the potential for resurgence in the highlands post elimination due to human migration is real and may present a threat to national elimination programmes of neighbouring countries, or countries with which PNG has close and frequent contact. Here we have presented a review of the present and historical state of malaria epidemiology and control in Papua New Guinea in

an effort to identify obstacles to the success of previous malaria control programmes and to inform current and future control and elimination strategies.

References

1 World Bank. *Papua New Guinea overview*, <http://www.worldbank.org/en/country/png> (2016).

2 Ravi, A. & Gupta, R. World Malaria Report 2012. *Australasian Medical Journal (Online)* **6**, 130 (2013).

3 Müller, I., Bockarie, M., Alpers, M. & Smith, T. The epidemiology of malaria in Papua New Guinea. *Trends in parasitology* **19**, 253-259 (2003).

4 Hetzel, M. W. *et al.* Insecticide-treated nets and malaria prevalence, Papua New Guinea, 2008-2014. *Bull World Health Organ* **95**, 695-705b, doi:10.2471/blt.16.189902 (2017).

5 Parkinson, A. Malaria in Papua New Guinea 1973. *PNG Med J* **17**, 8-16 (1974).

6 Riley, I. D. Population change and distribution in Papua New Guinea: an epidemiological approach. *Journal of Human Evolution* **12**, 125-132 (1983).

7 Sharp, P. T. Highlands malaria: malaria in Enga Province of Papua New Guinea. *P N G Med J* **25**, 253-260 (1982).

8 Koepfli, C. *et al.* Sustained Malaria Control Over an 8-Year Period in Papua New Guinea: The Challenge of Low-Density Asymptomatic Plasmodium Infections. *The Journal of infectious diseases* **216**, 1434-1443, doi:10.1093/infdis/jix507 (2017).

9 Hetzel, M. W. *et al.* Prevalence of malaria across Papua New Guinea after initial roll-out of insecticide-treated mosquito nets. *Tropical medicine & international health : TM & IH* **20**, 1745-1755, doi:10.1111/tmi.12616 (2015).

10 Cooper, R. D. *et al.* Malaria vectors of Papua New Guinea. *International journal for parasitology* **39**, 1495-1501, doi:10.1016/j.ijpara.2009.05.009 (2009).

11 Mueller, I. & Hetzel, M. W. Malaria in Papua New Guinea 2000-2013: back from the brink, but where to now? *Papua and New Guinea medical journal* **57**, 1-6 (2014).

12 World Health Organization. World malaria report 2016: summary. (2017).

13 Hofmann, N. E. *et al.* The complex relationship of exposure to new Plasmodium infections and incidence of clinical malaria in Papua New Guinea. *eLife* **6**, doi:10.7554/eLife.23708 (2017).

14 Cattani, J. A. *et al.* Small-area variations in the epidemiology of malaria in Madang Province. 1986. *P N G Med J* **48**, 95-101 (2005).

15 Hii, J. L. K. *et al.* Spatial and Temporal Variation in Abundance of Anopheles (Diptera: Culicidae) in a Malaria Endemic Area in Papua New Guinea. *Journal of Medical Entomology* **34**, 193-205, doi:10.1093/jmedent/34.2.193 (1997).

16 Burkot, T. R., Graves, P. M., Paru, R., Wirtz, R. A. & Heywood, P. F. Human malaria transmission studies in the Anopheles punctulatus complex in Papua New Guinea: sporozoite rates, inoculation rates, and sporozoite densities. *Am J Trop Med Hyg* **39**, 135-144 (1988).

17 Greenwood, B. M. 3. Impact of culture and environmental changes on epidemiology and control of malaria and babesiosis. The microepidemiology of malaria and its importance to malaria control. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **83**, 25-29, doi:10.1016/0035-9203(89)90599-3 (1989).

18 Burkot, T. R., Dye, C. & Graves, P. M. An analysis of some factors determining the sporozoite rates, human blood indexes, and biting rates of members of the Anopheles punctulatus complex in Papua New Guinea. *Am J Trop Med Hyg* **40**, 229-234 (1989).

19 Hairston, N., Bang, F. & Maier, J. Malaria in the natives of New Guinea. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **40**, 795-807 (1947).

20 Peters, W. Part VI. Unstable highland malaria—analysis of data and possibilities for eradication of malaria. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **54**, 542-548, doi:10.1016/0035-9203(60)90029-8 (1960).

21 Attenborough, R. D., Burkot, T. R. & Gardner, D. S. Altitude and the risk of bites from mosquitoes infected with malaria and filariasis among the Mianmin people of Papua New Guinea. *Trans R Soc Trop Med Hyg* **91**, 8-10 (1997).

22 Charlwood, J. D., Graves, P. M. & Alpers, M. P. The ecology of the Anopheles punctulatus group of mosquitoes from Papua New Guinea: a review of recent work. *P N G Med J* **29**, 19-26 (1986).

23 Malaria Atlas Project. <http://www.map.ox.ac.uk/explore/mosquito-malaria-vectors/bionomics> (2016).
24 Spencer, T., Spencer, M. & Venters, D. Malaria vectors in Papua New Guinea. *Papua New Guinea Medical Journal* **17**, 22-30 (1974).
25 Van Dijk, W. & Parkinson, A. Epidemiology of malaria in New Guinea. *PNG Med J* **16**, 17 (1974).
26 Cattani, J. A. *et al.* The epidemiology of malaria in a population surrounding Madang, Papua New Guinea. *Am J Trop Med Hyg* **35**, 3-15 (1986).
27 Cooper, R. D., Waterson, D. G., Frances, S. P., Beebe, N. W. & Sweeney, A. W. Speciation and distribution of the members of the Anopheles punctulatus (Diptera: Culicidae) group in Papua New Guinea. *J Med Entomol* **39**, 16-27 (2002).
28 Hii, J. L. *et al.* Spatial and temporal variation in abundance of Anopheles (Diptera:Culicidae) in a malaria endemic area in Papua New Guinea. *J Med Entomol* **34**, 193-205 (1997).
29 Mueller, I., Tulloch, J., Marfurt, J., Hide, R. & Reeder, J. C. Malaria control in Papua New Guinea results in complex epidemiological changes. *Papua New Guinea Medical Journal* **48**, 151 (2005).
30 Burkot, T. R., Graves, P. M., Paru, R. & Lagog, M. Mixed blood feeding by the malaria vectors in the Anopheles punctulatus complex (Diptera: Culicidae). *J Med Entomol* **25**, 205-213 (1988).
31 Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasites & vectors* **4**, 89 (2011).
32 Charlwood, J. D., Graves, P. & Alpers, M. The ecology of the Anopheles punctulatus group of mosquitoes from Papua New Guinea: a review of recent work. *PNG Med J* **29**, 19-26 (1986).
33 Reimer, L. J. *et al.* Malaria transmission dynamics surrounding the first nationwide long-lasting insecticidal net distribution in Papua New Guinea. *Malar J* **15**, 25, doi:10.1186/s12936-015-1067-7 (2016).
34 Koch R. Professor Koch's Investigations on Malaria: Second Report to the German Colonial Office. *British Medical Journal* **1**, 325-327 (1900).
35 Koch R. Professor Koch's Investigations on Malaria. *British Medical Journal* **1**, 1183-1186 (1900).
36 McMahon, J. A note on changing concepts of malaria control. *Papua New Guinea Medical Journal* **16**, 78-79 (1973).
37 Centers for Disease Control and Prevention. *Ross and the Discovery that Mosquitoes Transmit Malaria Parasites*, <http://www.cdc.gov/malaria/about/history/ross.html> (2015).
38 Koch R. Professor Koch's investigations on malaria. Fourth report to the colonial department of the German colonial office. *British Medical Journal* (1900).
39 Campos-Outcalt, D. Health services in Papua New Guinea. *Public health* **103**, 161-169 (1989).
40 Betuela, I. *et al.* Epidemiology of malaria in the Papua New Guinean highlands. *Tropical Medicine & International Health* **17**, 1181-1191 (2012).
41 Peters, W. & Christian, S. H. Studies on the epidemiology of malaria in New Guinea. 4. Unstable highland malaria--the clinical picture. *Trans R Soc Trop Med Hyg* **54**, 529-536 (1960).
42 Radford, A., Van Leeuwen, H. & Christian, S. Social aspects in the changing epidemiology of malaria in the highlands of New Guinea. *Annals of Tropical Medicine & Parasitology* **70**, 11-23 (1976).
43 Gunther, J. The early history of malaria control in Papua New Guinea. *Papua New Guinea Medical Journal* **17**, 4-7 (1973).
44 Hetzel, M. W. An integrated approach to malaria control in Papua New Guinea. *Papua New Guinea Medical Journal* **52**, 1 (2009).
45 Baird, J. K. Resistance to chloroquine unhinges vivax malaria therapeutics. *Antimicrobial agents and chemotherapy* **55**, 1827-1830 (2011).
46 Peters, W. Malaria Control in Papua and New Guinea. *Papua and New Guinea Medical Journal* **3**, 66-75 (1959).
47 Ewers, W. H. & Jeffrey, W. Parasites of man in Niugini. *Parasites of man in Niugini.* (1971).
48 McMahon, J. Malaria endemicity amongst the semi-nomadic people of the Karimui area in Papua New Guinea. *PNG Med J* **17**, 99-107 (1974).

49 Black RH. Malaria. *Papua New Guinea Medical Journal* **17**, 1-3 (1974).
50 Colbourne, M. J. & Stevenson, K. Report by WHO on the anti-malarial programme in Papua New Guinea, 8 November-28 December 1970. (1971).
51 Hetzel, M. W. *et al.* Progress in mosquito net coverage in Papua New Guinea. *Malaria journal* **13**, 242.241-242.214 (2014).
52 Hetzel, M. W. *et al.* Ownership and usage of mosquito nets after four years of large-scale free distribution in Papua New Guinea. *Malaria journal* **11**, 1 (2012).
53 Mueller, I. *et al.* Epidemic malaria in the highlands of Papua New Guinea. *Am J Trop Med Hyg* **72**, 554-560 (2005).
54 D'Acremont, V., Lengeler, C. & Genton, B. Malaria: Stop ambiguous messages on malaria diagnosis. *BMJ: British Medical Journal* **334**, 489 (2007).
55 Genton, B. *et al.* Malaria: how useful are clinical criteria for improving the diagnosis in a highly endemic area? *Transactions of the Royal Society of Tropical Medicine and Hygiene* **88**, 537-541 (1994).
56 Papua New Guinea National Department of Health. National malaria treatment protocol. (National Department of Health, Port Moresby, 2009).
57 Karunajeewa, H. A. *et al.* A trial of combination antimalarial therapies in children from Papua New Guinea. *The New England journal of medicine* **359**, 2545-2557, doi:10.1056/NEJMoa0804915 (2008).
58 Macfarlane, J. E. & Alpers, M. P. Treatment-seeking behaviour among the Nasioi people of Bougainville: choosing between traditional and western medicine. *Ethnicity & health* **14**, 147-168, doi:10.1080/13557850802546588 (2009).
59 Kurumop, S. F., Pulford, J., Mueller, I., Siba, P. M. & Hetzel, M. W. Diagnostic capacity and antimalarial availability in Papua New Guinea before the introduction of a revised national malaria treatment protocol. *P N G Med J* **57**, 59-67 (2014).
60 Maltha, J. *et al.* Evaluation of a rapid diagnostic test (CareStart Malaria HRP-2/pLDH (Pf/pan) Combo Test) for the diagnosis of malaria in a reference setting. *Malar J* **9**, 171, doi:10.1186/1475-2875-9-171 (2010).
61 Pulford, J., Kurumop, S., Mueller, I., Siba, P. M. & Hetzel, M. W. The impact of the scale-up of malaria rapid diagnostic tests on the routine clinical diagnosis procedures for febrile illness: a series of repeated cross-sectional studies in Papua New Guinea. *Malar J* **17**, 202, doi:10.1186/s12936-018-2351-0 (2018).
62 Zegers, d. B. C., Koenker, H., Obi E., Adegbe, E., Selby, R.A, Kilian, A. . *Distribution, delivery and allocation strategies for mass campaigns to achieve universal coverage with insecticide treated nets Which work best? A multi-country comparison*, <http://www.malariaconsortium.org/media-downloads/234> (2016).
63 Pulford, J. *et al.* Indifferent to disease: a qualitative investigation of the reasons why some Papua New Guineans who own mosquito nets choose not to use them. *Social science & medicine (1982)* **75**, 2283-2290, doi:10.1016/j.socscimed.2012.08.030 (2012).
64 Angwin, A., Hetzel, M. W., Mueller, I., Siba, P. M. & Pulford, J. A qualitative study of how affected individuals or their caregivers respond to suspected malaria infection in rural Papua New Guinea. *Papua New Guinea Medical Journal* **57** (2014).
65 Müller, I., Smith, T., Mellor, S., Rare, L. & Genton, B. The effect of distance from home on attendance at a small rural health centre in Papua New Guinea. *International journal of epidemiology* **27**, 878-884 (1998).
66 Bauze, A. E. *et al.* Equity and geography: the case of child mortality in Papua New Guinea. *PLoS One* **7**, e37861 (2012).
67 Waltmann, A. *et al.* Increasingly inbred and fragmented populations of Plasmodium vivax associated with the eastward decline in malaria transmission across the Southwest Pacific. *PLoS Negl Trop Dis* **12**, e0006146, doi:10.1371/journal.pntd.0006146 (2018).
68 Mueller, I., Kaiok, J., Reeder, J. C. & Cortes, A. The population structure of Plasmodium falciparum and Plasmodium vivax during an epidemic of malaria in the Eastern Highlands of Papua New Guinea. *Am J Trop Med Hyg* **67**, 459-464 (2002).

CHAPTER 4

Spatial ecology and predicted risk of malaria in Papua New Guinea:

A comparison of models from which to generate risk maps

CHAPTER 4. SPATIAL ECOLOGY AND PREDICTED RISK OF MALARIA IN PAPUA NEW GUINEA: A COMPARISON OF MODELS FROM WHICH TO GENERATE RISK MAPS

CONTEXT

Chapter 4 compares statistical models for generating national scale spatial predictions of malaria distribution in PNG, a country in the control stage of malaria elimination. In this chapter, the application of traditional frequentist GLM approaches and the emerging BDN modelling approach have been explored with respect to spatial prediction of malaria risk, and the predictive accuracy of both methods were quantified.

P. falciparum and *P. vivax* national point prevalence data, collected as part of a nationwide malaria indicator survey, were used in conjunction with ecological covariates associated with malaria transmission in PNG to predict the spatial distribution of malaria on a national scale. We explore environmental and climate drivers associated with transmission of both parasite species in PNG and discuss the difficulties in malaria risk prediction in this context, due to complexity of the disease ecology of malaria and spatial variation in malaria risk factors across PNG. Comparison of the accuracy of BDN models with a GLMs found a better predictive performance of the BDN models. Ensuring the predictive accuracy of statistical models in spatial risk prediction is integral to the success of malaria control programmes. This chapter is formatted for submission as a manuscript to Malaria Journal.

CHAPTER 4. SPATIAL ECOLOGY AND PREDICTED RISK OF MALARIA IN PAPUA NEW GUINEA: A COMPARISON OF MODELS FROM WHICH TO GENERATE RISK MAPS

Cleary, E[1]; Hetzel, M.W.[2,3]; Siba, P[4]; Lau, C[1]; Clements, A.C.A[1].

1. *Research School of Population Health, Australian National University, Canberra, Australia*

2. *Swiss Tropical and Public Health Institute, Basel, Switzerland*

3. *University of Basel, Basel, Switzerland*

4. *Papua New Guinea Institute of Medical Research, Goroka, PNG*

Abstract

Introduction

Considerable progress towards controlling malaria has been made in Papua New Guinea through the national malaria control programme's free distribution of long lasting insecticidal nets, improved diagnosis with rapid diagnostic tests and improved access to artemisinin combination therapy. Predictive risk maps can help to inform targeted interventions and monitor changes in malaria epidemiology over time as control efforts continue. This study aims to compare the predictive performance of risk maps generated using Bayesian decision network (BDN) models and multilevel logistic regression models for improved accuracy in malaria spatial risk prediction.

Methods

Multilevel logistic regression models and Bayesian decision network models were developed using 2010/2011 malaria prevalence survey data collected from 77 randomly selected villages, to determine associations of *Plasmodium falciparum* and *P. vivax* prevalence with precipitation, temperature, elevation, slope, enhanced vegetation index and distance to the

coast. Predictive performance of multilevel logistic regression and BDN models were compared by cross validation methods.

Results

Prevalence of *P. falciparum* was significantly associated with precipitation during the three driest months of the year, June to August (β= 0.015; 95% CI = 0.01 – 0.03), whereas *P. vivax* infection was associated with elevation (β = -0.26; 95% CI = -0.38 - -3.04), precipitation during the three driest months of the year (β = 0.01; 95% CI = -0.01 - 0.02) and slope (β = 0.12; 95% CI = 0.05 - 0.19). Compared with GLM model performance, BDNs showed improved accuracy in prediction of *P. falciparum* (AUC = 0.5681; AUC = 0.7502, respectively) and *P. vivax* (AUC = 0.6786; AUC = 0.7488, respectively) on cross-validation.

Conclusion

BDNs provide a more flexible modelling framework than generalized linear models and may have a better predictive performance when developing malaria risk maps due to the multiple, interacting factors that drive malaria risk in different geographical areas. When developing malaria risk maps, BDNs may be particularly useful in predicting risk where spatial variation in climate and environmental drivers of malaria transmission exists, as is the case in Papua New Guinea.

1. Introduction

Papua New Guinea (PNG), a Pacific island nation with a population of over 8 million people[1], has had a steady decline in malaria prevalence since 2004 when the national malaria control programme was awarded a Global Fund to Fight Aids, Malaria and Tuberculosis grant. This funding facilitated the free national distribution of long lasting insecticide treated nets (LLINs), improved diagnosis by rapid diagnostic tests (RDTs) and scaling up of artemisinin-based combination therapy (ACT) in all health facilities[2]. Consequently, *Plasmodium falciparum* and *P. vivax* prevalence has reduced from 3.4%[3] and 2.1% to 1.6% and 0.5% between 2010 and

2014 respectively[4,5]. PNG is currently classified as being in the control stage on the pathway towards malaria elimination[2]. Despite this decline in prevalence, PNG still has the highest incidence of malaria in the Asia Pacific Region, equal only in a global context to the highest burden countries in Sub-Saharan Africa[6,7].

The epidemiology of malaria varies considerably across the country and small-area spatial variations in malaria prevalence also exist[8], attributed in part to varied implementation of interventions including village LLIN use and the availability of antimalarials[8]. Environmental and climate factors associated with mosquito breeding sites and different vector dynamics, particularly between low lying coastal areas and the highlands, also contribute to the variation in the spatial distribution of malaria prevalence[9,10]. In the PNG lowlands, malaria transmission is perennial, with seasonal transmission only in coastal regions where rainy and dry seasons are distinguishable[11]. In highland regions, marked seasonality exists where transmission is lower and unstable. In these areas, which are prone to seasonal epidemics or outbreaks and where populations lack acquired immunity, morbidity and mortality can be more severe [11]. The spatial distribution of both *P. falciparum* and *P. vivax* spans the entire country, however in terms of relative contribution to disease, *P. falciparum* is responsible for a greater proportion of infections[3]. The relative abundance of *P. vivax* has recently been observed to be increasing however, concurrent with a decrease in *P. falciparum* prevalence[12].

Predictive risk maps based on spatial statistical models examining associations between environmental variables (often sourced using satellite remote sensing) and disease risk (often measured using surveys or surveillance data) are useful evidence-based decision tools for allocation of resources in control programmes[10,13]. These tools are of particular value in the context of constrained resources and in directing interventions to remote or difficult to access communities. Spatial risk maps can assist with surveillance and control interventions through revealing the geographical bounds of disease occurrence and variations in disease risk,

including spatial changes in prevalence in response to control intervention[14-17], providing a better understanding of the epidemiology of disease over various spatial scales[18].

Epidemiological risk maps are often generated using the results of generalized linear models (GLM) that include environmental, demographic and intervention-related covariates. Such models can be developed at a range of spatial scales from global to local[19,20]. However, challenges in using GLMs in the spatial prediction of malaria can be posed by spatial and temporal non-stationarity (where relationships between variables and correlation structures vary across a study area or time period), non-linear associations with predictive risk factors, spatial autocorrelation, complex causal pathways, and complex interactions between covariates, including collinearity[21]. All of these factors might limit the prediction accuracy of GLM based approaches.

In recent years, graphical model-based approaches, such as Bayesian decision networks (BDNs), have become more ubiquitous in infectious disease risk prediction, and used with good success[14,22]. BDNs are graphical representations of variables, or nodes, in a system linked together, and to the outcome of interest, to describe a network of complex interactions[23,19]. Such models can capture complex interactions of drivers of transmission and interacting nonlinear effects, and can provide quantitative representation of uncertainty in spatial predictions[14,19]. Variables are connected via directed arcs, indicating the direction of the association, with conditional probability tables quantifying the relationship between each variable[23-25,26]. BDNs can be structured using a machine algorithm to learn the model structure with available data or structured using expert knowledge[26-28]. BDNs have been shown to have good prediction accuracy for malaria at high temporal and spatial resolution, although previous research has focused on malaria predictions at village level only and not at a broader spatial surface[21].

Spatially explicit BDNs can provide visual representations of variables of interest, such as predictive risk of infectious disease[14,19,24,29], and have become increasingly popular for modelling of ecological and environmental systems with improved predictive accuracy compared to traditional methods (such as GLMs) [25,30]. Predictive accuracy, and the ability to demonstrate uncertainty in predictions, is beneficial for appropriately allocating resources when deciding where to disseminate control interventions. In this paper, we aim to produce national risk maps for *P. falciparum* and *P. vivax* infection in PNG, and to compare the predictive accuracy of GLM and BDN based methods for generating malaria risk maps in a complex environment.

2. Materials and Methods

2.1 Infection data

Data were collected as part of the national malaria indicator survey (MIS), which was conducted by the Papua New Guinea Institute of Medical Research, in 2010 and 2011. The survey was conducted in five villages randomly sampled from a geo-referenced village database in 17 of the 20 provinces of PNG, representing 77 villages in total. In each village, 30 households were randomly selected for inclusion and all present, consenting household members over six months of age were included as eligible for participation in the survey. Datasets collected included information on household use of LLINs, a treatment-seeking behaviour survey relating to recent febrile illness and collection of capillary blood samples for determination of parasite prevalence and species by microscopy. Malaria diagnosis was determined by RDT and treatment was provided for any participants with detected cases encountered during the course of the survey[5]. Village GPS coordinates and elevation above sea level were also recorded, and village level prevalence of *P. falciparum* and *P. vivax* was

determined based on survey results. More detailed information pertaining to data collection can be found in the survey report[5].

2.2 Data on the physical environment

Average monthly precipitation and temperature data, aggregated over a 50 year period from 1950 to 2000, at 1km^2 resolution were downloaded from the WorldClim website[31]. Elevation and slope data were extracted from a global digital elevation model (GDEM) obtained from the National Aeronautics and Space Administration (NASA) online repository of remote sensing image data, collected by the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) aboard the Terra satellite[32]. Enhanced vegetation index (EVI) data were derived from the remote sensing images collected by the Moderate Resolution Imaging Spectroradiometer (MODIS), also aboard Terra. Distance to the coast was calculated using geographic information system (GIS) software by defining a coastline polygon and calculating the Euclidean distance from each cell on the map to the coast. All covariate data processing was carried out using ArcGIS software version 10.3 (ESRI, Redlands, California).

2.3 Univariate analysis and variable selection

Maps of observed *P. falciparum* (Figure 1.) and *P. vivax* (Figure 2.) prevalence across the 77 surveyed villages in PNG were generated in ArcGIS and overlain with climate and environmental raster layers. Median values for temperature during the three hottest and coldest months (December to February, and June to August, respectively), and precipitation during the wettest and driest three months (January to March, and June to August, respectively), EVI during the hottest (January) and coldest (July) months of the year, and slope and elevation data were extracted to 5km and 10km buffer zones around the centre point of each survey village location. The Euclidean distance from each centre village point to the coastline of PNG was also calculated and values extracted. All data management and extraction was carried out using

ArcGIS software. Univariate analysis of associations between prevalence of *P. falciparum* and *P. vivax* with environmental and climate data, using both 5km and 10km buffers, was carried out using multilevel logistic regression models that accounted for clustering at the village level. Variables were selected for inclusion in further analyses based on a *p*-value of < .05 and lowest value of the Akaike Information Criterion (AIC). Collinearity of variables between environmental and climate variables was assessed prior to inclusion in models using a tolerance cut point of <.02 and VIF cut-off value of >5. The only variables found to exhibit collinearity were temperature and elevation and the most appropriate variable for inclusion in multivariate regression models were selected based on lowest AIC value, obtained on univariate analysis. All univariate and multivariate regression analyses were carried out using Stata statistical software version 14 (StataCorp, College Station, Texas).

2.4 Multivariable generalised linear regression models

Multilevel mixed-effects logistic regression models were developed for *P. falciparum* and *P. vivax* using selected variables, with proportion of bednet ownership of each village (survey data included both insecticide treated net (ITN) and LLIN coverage), age, gender, wealth quintile and annual quarter during which the survey was carried out included in the models to adjust for confounding of associations with environmental variables. Separate multivariable models were built for both *Plasmodium* species and final models for each were selected based on the lowest Akiake Information Criterion (AIC) value. To identify spatial autocorrelation, semivariograms of the regression model residuals were plotted using the R open source software version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). There was no evidence of spatial autocorrelation from the semivariograms and we were unable to fit spatial GLMs using model-based geostatistics, therefore spatially explicit GLMs were not developed further.

2.5 Spatial risk prediction and generation of risk maps using generalized linear modelling results

Spatial risk predictions were made, using environmental and climate fixed effects only, by multiplying values for each cell of the environmental variable raster layers in the model by the corresponding covariate coefficient from the multilevel regression models, adjusted for socio-demographics. The resulting raster layer values were summed (together with the intercept) and the logit calculated using the map algebra tool in ArcGIS. Although not all variables in *P. falciparum* and *P. vivax* multilevel mixed-effects models were found to be significant upon regression analysis, all were retained for generation of predicted risk maps. The equations for spatial risk prediction (*p*) in each location (*i*) of *P. falciparum* and *P. vivax* are as follows:

P. falciparum: Logit (p_i) =-59.467 – .005 * Enhanced vegetation index in July$_i$ – .073 * Distance to coast$_i$ + 4.77 * Maximum temperature (December to February)$_i$ – .09 * Maximum temperature (December to February)$_i^2$ + .015 * Precipitation (June to August)$_i$

P. vivax: Logit (p_i) =-5.941 + .038 * Enhanced vegetation index (January)$_i$ – 1.33 * Distance to coast$_i$ – .26 * Elevation$_i$ + .01 * Precipitation (June to August)$_i$ + .12 * Slope$_i$

2.6 Bayesian decision network models

A machine-learned BDN, and a BDN model structured based on the biological assumptions underlying malaria transmission in PNG, and associations with ecological covariates obtained upon initial univariate analysis (an expert structured model), were compiled using Netica software version 5.24 (Norsys Software Corp., Vancouver, Canada) and the bnlearn package[33] in R statistical software. For the expert structured model, the variables found to have strongest associations with the outcome, based on AIC criterion, were placed closest to the parent node and sensitivity to findings analyses were conducted in Netica software to verify appropriate positioning of variables in the network. Sensitivity to findings analysis verify which nodes in the model are most informative in making predictions for the outcome of interest. Machine-

learned models were defined and optimal model structure determined based upon the hill-climbing algorithm in the bnlearn package in R statistical software. Continuous predictor variables in the model were discretised using the equal interval method for discretisation in R software. Climate and environmental variables used in the BDN models were, for the most part, the same predictor variables as those used for the GLM models. There were however a few exceptions, described as follows.

Whereas a measure of maximum temperature during the three hottest months of the year, together with a corresponding quadratic term was used in the *P. falciparum* GLM model, it was not considered appropriate to use a discretised quadratic term in the BDN model as the non-linear relationship with *P. falciparum* would only be captured using a continuous variable. Instead, the variable with the next strongest association to the outcome, minimum temperature during the three hottest months, was used. Altitude was also included in this model, which could not be included in the GLM model due to collinearity. Likewise, temperature (minimum temperature during the three hottest months of the year) was included in the *P. vivax* model, which was previously excluded from the GLM model due to collinearity.

Network models were structured with *P. falciparum* and *P. vivax* infection status as parent nodes and explanatory variables as child nodes with directed arcs connecting explanatory variables in the model to each other and to the outcome variable of interest. Networks were structured using environmental variables only, in order to ensure that 1) conditional probability tables (CPTs) could be generated quantifying associations between the states of all variables in the model given the number of observations in the dataset, 2) to ascertain prediction accuracy using environmental variables alone for comparison with regression model spatial predictions which were validated against demographic adjusted village level prevalence and 3) spatial predictions from the models would be based on environmental variables only, given that these were the variables for which all non-sampled locations had observations.

CPTs were generated quantifying the relationships between explanatory variables and the outcome variable. CPTs and predicted probability of the outcome[34] were based on data entered into the model and a-priori beliefs were updated through belief propagation using Bayes' Theorem (posterior = likelihood * prior / probability of evidence)[23,26]. A-priori beliefs relate to the logical structure of explanatory nodes in the expert structured model and a-priori probabilities are updated as new knowledge about the systems is obtained (observational data on which the model is learned and CPTs are produced) making them posterior beliefs[35].

2.7 Spatial risk prediction and generation of risk maps using Bayesian decision network model results

Prediction probabilities for *P. falciparum* and *P. vivax* were determined according to conditional probability tables from the BDN model and predictions were made for each spatial point on a continuous gridded vector layer of environmental and climate measures for PNG using the bnspatial package in R. The spatial distributions of these predicted probabilities were plot in ArcGIS and the resulting gridded point maps smoothed using the inverse weighting function. Maps showing degree of spatial entropy representing uncertainty for both models were also produced in the bnspatial package.

2.8 Cross validation of GLMS and BDNs

The predictive accuracy of the *P. falciparum* and *P. vivax* models was assessed by cross-validation of predicted prevalence (p_i) values against observed adjusted prevalence, generating Receiver operating characteristic (ROC) curves and obtaining Area under the curve (AUC) values. Observed prevalence values adjusted for age, gender, wealth quintile, bednet use and season during which the surveys were carried out were calculated using the dstdize command in Stata statistical software. Cross validation of the predictive accuracy of the spatial risk prediction of the models was carried out by predicting spatial distribution of *P. falciparum* and

P. vivax based on the results of regression models using 75% of the survey data (training dataset), and extracting predicted median values to a 5km buffer around each village point for the remaining 25% of villages (test dataset) in ArcGIS. Cross validation to assess predictive accuracy of the models was carried out by plotting predicted values test datasets against observed adjusted prevalence using ROC curves, from which AUC values were obtained.

The predictive accuracy of machine-learned BDN models and expert structured models was assessed using cross validation methods. Cross validation was carried out by defining a training dataset (75% of the full dataset) from which CPTs in the machine-learned and expert structured models were defined. Predictions were then made from these models on a test dataset (the remaining 25% of the full dataset) and ROC curves and AUC values showing results of cross-validation generated. Two iterations of cross-validation were carried out for each *P. falciparum* and *P. vivax* machine-learned and expert structured models using two sets of randomly generated training and test datasets. The predictive accuracy of machine-learned models was compared with expert structured models, and the latter, which demonstrated the most accurate predictive performance on cross validation, were selected for spatial prediction of malaria risk.

3. Results

3.1 Demographics

Survey results showed that observed *P. falciparum* prevalence (2.83%; 95% CI: 2.51% - 3.16%) was slightly higher than *P. vivax* prevalence (2.07%; 95% CI: 1.79% - 2.35%) (Table1). Geographically, the highest prevalence of both *P. falciparum* (Figure 1) and *P. vivax* (Figure 2) was observed in the islands of East New Britain and New Ireland, the north coast and on the Papuan Peninsula in the east of the country. Prevalence of *P. falciparum* was highest among children aged between six months and five years of age (5.59%) and males (3.11%). The highest observed prevalence of *P. falciparum* at village level was 27.6%, whereas the highest

prevalence of *P. vivax* at the village level was 15.8%. Prevalence of both *P. falciparum* and *P. vivax* was slightly higher among LLIN users (3.02%; 2.25%, respectively), possibly suggesting higher LLIN usage among populations at known higher risk.

Table 1

Plasmodium falciparum and *Plasmodium. vivax* prevalence of 10,028 individuals surveyed in the 2010/2011 National Malaria Control Intervention and Prevalence of Parasitaemia Household Survey

	P. falciparum	*P. vivax*
Variable	N (%)	N (%)
Disease Prevalence	284 (2.83; 95% CI: 2.52 – 3.18)	208 (2.07; 95% CI: 1.80 – 2.37)
Gender		
Female	135 (2.56)	103 (1.95)
Male	147 (3.11)	104 (2.20)
Age		
0 – 5	99 (5.59)	88 (4.97)
6 – 18	117 (3.76)	74 (2.38)
19 – 100	68 (1.32)	46 (0.90)
Bednet use		
No	122 (2.62)	87 (1.87)
Yes	162 (3.02)	121 (2.25)
Total	284 (2.83)	208 (2.07)

3.2 Generalized linear modelling analysis

Among the environmental predictor variables, *P. falciparum* was found only to be significantly associated with precipitation during the three driest months of the year, June to August (β= 0.015; 95% CI = 0.01 – 0.03) in the final multivariable model (Table 2), whereas *P. vivax* infection at village level was associated with elevation (β = -0.26; 95% CI = -0.38 - -3.04), precipitation during the three driest months of the year (β = 0.01; 95% CI = -0.01 - 0.02) and slope (β = 0.12; 95% CI = 0.05 - 0.19). In terms of demographics, the highest wealth quintile was negatively associated with *P. falciparum* prevalence (β = -0.89; 95% CI = -1.62 - -0.016) as was age, with participants aged between five and 18 years of age at lower risk compared

with children under five (β = -0.33; 95% CI = -0.53 - -0.12), as well as those over 18 years of age (β = -1.27; 95% CI = -1.62, -0.92). *P. falciparum* prevalence was also associated with season during which the survey was conducted, with a higher risk between March and May, compared with November to February (β = 1.11; 95% CI = 0.33 – 1.88). *P. vivax* was associated with age, with lowest risk among the participants over 18 years of age (β = -1.55; 95% CI = -2.08 - -1.02). Maps generated based on the results of these models show highest *P. falciparum* predicted risk (0.03% – 0.06%) in New Britain, New Ireland, Sandaun and Milne Bay (Figure 3), consistent with observations from the national malaria survey. The spatial distribution of *P. vivax* (Figure 4.) generated based on the results of the GLM show highest predicted prevalence in Western, Sanduan and East Sepik provinces with a low predicted prevalence in the highlands and islands regions, which is contrary to expected findings given the observational results obtained from the parasite prevalence survey.

Table 2

Results of *Plasmodium. falciparum* and *Plasmodium vivax* generalised linear multivariable regression models

P. falciparum		*P. vivax*	
Variable	Beta Coefficient (95% CI)	Variable	Beta Coefficient (95% CI)
Enhanced Vegetation Index Jul	-0.005 (-0.07 - 0.06)	Enhanced vegetation Index Jan	0.038 (-0.02 - 0.09)
Distance to the coast	-0.73 (-2.73 - 0.77)	Distance to the coast	1.33 (-0.01 – 2.98)
Tmax Dec to Feb (hottest)	4.77 (-1.10 – 9.97)	**Elevation**	-0.26 (-0.38 - -3.04)
Tmax sq	-0.09 (-0.19 - 0.02)	**Precipitation Jun to Aug (driest)**	0.01 (-0.01 - -0.02)
Precipitation Jun to Aug (driest)	0.01 (0.01 - 0.03)	**Slope**	0.12 (0.05 - 0.19)
Bednet ownership proportion village	0.91 (-0.23 – 2.11)	Bednet ownership proportion village	-0.45 (-1.78 – 0.88)
Female	-0.14 (-0.37 - 0.09)	Female	-0.04 (-0.40 - 0.31)
Wealth quintile 2	0.30 (-0.15 - 0.75)	Wealth quintile 2	-0.06 (-0.39 - 0.50)
Wealth quintile 3	-0.06 (-0.57 - 0.45)	Wealth quintile 3	0.16 (-0.41 - 0.74
Wealth quintile 4	-0.36 (-0.83 - 0.09)	Wealth quintile 4	-0.24 (-0.81 - 0.33)
Wealth quintile 5	-0.89 (-1.62 - -0.16)	Wealth quintile 5	-0.79 (-1.57 - -0.15)
Age >5 - 18	-0.33 (-0.53 - -0.12)	**Age >5 – 18**	-0.67 (-1.10 - -0.25)
Age >18 +	-1.27 (-1.62 - -0.92)	**Age >18 +**	-1.55 (-2.08 - -1.02)
Season Mar - May	1.11 (0.33 – 1.88)		
Season Jun - Aug	0.74 (-0.11 – 1.59)		

3.3 Bayesian decision network spatial predictions

Expert BDNs for both *P. falciparum* (Figure 5) and *P. vivax* (Figure 6.) were structured with EVI, region and distance to the coastline variables positioned with arcs directly related to disease prevalence, as these variables were found to be the strongest predictors of both *P. falciparum* and *P. vivax* risk. Spatial risk maps showing the predicted distribution of *P. falciparum* (Figure 7) based on the results of the BDN models predict the probability of *P. falciparum* prevalence to be highest in the island provinces of PNG, New Ireland and New Britain (.03 to 0.12), consistent with the results of the observed prevalence collected in the household survey. High predicted prevalence was also seen along the northern coast in the provinces of Sandaun, East Sepik and Madang (0.03 to 0.12). Average predicted probability was lower in the highland provinces (0.001 to 0.03), along the south coast (0.001 to 0.03), where population density is sparser, and in Milne bay (0.001 to 0.03). The predicted probability of *P. vivax* (Figure 8) was also highest in the islands ranging from 1.0% to 3.0% and 6.0% to 8.0%. The highest predicted probability of *P. vivax* prevalence along the north coast where highest observed prevalence was observed ranged between 3.0% and 6.0%. Predicted probability was lowest along the south coast, similar to patterns observed for *P. falciparum* (0.01 to 0.03). The spatial pattern of entropy, or uncertainty in risk prediction, had a similar distribution to the spatial distribution of highest predicted probability of *P. falciparum* (Figure 9) and *P. vivax* (Figure 10), reflecting higher standard errors for higher predicted prevalences.

3.4 Comparison of prediction accuracy

For the multivariable, multilevel GLMs, cross validation (Table 3) showed unsatisfactory agreement for *P. falciparum* (Figures 11a & 11b), with the predicted risk not performing much better than random allocation of status relative to any of the three cut-points (AUC at 1% = 0.5681; AUC at 2.5% = 0.5927; AUC at 5% = 0.5634) and somewhat satisfactory agreement

with observed data for *P. vivax* (Figures 11c & 11d), at 1% and 2.5% prevalence cut-off values (AUC at 1% = 0.6786; AUC at 2.5% = 0.5739), but not at a prevalence cut-off value of 5% (AUC at 5% = 0.4988). Validation of expert structured BDN models for spatial prediction probability of *P. falciparum* and *P. vivax* showed better prediction accuracy for *P. falciparum* (BDN test data AUC = 0.74502; Figure 12) and *P. vivax* (BDN test data AUC = 0.7623**;** figure 13) compared with the GLMs, and machine learned models, on both iterations of cross validation.

Table 3.

Results of cross validation for generalised regression models and Bayesian decision network models

	Area under receiver operating characteristic curve (AUC)							
	P. falciparum				*P. vivax*			
GLM cross-validation	1% cut-off	2.5% cut-off	5% cut-off		1% cut-off	2.5% cut-off	5% cut-off	
Village level	0.5681	0.5927	0.5634		0.6786	0.5739	0.4988	
BDN cross validation	1st iteration		2nd iteration		1st iteration		2nd iteration	
Village level expert structured	Train	Test	Train	Test	Train	Test	Train	Test
	0.7412	0.7502	0.7412	0.7502	0.7448	0.7623	0.7413	0.7769
BDN cross validation	1st iteration		2nd iteration		1st iteration		2nd iteration	
Village level machine structured	Train	Test	Train	Test	Train	Test	Train	Test
	0.7407	0.6978	0.7252	0.7442	0.5	0.5	0.6843	0.7092

BDN cross validation: The training and test datasets were split twice and cross validation carried out. 1st and 2nd iterations are the results of the cross validation each time the training and test datasets were split

4. Discussion

The results of this analysis show a better accuracy in the spatial prediction of malaria in PNG when using BDN models compared with the more commonly used GLM approach. The reasons for this improvement in prediction accuracy may lie in the ability of BDN models to retain collinear variables and incorporate complex interactions between explanatory variables in the model[30], meaning that more information is available for estimating the outcome. Our findings

are consistent with other studies that have shown BDNs to have improved prediction accuracy for levels of *E. coli* in recreational water sites in New Zealand compared with several alternative methods[36], and in assessing risk of Murray Valley encephalitis virus in Western Australia[37].

Predicted risk of *P. falciparum* was found to be highest along the northern coast of PNG in Sandaun, East Sepik, Madang provinces, which is consistent with the higher average temperatures (Appendix 4.1) observed in these provinces of between 27°C and 32°C. Lower average risk was predicted in Morobe where average temperatures fall to between 11°C and 24°C, and along the southern coast in provinces where population density is much lower than those on the northern coast. Predicted spatial prevalence in the Western Highlands province, Chimbu and Enga was low, consistent with survey results, and concordantly, predicted risk was highest in the island provinces of New Britain and New Ireland. The predominant predicted risk in West New Britain was slightly lower than in East New Britain, where elevation and vegetation index values are lower (Appendix 4.2 & 4.3). The spatial distribution of *P. vivax* predicted risk, based on the results of the expert structured BDN model had a similar distribution pattern to *P. falciparum* predicted risk.

As stated previously, drivers of malaria transmission across PNG vary spatially, and therefore a single, stationary model of environmental and climate predictors (such as the multilevel GLM presented here) does not seem to be appropriate for prediction of malaria risk. The drivers of malaria transmission in PNG appear to vary spatially. For example, while temperature may be a significant driver of transmission in the lowlands or coastal areas, altitude may be a better predictor of malaria risk in the highlands. To some extent, this variation was captured in the BDN using the region variable, but other solutions could include non-stationary GLMs with different covariate effects for different regions, or geographically weighted regression approaches.

An additional benefit of BDNs is that they can be used to define scenarios by which specific states of explanatory variables are selected, such as lowest defined temperature range and highest range of values for rainfall, and the probability of malaria under those specific environmental conditions can be predicted by the model. This could be useful for predicting the spatial distribution of malaria under different climate and environmental scenarios, or by incorporating intervention variables and scenarios tested to model the effects of different interventions on disease distributions.

Prediction accuracy of the expert structured BDN model was better than the machine learning hill climbing algorithm. However, for expert structured BDNs, the way in which explanatory variables are structured is subject to interpretability and inconsistencies between different models of the same system or disease are likely to arise[37]. Coupling expert opinion with statistical inference to weigh the importance of explanatory variables with regard to the outcome should improve consistency in the way in which relationships in the model are structured[28]. While machine learning algorithms can be used to structure models[21], spurious associations may exist that arise due to chance, as was observed here with associations defined by the machine learned models (Appendix 4.3 & 4.4). Cyclical arcs defined between nodes in the machine-learned models may violate conditional independence assumptions of the model, creating difficulty in generating CPTs and possibly contributing to the poor predictive performance of machine structured models observed here. Defining too many associations between explanatory variables, when insufficient observations exist to support these associations, can prevent estimation of CPTs making spatial predictions based on the results of these models difficult.

Using BDN models allows graphical representations of the complex interactions of demographic and environmental covariates associated with infectious disease transmission, and ecology of disease, to be generated which can be of particular benefit when communicating

information to the public, stakeholders, national and local control programmes and funding authorities[13,38,39]. The inherent ability of BDN models to represent uncertainty associated with spatial prediction risk is also valuable for portraying the reliability of risk maps to control programmes. National malaria control interventions need to ensure that disease prevention and control interventions are delivered to areas where risk is highest, and being able to visually represent the accuracy of risk maps can help guide decisions about efficient and cost-effective targeting of vector control interventions[40].

Generating spatial risk maps using the results of models from which we can represent this uncertainty in predictions, as well as carrying out cross validation on model predictions, make BDN models valuable epidemiological tools for guiding interventions and surveillance[41]. The visual nature of BDNs lend themselves to being easily interpretable in population health communication and in demonstrating different explanations of the outcome[26,37]. In Vietnam, for example, BDNs have been used for communication of mitigation and public health strategies to farmers on complex interactions of various factors involved in small-scale agriculture which can impact levels of *E.coli* in drinking water[42]. In PNG, evidence suggests that indifference, due to perceived low risk of malaria and absence of mosquitoes, are barriers to high coverer age of LLIN use[43]. In education and behavioural change programmes, the improved visualisation of novel tools such as BDNs showing how risk may vary between populations may improve coverage and uptake of vector control interventions[38,44].

P. vivax may present a particular challenge to malaria control programmes due to the high number of infections in PNG attributed to recrudescence[45] and novel epidemiological tools for improved risk prediction and insight into environmental contributions to *P. vivax* transmission will be useful for informing control and elimination programmes[46]. The high recrudescence rate complicates the development of ecological models of transmission due to introducing a source of error in estimating covariate effects, and may limited the prediction accuracy of such

models. In the future it would be useful to model the impact of LLIN coverage across PNG, and to generate national predictions of parasite prevalence using national survey data collected subsequent to the data used here to examine, and validate, how well BDN models perform at predicting parasite response to intervention coverage.

Limitations

Explanatory variables with continuous data must first be discretised before being used in the BDN approach demonstrated here, leading to a potential loss of information and subjective decisions regarding the discretization threshold[26]. Our approach did not incorporate spatial autocorrelation in the models, which would make them unsuitable for data in which there is spatial dependency[21], and methods need to be developed to accommodate this issue.

Conclusions

Results obtained from our comparative analysis examining the predictive accuracy of BDNs and GLMs found BDNs to perform better in terms of prediction accuracy for malaria in PNG. More work needs to be done to develop spatial BDN approaches and to make them more accessible to epidemiologists and disease control personnel.

References

1 World Bank. *Papua New Guinea overview*, <http://www.worldbank.org/en/country/png> (2016).

2 Hetzel, M. W. *et al.* Evaluation of the global fund-supported national malaria control program in Papua New Guinea, 2009–2014. *PNG Med J* **57**, 2009-2014 (2014).

3 Hetzel, M. W. *et al.* Prevalence of malaria across Papua New Guinea after initial roll-out of insecticide-treated mosquito nets. *Tropical medicine & international health : TM & IH* **20**, 1745-1755, doi:10.1111/tmi.12616 (2015).

4 Hetzel, M., Pulford, J, Gouda, H, Hodge, A, Siba, P, Mueller, I. The Papua New Guinea national malaria control program: primary outcome & impact indicators, 2009-2014. (2014).

5 Hetzel, M., Pulford, J, Ura, Y, Robinson, L, Morris, H, Mueller, I, Siba, PM. Country-wide household survey 2010/11: malaria control intervention coverage and prevalence of parasitaemia. (2012).

6 Murray, C. J. *et al.* Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384**, 1005-1070 (2014).

7 Mueller, I. *et al.* The epidemiology of malaria in the Papua New Guinea highlands: 1. Western Highlands Province. *Papua and New Guinea medical journal* **46**, 16-31 (2002).

8 Cattani, J. *et al.* Small-area variations in the epidemiology of malaria in Madang Province. *Papua New Guinea Medical Journal* **29**, 11-17 (1986).

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

10 Dalrymple, U., Mappin, B. & Gething, P. W. Malaria mapping: understanding the global endemicity of falciparum and vivax malaria. *BMC medicine* **13**, 1 (2015).

11 Müller, I., Bockarie, M., Alpers, M. & Smith, T. The epidemiology of malaria in Papua New Guinea. *Trends in parasitology* **19**, 253-259 (2003).

12 Koepfli, C. *et al.* Sustained Malaria Control Over an 8-Year Period in Papua New Guinea: The Challenge of Low-Density Asymptomatic Plasmodium Infections. *The Journal of infectious diseases* **216**, 1434-1443, doi:10.1093/infdis/jix507 (2017).

13 Mihretie, A. *et al.* Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. *Malaria journal* **16**, 89 (2017).

14 Bhatt, S. *et al.* Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of The Royal Society Interface* **14**, 20170520 (2017).

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

16 Magalhães, R. J. S. *et al.* Geographical distribution of human Schistosoma japonicum infection in the Philippines: tools to support disease control and further elimination. *International journal for parasitology* **44**, 977-984 (2014).

17 Sturrock, H. J. *et al.* Mapping Malaria Risk in Low Transmission Settings: Challenges and Opportunities. *Trends in parasitology* (2016).

18 Pigott, D. M. *et al.* Prioritising infectious disease mapping. *PLoS neglected tropical diseases* **9**, e0003756 (2015).

19 Semakula, H. M., Song, G., Achuu, S. P. & Zhang, S. A Bayesian belief network modelling of household factors influencing the risk of malaria: a study of parasitaemia in children under five years of age in sub-Saharan Africa. *Environmental Modelling & Software* **75**, 59-67 (2016).

20 Onyiri, N. Estimating malaria burden in Nigeria: a geostatistical modelling approach. *Geospatial health* **10** (2015).

21 Haddawy, P. *et al.* Spatiotemporal Bayesian networks for malaria prediction. *Artificial intelligence in medicine* **84**, 127-138 (2018).

22 Bhatt, S. *et al.* The global distribution and burden of dengue. *Nature* **496**, 504-507 (2013).

23 Korb, K. B. & Nicholson, A. E. *Bayesian artificial intelligence*. (CRC press, 2010).

24 Haddawy, P. *et al.* Spatiotemporal Bayesian Networks for Malaria Prediction: Case Study of Northern Thailand. *Studies in health technology and informatics* **228**, 773 (2016).

25 Chee, Y. E. *et al.* Modelling spatial and temporal changes with GIS and Spatial and Dynamic Bayesian Networks. *Environmental Modelling & Software* **82**, 108-120 (2016).
26 Chen, S. H. & Pollino, C. A. Good practice in Bayesian network modelling. *Environmental Modelling & Software* **37**, 134-145 (2012).
27 Stefanini, F. M. in *Proceedings of the Fourth European Workshop on Probabilistic Graphical Models.*
28 Wijesiri, B., Deilami, K., McGree, J. & Goonetilleke, A. Use of surrogate indicators for the evaluation of potential health risks due to poor urban water quality: A Bayesian Network approach. *Environmental Pollution* **233**, 655-661 (2018).
29 Scutari, M. Learning Bayesian networks with the bnlearn R package. *arXiv preprint arXiv:0908.3817* (2009).
30 Beresniak, A. *et al.* A Bayesian network approach to the study of historical epidemiological databases: modelling meningitis outbreaks in the Niger. *Bulletin of the World Health Organization* **90**, 412-417a (2012).
31 WorldClim. *Global climate data.* , <http://www.worldclim.org/> (2019).
32 NASA. *Earthdata*, <https://earthdata.nasa.gov/> (2016).
33 Scutari, M. Learning Bayesian Networks with the bnlearn R Package. . *Journal of Statistical Software* **35**, 1-22 (2010).
34 Wood, J. *et al.* The genetic demography of the Gainj of Papua New Guinea. I. Local differentiation of blood group, red cell enzyme, and serum protein allele frequencies. *American journal of physical anthropology* **57**, 15-25 (1982).
35 Castelletti, A. & Soncini-Sessa, R. Bayesian Networks and participatory modelling in water resource management. *Environmental Modelling & Software* **22**, 1075-1088 (2007).
36 Avila, R., Horn, B, Moriarty, E, Hodson, R, Moltchanova, E. Evaluating statistical model performance in water quality prediction. *J Environ Manage* **206**, 910-919 (2017).
37 Ho, S. H., Speldewinde, P. & Cook, A. A Bayesian Belief Network for Murray Valley encephalitis virus risk assessment in Western Australia. *International journal of health geographics* **15**, 6 (2016).
38 Fagerlin, A. *et al.* Communicating infectious disease prevalence through graphics: Results from an international survey. *Vaccine* (2017).
39 Spiegelhalter, D., Pearson, M. & Short, I. Visualizing uncertainty about the future. *science* **333**, 1393-1400 (2011).
40 World Health Organization. *Global technical strategy for malaria 2016-2030.* (World Health Organization, 2015).
41 Tatem, A. J. *et al.* Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria journal* **13**, 52 (2014).
42 Hall, D. C. & Le, Q. B. Use of Bayesian networks in predicting contamination of drinking water with E. coli in rural Vietnam. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **111**, 270-277 (2017).
43 Hetzel, M. W. *et al.* Progress in mosquito net coverage in Papua New Guinea. *Malaria journal* **13**, 1 (2014).
44 Pulford, J. *et al.* Indifferent to disease: a qualitative investigation of the reasons why some Papua New Guineans who own mosquito nets choose not to use them. *Social Science & Medicine* **75**, 2283-2290 (2012).
45 Lin, E. *et al.* Differential patterns of infection and disease with P. falciparum and P. vivax in young Papua New Guinean children. *PLoS One* **5**, e9047, doi:10.1371/journal.pone.0009047 (2010).
46 Wells, T. N., Burrows, J. N. & Baird, J. K. Targeting the hypnozoite reservoir of Plasmodium vivax: the hidden obstacle to malaria elimination. *Trends in parasitology* **26**, 145-151 (2010).

Tables and Figures

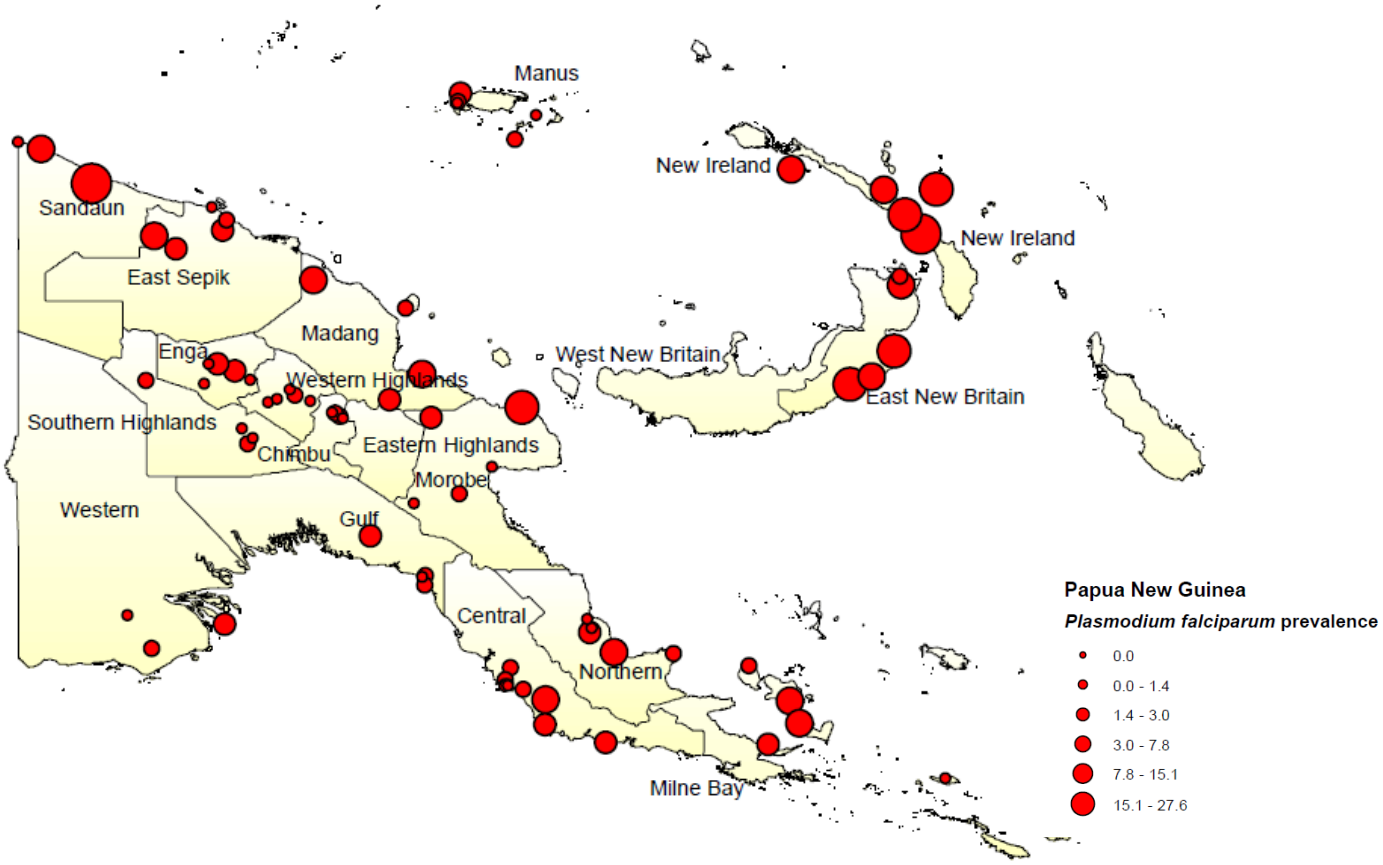


Figure 1. *Plasmodium falciparum* prevalence among 77 survey villages in Papua New Guinea , 2010/2011

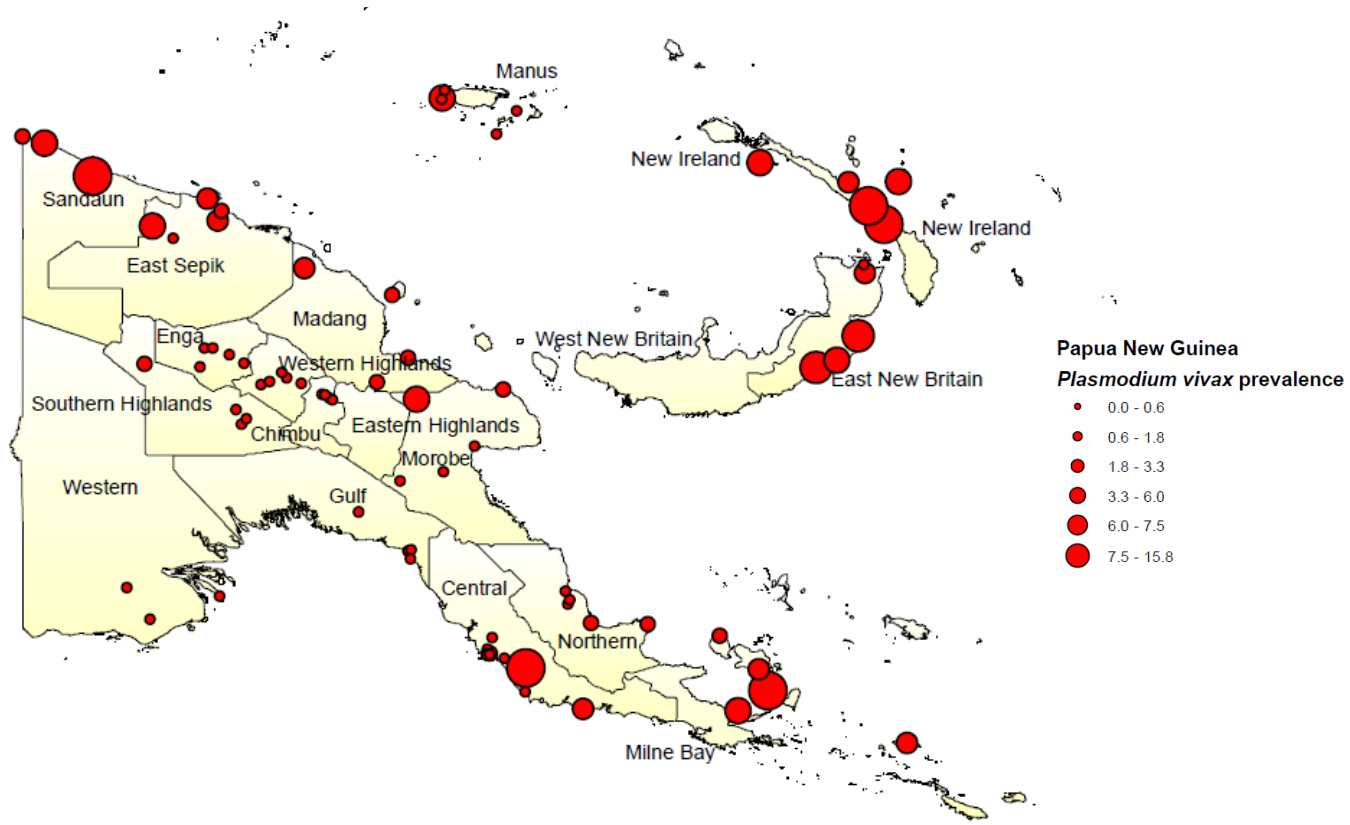


Figure 2. *Plasmodium vivax* prevalence among 77 survey villages in Papua New Guinea, 2010/2011

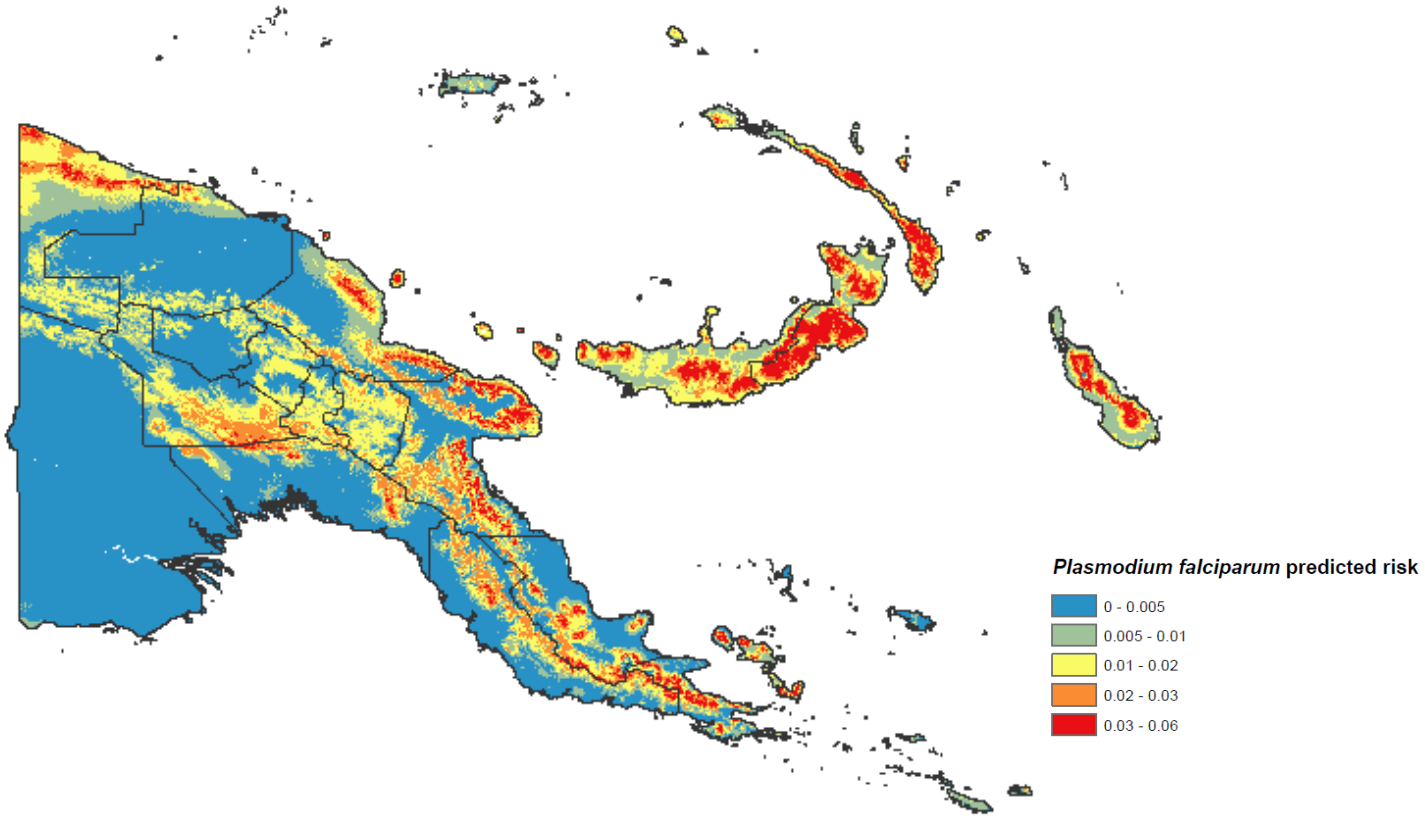


Figure 3. *Plasmodium falciparum* predicted risk spatial distribution based on the results of a generalised linear multivariate model in Papua New Guinea

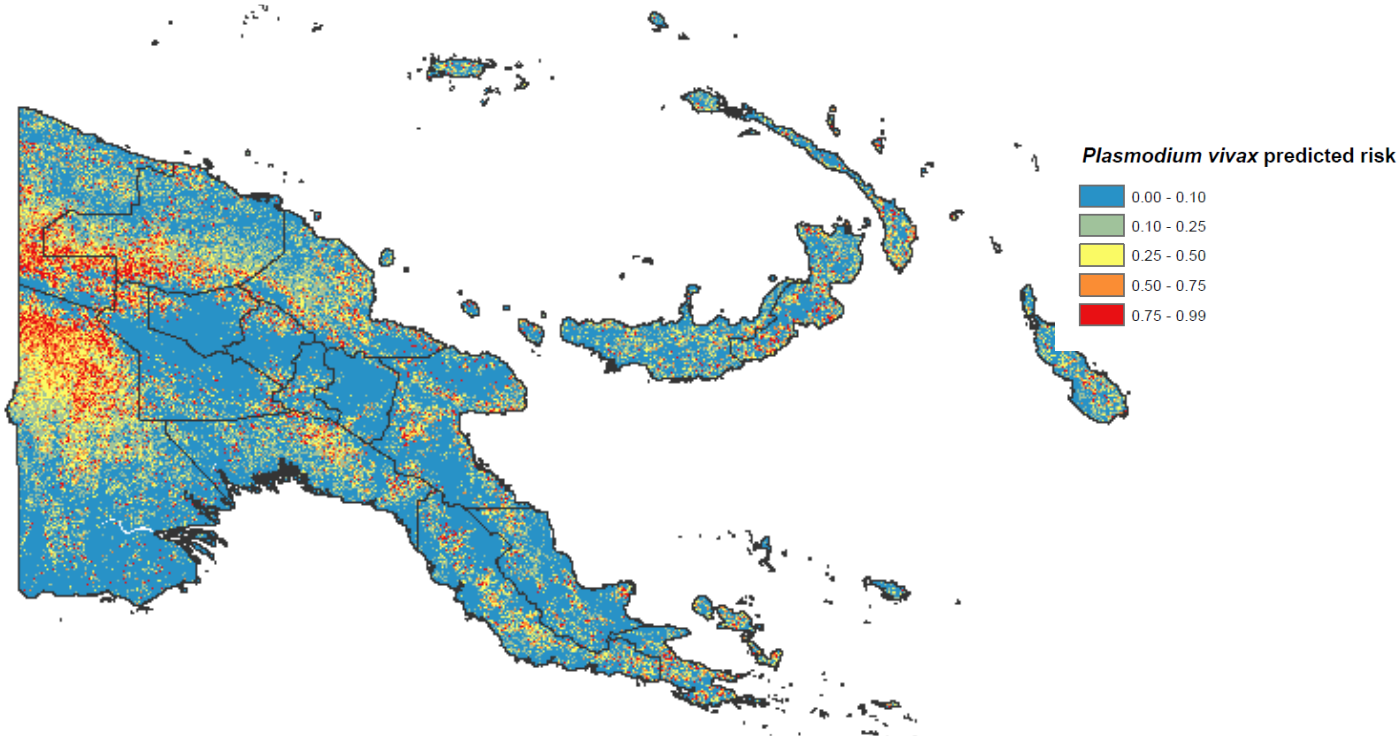


Figure 4. *Plasmodium vivax* predicted risk spatial distribution based on the results of a generalised linear multivariate model in Papua New Guinea

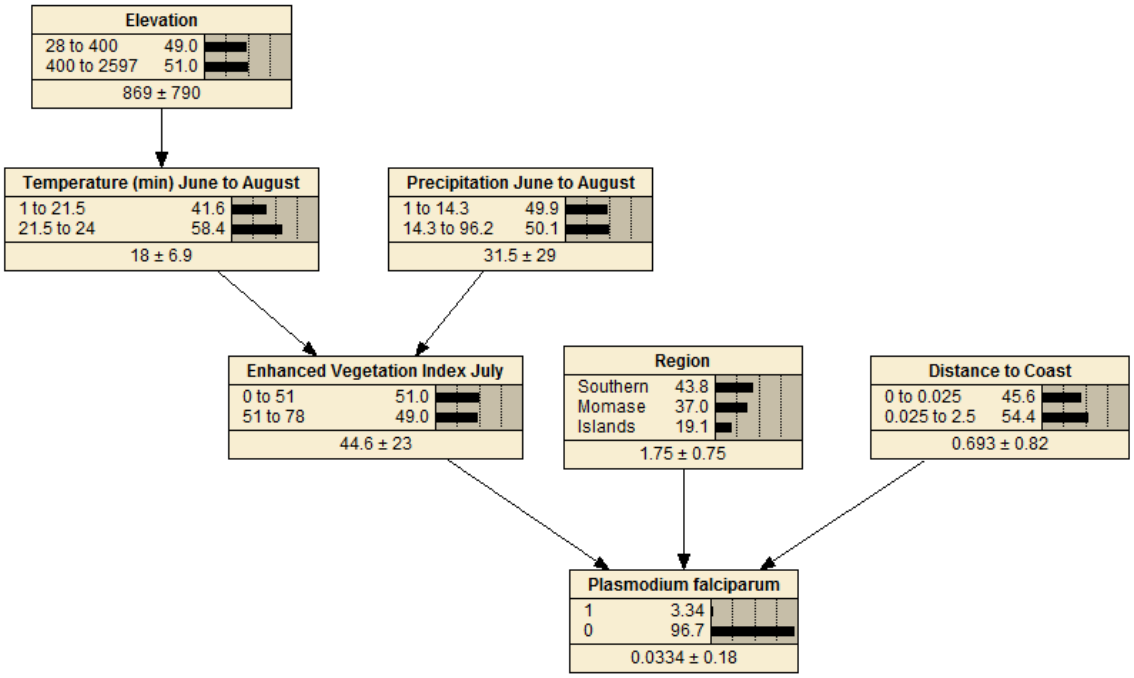


Figure 5. Expert structured Bayesian decision network showing associations of environmental variables with *Plasmodium falciparum* prevalence in Papua New Guinea.

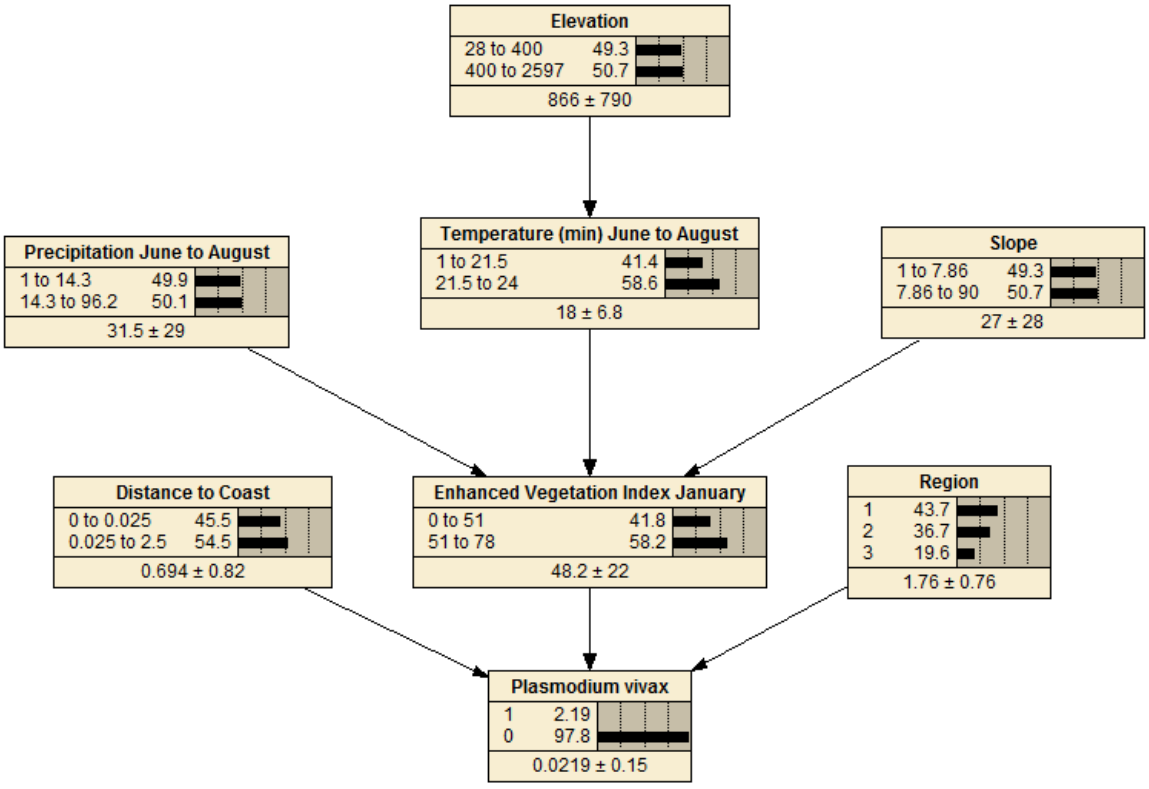


Figure 6. Expert structured Bayesian decision network showing associations of environmental variables with *Plasmodium vivax* prevalence in Papua New Guinea.

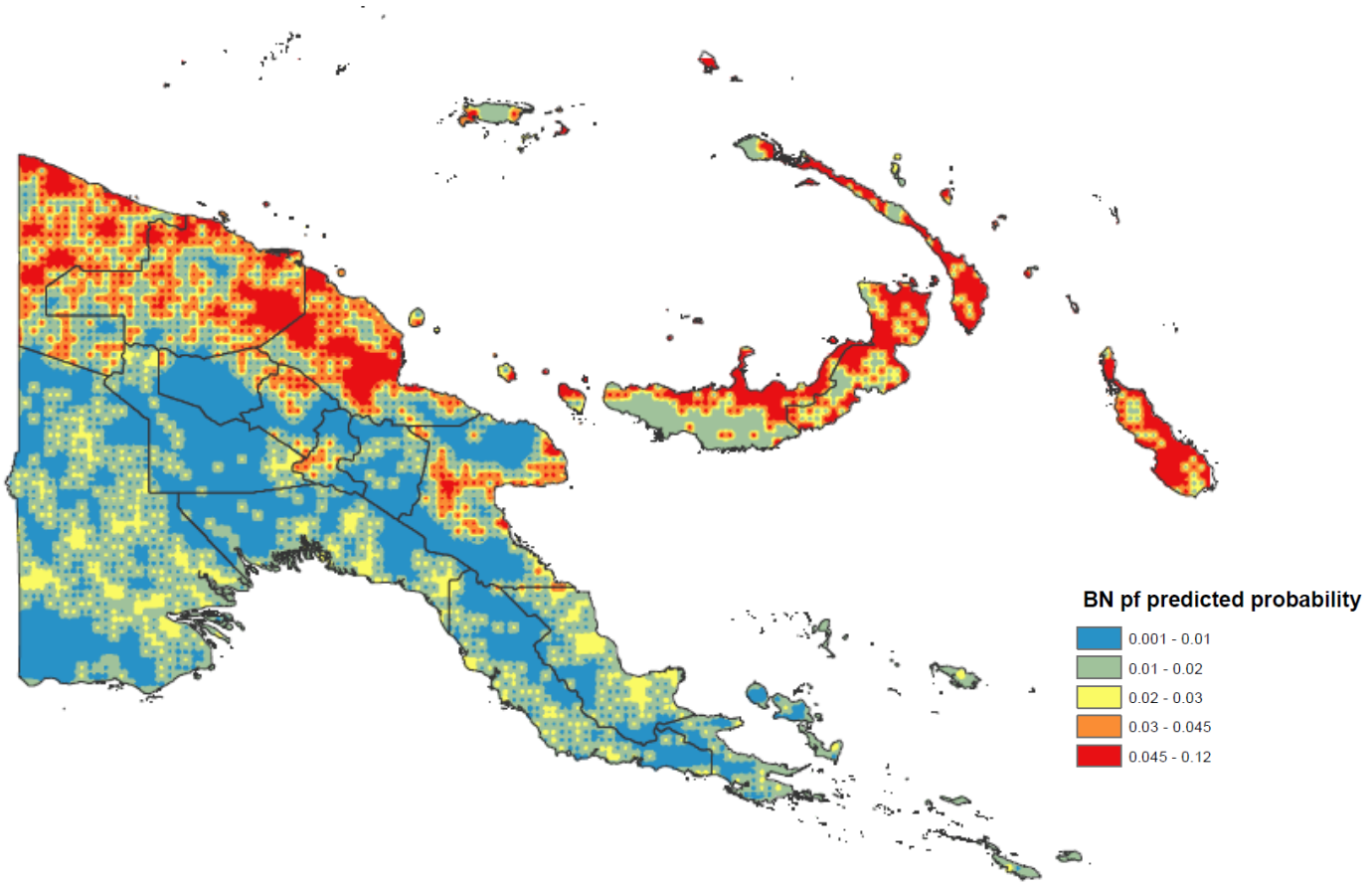


Figure 7. Predicted spatial distribution of *Plasmodium falciparum* risk in Papua New Guinea based on a Bayesian decision network model

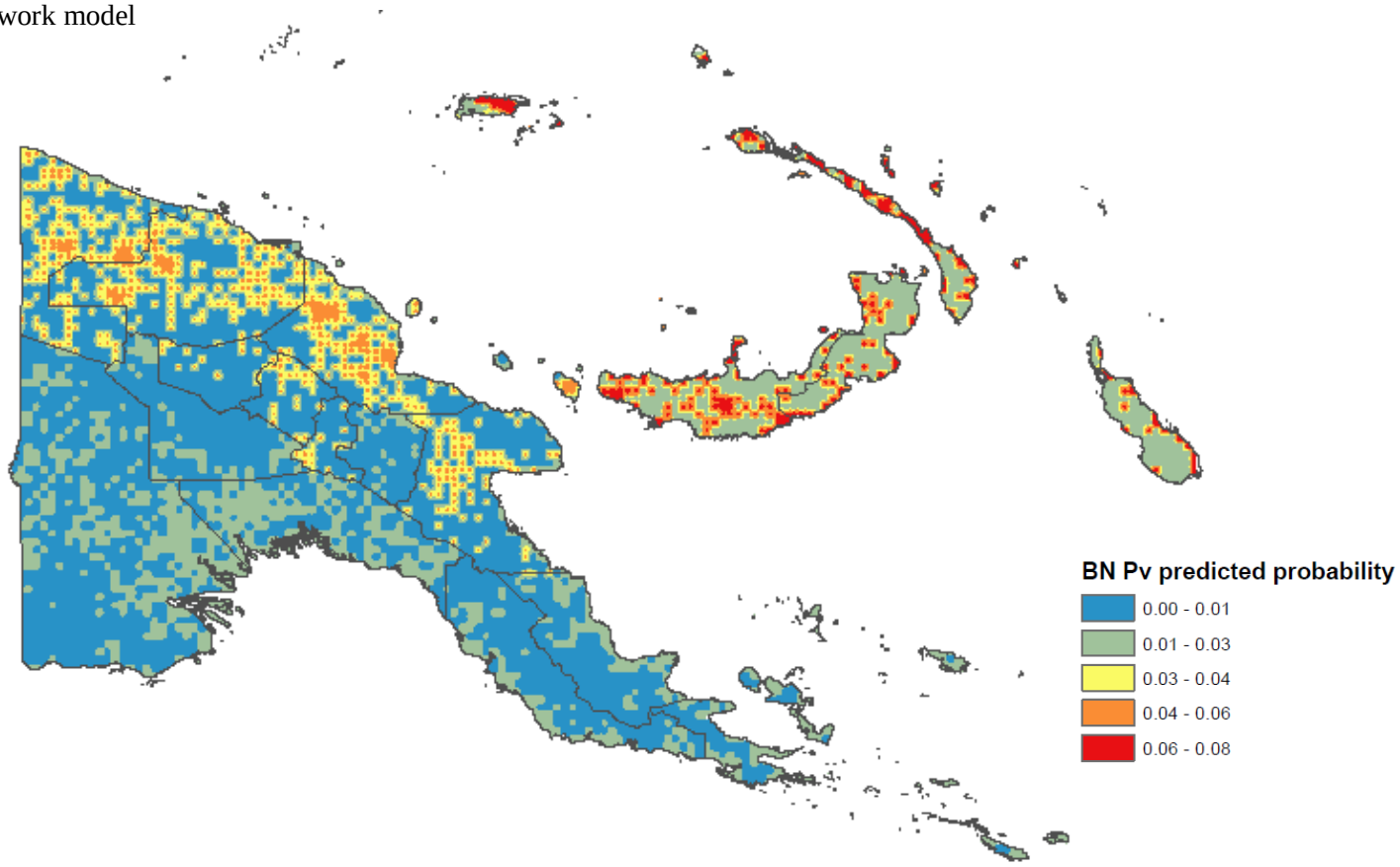


Figure 8. Predicted spatial distribution of *P. vivax* risk in Papua New Guinea based on a Bayesian decision network model

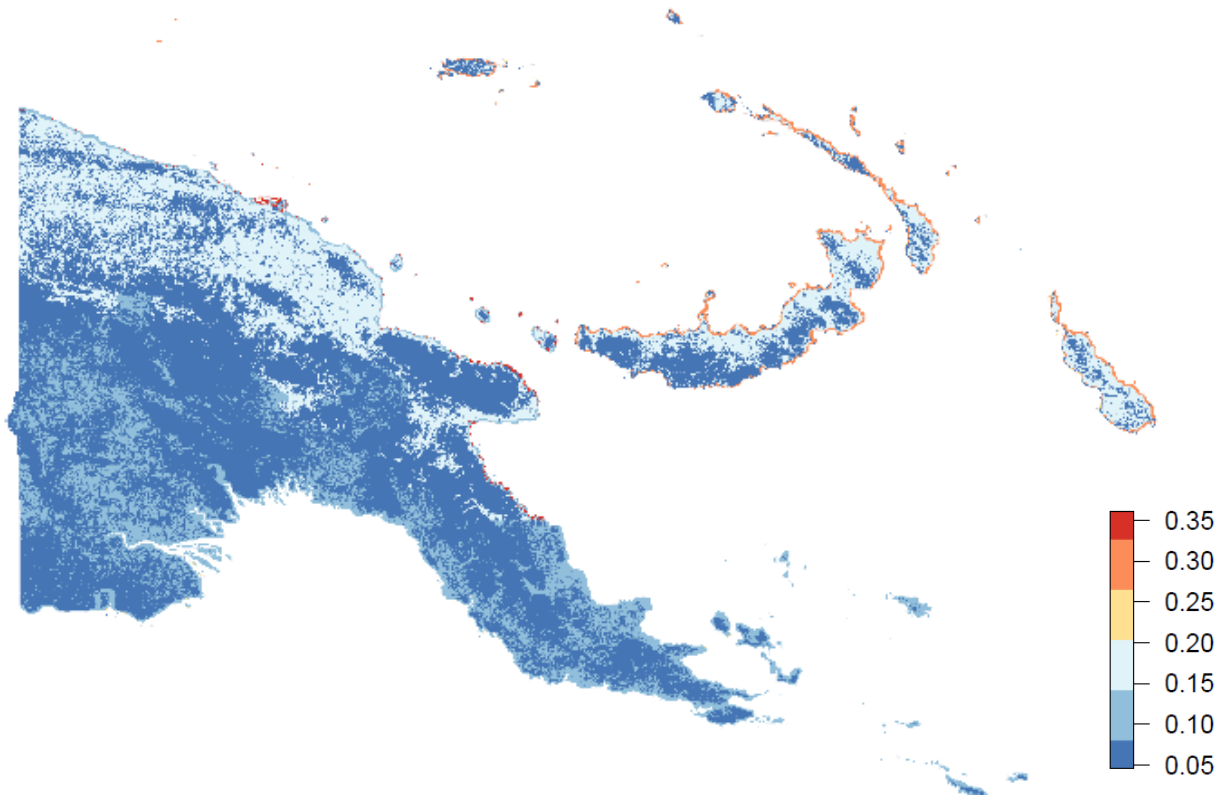


Figure 9. Spatial distribution of Shannon Index measure of entropy or uncertainty for *Plasmodium falciparum* predictions made using Bayesian decision network ecological model.

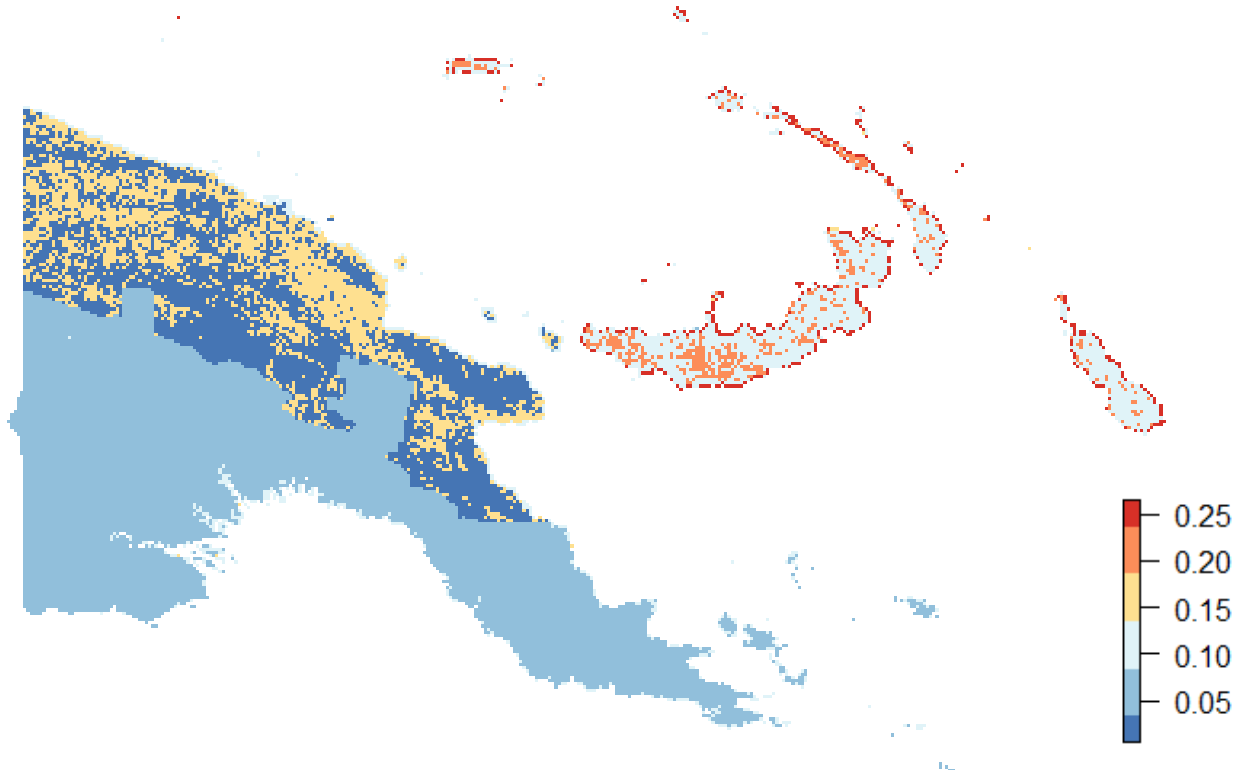


Figure 10. Spatial distribution of Shannon Index measure of entropy or uncertainty for *Plasmodium vivax* predictions made using Bayesian decision network ecological model

Figure 11 a – d. Receiver operating characteristic curves from results of *Plasmodium falciparum* and *Plasmodium vivax* generalised linear model cross validation

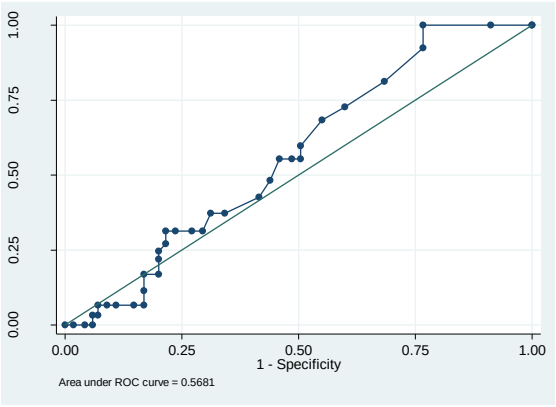


Figure 11a. *Plasmodium falciparum* cross validation 0.01 demographic adjusted village level prevalence cut-off against median village prevalence extracted from cross validation maps. AUC = 0.5681

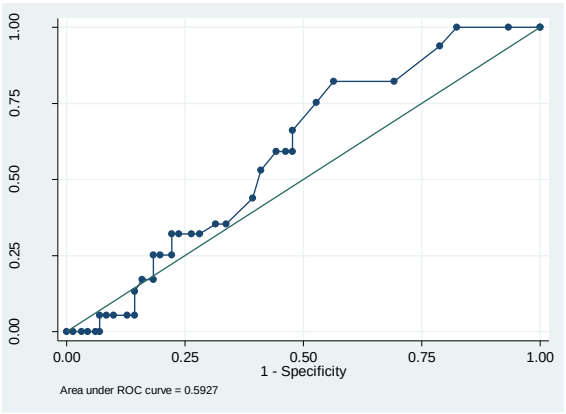


Figure 11b. *Plasmodium falciparum* cross validation 0.025 demographic adjusted village level prevalence cut-off against median village prevalence extracted from cross validation maps. AUC = 0.5927

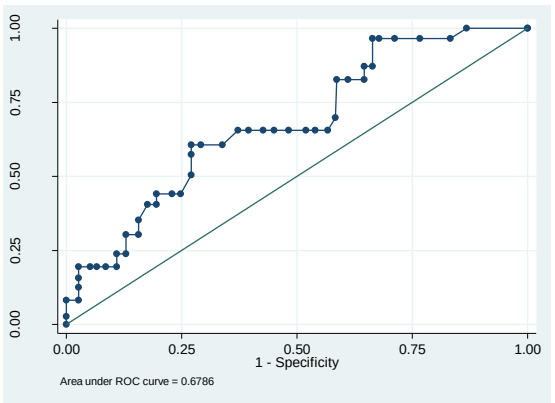


Figure 11c. *Plasmodium vivax* cross validation 0.01 demographic adjusted village level prevalence cut-off against median village prevalence extracted from cross validation maps. AUC = 0.6786

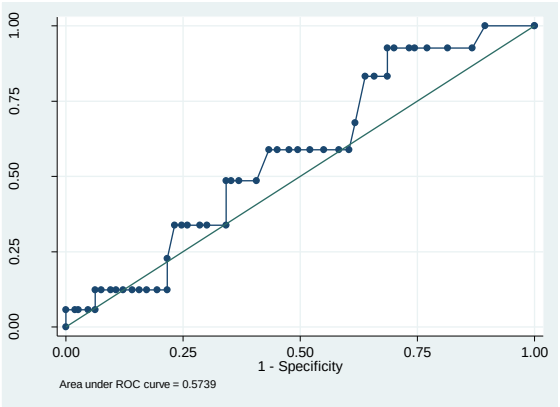
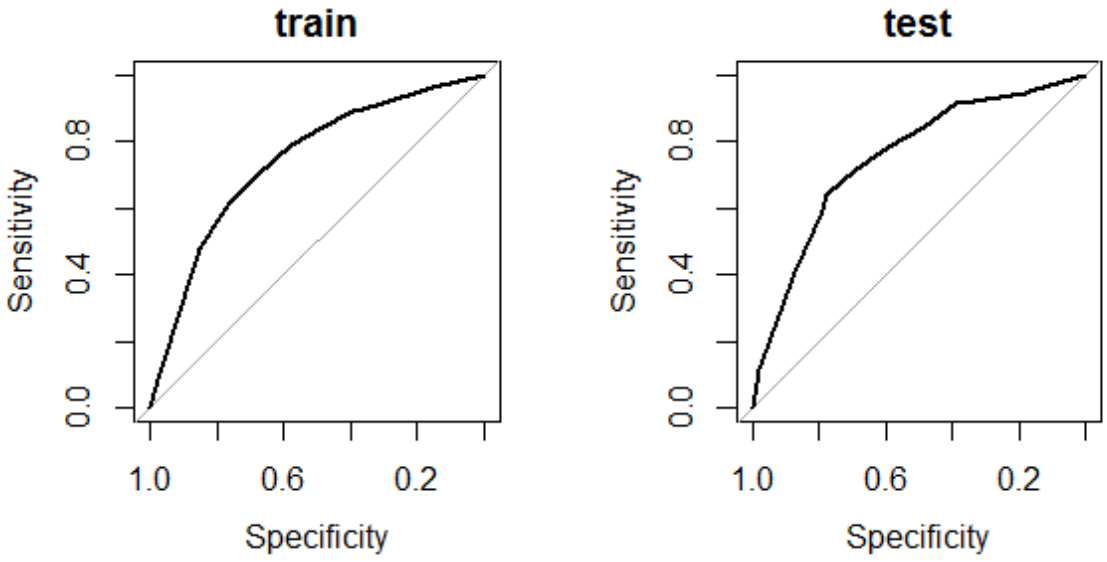
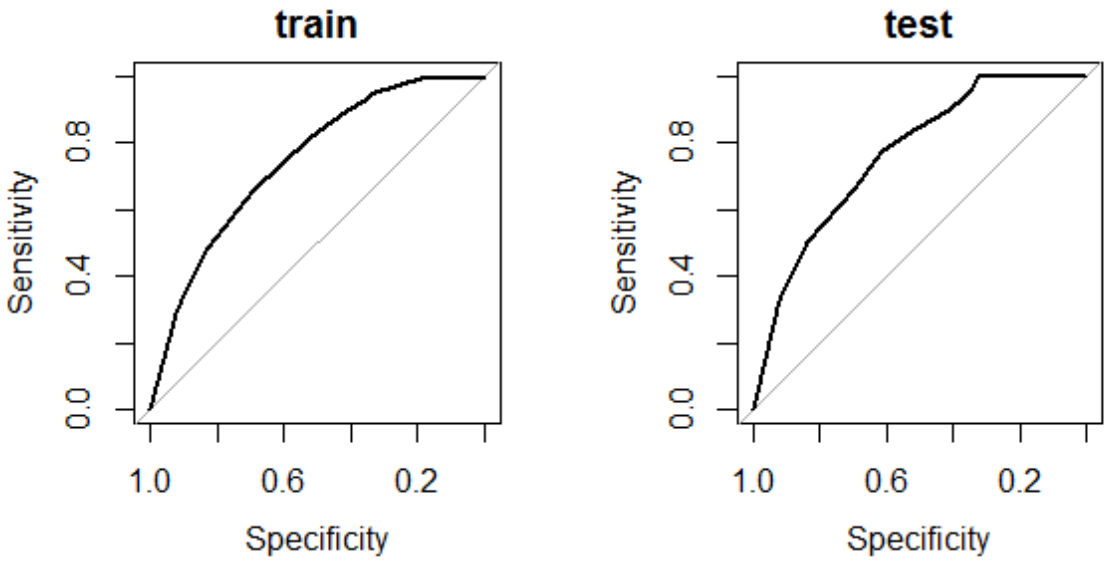


Figure 11d. *Plasmodium vivax* cross validation 0.025 demographic adjusted village level prevalence cut-off against median village prevalence extracted from cross validation maps. AUC = 0.5739



AUC = 0.7412 AUC = 0.7502

Figure 12. Receiver operating characteristic curves from results of *Plasmodium falciparum* Bayesian decision network model cross validation



AUC = 0.7448 AUC = 0.7623

Figure 13. Receiver operating characteristic curves from results of Bayesian decision network *Plasmodium vivax* model cross validation

Appendix 4

Supplementary material

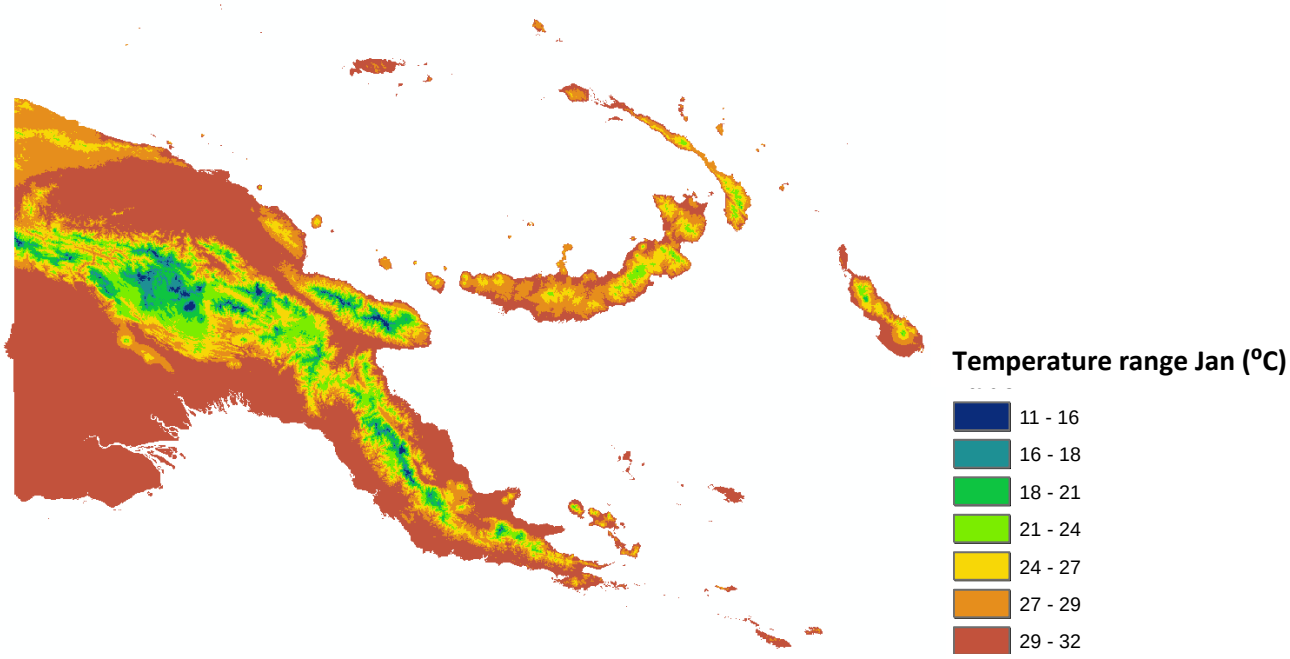


Figure S1 Average Temperature range in Papua New Guinea in January. Average temperature data aggregated over 50 years from 1950 to 2000

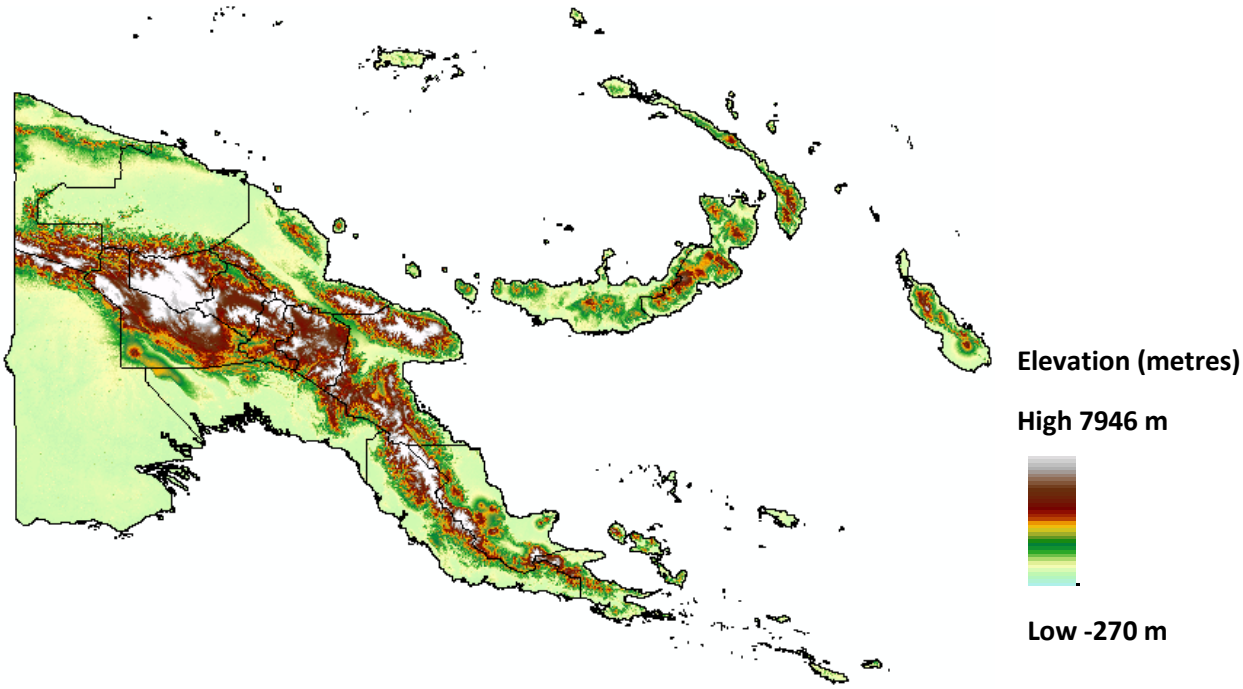


Figure S2 Elevation range Papua New Guinea

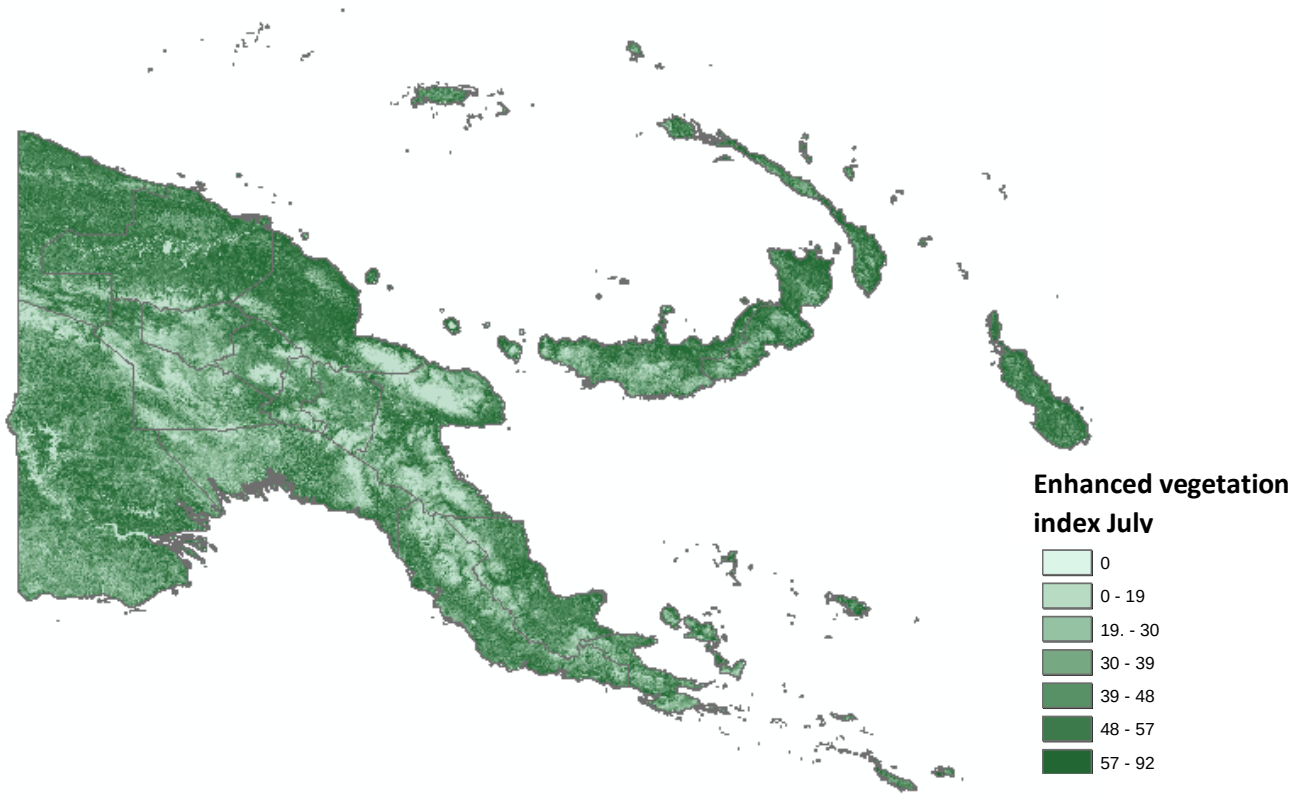


Figure S3 Enhanced Vegetation Index in Papua New Guinea in July 2011

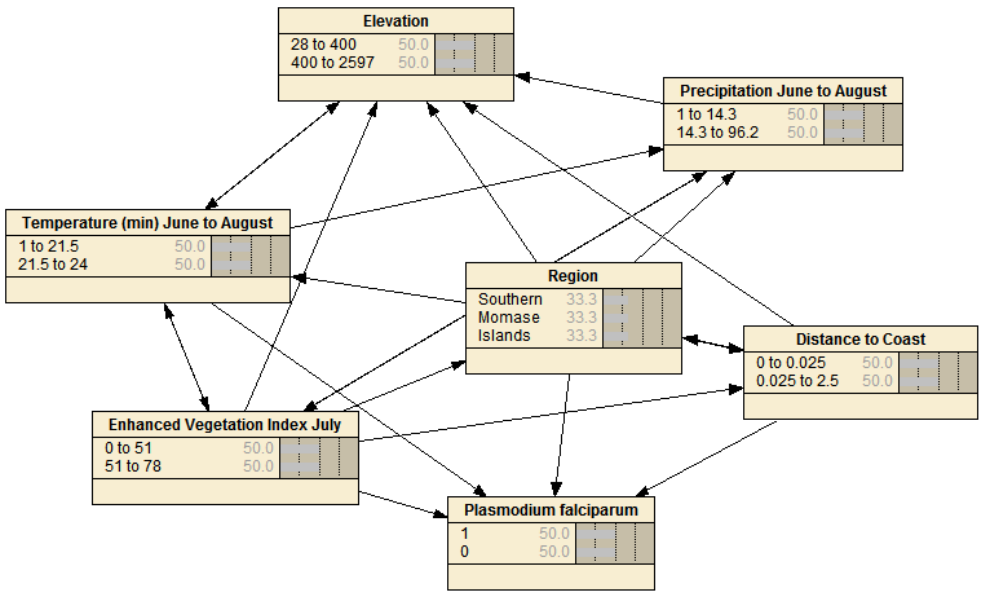


Figure S4 *Plasmodium falciparum* machine-learned Bayesian decision network model

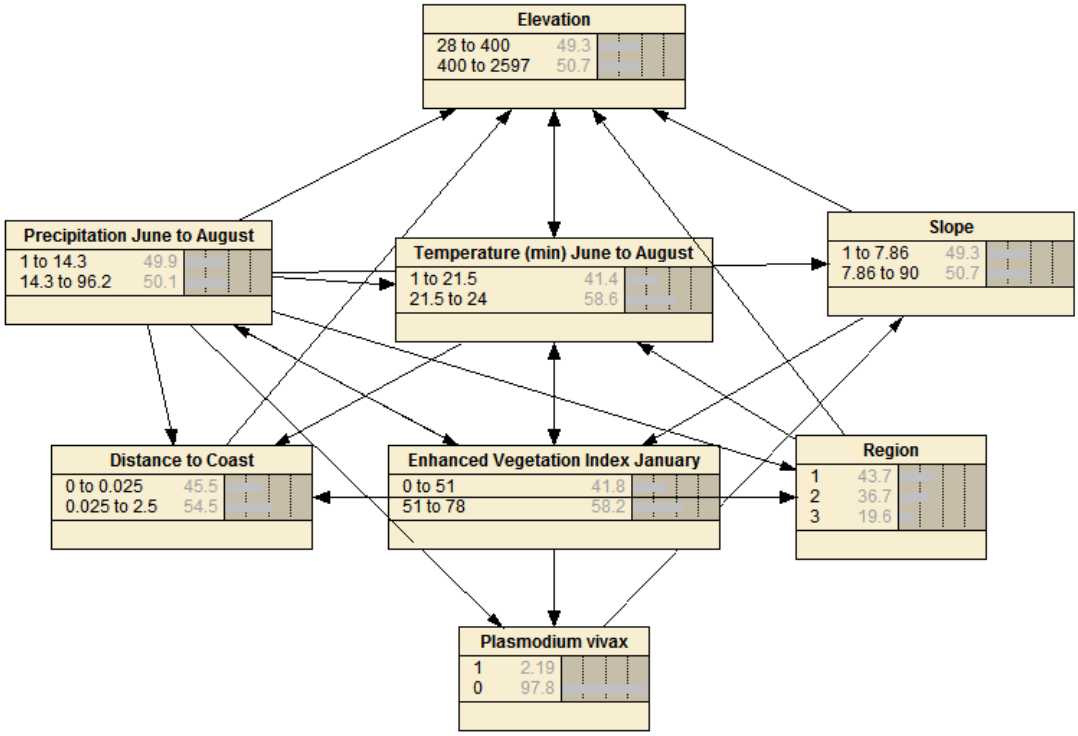


Figure S5 *Plasmodium vivax* machine-learned Bayesian decision network model

CHAPTER 5

Landscape genomics of Plasmodium falciparum in Papua New Guinea reveals major population subdivisions associated with ecological niches and routes of malaria transmission

CHAPTER 5. LANDSCAPE GENOMICS OF *PLASMODIUM FALCIPARUM* IN PAPUA NEW GUINEA REVEALS MAJOR POPULATION SUBDIVISIONS ASSOCIATED WITH ECOLOGICAL NICHES AND ROUTES OF MALARIA TRANSMISSION.

CONTEXT

In chapter 5, we propose a novel method for examining how parasite populations from distinct geographical areas are connected by human mobility among endemic areas. Determining parasite population connectivity of geographic areas through human mobility is important for informing malaria elimination programmes, as without identifying potential sources and routes of malaria transmission, elimination efforts in connected areas may be hampered.

This chapter uses a Dirichlet regression model to examine associations between distinct population clusters determined through *Plasmodium falciparum* genotype data and landscape ecology data to predict predominance of distinct parasite genotypes across Papua New Guinea (PNG). By examining the predicted spatial distribution of these distinct parasite populations or 'demes', we were able to infer how certain land areas of PNG permit parasite migration, or have been connected historically, in terms of malaria transmission. We extrapolate that these corridors of transmission have arisen via human migration. In contemporary malaria spatial epidemiology, addressing the impact of human mobility on malaria transmission, and examining methods for measuring human mobility, is one of the key areas of research currently receiving attention and has important implications for planning malaria control and containment of drug resistance. The research method in the chapter presented here is a novel interpretation of how parasite genetic data can be used to infer patterns of human mobility, together with its impact on malaria transmission dynamics, and informing elimination

programmes in such a manner as to prevent resurgence once elimination in specific areas within a country has been achieved.

This chapter has been formatted for submission as a manuscript to the j*ournal* of the Nature Communications.

CHAPTER 5. LANDSCAPE GENOMICS OF *PLASMODIUM FALCIPARUM* IN PAPUA NEW GUINEA REVEALS MAJOR POPULATION SUBDIVISIONS ASSOCIATED WITH ECOLOGICAL NICHES AND ROUTES OF MALARIA TRANSMISSION.

Authors:

Eimear Cleary[1], G.L. Abby Harrison2,3, Stuart Lee2,3, Livingstone Tavul4, Manuel Hetzel6,7, Ivo Mueller2,3,5, Melanie Bahlo2,3, Alyssa Barry[2,3], Archie Clements[1]

Affiliations:

1. Department of Global Health, the Research School of Population Health, the Australian National University, Canberra
2. Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, AUSTRALIA
3. Department of Medical Biology, University of Melbourne, Parkville, Victoria, AUSTRALIA
4. Vector Borne Diseases Unit, Papua New Guinea Institute of Medical Research, Madang, PAPUA NEW GUINEA
5. Parasites and Vectors Division, Institut Pasteur, Paris, FRANCE
6. Swiss Tropical and Public Health Institute, Basel, SWITZERLAND
7. University of Basel, Basel, SWITZERLAND

Abstract:

As malaria transmission declines and recedes to small geographic foci, gene flow becomes more contained geographically, and genetically distinct pathogen populations may occupy distinct niches with differing transmission dynamics. This study aimed to predict the spatial distribution of distinct genotype clusters of *Plasmodium falciparum* across Papua New Guinea (PNG), a country with highly variable malaria transmission. A total of 708 parasite genotypes based on a 154 single nucleotide polymorphism (SNP) barcode were obtained from 27 survey locations across all malaria-endemic provinces of PNG. Eight genetically distinct sub-populations (clusters) were determined based on ancestry co-efficients derived from Bayesian cluster analysis. A Dirichlet multivariate regression model examined associations of the eight *P. falciparum* genotype clusters with latitude and longitude, elevation, human population density and Euclidean distance from the coastline. The regression model was used to predict the spatial distribution of each genotype cluster across the country based on associations with these ecological covariates. Statistically significant associations were found with latitude and longitude for 6 of the 8 genetic clusters. Four clusters were associated with distance to the coastline and population density, but only two clusters were associated with elevation. Based on associations of spatial *P. falciparum* genotype data with environmental variables, distinctive geographical niches were predicted, including at least two distinct populations on the mainland that differ from those of the outlying islands. The results identify potential routes of parasite migration (gene flow) between neighbouring areas with high transmission. Mapping these geographical niches and parasite population connectivity can guide targeted control and intervention programmes for prevention of transmission between distinct geographic areas which are connected in terms of malaria transmission.

One Sentence Summary:

The geographical distribution of distinct *Plasmodium falciparum* genotype clusters was determined using a Dirichlet regression model and the results were used to predict the relative predominance of each genotype cluster across Papua New Guinea.

Main text:

5.1 Introduction

A concerted international commitment to malaria control has resulted in a reduction in the global burden of malaria of over 30% over the past two decades[1,2]. Despite this, the disease remains a major global health problem causing an estimated 216 million clinical cases and 435,000 deaths in 2017[2]. As endemic countries progress towards malaria elimination and malaria parasite populations shrink, major challenges to achieving malaria elimination will include the adaptation of the parasite in response to control interventions, and tailoring surveillance and intervention programmes to track and accommodate these changes. Particularly challenging to elimination will be the potential resurgence, and sustained transmission, of malaria facilitated by human migration[3], including that of emergent anti-malarial resistant parasite strains[4,5,6].

Parasite population genomics can provide deep insights into transmission dynamics that could be harnessed to accelerate malaria control and elimination[7-9]. In areas of high transmission, parasite populations tend to be genetically diverse, with little genetic differentiation between different geographic areas[10,11]. This is mainly as a result of extensive co-infection of strains, providing ample opportunity for meiotic crossover between genetically distinct parasites upon transmission to the mosquito vector and population mixing (gene flow) as a result of parasite migration between different geographic locations[12-14]. As transmission declines and recedes to small geographic pockets or hotspots, focal parasite inbreeding increases and gene flow is

contained to defined geographic areas, such that parasite populations differentiate into genetically distinct clusters[8,14]. Emergence of population structure may arise because of a reduction in the number of distinct strains and more focal clustering of infections due to effective control programmes[10,11,13-16]. A reduction in local parasite diversity may also arise as a result of limited migration between distinct populations. This potentially leads to a highly structured population demonstrative of marked interruption to local transmission[17].

Examining the extent to which *Plasmodium* parasite populations are structured in relation to human mobility can help to infer to what extent human population movement impacts on malaria transmission in a given setting[11,18]. Genomic data can also provide high precision accuracy in determining the source of infections in low transmission areas where outbreaks may be the result of imported or local residual transmission[19-21]. However, a major barrier to the uptake of genomic surveillance by malaria control programs is the complex output resulting from population genetic analyses. New approaches need to be developed that allow this information to be translated into easily understood data that could guide malaria policy and interventions.

Landscape genomics is an emerging field combining genetics, spatial statistics and landscape ecology[22] that allows parasite genome variation attributed to geographic processes to be studied in detail[8], and identify genetic changes associated with environmental variation[23]. Statistical models incorporating georeferenced population genomics data with spatial environment data can help to examine the geographic distribution of parasite population genetic variation, gene variants that drive adaptation[23] and variation in the ecology of distinct or related parasite lineages[24]. Genomic data have not yet been effectively employed in malaria surveillance[25], and no previous research has utilised genomic data in predicting the geographic distribution of *Plasmodium* parasites, or their spatial relationships with the physical environment[24].

We have generated a high density spatially referenced parasite population genetic dataset through genomic profiling (also known as "barcoding") of *P. falciparum* isolates from throughout Papua New Guinea (PNG). In 2008, the PNG National Malaria Control Program conducted an extensive malaria indicator survey collecting blood samples from more than 10,000 individuals living in all malaria endemic regions of the country (17/20 provinces)[26]. At this time, prevalence of *P. falciparum* ranged from 0 - 47% throughout the country[27] and analysis of the complexity of infection and genetic diversity of *P. falciparum* populations (based on *msp2* genotyping) was consistent with a wide range of transmission intensities [28].

To evaluate genomic surveillance as a tool for malaria control and elimination, we determined *P. falciparum* population structure by barcoding more than 700 isolates using a panel of 154 neutral genome wide single nucleotide polymorphisms (SNPs) (Harrison et al. manuscript in prep). Population genetic analysis of this data revealed a subdivided parasite population comprised of eight distinct genetic "clusters" distributed unevenly throughout the malaria endemic areas of PNG (Figure 1, Harrison et al, in prep). Together with remote sensing data, this dataset has provided a unique opportunity to explore parasite genetic data using spatial risk models. The aim of the current study is to use these datasets to produce predictive maps of *P. falciparum* population structure (specifically, the relative predominance of the eight genotype clusters) across PNG, based on a Dirichlet regression model.

5.2 Results

P. falciparum SNP barcoding reveals geographic subdivision of the PNG parasite population.

Genetic clusters identified through STRUCTURE analyses were asymetrically distributed throughout the country (Figure 2). This suggested substantial population structure in some regions (e.g. Islands, Milne Bay, Manus) and mixing between others (Sepik, Madang, Morobe).

Some clusters were represented across broad regions of the country (clusters 1, 2) while others were restricted to particular geographic regions (3= Islands, 5=Sepik, 6=mainland, 7=Milne Bay) and some were restricted to a small number of isolates in a single geographic region suggesting they may represent a clonal expansion of a recently imported infection or samples from relatively isolated villages (4 = Sepik, 8 = Morobe). Notably, the Manus population consisted of isolates with high predominance of clusters 1 and 2. Supplementary table 1 contains summaries of the proportion of isolates within each geographic region that have >=0.75 ancestry in each cluster.

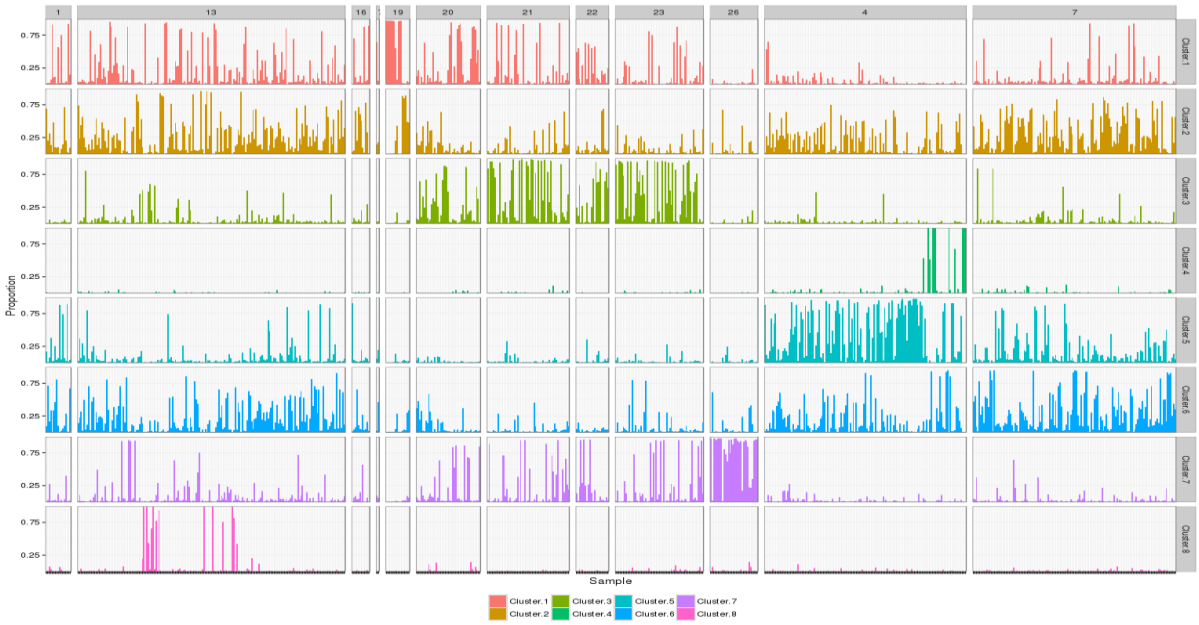


Figure 1. Results of Bayesian cluster analysis of 708 *P. falciparum* isolates barcoded with 154 SNPs for eight genetic clusters (K=8). The cleaned SNP barcode dataset was subject to analysis using STRUCTURE software[29] by running the analysis for K=1-20 and 20 runs, with an MCMC burnin of 5000 and total MCMC iterations up to 50,000. The data were visualised using the Starmie package[30] which separates the admixture co-efficients for each K. The parasite populations were subdivided according to provincial layers which accounts for geographic proximity of village clusters and/or known routes of human migration (1=West Sepik/Papua border, 13=Morobe/Eastern highlands fringe, 16=Northern (Oro), 17 = Central, 19= Manus, 20 = West New Britain (north), 21 = West New Britain (south), 22=East New Britain, 23= New Ireland, 26 = Bougainville/Milne Bay, 4 = Sepik (coast), 7 = Madang/W Highlands fringe).

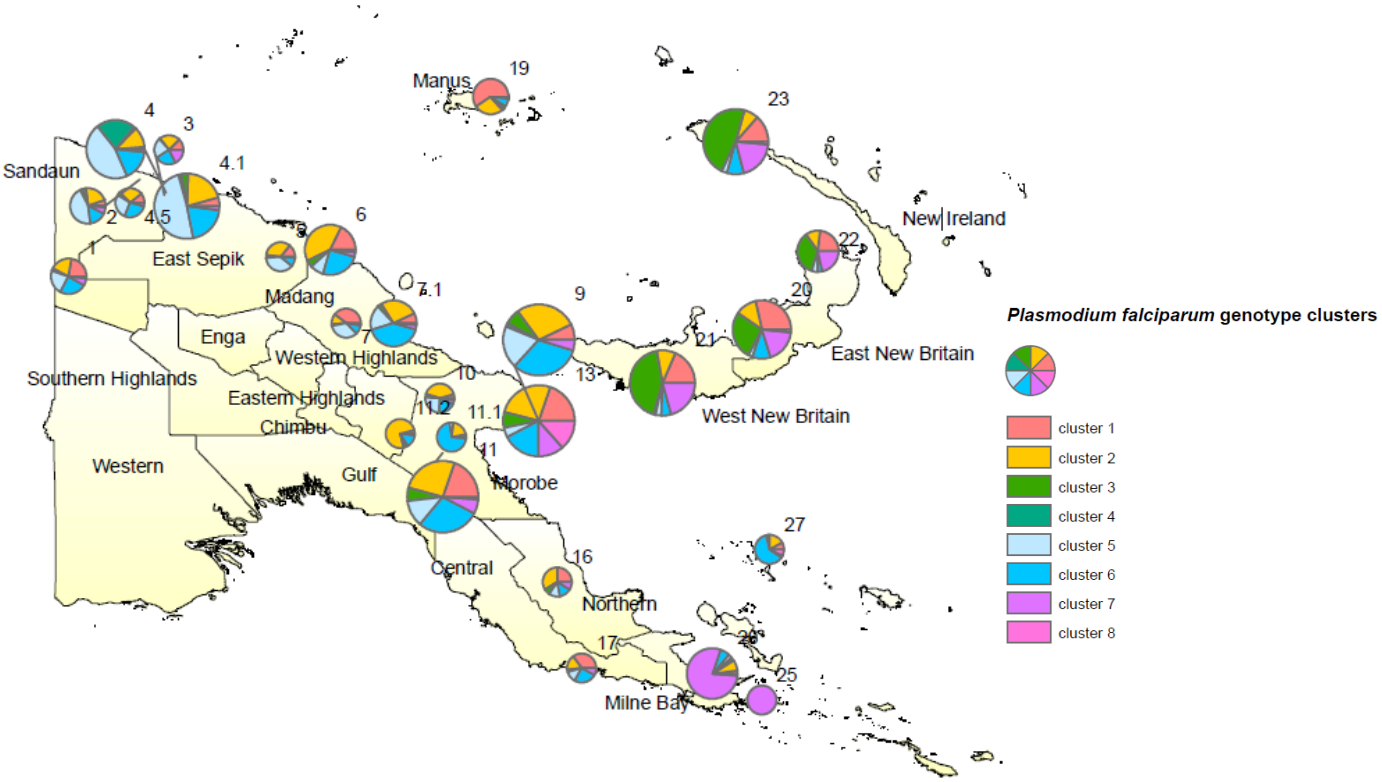


Figure 2. Map of Papua New Guinea showing sampling locations and mean ancestry coefficients for each of 27 geographic areas.

Statistical models reveal that genotype clusters were strongly associated with geographic provenance

Most *P. falciparum* genetic clusters were found to demonstrate spatial trends captured by their associations with latitude and/or longitude, and their quadratic terms, with the exception of clusters 4 and 8 (Table 1). Cluster 4 accounted for only a small proportion of genetic ancestry in the majority of geographic locations (.001 to .217), with the exception of Ilahup, East Sepik Province where 21.7% of genetic ancestry was attributed to cluster 4. Likewise, cluster 8 had a relatively small contribution to genetic ancestry in all geographic locations from where samples were taken, with a range of .001 in Wonera, Eastern Highlands Province to .134 in Morobe Province. These were represented by a small number of samples (Figure 2, Table S2) and may indicate relatively isolated villages with highly focal transmission or cases that have emerged from a single origin importation event.

Clusters 1 ($p < .0001$), 2 ($p < .001$) and 6 ($p < .001$) were significantly associated with Euclidean distance to the coast, with cluster 1 exhibiting the highest proportion of genetic ancestry in the coastal province of Madang (0.312), Manus Island (0.590) and West New Britain (0.282) and cluster 2 exhibiting the highest genetic ancestry in Mandang and Morobe (0.390, 0.408). Cluster 6 had highest proportion of genetic ancestry in Madang province (0.402), Morobe (0.702), the coastal highland provinces of East and West Sepik (0.230, 0.269) and Central province on the south eastern coast. Clusters 1 ($p < .001$), 2 ($p < .0001$), 6 ($p < .001$) and 7($p < .01$) were significantly associated with human population density.

Spatial predictions reveal the predicted range of genetic clusters

Spatial predominance of each distinct *P. falciparum* genotype cluster was ascertained by making predictions from the Dirichlet regression model to a 3768 point gridded dataframe of measures of population density, elevation, distance to the coastline and latitude and longitude coordinates across PNG (Figure 3). Cluster 1 shows the greatest predominance (20% – 90%) in the highlands of PNG, the outer islands, Manus Island, New Hanover in New Ireland, Mussau Island one of the St. Matthias islands, situated to the north west of New Ireland, and Port Moresby. Cluster 2 was predicted to dominate only in the highlands (25% - 40%). Cluster 3 was predicted to dominate in the east of New Britain and New Ireland and in the west of Bougainville (25% – 35%).

Areas where clusters 4 (7% - 10%) and 5 (25% to 40%) were predominant were in the South Western and Sepik provinces. Cluster 6 was predicted to be more dominant in the highlands, extending into the north coastal provinces of Madang and East Sepik (25% - 35%), with substantial overlap with areas where cluster 2 was predominant. Cluster 7 exhibited greater dominance in the archipelago off Milne Bay, with moderate predominance (40% - 70%) on the PNG mainland at Milne Bay. Cluster 8 was found to be mainly coastal in the south and north

of the country and on the coast of New Britain, although with a generally low degree of predominance (7% - 8%). Upon cross validation of model results, spatial predictions generated using a training subset of the data and applied to a validation subset were found to perform well for clusters 1 to 7 but poorly for genotype cluster 8 (AUC test dataset range = 0.505 – 0.652; Table 2).Predictions for cluster 3 were found to perform the best overall (AUC test dataset range = 0.695 – 0.908).

5.3 Discussion

Spatially referenced parasite genetic surveys provide data that can help to identify routes of transmission, emergence of genetic traits such as anti-malarial resistance, and changes in malaria epidemiology in response to control interventions[7,16,31,32]. Examining the spatial distribution of parasite population structure[19] can help to identify changing transmission dynamics in different areas and evaluate the efficacy of control and elimination strategies[33,10,15]. Defining patterns of gene flow and spatial variance within parasite population structures can help countries decide when to switch from broad-ranging control efforts to targeted control, to define relatively isolated and fragmented populations for targeted elimination, and to map migration patterns between distinct epidemiological areas to assess the risk of reintroduction and the spread of drug resistant parasites[19]. Models comparing genetic data with cell phone data for quantifying movement of people find related information about the patterns of geographic spread of parasites and impact of human mobility on malaria transmission[34,35].

Bayesian cluster analysis initially revealed eight genetically distinct clusters that were unevenly distributed amongst the geographic areas of PNG. Parasites from all endemic areas had common ancestry in several of these clusters, which suggests a common founder population and/or historically high levels of gene flow between endemic areas. However, specific clusters were dominant in the eastern group of Islands (New Britain, New Ireland, Bougainville), Milne

Bay and Manus Island. In addition, a number of clusters were more dominant on the Mainland and relatively underrepresented in the Islands and Milne Bay. This is consistent with low gene flow between the major endemic regions of the Mainland, the Islands and Milne Bay, suggesting that these regions harbour distinct parasite populations or "demes" that could be independently targeted by control interventions. The way in which population genetic data have been presented to date (e.g. ancestry co-efficient plots) fail to capture the spatial distribution of the genetic clusters or demes, making the information and spatial patterns difficult for non-experts to interpret. Here, using spatial data on *P. falciparum* genotypes and related environmental variables, we were able to predict the ecological niches of eight genetically distinct *P. falciparum* genotype clusters in PNG with a high degree of accuracy for some clusters.

The spatial distribution of cluster 1 implies substantial gene flow between Port Moresby, the Highland region, Manus Island and western New Ireland, possibly due to recent human population movement. The distribution pattern implies that parasite migration could be occurring via the capital city Port Moresby to highland areas and the islands, possibly as a result of movement of people by air, given that Port Moresby is the commercial airline flight hub of PNG through which all domestic flights connect. However, *P. falciparum* transmission is extremely low in the southern regions of PNG[27] (including where the capital, Port Moresby, is located), thus the dataset excluded samples from this region, and we were unable to confirm whether parasites from this region were locally transmitted or imported from other endemic areas. However, these results closely mirror those of a recent study of the other major human malaria parasite in the region, *Plasmodium vivax*, where the endemic areas nearby to Port Moresby were found to harbour parasites with both Island and Mainland ancestry, suggesting it is a sink of malaria transmission[28] (Fola et al. 2018 MEEGID). This suggests that Port

Moresby airport could be a point of intervention following subnational elimination to prevent reintroduction to other regions of PNG.

Also of note was the predicted localised transmission in the Highland region of clusters 2 and 6, indicating co-circulation and localised transmission of these distinct *P. falciparum* genotypes, albeit at low levels as indicated by the low predicted predominance in this region (0.25 to 0.4 & 0.25 – 0.35, respectively). Localised transmission of one parasite genotype in the highlands suggest that control and prevention of reintroduction interventions in highland provinces could be successful for targeted elimination. The higher predicted predominance of cluster 3 in the spatially proximal locations of East New Britain, New Ireland and western Bougainville suggests frequent population movement and localised transmission between these islands. Travel between these islands is likely by sea and possibly to and from several ports on each island, which may pose a challenge to surveillance and interventions aimed at these mobile populations.

The dominance of cluster 4 as a distinct parasite population in the western most parts of the country, South Western and Sepik provinces, suggest a cross-border transmission zone including West Papua. The spatial prevalence of cluster 7 in the archipelago off Milne Bay, and to a lesser extent on the Milne Bay mainland, suggest gene flow between the mainland and the islands of Milne Bay. Interestingly, spatial predictions of any of the other 7 clusters were not dominantly exhibited in this peninsula, which may suggest that transmission remains strongly localised in this area and that this population is disconnected from other parts of PNG, offering a good target for spatial malaria elimination. We do acknowledge that the dataset was spatially sparse in some areas, with spatial predictions of relative predominance of each cluster being informed by observations from distant locations. This was mainly an issue with predictions in the Highlands that were not informed by any data from that region.

Genetic diversity of malaria parasites is indicative of the capacity for adaptation to selective, evolutionary and environmental pressures, as has been observed with the emergence of antimalarial resistant parasite strains[8,18]. However, there is also evidence to suggest that occupation of geographical niches by distinct *Plasmodium* genotypes may be driven, in part, by selective pressure and genetic adaptation exerted on *Plasmodium* parasite strains by different anopheline species[36,37]. The three main anopheles species involved in human malaria transmission in PNG occupy distinct habitats and as a result there is marked heterogeneity in the geographical distribution of these species[38,24]. Specifically, different species predominate in the lowlands, inland and coastal regions, and island provinces. Furthermore, the PNG human population shows spatial genetic heterogeneity with differing prevalence of genetic polymorphisms in different geographic areas, including those associated with malaria susceptibility[39,40] . Thus, the distinct host species in different areas might explain some of the spatial variation in predominant *Plasmodium* genotypes.

Local circulation of parasites where population movement is limited may also explain distinct spatial separation of parasite clusters. The human population of PNG is predominantly rural and sometimes isolated, separated geographically and culturally[41] by high mountain ranges, forests, large rivers[42], and poor road infrastructure, which may have resulted in limited *Plasmodium* gene flow between areas. This, together with differences in intensity of transmission and control interventions across the country may explain the highly structured parasite population, high levels of genetic differentiation and variable levels of *P. falciparum* genetic diversity between distinct populations and geographic areas within PNG[15,42].

In planning interventions, the genetic diversity and structure of malaria parasite populations is an important factor to consider in how well control strategies may work in different geographic locations[42]. Genetic diversity is an indicator of how robust the parasite population may be when faced with environmental pressures, for example, how susceptible it may be to targeted drug

treatment[42]. If these parasites also diverge at antigen loci, increasing mobility of human populations may result in infection by genotypes that people have not formerly been exposed to, and consequently, lead to more severe clinical infections. Understanding the parasite population structure and origin of infection may be beneficial for allocating resources and planning treatment and interventions for infection with non-indigenous parasite genotypes that might result in more severe clinical cases. Identifying introduced infection when estimating small area incidence is important for determining whether cases are attributed to local disease ecology or importation for targeting appropriate control responses[43]. Recent genotyping of malaria parasitic infections in Eswatini however found high genetic diversity in a low transmission area, with similar levels of genetic diversity among imported and local cases, with difficulty in accurately discriminating between these two groups of infection[44].

The impact of human population movement on malaria transmission will present a major barrier to elimination[18,19] and quantifying patterns of human movement and migration is an area of research that is becoming more widely used in determining transmission patterns of infectious disease[6,34,45-47]. For example, using mobile phone tracking data to inform patterns of human movement is becoming increasingly popular. Mobile phone data captures only population level migratory patterns however and does not include information on movement of high risk or infected populations. Population genetics of malaria parasites, may provide novel methods for inferring patterns of human movement by examining the distribution of geographically distinct parasitic genotype clusters and connected catchments[48] or genetic relatedness based on identity by descent[49]. Understanding how distinct populations are connected in terms of malaria transmission, together with information on population and infection size in areas of genetically distinct infection clusters allows for targeted control efforts in transmission sources and sinks. However, it is difficult, given the lack of other indicators of migration including mobile phone records and travel history[34] to explore the validity of these

maps as inference for human movement patterns within PNG, and this work would benefit by comparing the results obtained here with human movement patterns obtained using mobility data and movement models[35].

Here we use the novel approach of combining genetic data with spatial epidemiological methods to examine the predicted spatial distribution of *Plasmodium* parasite subpopulations. In examining the distinct geographic areas where malaria transmission is most likely attributed to specific plasmodium parasite genotypes, and inferring connectivity between these geographic locations in terms of malaria transmission, we attempt to better understand how malaria transmission may be facilitated by human movement. This approach could lead to the development of indicators of the success of control efforts in limiting onward transmission of malaria[10] and assessing the risk of re-introduction of malaria from reservoirs of infection in neighbouring villages and countries after local transmission has ceased[50]. Examining the geographic niches of *P. falciparum* genotypes also allows us to identify areas in which distinct genotypes are circulating enabling targeted elimination in areas where transmission is localised.

5.4 Conclusion

Eight distinct *Plasmodium falciparum* subpopulations were found to inhabit distinctive geographical niches in PNG and statistically significant correlations were found between parasite genotype cluster relative predominance and geolocation, distance from the coastline, elevation, and human population density. Parasite genomic data may prove useful in developing tools for risk assessment in conjunction with spatial epidemiology and as an indicator for understanding how human migration may be contributing to onward transmission of malaria.

5.5 Methods

Study site and samples

A total of 8936 samples were collected during a national malaria indicator survey of participants from randomly selected households in 49 villages from 16 provinces in PNG between October 2008 and August 2009[51]. Further details of the survey methodology are published elsewhere[51,52]. For the genotyped samples, we assigned the villages to 27 catchment areas predicted to harbour distinct parasite populations on the basis of the close proximity of villages (< 4.5° lat long), the surrounding topography, including elevation, and predicted human movement based on our knowledge about transport networks within PNG. Genomic DNA was extracted from whole-blood samples using the QiaAmp DNA Extraction Kit (Qiagen, Chadstone, Victoria, Australia) or the Favorprep™ genomic DNA extraction kit (Favorgen, Taiwan). Light microscopy (LM) and ligase detection reaction–fluorescent microsphere assays (LDR-FMA) were performed to identify samples infected with different *Plasmodium* species[53].

SNP genotyping and population genetic analysis

All *P. falciparum* positive samples from the nationwide survey (n = 1513) were initially genotyped using the highly polymorphic *Pfmsp2* marker as previously described to determine multiplicity of infection (MOI)[54]. A total of 722 isolates with a single clone infection (MOI=1) and a subset of samples with two clone infections (MOI=2) were selected for this study to supplement the sample sizes in some geographic areas. Isolates were then genotyped for 191 SNPs using the Fluidigm BioMark platform as described elsewhere (Harrison et al. in prep). After excluding 37 SNPs and 14 samples with high levels of missing data, a total of 708 isolates genotyped at 154 SNPs with sample sizes ranging between 2– 87 in the different geographic areas were used to explore population structure (Table S1, Dataset 1). Genetic ancestry was defined using the Bayesian clustering algorithm implemented in the software STRUCTURE[29].

STRUCTURE was run for 1-20 genetic clusters (K) repeating the runs 20 times with different seeds, an MCMC burn-in period of 5000 iterations and total MCMC iterations up to 50,000. The optimal K was determined using the Evanno method[55]. We also checked MCMC diagnostics by looking at the change in admixture parameter over the MCMC runs. All of the analysis and visualisations were performed in a new R package known as Starmie[30].

Data on the Physical Environment

Elevation data, distance of each of the 27 village locations to the coastline and population density were obtained from open-source remote sensing image data repositories. Elevation raster files for each tile covering PNG were obtained from a Global Digital Elevation Model downloaded from NASA's Earth Observing System Data and Information System, Reverb[56] (download date: December, 2015). Spatial population density data were obtained from the WorldPop online repository[57](download date: May 2017). A coastline polyline was defined using the PNG shapefile coastline and distance calculated from the centre point of each of the 27 geo-referenced village locations from which survey data were collected. Spatial averages of all environmental covariates were extracted from a 5km buffer around each of the 27 village locations. All spatial data processing was carried out using ArcGIS software version 10.3 (ESRI, Redlands, California).

Statistical Analysis

The aim of the statistical analysis was to 1) quantify the association between explanatory variables relating to the physical environment and the relative predominance of each genotype cluster in each location, and 2) to predict the relative predominance of each genotype cluster throughout the land surface of PNG. For multivariable regression, a Dirichlet distribution was chosen due to the continuous multinomial distribution of the dependent variables. There were eight dependent variables, the value of each of which was a proportion of 1; these dependent

variables were the proportions of each *P. falciparum* genotype cluster attributed to the malaria infection of each of the 789 survey participants.

Explanatory variables were latitude and longitude coordinates, elevation, distance to the coastline, and human population density. Quadratic terms for latitude and longitude were included to allow for spatial prediction of non-linear trends across the land surface of PNG. Location of survey locations was selected as an explanatory variable based on the assumption that geographic proximity was an important predictor of the spatial predominance of distinct parasite genotypes. Elevation and distance to coastline were selected as they are associated with the habitats of distinct *Anopheles* species in PNG, and with human movement as topography and elevation i.e. presence of mountain ranges may be barriers to human human mobility. Population density was included to attempt to exclude plasmodium genotype predictions being made for ecologically suitable areas, based on results of the model, but where population density is low and high spatial distribution of plasmodium genotypes unlikely. All variables were added to the model without implementing a variable reduction strategy. Statistical analysis was carried out using R open source software version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). Spatial predictions were made to a data file with 3769 data points representing a grid of georeferenced points covering the land mass of PNG. Elevation, population density and Euclidean distance to the coast, were extracted at each grid point. Spatial predictions of the proportion of each genotype cluster found at each grid point location were done using the fitted parameters of the Dirichlet regression model.

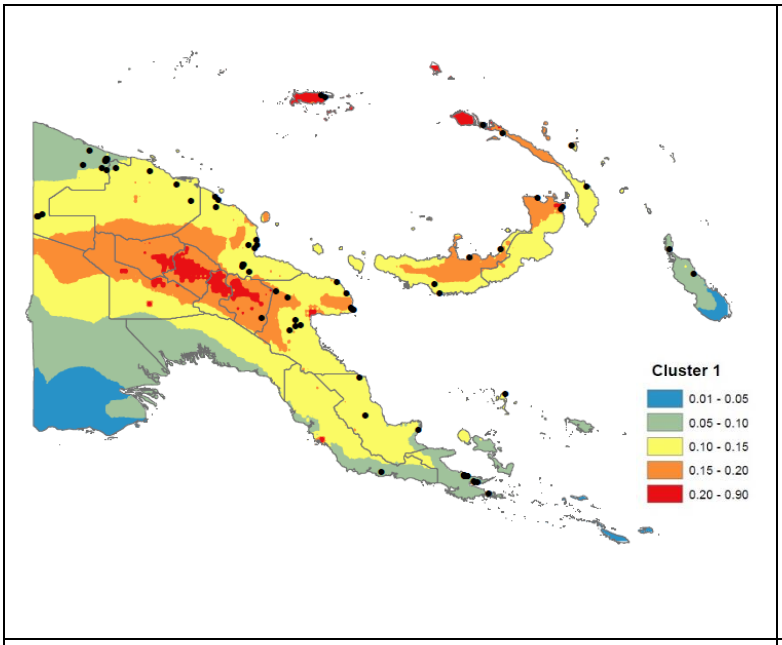
Model Validation

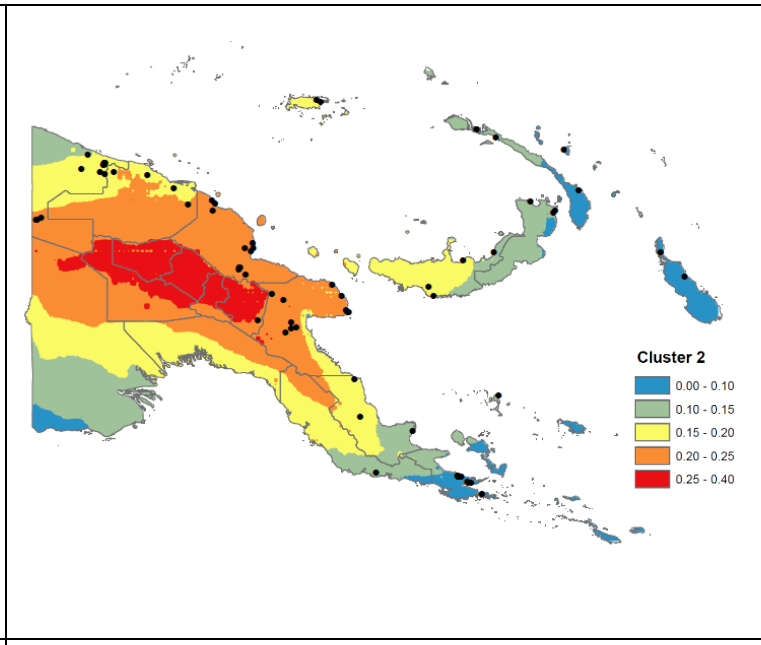
Cross validation methods using R statistical software were used to validate spatial predictions of the Dirichlet regression model for each genetically distinct cluster (Table 2). The dataset was split into a training subset (containing 70% of observations) and a test subset (containing

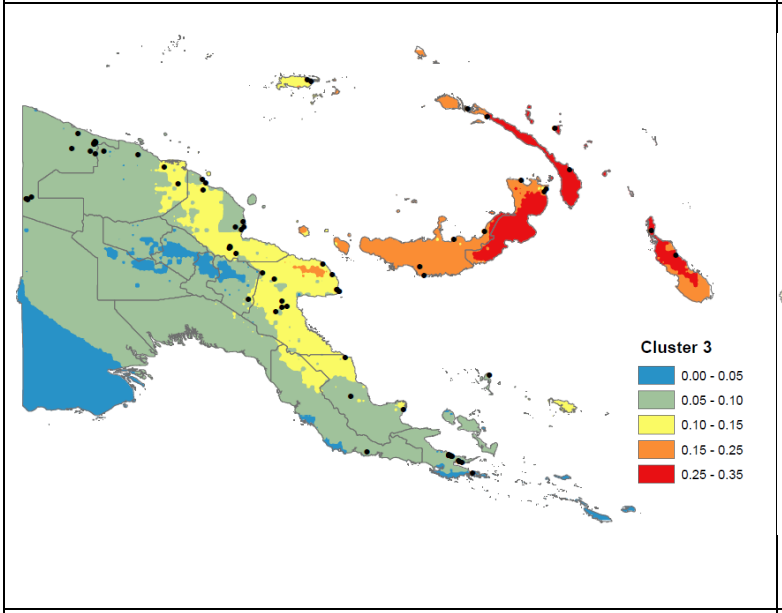
30% of observations) using random allocation. Dirichlet regression was carried out on the training dataset and predictions from the Dirichlet regression models were made for the test dataset and compared to the observed proportion of each genotype dichotomized relative to four different cut-points, 0.1, 0.25, 0.5 and median predominance. For clusters 4 and 8, the model coefficients were too small to use 0.25 and 0.5 as cut off points for cross validation and so these values were reduced by a magnitude of ten to produce reasonable validation results. Receiver operating characteristic (ROC) curves were then estimated to test the discriminatory performance of the predictions.

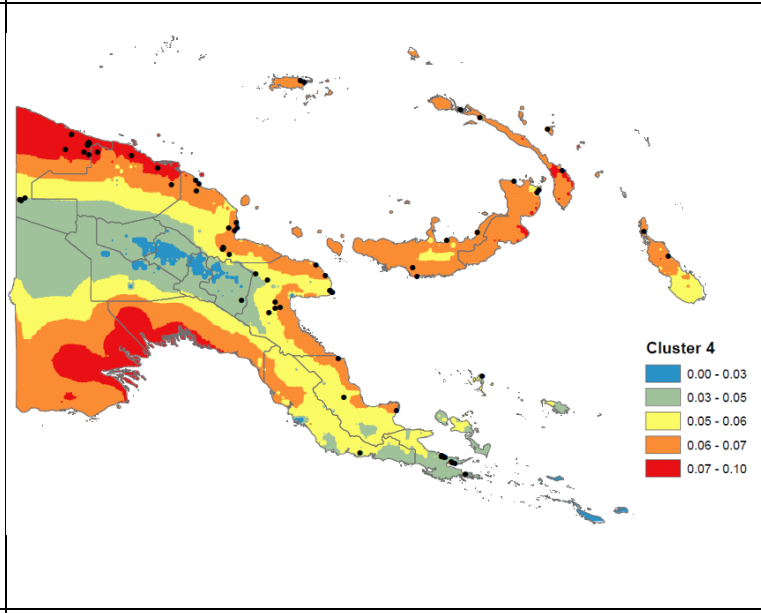
Table 1. Associations of *Plasmodium falciparum* genetic clusters with ecology covariates

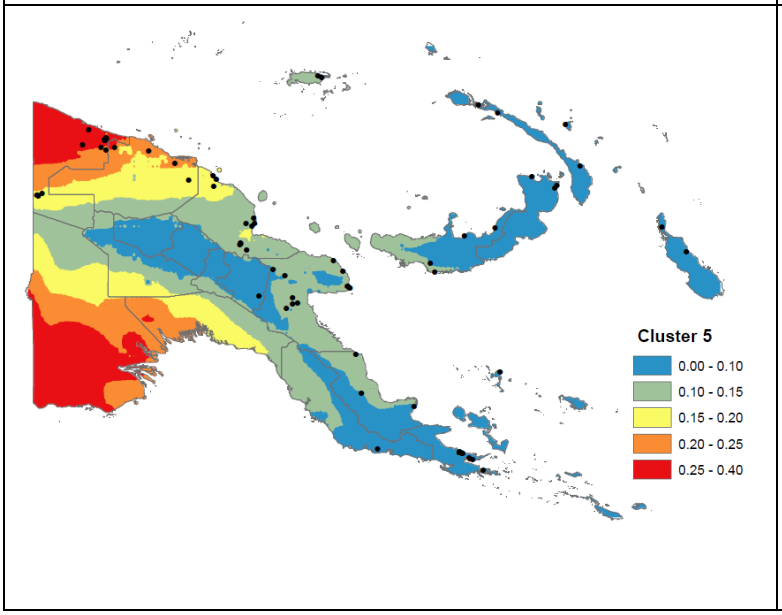
Cluster Coefficient								
Covariates	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Longitude	0.258***	-0.110*	0.485***	-0.071	-0.436***	-0.155**	0.122*	0.015
Latitude	0.156***	-0.020	0.135**	-0.016	-0.061	-0.066	-0.319***	-0.056
Longitude^2	0.018	-0.032	-0.074**	0.002	-0.013	-0.045	0.160***	-0.004
Latitude^2	-0.212***	-0.221***	-0.057	0.044	-0.099*	-0.163**	0.215***	-0.004
Elevation	0.001	0.001	0.001	0.001	0.001	0.001’	0.001	-0.001
Dist Coast	0.819***	0.305**	0.281	-0.141	-0.223	0.481**	0.090	0.086
Population	0.008**	0.009***	-0.011***	-0.003	-0.004	0.005**	0.007*	0.004
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’’0.1								

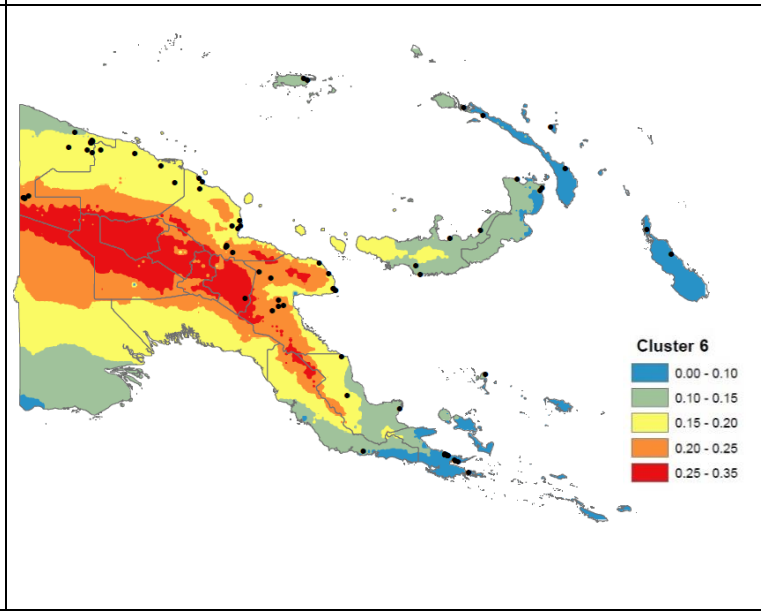

Cluster 1
0.01 - 0.05
0.05 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.90

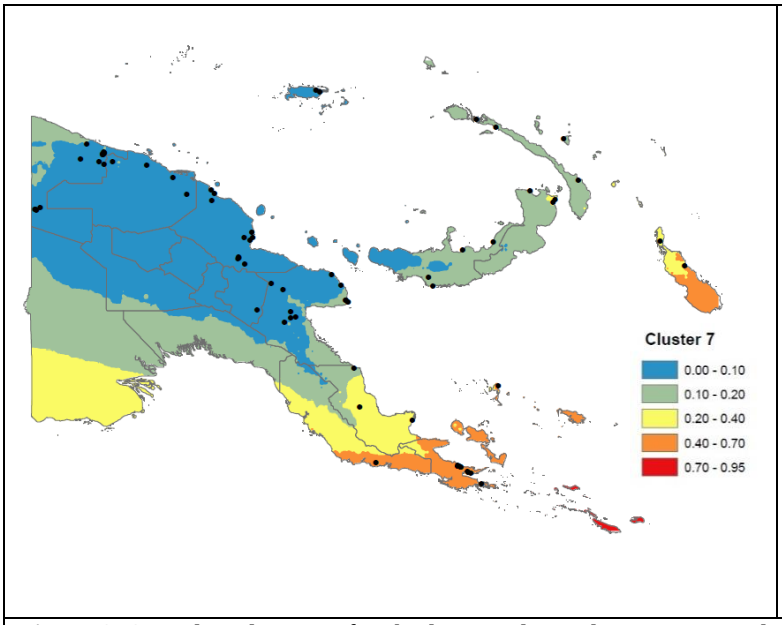

Cluster 2
0.00 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.25
0.25 - 0.40


Cluster 3
0.00 - 0.05
0.05 - 0.10
0.10 - 0.15
0.15 - 0.25
0.25 - 0.35


Cluster 4
0.00 - 0.03
0.03 - 0.05
0.05 - 0.06
0.06 - 0.07
0.07 - 0.10


Cluster 5
0.00 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.25
0.25 - 0.40


Cluster 6
0.00 - 0.10
0.10 - 0.15
0.15 - 0.20
0.20 - 0.25
0.25 - 0.35



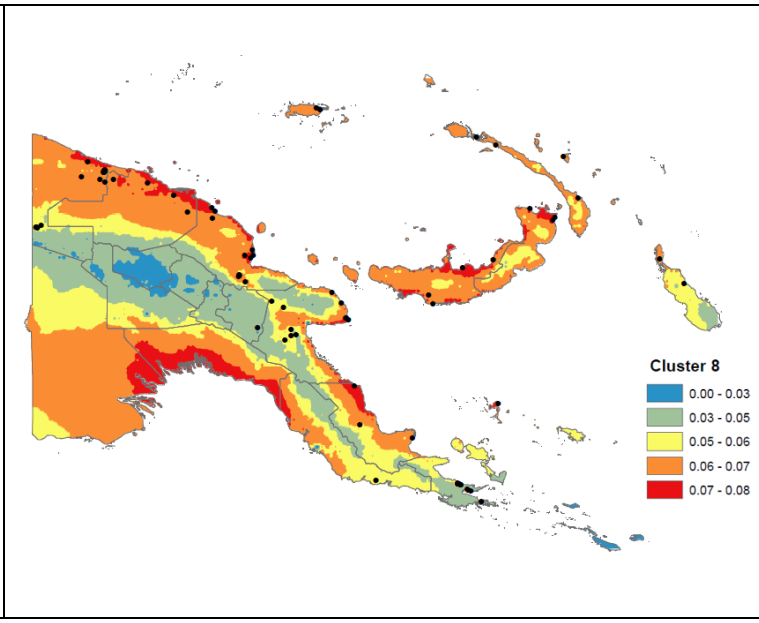


Figure 3. Spatial predictions of eight distinct plasmoduim genotype clusters based on the results of a Dirichlet regression model, Papua New Guinea.

Table 2. AUC values obtained from cross validation of eight *Plasmodium genotype* model coefficients against 4 different cut-off values

Training dataset: 70% of full dataset

Test dataset: Full dataset – training dataset = 30% of full dataset

	.1 cut off value AUC		.25 cut off//.025 cut off AUC		.5 cut off//.05 cut off AUC		Median cut off AUC	
	Train	Test	Train	Test	Train	Test	Train	Test
Cluster 1	.685	.639	.701	.649	.698	.682	.66	.647
Cluster 2	.697	.662	.685	.662	.65	.632	.695	.665
Cluster 3	.786	.855	.832	.908	.683	.908	.701	.695
Cluster 4*	.63	.601	.703	.706	.731	.822	.548	.52
Cluster 5	.776	.812	.817	.822	.864	.845	.765	.753
Cluster 6*	.742	.717	.721	.77	.723	.798	.73	.791
Cluster 7	.741	.781	.813	.821	.841	.905	.657	.719
Cluster 8	.589	.505	.597	.593	.639	.652	.537	.533

*Due to the small size of predicted predominance coefficients of clusters 4 and 8, the cut off values for model prediction validation were reduced by a fator of 10. The .25 and .5 cut off values for each other cluster were therefore changed to .025 and .05. The .1 cut off value remained unchanged.

References

1 World Health Organisation. *Rolling Back Malaria*, <https://www.who.int/whr/1999/en/whr99_ch4_en.pdf> (1999).

2 World Health Organization. World Malaria Report 2018. (Geneva, 2018).

3 Sturrock, H. J., Roberts, K. W., Wegbreit, J., Ohrt, C. & Gosling, R. D. Tackling imported malaria: an elimination endgame. *The American journal of tropical medicine and hygiene* **93**, 139-144 (2015).

4 Nyunt, M. H. *et al.* Molecular surveillance of artemisinin resistance falciparum malaria among migrant goldmine workers in Myanmar. *Malar J* **16**, 97, doi:10.1186/s12936-017-1753-8 (2017).

5 Roberts, L. Drug-resistant malaria advances in Mekong. *American Association for the Advancement of Science* **358**, 155-156 (2017).

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

7 Mobegi, V. A. *et al.* Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity. *Molecular biology and evolution* **31**, 1490-1499 (2014).

8 Kwiatkowski, D. Malaria genomics: tracking a diverse and evolving parasite population. *International health* **7**, 82-84 (2015).

9 Hemingway, J. *et al.* Tools and strategies for malaria control and elimination: what do we need to achieve a grand convergence in malaria? *PLoS biology* **14**, e1002380 (2016).

10 Barry, A. E., Waltmann, A., Koepfli, C., Barnadas, C. & Mueller, I. Uncovering the transmission dynamics of Plasmodium vivax using population genetics. *Pathogens and global health* **109**, 142-152 (2015).

11 Lo, E. *et al.* Frequent Spread of Plasmodium vivax Malaria Maintains High Genetic Diversity at the Myanmar-China Border, Without Distance and Landscape Barriers. *The Journal of Infectious Diseases*, jix106 (2017).

12 Razak, M. R. M. A. *et al.* Genetic diversity of Plasmodium falciparum populations in malaria declining areas of Sabah, East Malaysia. *PloS one* **11**, e0152415 (2016).

13 Daniels, R. *et al.* Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PloS one* **8**, e60780 (2013).

14 Hastings, I. & Wedgwood-Oppenheim, B. Sex, strains and virulence. *Parasitology Today* **13**, 375-383 (1997).

15 Auburn, S. & Barry, A. E. Dissecting malaria biology and epidemiology using population genetics and genomics. *International journal for parasitology* **47**, 77-85 (2017).

16 Khaireh, B. A. *et al.* Population genetics analysis during the elimination process of Plasmodium falciparum in Djibouti. *Malaria journal* **12**, 201 (2013).

17 Auburn, S. *et al.* Genomic analysis of a pre-elimination Malaysian Plasmodium vivax population reveals selective pressures and changing transmission dynamics. *Nature communications* **9**, 2585 (2018).

18 Nabet, C. *et al.* Genetic diversity of Plasmodium falciparum in human malaria cases in Mali. *Malaria journal* **15**, 353 (2016).

19 Preston, M. D. *et al.* A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. *Nature communications* **5** (2014).

20 Brasil, P. *et al.* Outbreak of human malaria caused by Plasmodium simium in the Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation. *The Lancet Global Health* **5**, e1038-e1046 (2017).

21 Wong, V. K. *et al.* An extended genotyping framework for Salmonella enterica serovar Typhi, the cause of human typhoid. *Nature communications* **7**, 12827 (2016).

22 Mdladla, K., Dzomba, E. & Muchadeyi, F. Landscape genomics and pathway analysis to understand genetic adaptation of South African indigenous goat populations. *Heredity* **120**, 369 (2018).

23 Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M. & Holderegger, R. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* **24**, 4348-4370 (2015).
24 Cornuault, J. *et al.* The role of ecology in the geographical separation of blood parasites infecting an insular bird. *Journal of Biogeography* **40**, 1313-1323 (2013).
25 Escalante, A. A. *et al.* Malaria molecular epidemiology: lessons from the International Centers of Excellence for Malaria Research Network. *The American journal of tropical medicine and hygiene* **93**, 79-86 (2015).
26 Hetzel, M. W. *et al.* Progress in mosquito net coverage in Papua New Guinea. *Malaria journal* **13**, 242 (2014).
27 Hetzel, M. W. *et al.* Prevalence of malaria across Papua New Guinea after initial roll-out of insecticide-treated mosquito nets. *Tropical Medicine & International Health* **20**, 1745-1755 (2015).
28 Fola, A. A. *et al.* Higher complexity of infection and genetic diversity of Plasmodium vivax than Plasmodium falciparum across all malaria transmission zones of Papua New Guinea. *The American journal of tropical medicine and hygiene* **96**, 630-641 (2017).
29 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).
30 starmie: Population Structure Model Inference and Visualisation. v. R package version 0.1.2. (2016).
31 Manske, M. *et al.* Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **487**, 375 (2012).
32 Teboh-Ewungkem, M. I., Mohammed-Awel, J., Baliraine, F. N. & Duke-Sylvester, S. M. The effect of intermittent preventive treatment on anti-malarial drug resistance spread in areas with population movement. *Malaria journal* **13**, 428 (2014).
33 Daniels, R. F., Rice, B. L., Daniels, N. M., Volkman, S. K. & Hartl, D. L. The utility of genomic data for Plasmodium vivax population surveillance. *Pathogens and global health* **109**, 153-161 (2015).
34 Chang, H.-H. *et al.* The geography of malaria elimination in Bangladesh: combining data layers to estimate the spatial spread of parasites. *bioRxiv*, 421578, doi:10.1101/421578 (2018).
35 Tessema, S. *et al.* Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. *Elife* **8**, e43510 (2019).
36 Joy, D. A. *et al.* Early origin and recent expansion of Plasmodium falciparum. *science* **300**, 318-321 (2003).
37 Molina-Cruz, A. & Barillas-Mury, C. The remarkable journey of adaptation of the Plasmodium falciparum malaria parasite to New World anopheline mosquitoes. *Memórias do Instituto Oswaldo Cruz* **109**, 662-667 (2014).
38 Cooper, R. *et al.* Malaria vectors of Papua New Guinea. *International journal for parasitology* **39**, 1495-1501 (2009).
39 Patel, S. S., King, C. L., Mgone, C. S., Kazura, J. W. & Zimmerman, P. A. Glycophorin C (Gerbich antigen blood group) and band 3 polymorphisms in two malaria holoendemic regions of Papua New Guinea. *American journal of hematology* **75**, 1-5 (2004).
40 Wood, J. *et al.* The genetic demography of the Gainj of Papua New Guinea. I. Local differentiation of blood group, red cell enzyme, and serum protein allele frequencies. *American journal of physical anthropology* **57**, 15-25 (1982).
41 Ross, M. Pronouns as a preliminary diagnostic for grouping Papuan languages. *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*, 15-65 (2005).
42 Schultz, L. *et al.* Multilocus haplotypes reveal variable levels of diversity and population structure of Plasmodium falciparum in Papua New Guinea, a region of intense perennial transmission. *Malaria journal* **9**, 336 (2010).
43 Chenet, S. M., Schneider, K. A., Villegas, L. & Escalante, A. A. Local population structure of Plasmodium: impact on malaria control and elimination. *Malaria journal* **11**, 412 (2012).
44 Roh, M. E. *et al.* High Genetic Diversity of Plasmodium falciparum in the Low-Transmission Setting of the Kingdom of Eswatini. *The Journal of infectious diseases* **220**, 1346-1354 (2019).

45 Pindolia, D. K. *et al.* Human movement data for malaria control and elimination strategic planning. *Malaria journal* **11**, 205 (2012).
46 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).
47 Wesolowski, A. *et al.* Mapping malaria by combining parasite genomic and epidemiologic data. *BMC medicine* **16**, 190 (2018).
48 Crellen, T. *et al.* Whole genome resequencing of the human parasite Schistosoma mansoni reveals population history and effects of selection. *Scientific reports* **6**, 20954 (2016).
49 Taylor, A. R. *et al.* Quantifying connectivity between local Plasmodium falciparum malaria parasite populations using identity by descent. *PLoS genetics* **13**, e1007065 (2017).
50 Noor, A. M. *et al.* Mapping the receptivity of malaria risk to plan the future of control in Somalia. *BMJ open* **2**, e001160 (2012).
51 Hetzel, M. W. *et al.* Prevalence of malaria across Papua New Guinea after initial roll-out of insecticide-treated mosquito nets. *Tropical medicine & international health : TM & IH* **20**, 1745-1755, doi:10.1111/tmi.12616 (2015).
52 Hetzel, M. W. *et al.* Ownership and usage of mosquito nets after four years of large-scale free distribution in Papua New Guinea. *Malar J* **11**, 192, doi:10.1186/1475-2875-11-192 (2012).
53 McNamara, D. T. *et al.* Diagnosing infection levels of four human malaria parasite species by a polymerase chain reaction/ligase detection reaction fluorescent microsphere-based assay. *Am J Trop Med Hyg* **74**, 413-421 (2006).
54 Falk, N. *et al.* Comparison of PCR-RFLP and Genescan-based genotyping for analyzing infection dynamics of Plasmodium falciparum. *Am J Trop Med Hyg* **74**, 944-950 (2006).
55 Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology* **14**, 2611-2620 (2005).
56 NASA. *Earthdata*, <https://earthdata.nasa.gov/> (2017).
57 Worldpop. <https://www.worldpop.org/> (2017).

Appendix 5

Supporting files

Supplementary table S1 contains summaries of the proportion of isolates within each geographic region that have >=0.75 ancestry in each cluster.

Table S1. Prevalence of isolates with unmixed ancestry in each geographic region. Shading indicates the prevalence level (darker colours = higher).

			Proportion of isolates with ancestry >=0.75								
Geographic area	Labels	n	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Total
West Sepik/Papua	1	17	0.18	0.00	0.00	0.00	0.12	0.06	0.00	0.00	0.35
West Sepik inland (YAM,TAB)	4.5	11	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.09
East Sepik (DRE,Sentinel site)	4.1	56	0.00	0.02	0.00	0.00	0.38	0.04	0.00	0.00	0.43
East Sepik inland (Maprik WGS)	4	47	0.00	0.00	0.00	0.17	0.40	0.11	0.00	0.00	0.68
West Sepik coast (SIA)	2	15	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.27
East Sepik north coast (PAN,WIA)	3	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
East Sepik south coast (KAR,ORE)	5	5	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.20
Madang north coast (ZOG,WAZ)	6	33	0.09	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.24
Madang (ORD,BAF)	7	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Madang town (WGS)	7.1	24	0.00	0.08	0.00	0.00	0.00	0.17	0.00	0.00	0.25
Madang valley (KES,MAO)	9	78	0.00	0.03	0.03	0.00	0.09	0.14	0.00	0.00	0.28
EHP (WON)	11.2	1	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
Morobe valley/EHPborder (WAR,NGA,ABO)	10	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Morobe (MUM,Sentinel site)	11	87	0.06	0.02	0.01	0.00	0.06	0.03	0.00	0.00	0.18
Morobe (BUN)	11.1	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Morobe (FIN,Sentinel site;GIT,AGO,SIU,GOD	13	87	0.13	0.14	0.00	0.00	0.01	0.03	0.06	0.11	0.48
Northern/Oro (KEN,FOR,MAR)	16	12	0.17	0.08	0.00	0.00	0.08	0.00	0.00	0.00	0.33
Central	17	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Manus Island	19	16	0.56	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.75
WNB north coast (BNL,EWA)	20	43	0.19	0.00	0.12	0.00	0.00	0.00	0.09	0.00	0.40
WNB south coast (SIM,KUL)	21	55	0.13	0.00	0.36	0.00	0.00	0.00	0.09	0.00	0.58
ENB	22	22	0.00	0.00	0.14	0.00	0.00	0.00	0.14	0.00	0.27
New Ireland	23	57	0.05	0.00	0.37	0.00	0.00	0.04	0.14	0.00	0.60
Bougainville	24	2	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.50
Milne Bay	25	1	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00
Milne Bay (Alotau WGS)	26	30	0.00	0.00	0.00	0.00	0.00	0.00	0.73	0.00	0.73
Milne Bay outer islands (MUT)	27	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table S2 contains data on the geolocation, environmental covariate data and population size of the village locations where the survey was carried out, as well as the number of people in each village who were recruited as part of the survey (sample no.)

Table S2. Summary of sample numbers, geographic location, GPS co-ordinates and environmental data for each village location where survey was carried out from 16 provinces in PNG between October 2008 and August 2009

Province	Village	Latitude	Longitude	Distance to coast	Elevation	Population size	Sample no.
Central	Abau	-10.11	148.46	0.086	32	5	2
East_New_Britain	Vumarita	-4.21	151.80	0.002	75	25	22
East_Sepik	Ilahup	-3.39	142.57	0.145	705	8	47
	Kamakor	-3.42	142.55	0.183	858	8	2
	Dreikikir	-3.58	142.77	0.268	322	3	56
	Oremai	-4.29	144.37	0.314	15	2	5
Eastern Highlands Province	Wonera	-6.80	145.89	1.047	1775	10	2
Madang	Zogari	-4.20	144.90	0.001	16	25	33
	Matupi	-5.23	145.79	0.003	12	52	24
	Bafulu	-5.23	145.61	0.160	80	27	2
Manus	Warabei	-2.03	147.17	0.031	83	12	16
Milne_Bay	Ahioma	-10.19	150.31	0.005	374	13	30
	Doma	-10.56	150.75	0.001	11	10	2
	Mutawa	-8.43	151.12	0.011	14	32	2
Morobe	Ngariawang	-6.35	146.45	0.598	632	7	3
	Finschhafen	-6.61	147.85	0.002	202	33	78
	Bundun	-6.84	146.62	0.326	835	6	2
	Gwasak	-7.07	146.49	0.467	913	7	87
New_Ireland	Butei	-2.67	150.64	0.003	16	7	57
Northern_Oro	Kendata	-8.89	148.11	0.325	546	6	12
West_New_Britain	Ewasse	-5.32	151.01	0.004	52	21	43
	Simimla	-6.08	149.60	0.127	311	4	55
West_Sepik	Siaute_No2	-3.20	142.20	0.094	86	8	15
	Tabale	-3.53	142.07	0.441	381	8	11
	Skonga	-4.57	141.20	1.735	386	2	17

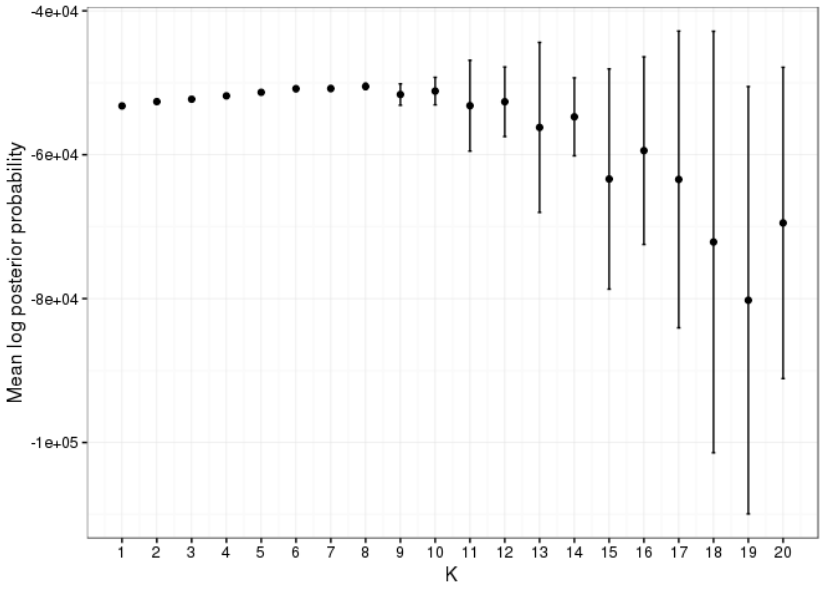


Figure S1. Mean log posterior probability for each K in the STRUCTURE analysis.

We estimated optimal K by simply plotting the log posterior probability of the data for each K and look for the inflection point in the curve. STRUCTURE computes the estimated log posterior probability by taking the ratio of the mean estimated log-likelihood of the data and the estimated variance of the log-likelihood of the data over all MCMC chains.

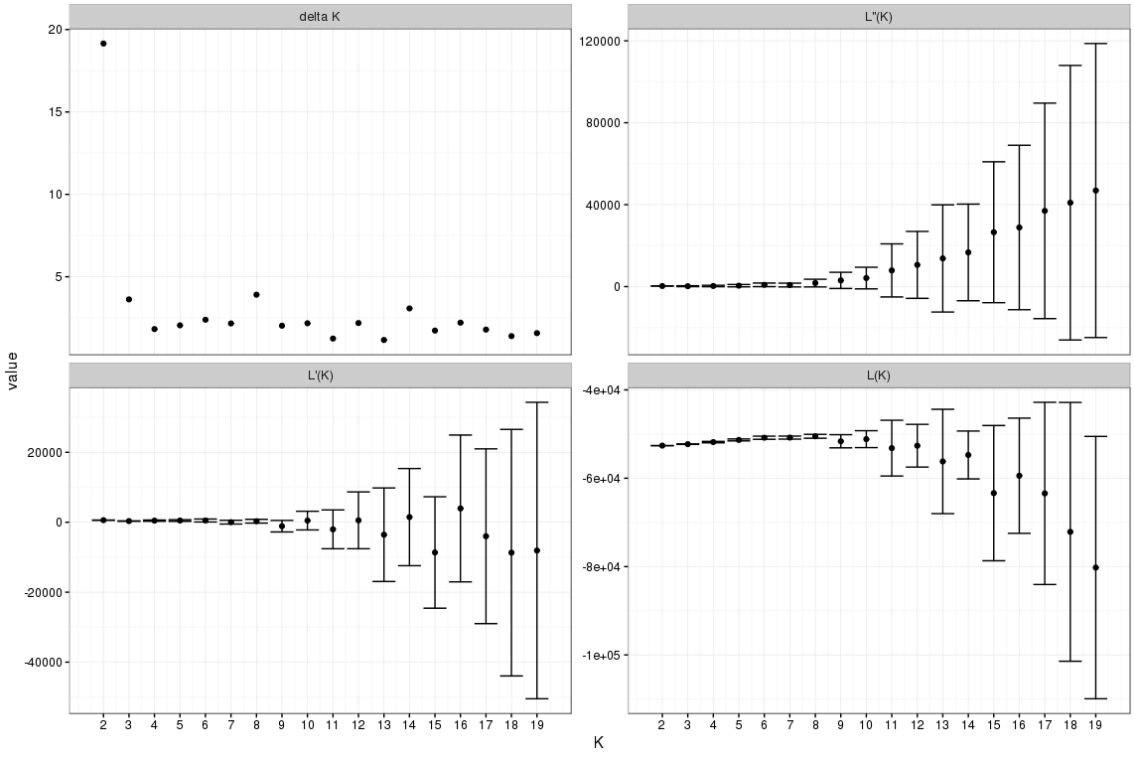


Figure S2. DeltaK estimates for each K in the STRUCTURE analysis.

The Evanno method[55] (REF) was used to measure the change in log-likelihood for each K over all runs. Note the peak at K=8.

CHAPTER 6

Examining the impact of human mobility on malaria resurgence using a Ross-Macdonald meta-population model and the example of Solomon Islands

CHAPTER 6: EXAMINING THE IMPACT OF HUMAN MOBILITY ON MALARIA RESURGENCE USING A ROSS MACDONALD META-POPULATION MODEL AND THE EXAMPLE OF SOLOMON ISLANDS

CONTEXT

In chapter 6 we examine how human mobility and sustained or relaxed vector control through LLIN use and IRS may impact malaria infection resurgence in Nggela, a small island group in Solomon Islands, which is well connected to Guadalcanal, a larger island where malaria is endemic. Estimating time to resurgence, and magnitude of resurgent infection prevalence under sustained or relaxed vector control in an eliminating area that remains connected to an area endemic for malaria is important for malaria programmes to determine which vector control interventions should be prioritised for use post-elimination. Estimating likelihood of resurgence after cessation of control efforts for guiding surveillance operations is also important for guiding surveillance efforts in areas where local transmission has ceased but where importation of infection remains a risk.

This chapter uses a Ross-Macdonald model parameterised with data on malaria infection from Guadalcanal, assuming local transmission in Nggela has ceased, simulating chance of resurgence given continued migration between both islands. We hypothesised that malaria transmission between both islands may be occurring, facilitated by human mobility, given that malaria infections attributed to clonal (i.e. identical) parasites were detected during a malaria indicator survey conducted in 2008/2009. Model simulations were run under different scenarios of migration rate, infection prevalence in Guadalcanal, human biting rate and vector abundance. The outcomes assessed were estimated time to, and magnitude, of resurgent infection peak in Nggela. From model simulations we were able to determine which model parameters had greatest impact on infection peak and from this we inferred which interventions

may be most beneficial in the prevention of resurgence of malaria in Nggela, should elimination be achieved while transmission is ongoing in Guadalcanal. This manuscript is formatted for submission to the journal Epidemics.

CHAPTER 6: EXAMINING THE IMPACT OF HUMAN MOVEMENT ON MALARIA RESURGENCE USING A ROSS MACDONALD META-POPULATION MODEL AND THE EXAMPLE OF SOLOMON ISLANDS

Cleary, E[1]; Alyssa Barry[2,3] ; McCaw J[4,5,6]; Yakob L[7], Clements, A.C.A[1,8].

1. Department of Global Health, the Research School of Population Health, the Australian National University, Canberra, Australia
2. Population Health and Immunity Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
3. Department of Medical Biology, University of Melbourne, Parkville, Victoria, Australia
4. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Australia
5. School of Mathematics and Statistics, The University of Melbourne, Parkville, Australia.
6. Victorian Infectious Diseases Reference Laboratory Epidemiology Unit, Peter Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne, Parkville, Australia
7. Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine
8. Faculty of Health Sciences, Curtin University, Bentley, Western Australia, Australia

Abstract

Introduction

As progress is made toward malaria elimination, connectivity between areas of different levels of transmission via human mobility needs to be better understood to avoid resurgence of infection where and when elimination has been achieved. Here we perform a model based assessment on the impact of population mobility between a location where malaria

transmission has been interrupted, represented by a small island group in Solomon Islands, and another location (a larger island), where transmission is ongoing, on risk of malaria resurgence in the malaria-eliminated location.

Methods

A Ross-Macdonald metapopulation model was parameterised using malaria prevalence data from Guadalcanal Island and data pertaining to vector biology and malaria transmission dynamics from published research. Simulations of the model were run with varying estimates of rate of human mobility, human biting rate and vector abundance to elucidate which model parameters would have greatest impact on malaria resurgence in Nggela (connected to Guadalcanal by human mobility), in the event that local transmission has been interrupted in Nggela.

Results

An increasing rate of human migration between Guadalcanal and Nggela had the greatest impact on malaria resurgence in Nggela. Increasing biting rate, vector abundance and infection prevalence in Guadalcanal also increased the number of infections in Nggela at infection peak indicating that maintaining vector control interventions whilst malaria transmission is ongoing in Guadalcanal is necessary to prevent malaria resurgence in Nggela.

Conclusion

In order for countries to maintain their malaria-free status post elimination, surveillance aimed towards detecting imported infections and vector control interventions must be maintained while post-elimination areas remain connected to areas where transmission is ongoing.

6.1 Introduction

One of the major obstacles to achieving global malaria elimination will be understanding the role that human mobility and migration plays in facilitating ongoing transmission and reintroduction of malaria post-elimination[1]. Introduction of malaria parasites into populations where vectorial capacity remains, but background immunity has waned following extended time periods with no exposure, carries the potential for resurgence in a relatively short period of time[2,3]. Parasite re-introduction may occur via migration of asymptomatic but infectious individuals to an area where local transmission has ceased, or migration of susceptible individuals to a location where transmission is ongoing, before subsequently returning to the area vulnerable to reintroduction[4,5,6]. Indeed, many pre-elimination and eliminating countries have reported a decrease in locally acquired malaria infection but a stable rate of imported cases as they approach elimination[1,7,8]. Understanding the contribution of human movement to sustained malaria transmission between endemic and eliminating areas will be crucial for planning malaria surveillance operations and allocation of vector control resources, such as long lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS), in the prevention of resurgent malaria epidemics[9-11]. Assessing the risk of resurgence and determining when to cease vector control following elimination, and which interventions to prioritize in the post-elimination period, can however be challenging[12].

In the Nggela Islands, a small group of islands in the Central Province of Solomon Islands, transmission of malaria is low with presence of some asymptomatic *P. falciparum* infections on the island[13]. It is unclear whether local transmission of *Plasmodium falciparum* has ceased or if parasites are being re-introduced by incoming travellers or returning residents from areas with higher *P. falciparum* burden[13]. Parasite phylogeny carried out on samples collected in

2008/2009 (unpublished data, Harrison in prep) show transmission attributed predominantly to a clonal parasite genotype on the island, which has been found by phylogenetic analysis to be present in the much larger, neighbouring island, Guadalcanal, where malaria is endemic and transmission is ongoing. Presence of this clonal parasite on both islands provided justification for the hypothesis of malaria transmission occurring between Nggela and Guadalcanal as a result of human mobility between both islands. Guadalcanal has a high incidence of malaria relative to many of the other islands in the archipelago, with three-quarters of infections found in the northern part of the island, most proximal to Nggela[14]. Guadalcanal is well connected to Nggela by a ferry service and private motorized boats[13]. Presence of this same parasite genotype on both islands suggests malaria transmission occurring between both Nggela and Guadalcanal, facilitated by human movement, although it is not clear from observational data in which direction transmission may be occurring.

In this study, we aimed to examine the impact of connectivity via human migration between Nggela and Guadalcanal on malaria transmission in Nggela by implementing a bi-directional metapopulation Ross-Macdonald model based on the model developed by Acevedo et al[15]. The metapopulation model simulates transmission dynamics in two patches which are connected because of contact or mobility between distinct populations[2,15]. The Ross-Macdonald model describes the cyclical transmission of the *Plasmodium* parasite between mosquitoes and humans. The model incorporated parameters relating to mosquito biting rate, the proportion of bites that produce an infection in humans and mosquitoes, and the extrinsic incubation period of the parasite within the mosquito. Interventions can be introduced into the model, whereby they reduce the basic reproduction number, R_0, which represents the average number of new cases arising due transmission of infection from an index case[16,17] .

We simulated human mobility under varying human biting rates, relative vector abundance, infection prevalence in Guadalcanal and rates of human migration between both islands, under

the assumption that local transmission in Nggela has ceased and initial infection prevalence was zero, in order to gain insight into how movement of people between Guadalcanal and Nggela may affect malaria transmission dynamics in Nggela once elimination is achieved. In doing so, we aimed to examine the conditions under which human mobility impacts malaria transmission in elimination settings, particularly those that are vulnerable to (re-)introduction due to close proximity to high-transmission areas. The specific objectives of this paper were to estimate:

1. To what extent human migration impacts on risk of (re)introduction of malaria into a setting in Solomon Islands where transmission has been interrupted.
2. What length of time might it take for resurgence to occur if vector control interventions in Nggela have ceased.
3. The effect that different rates of human migration between Nggela and Guadalcanal may have on malaria transmission.
4. To what extent different levels of vector control and vector abundance impact malaria transmission in Nggela, given sustained human migration between both islands.

6.2 Methods

Study site

Solomon Islands is an archipelago of islands situated to the north east of Australia, bordering Papua New Guinea (PNG) to the west. It is made up of nine island provinces, with the capital, Honiara, located in the northern part of Guadalcanal, and Nggela islands situated in Central province. Guadalcanal and Nggela are designated the terms patch 1 and patch 2 when describing the mathematical model (Figure 1)[18].

Ross-Macdonald metapopulation model

We used a Ross-Macdonald two-patch metapopulation model to examine the impact of frequent human migration between the islands of Nggela (patch 1) and Guadalcanal (patch 2; Figure 1). Varying values were used for relative vector abundance and human biting rate, reflecting different degrees of vector control, as well as differing rates of migration and infection prevalence in patch 2.

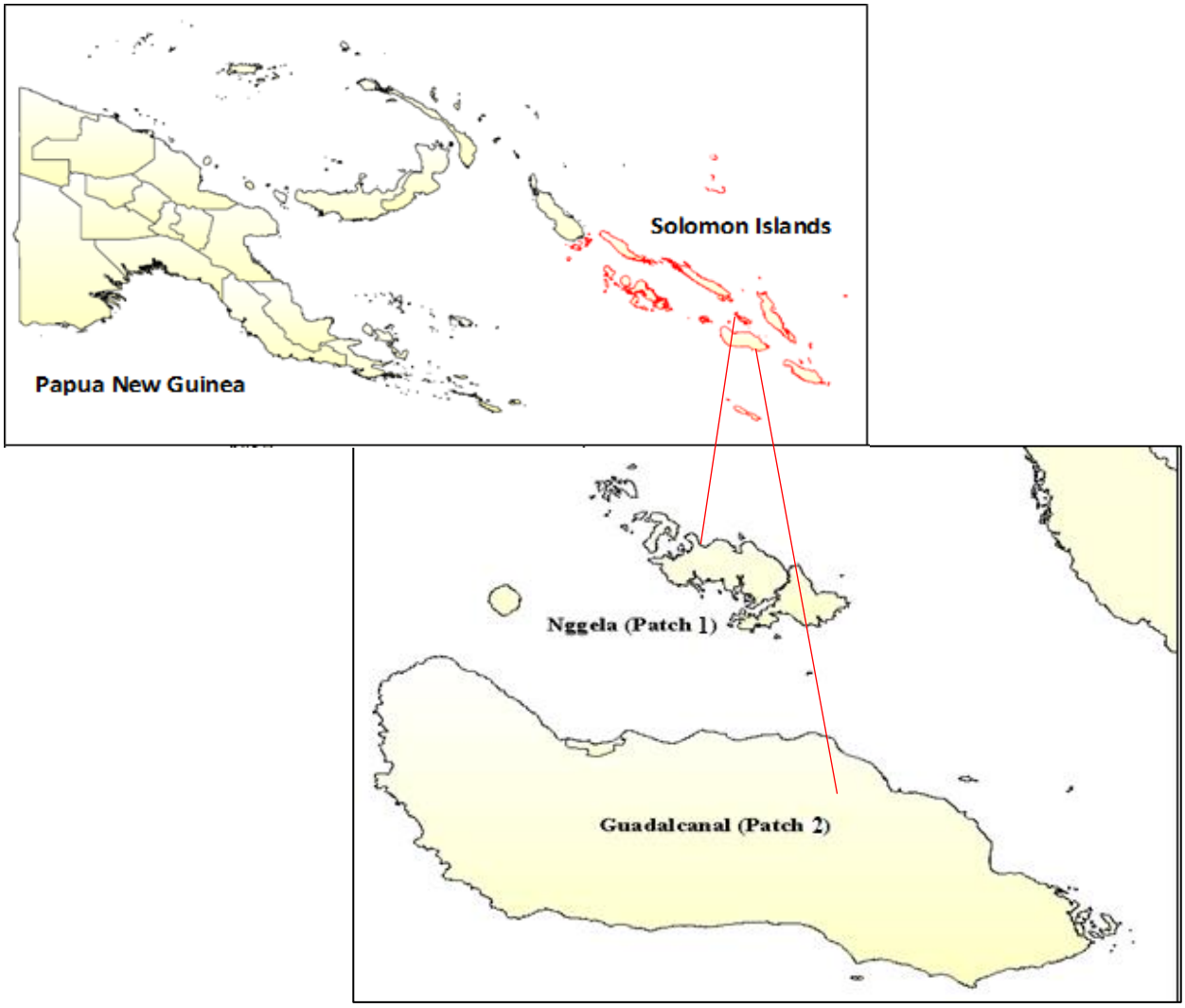


Figure 1. Geographic location of Solomon Islands, Guadalcanal (patch 1) and Nggela (patch 2).

We assumed heterogeneous transmission of malaria between the two distinct patches[15], but that characteristics of individuals and likelihood of infection in both populations was homogeneous, consistent with previous metapopulation modelling work[15,19,20]. Within each patch, we also

assumed that the population was well mixed, and that each individual had an equal chance of coming into contact with infectious vectors[21,22].

The models implemented to run these simulations were based on models developed by Acevedo et al.[15], with infection parameters adjusted to local transmission dynamics in Solomon Islands. In each patch, susceptible individuals (S_h) are infected by infectious vectors (I_v) and susceptible vectors (total vector population, N_v – infectious vectors, I_v) are infected by infectious humans (I_h). The subscripts $_1$ and $_2$ denote humans and vectors in patch 1 and 2 respectively, while K_{12} and K_{21} denote travel between each patch (1 to 2 and vice versa). The model also assumes that susceptible vectors in patch 1 are infected by infectious individuals who travel from patch 2 to 1, and susceptible humans who travel from patch 1 to 2 are infected by infectious vectors in patch 2. The differential equations used to estimate transition between different compartments in the model are as follows:

Differential equations for the Ross-McDonald metapopulation model

Nggela Patch 1 human (*h*) & vector (*v*) model

$$dI_{v1}/dt = ac(I_{h1} + k_{21}I_{h2})(exp(-\mu n))((N_{v1}-I_{v1})/N_{v1})) - \mu I_{v1};$$

$$dS_{h1}/t = -mabI_{v1}(S_{h1}/N_{h1}) + (\Upsilon I_{h1});$$

$$dI_{h1}/dt = mabI_{v1}(S_{h1}/N_{h1}) - (\Upsilon I_{h1});$$

Guadalcanal Patch 2 human (*h*) & vector (*v*) model

$$dI_{v2}/dt = ac(I_{h2} + k_{12}I_{h1})(exp(-\mu n))((N_{v2}-Iv2)/N_{v2}) - \mu I_{v2};$$

$$dS_{h2}/dt = -mabI_{v2}(S_{h2}/N_{h2}) + (\Upsilon I_{h2});$$

$$dI_{h2}/dt = mabI_{v2}(S_{h2}/N_{h2}) - (\Upsilon I_{h2});$$

Table 1. Ross-Macdonald metapopulation model parameters		
a	Human biting rate[15]	0.1 (bites per mosquito per day)

b	Proportion of bites that produce an infection in humans[15,27]	0.1 (probability)
c	Proportion of bites that produce an infection in mosquitoes[15,28]	0.214 (probability)
μ	Per capita rate of mosquito mortality[15,29]	0.167 (probability of mosquito dying per day)
n	Extrinsic incubation period of parasite in mosquitoes[15,25]	10 days
Υ	Average human recovery rate[15]	0.0067 $days^{-1}$
m	Ratio of female mosquitos to humans[15]	10 (relative vector abundance)
I_{h1}	Proportion of infectious humans, Nggela (patch 1)	0.0 (0% prevalence)
I_{h2}	Proportion of infectious humans, Guadalcanal (patch 2) [13]	0.096 (9.6% prevalence)
N_{h1}	Total human population size, Nggela (patch 1) [24]	26051 people
N_{h2}	Total human population size, Guadalcanal (patch 2) [24]	158222 people
K_{12}	Proportion of total population travelling from patch 1 to 2 who are likely to be infections	0.1 (infectious migratory population)
K_{21}	Proportion of total population travelling from patch 2 to 1 who are likely to be infections	0.1 (infectious migratory population)

We modelled the effect of migration between Nggela (patch 1) and Guadalcanal (patch 2) under the assumption that local transmission in Nggela has ceased and that the percentage of infectious humans is zero in order to estimate the level of vector control which needs to be maintained in Nggela in order to prevent resurgence of transmission. Fixed values for model parameters pertaining to vector and disease biology were used in the model (Table 1). All simulations were run using the deSolve package[23] in R open source software version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria).

Simulations from the model were run under the following scenarios:

1. Zero infection prevalence in patch 1 with connectivity between patch 1 and patch 2 quantified using varying migration rates.
2. Zero infection prevalence in patch 1 with varying migration rates between patch 1 and patch 2 and varying values of human biting rate.

3. Zero infection prevalence in patch 1 with varying migration rates between both islands and varying vector abundance in both patches.
4. Zero infection prevalence in patch 1 with varying migration rates between both islands and varying infection prevalence in patch 2.

Fixed model parameters

Parameters related to vector biology and disease transmission dynamics were obtained from methods published by Acevedo et al.[15]. The total human population (N_{h1}) in Nggela (patch 1) was 26,051 and the total human population size (N_{h2}) in Guadalcanal (patch 2) was 158,222 in accordance with data recorded in the 2009 Solomon Islands census[24] (noting that the current population estimate for both islands is significantly higher, but that 2009 provides the last accurate and verifiable estimate). In both islands, the extrinsic incubation period of the *P. falciparum* parasite within the mosquito (n) was estimated to be 10 days[15,25], the proportion of bites producing parasite infection in mosquitoes (c) was estimated at 0.214, and the proportion of bites estimated to produce infection in humans (b) was 0.1. The average human recovery rate (γ), assumed to be the same on both islands, was 0.67% of infected population per day[15] and the average mosquito mortality rate (μ) was 16.7% per day. In this model, both susceptible and infectious mosquito mortality rates are included to account for a proportion of the mosquito population dying before becoming infectious.

Model simulations with variable model parameters

P. falciparum infection prevalence (I_{h2}) in Guadalcanal (patch 2) was initially set at 96 infections per 1000 population[13] (9.6%). In Nggela (patch 1) infection prevalence (I_{h1}) was set at 0% to examine the likelihood of resurgence under the assumption that local transmission has ceased. Human biting rate (a) was initially set at 0.1 bites per mosquito per day and initial relative vector abundance (m) was set at a ratio of 10 female mosquitoes to humans. Migration

rates were expressed as the proportion of the population travelling daily between each patch who are likely to be infectious. As the proportion of asymptomatic infections in areas where malaria remains endemic is high such as patch 2, it is likely that asymptomatic yet infectious individuals are likely to continue to travel. In low transmission areas such as patch 1, parasitemia is more likely to result in clinical infection and therefore infectious individuals are less likely to travel[26]. Model simulations were run with the parameter settings as described above.

Sensitivity analysis with variable model parameters

The sensitivity of time to infection peak, magnitude of infection peak, and number of people infected 100 days, 250 days and 500 days following introduction of infection to Nggela was assessed under varying rates of model parameters migration rates, human biting rates, vector abundance and infection prevalence in Guadalcanal. The impact on infection in Nggela, was examined using infection prevalence in Guadalcanal ranging from I_{h2} = 0.096 to I_{h2} = 0.2 and migration rate between patch 1 and 2 varying from 0.1 to 0.5.

Simulations were run with relative vector abundance (m) varying from a ratio of female mosquitoes to humans of m = 10 to m = 15 to m = 20 in both patches, to examine the impact of IRS control interventions on mosquito abundance on infection prevalence in Nggela. Sensitivity analyses also examined the impact on infection when the model was parameterised with a 50% increase in human biting rate (a = 0.15 bites per mosquito per day) and a 50% decrease in human biting rate (a = 0.05 bites per mosquito per day) to examine the impact of intensified and relaxed LLIN use in Nggela, respectively. Further sensitivity analyses were also run examining impact of varying the human biting rate from a = 0.01 to a = 0.3, and migration rate from 0.01 to 0.5 on magnitude of peak infection, to examine impact of biting rate and migration rate with all other model parameters fixed.

6.3 Results

Under rates of migration between Nggela and Guadalcanal of k_{12} = 0.1 and k_{21} = 0.1, with human biting rate a = 0.1, vector abundance of m = 10 in both locations, and infection prevalence in Guadalcanal I_{h2} = 0.096, 1.15% of the population were infected within approximately eight months of initial infection importation to Nggela (infection peak = 301 people at day 238; Figure 2). A 50% increase in vector abundance (m = 15) in both Nggela and Guadalcanal resulted in a 100% increase in number of people infected at infection peak in Nggela (Figure 3), within a slightly longer timeframe (infection peak = 602 people at day 322).

A 50% increase in biting rate (a=.015) also resulted in an increase in number of people infected at peak infection and increase in time to infection peak (infection peak = 1706 people at day 500; Figure 4). Increasing the migration rate between Nggela and Guadalcanal by a factor of five (k_{12} = 0.5; k_{21} = 0.5) resulted in an increase in the number of people infected at infection peak to 1453, with a slight decrease in time to infection peak (237 days; Figure 5). Increasing infection prevalence in Guadalcanal to 20% (I_{h2} = 0.2) at the higher rate of migration (k_{12} = 0.5; k_{21} = 0.5) substantially increased the number of people infected in Nggela at infection peak and decreased time to peak infection (infection peak = 2,810 people at day 225; Figure 6).

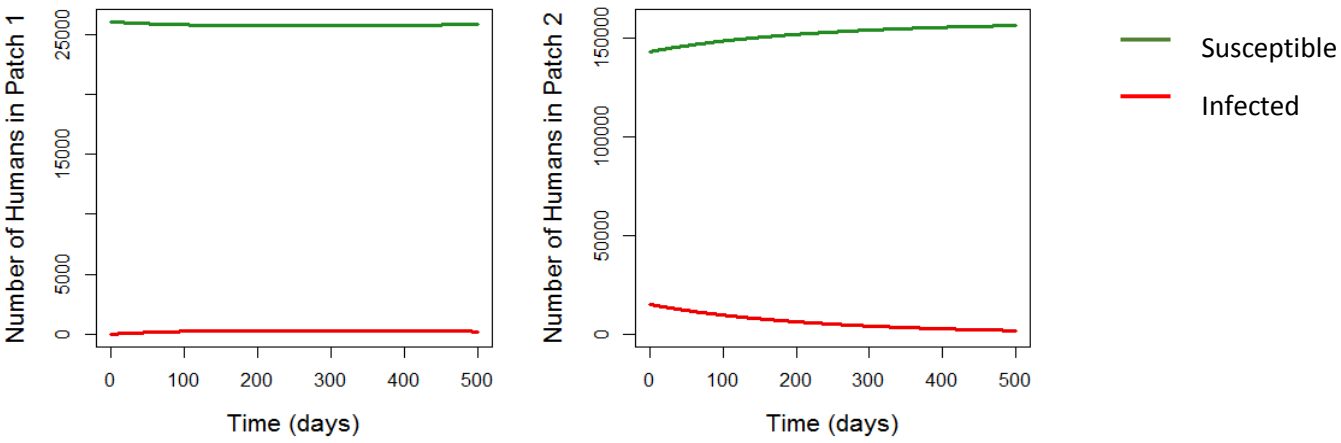


Figure 2. Model simulation showing time to malaria resurgence in patch 1 with migration rate between patch 1 and 2 = 0.1 and migration rate from patch 2 to 1 = 0.1. Human biting rate = 0.1; infection prevalence in patch 2 = 0.096; relative mosquito to human population abundance = 10.

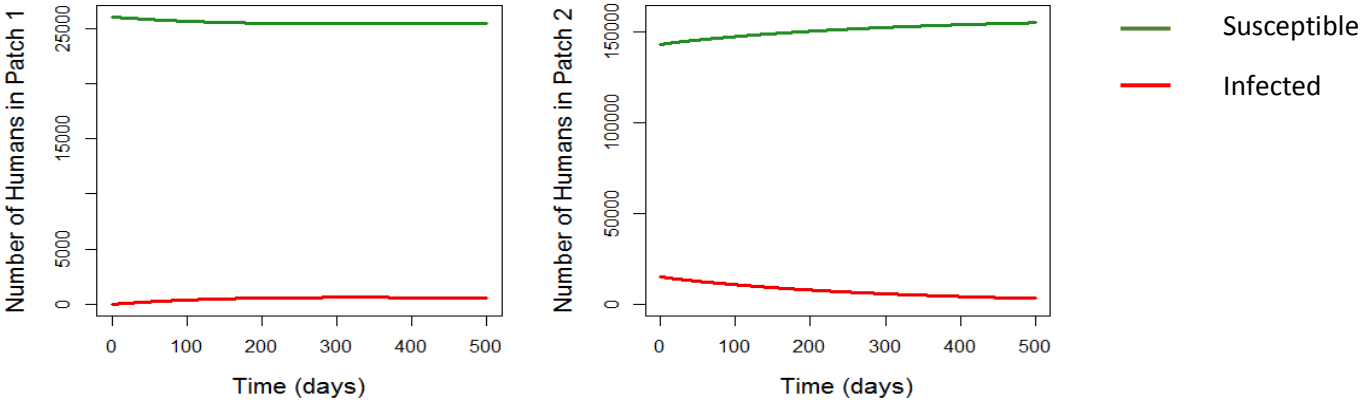


Figure 3. Model simulation showing time to malaria resurgence in patch 1 with migration rate from patch 1 and 2 = 0.1, migration rate from patch 2 to 1 = 0.1 and 50% increase in vector abundance in patch 1 and 2. Human biting rate = 0.1; infection prevalence in patch 2 = 0.096; relative vector abundance = 15.

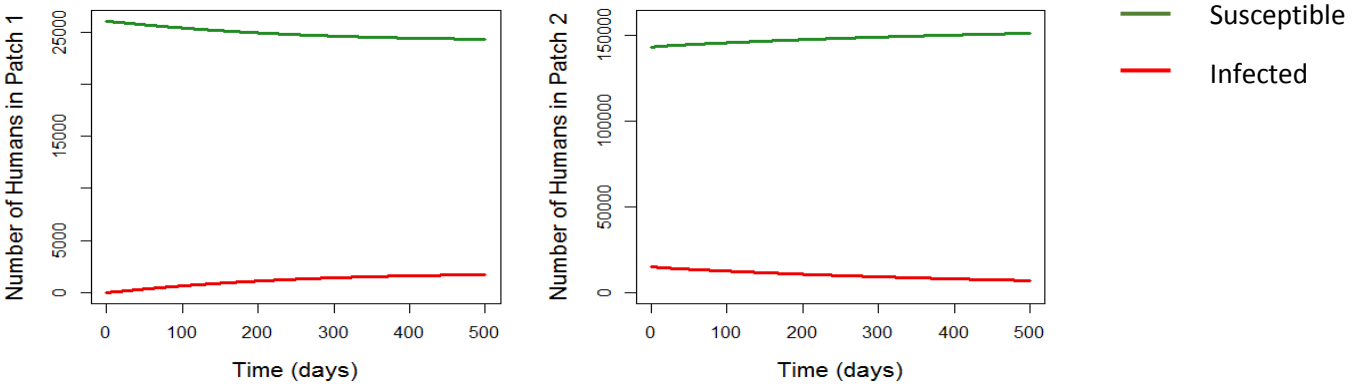


Figure 4. Model simulation showing time to malaria resurgence in patch 1 with migration rate from patch 1 and 2 = 0.1, migration rate from patch 2 to 1 = 0.1 and a 50% increase in human biting rate. Human biting rate = 0.15; infection prevalence in patch 2 = 0.096; relative vector abundance = 10.

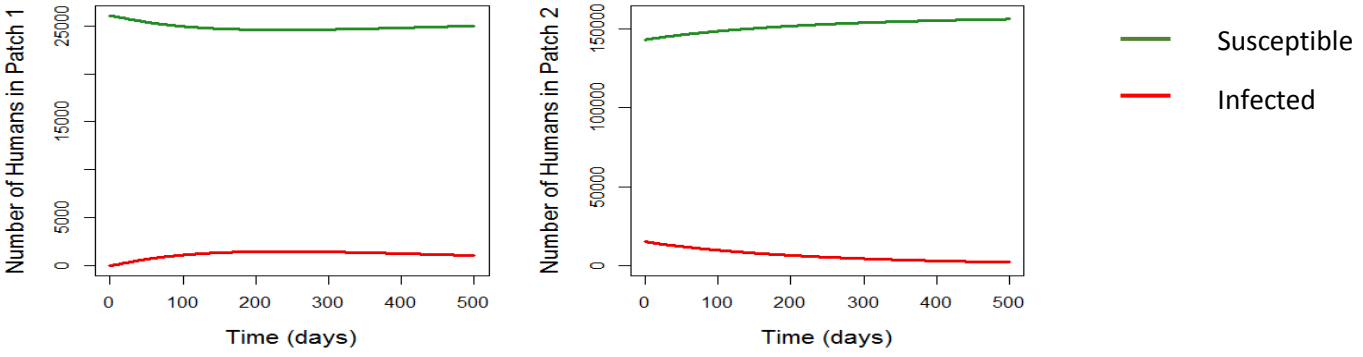


Figure 5. Model simulation showing time to malaria resurgence in patch 1 with ten-fold increase in migration between patch 1 and 2. Migration rate from patch 1 to 2 = 0.5; migration rate from patch 2 to 1 = 0.5; human biting rate = 0.1; infection prevalence in patch 2 = 0.096; relative vector abundance = 10.

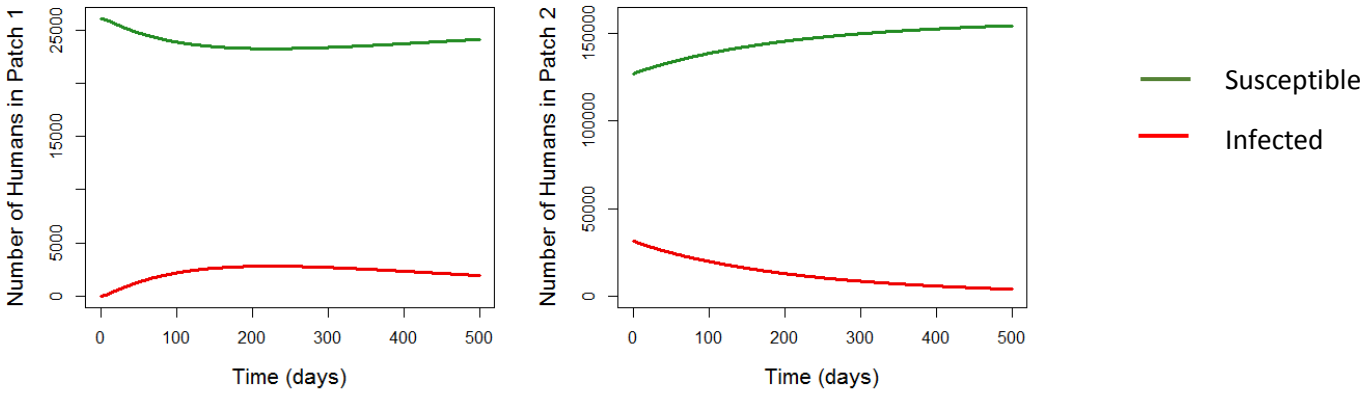


Figure 6. Model simulation showing time to malaria resurgence in patch 1 with migration rate between patch 1 and 2 increased by a factor of 10 and increase in infection prevalence in patch 2. Migration rate from patch 1 to 2 = 1.0; migration rate from patch 2 to 1 = 1.0; human biting rate=0.1; infection prevalence in patch 2 = 0.2; relative vector abundance = 10.

Varying vector abundance parameters also had an observable effect on infection peak, which increased with increasing migration rate between Nggela and Guadalcanal. Doubling the relative vector abundance (m = 20) at the lowest migration rate (k_{12}=0.1*;* k_{21} = 0.1) resulted in an infection peak of 1196 people at day 493 Table 2). A five-fold increase in migration rate coupled with an increase in vector abundance however resulted in an infection peak of 5232 people at day 500. The largest increase in infection peak (infection peak = 9535 people at day 441) was observed under increased rates of migration (k_{12} = 0.5; k_{21} = 0.5) biting rate (a = 0.15) and infection prevalence in Guadalcanal (I_{h2}).

An increase in biting rate (a = 0.15) and migration rate (k_{12} = 0.5; k_{21} = 0.5) alone also had a substantial impact on infection peak with 6932 people infected in Nggela after 500 days. A 50% decrease in biting rate (a = 0.5) at the lowest migration rate (k_{12} = 0.1; k_{21} = 0.1) however resulted in a substantial decrease in infection peak to 55 people at day 170. The impact of a 50% decrease in biting rate was also observed when the model was parameterised using higher rates of migration (k_{12} = 0.5; k_{21} = 0.5) and infection prevalence in Guadalcanal (I_{h2} = 0.2) with peak infection of 560 people at day 169. Sensitivity analysis conducted to examine the impact

Table 2. Sensitivity analysis model simulations run at different values for vector abundance, biting rate and migration rate between Nggela (Patch 1) and Guadalcanal (Patch 2)

Infection parameters					**Patch 1 results**				
Human biting rate (a)	Migration rate patch 1 to 2 (k12)	Migration rate patch 2 to 1 (k21)	Infectious humans patch 2 (Ih2)	Vector abundance (m)	Peak Infection (people)	Peak Infection (day)	Infection Day 100 (people)	Infection Day 250 (people)	Infection Day 500 (people)
Transmission dynamics with no migration between patch 1 and patch 2									
0.1	0	0	0.1	10	0	0	0	0	0
Transmission dynamics at varying migration rates and vector abundance with infection prevalence in patch 2 set at 0.1									
0.1	0.1	0.1	0.096	10	301	238	222	300	208
0.1	0.1	0.1	0.096	15	601	322	364	583	538
0.1	0.1	0.1	0.096	20	1196	493	532	998	1196
0.1	0.2	0.2	0.096	10	595	237	441	595	414
0.1	0.2	0.2	0.096	15	1177	319	723	1145	1053
0.1	0.2	0.2	0.096	20	2268	481	1051	1931	2266
0.1	0.5	0.5	0.096	10	1453	237	1085	1451	1032
0.1	0.5	0.5	0.096	15	2810	324	1761	2735	2579
0.1	0.5	0.5	0.096	20	5232	500	2531	4462	5232
Transmission dynamics at varying migration rates and fixed vector abundance with infection prevalence in patch 2 set at 0.2									
0.1	0.1	0.1	0.2	10	607	233	455	606	415
0.1	0.2	0.2	0.2	10	1189	230	899	1185	810
0.1	0.5	0.5	0.2	10	2810	225	2175	2795	1933
Transmission dynamics at varying migration rates, fixed vector abundance, 50% increase in human biting rate and infection prevalence in patch 2 set at 0.096 and 02.									
0.15	0.1	0.1	0.096	10	1706	500	626	1266	1706
0.15	0.2	0.2	0.096	10	3155	500	1233	2425	3155
0.15	0.5	0.5	0.096	10	6932	500	2950	5485	6933
0.15	0.1	0.1	0.2	10	2950	500	1256	2387	2951
0.15	0.2	0.2	0.2	10	5068	484	2431	4376	5067
0.15	0.5	0.5	0.2	10	9535	441	5542	8852	9508
Transmission dynamics at varying migration rates, fixed vector abundance, 50% decrease in human biting rate and infection prevalence in patch 2 set at 0.096 and 02. and infection prevalence in patch 2 with a 50% decrease in human biting rate									
0.05	0.1	0.1	0.096	10	55	170	48	50	22
0.05	0.2	0.2	0.096	10	109	170	96	100	44
0.05	0.5	0.5	0.096	10	272	170	239	249	110
0.05	0.1	0.1	0.2	10	113	170	100	104	46
0.05	0.2	0.2	0.2	10	226	169	199	206	91
0.05	0.5	0.5	0.2	10	560	169	494	510	226

of varying biting and migration rates on infection transmission found a linear associated between increased migration and infection peak and a monotonic, sigmoidal increase in infection peak associated with an increase in biting rate (Figure 7).

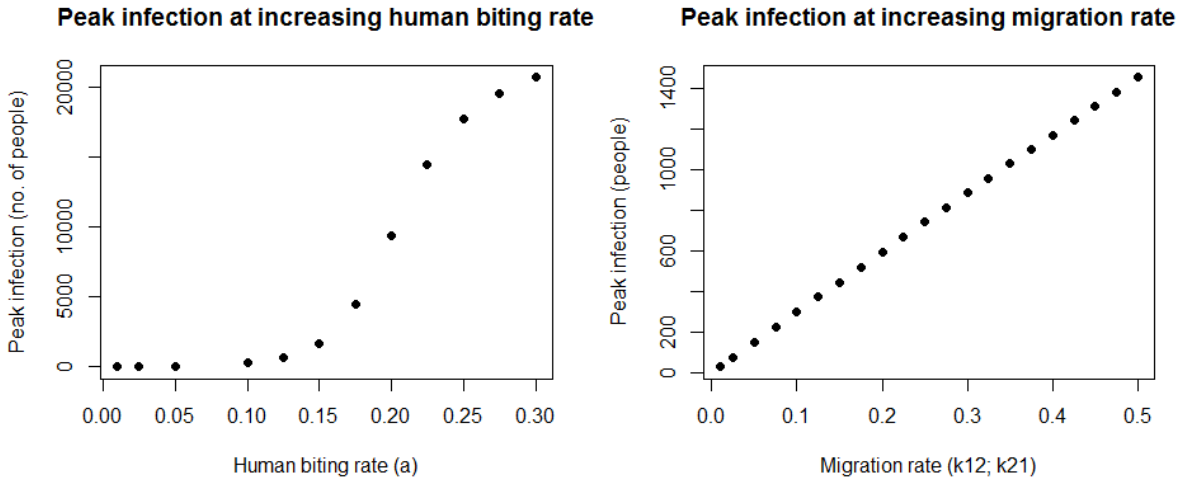


Figure 7. Results of sensitivity analysis showing increasing infection prevalence in patch 1 at increasing biting rate (a) and increasing migration rate (k_{12}; k_{21}).

6.4 Discussion

Results from model simulations carried out in this study show that an increased rate of daily migration is likely to result in an increase in imported infection and a resurgence in malaria transmission in a relatively short period of time (within one year). Parameterising the model with a low migration rate results in a relatively small resurgent infection peaks in Nggela, when compared with results of model simulations run using a five-fold increased migration rate. Migration rates used in this study were estimations of likely movement of infectious populations between both islands, and may not be reflective the actual mobility rates. An increased estimate of infection prevalence in Guadalcanal is also likely to result in an increase in infection incidence and imported infection cases, supporting a collaborative control strategy

between both islands which maintains vector control in both elimination and endemic areas until elimination has been achieved in both locations[30].

In terms of vector control, results from model simulations run in this study suggest that vector abundance and biting rate under varying rates of migration between both locations have a substantial impact on likelihood of infection resurgence in Nggela. Model simulations found that a 50% decrease in biting rate, even at higher infection prevalence in Guadalcanal and higher migration rate between Guadalcanal and Nggela resulted in an 80% decrease in infection peak compared with baseline model parameters. These results provide evidence supporting the continued use of vector control interventions in Nggela while transmission in Guadalcanal is ongoing for achieving reductions in biting rate and vector abundance as well as an increase in vector mortality rate through increased use and distribution of LLINs[31].

Our model results correspond with the results of other modelling work which concluded that the parameters most significant for control were human biting rate (a) and vector mortality rate (μ)[32]. The work carried out here also supports previous findings that where vectorial capacity and potential for transmission remains, removal of vector control carries a high risk or resurgence[34]. Continuing use of LLINs, which may already have a high coverage in elimination populations, until the threat of importation through human movement from adjacent endemic populations has ceased, may be an effective strategy for minimising the risk and scale of resurgence[26]. Partial loss of immunity among populations where transmission has ceased means that even more intensified coverage of vector control interventions may be necessary post elimination, rather than any kind of relaxation of interventions, as the population is more vulnerable than before[35].

The results of the phylogenetic analysis of samples from Nggela and Guadalcanal was indicative of either: clonal expansion, which arises as a result of importation of a genetically

distinct parasite into a new environment, and colonisation and proliferation of the unrepresentative gene in the new population[36] (i.e. infections in Nggela arose from importation of infection from Guadalcanal), or; reduction of local transmission in Nggela to a point where one parasite genotype remained and re-established in the local population. Outbreaks have resulted from clonal expansion of *Plasmodium* parasites imported via cross-border migration from neighbouring countries[37-39], and into low transmission areas, as has been observed in the highlands of neighbouring PNG[40], and epidemics caused by clonal expansion have the potential to spread antimalarial resistant malaria in a relatively short space of time[41].

While it is not clear whether or not clonal expansion of the *P. falciparum* genotype in Nggela is as a result of imported infection from Guadalcanal, presence of this genotype in both islands is nevertheless indicative of transmission occurring between both islands, likely as a result of human mobility. This inter-island transmission presents a risk of resurgence if elimination is achieved on one island while malaria transmission remains endemic on the other. In addition to this, the cases detected in Nggela were asymptomatic. Asymptomatic infections may be an important source of sustained local transmission and are less likely to present to health care facilities or be detected by passive surveillance systems[13].

The impact of a higher infection prevalence in Guadalcanal on greater infection resurgence on Nggela, suggests a mutually beneficial impact of malaria reduction in both locations, affirming the strategy of national and regional collaboration for overall reductions in malaria prevalence[30,42]. Results from other mathematical model simulations also predict that the success of elimination is dependent on preventing importation of infection and regional scale elimination strategies[30]. Surveillance to detect imported infections will be needed to inform national policy in the pre-elimination and elimination stages and estimate risk of resurgence while areas remain vulnerable and receptive to malaria tranmsission[8,35]. Future malaria surveillance in Solomon Islands should incorporate travel history to try to determine source of

infection.[43] Additional information collected should determine whether cases detected in Nggela are as a result of local transmission or inter-island travel[13].

Future work focusing on collecting more detailed migration data and parasite prevalence among migratory populations may give better estimates of the risk of inter-island malaria transmission. Using cell phone data[43-45], malaria indicator surveys incorporating travel history[46], or travel surveys incorporating diagnosis of infection by rapid diagnostic tests, may provide an improved measure of mobility rates and movement of malaria parasites[4]. Cell phone data may be useful in capturing more representative information about general migration patterns, and travel surveys might identify a greater proportion of the migratory population with asymptomatic malaria infection,[6] providing higher resolution data with which to parameterise models. Travel survey data could also assist in identifying sinks of transmission and common sources of infection[47,48], [49] and demographics of migratory populations for targeting control and surveillance operations to mitigate the risk of malaria importation[6,20,48].

6.5 Limitations

Some limitations are inherent in the model we used for this research. Firstly, the model does not address risk among mobile individuals or the probability of travel given infection[51] and so may not estimate correctly the rate of migration of infected individuals. In addition to this, the assumption of a well-mixed population between patch 1 and 2 through migration between Nggela and Guadalcanal results in an immediate non-zero prevalence in Nggela, whereas actual time to importation of infection to Nggela would likely be longer, dependent on likelihood of travelling with infection, or chance of onward transmission given travel of people with asymptomatic infection.

The model also assumes a well-mixed population with equal likelihood of infection based on interaction between human and vector populations in both patches. In reality, the probability

of infection may differ depending on geographic location and demography. Human mobility dynamics are also likely to be associated with demography, which in turn is associated with infection risk, rather than random. Model simulations were run under a simplified set of assumptions however to estimate the relative impact of migration on malaria resurgence in Nggela under varying levels of vector control. The assumption of a well-mixed population with equal risk of infection and likelihood of migration was therefore considered appropriate within this context of this research.

6.6 Conclusions

Increased rate of migration between Nggela and Guadalcanal, infection prevalence in Guadalcanal, human biting rate and vector abundance resulted in resurgent malaria in Nggela within an average of one year time period. Inter-island transmission of malaria through human migration may be a major obstacle to achieving sub-national malaria elimination in Solomon Islands, and the success of elimination programmes will be strongly dependent on the prevention of infection importation and resurgence through sustained surveillance and vector control.

References

1 Liu, Y., Sturrock, H. J., Yang, H., Gosling, R. D. & Cao, J. The challenge of imported malaria to eliminating countries. *The Lancet Infectious Diseases* **17**, 141 (2017).

2 Adams, B. & Kapan, D. D. Man bites mosquito: understanding the contribution of human movement to vector-borne disease dynamics. *PloS one* **4**, e6763 (2009).

3 Greenwood, B. M. Control to elimination: implications for malaria research. *Trends in parasitology* **24**, 449-454 (2008).

4 Edwards, H. M. *et al.* Novel cross-border approaches to optimise identification of asymptomatic and artemisinin-resistant Plasmodium infection in mobile populations crossing Cambodian borders. *PLoS One* **10**, e0124300 (2015).

5 Marangi, M. *et al.* Prevalence of Plasmodium spp. in malaria asymptomatic African migrants assessed by nucleic acid sequence based amplification. *Malaria journal* **8**, 12 (2009).

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

7 Routledge, I. *et al.* Estimating spatiotemporally varying malaria reproduction numbers in a near elimination setting. *Nature communications* **9**, 2476 (2018).

8 Velarde-Rodríguez, M. *et al.* Origin of malaria cases: a 7-year audit of global trends in indigenous and imported cases in relation to malaria elimination. *Global health action* **8**, 29133 (2015).

9 Mihretie, A. *et al.* Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. *Malaria journal* **16**, 89 (2017).

10 Griffin, J. T. *et al.* Potential for reduction of burden and local elimination of malaria by reducing Plasmodium falciparum malaria transmission: a mathematical modelling study. *The Lancet Infectious Diseases* **16**, 465-472 (2016).

11 Garcia, A. J., Pindolia, D. K., Lopiano, K. K. & Tatem, A. J. Modeling internal migration flows in sub-Saharan Africa using census microdata. *Migration Studies* **3**, 89-110 (2015).

12 Li, X. H. *et al.* A Historical Review of WHO Certification of Malaria Elimination. *Trends in parasitology* (2019).

13 Waltmann, A. *et al.* High rates of asymptomatic, sub-microscopic Plasmodium vivax infection and disappearing Plasmodium falciparum malaria in an area of low transmission in Solomon Islands. *PLoS neglected tropical diseases* **9**, e0003758 (2015).

14 Smith, J. *et al.* Malaria early warning tool: linking inter-annual climate and malaria variability in northern Guadalcanal, Solomon Islands. *Malaria journal* **16**, 472 (2017).

15 Acevedo, M. A. *et al.* Spatial heterogeneity, host movement and mosquito-borne disease transmission. *PloS one* **10**, e0127552 (2015).

16 Smith, D. L. *et al.* Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. *PLoS pathogens* **8**, e1002588 (2012).

17 Smith, D. L. & McKenzie, F. E. Statics and dynamics of malaria infection in Anopheles mosquitoes. *Malaria journal* **3**, 13 (2004).

18 Government, S. I. REPORT ON 2009 POPULATION & HOUSING CENSUS 2009. (2011).

19 Hickson, R. I., Mercer, G. N. & Lokuge, K. M. A metapopulation model of tuberculosis transmission with a case study from high to low burden areas. *PloS one* **7**, e34411 (2012).

20 Ada W. C. Yan, A. J. B., James M. McCaw, Nicolas Rebuli, Joshua V. Ross, Annalisa J. Swan, Roslyn I. Hickson. The distribution of the time taken for an epidemic to spread between communities. *In preparation* (2018).

21 Segel, L. A. & Edelstein-Keshet, L. *A primer on mathematical models in biology*. (SIAM, 2013).

22 Ross, R. Some a priori pathometric equations. *British medical journal* **1**, 546 (1915).

23 Soetaert, K., Petzoldt, T. & Setzer, R. W. Solving differential equations in R: package deSolve. *Journal of Statistical Software* **33** (2010).

24 Government, S. I. Population and Housing Census; Report on Migration and Urbainisation. (Solomon Islands National Statistical Office, Honiara, Solomon Islands, 2009).

25 Bekessy, A., Molineaux, L. & Storey, J. Estimation of incidence and recovery rates of Plasmodium falciparum parasitaemia from longitudinal data. *Bulletin of the World Health Organization* **54**, 685 (1976).
26 Thanh, P. V. *et al.* Epidemiology of forest malaria in Central Vietnam: the hidden parasite reservoir. *Malaria journal* **14**, 86 (2015).
27 Loyola, E. *et al.* Anopheles albimanus (Diptera: Culicidae) host selection patterns in three ecological areas of the coastal plains of Chiapas, southern Mexico. *Journal of medical entomology* **30**, 518-523 (1993).
28 Beier, J. C., Davis, J. R., Vaughan, J. A., Noden, B. H. & Beier, M. S. Quantitation of Plasmodium falciparum sporozoites transmitted in vitro by experimentally infected Anopheles gambiae and Anopheles stephensi. *The American Journal of Tropical Medicine and Hygiene* **44**, 564-570 (1991).
29 Graves, P., Burkot, T., Saul, A., Hayes, R. & Carter, R. Estimation of anopheline survival rate, vectorial capacity and mosquito infection probability from malaria vector infection rates in villages near Madang, Papua New Guinea. *Journal of Applied Ecology* **27**, 134-147 (1990).
30 Gerardin, J. *et al.* Effectiveness of reactive case detection for malaria elimination in three archetypical transmission settings: a modelling study. *Malaria journal* **16**, 248 (2017).
31 Reimer, L. J. *et al.* Malaria transmission dynamics surrounding the first nationwide long-lasting insecticidal net distribution in Papua New Guinea. *Malar J* **15**, 25, doi:10.1186/s12936-015-1067-7 (2016).
32 Beretta, E., Capasso, V. & Garao, D. G. A mathematical model for malaria transmission with asymptomatic carriers and two age groups in the human population. *Mathematical biosciences* **300**, 87-101 (2018).
33 Corbel, V. *et al.* Combination of malaria vector control interventions in pyrethroid resistance area in Benin: a cluster randomised controlled trial. *The Lancet infectious diseases* **12**, 617-626 (2012).
34 Yukich, J. O. & Chitnis, N. Modelling the implications of stopping vector control for malaria control and elimination. *Malaria journal* **16**, 411 (2017).
35 Organization, W. H. *Global technical strategy for malaria 2016-2030.* (World Health Organization, 2015).
36 Hastings, I. & Wedgwood-Oppenheim, B. Sex, strains and virulence. *Parasitology Today* **13**, 375-383 (1997).
37 Auburn, S. & Barry, A. E. Dissecting malaria biology and epidemiology using population genetics and genomics. *International journal for parasitology* **47**, 77-85 (2017).
38 Obaldia III, N. *et al.* Clonal outbreak of Plasmodium falciparum infection in eastern Panama. *The Journal of infectious diseases* **211**, 1087-1096 (2014).
39 Sáenz, F. E. *et al.* Clonal population expansion in an outbreak of Plasmodium falciparum on the northwest coast of Ecuador. *Malaria journal* **14**, 497 (2015).
40 Mueller, I., Kaiok, J., Reeder, J. C. & Cortés, A. The population structure of Plasmodium falciparum and Plasmodium vivax during an epidemic of malaria in the Eastern Highlands of Papua New Guinea. *The American journal of tropical medicine and hygiene* **67**, 459-464 (2002).
41 Khaireh, B. A. *et al.* Population genetics analysis during the elimination process of Plasmodium falciparum in Djibouti. *Malaria journal* **12**, 201 (2013).
42 Ejov, M., Davidyants, V. & Zvantsov, A. Regional framework for prevention of malaria reintroduction and certification of malaria elimination 2014–2020. (2014).
43 Tatem, A. J. *et al.* Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria journal* **13**, 52 (2014).
44 Deville, P. *et al.* Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences* **111**, 15888-15893 (2014).
45 Wesolowski, A., Buckee, C. O., Engø-Monsen, K. & Metcalf, C. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of Infectious Diseases* **214**, S414-S420 (2016).
46 Pindolia, D. K. *et al.* Human movement data for malaria control and elimination strategic planning. *Malaria journal* **11**, 205 (2012).

47 Sorichetta, A. *et al.* Mapping internal connectivity through human migration in malaria endemic countries. *Scientific data* **3**, 160066 (2016).
48 Tatem, A. J. & Smith, D. L. International population movements and regional Plasmodium falciparum malaria elimination strategies. *Proceedings of the National Academy of Sciences* **107**, 12222-12227 (2010).
49 Stoddard, S. T. *et al.* The role of human movement in the transmission of vector-borne pathogens. *PLoS neglected tropical diseases* **3**, e481 (2009).
50 Tatem, A. J. Mapping population and pathogen movements. *International health* **6**, 5-11 (2014).
51 Ruktanonchai, N. W. *et al.* Census-derived migration data as a tool for informing malaria elimination policy. *Malaria journal* **15**, 273 (2016).

CHAPTER 7

Discussion

CHAPTER 7. DISCUSSION

7.1 Introduction

At this current point of progression toward malaria elimination, and as advancements are made by regional and national control and elimination programs, several challenges to sustained control and effective elimination of malaria exist. In countries aiming for control, heterogeneities in malaria prevalence may present challenges in the allocation of vector control interventions and resources to areas where they will have most impact. As national and regional malaria prevalence decreases, changes in the demographic profile to a higher risk among older adults, occupational groups such as forest goers, and mobile and cross-border migratory populations may be observed[1,2]. This requires a reorientation of control interventions and surveillance operations toward these groups[3].

In low-transmission settings, the success of elimination programmes will be strongly dependent on identifying and targeting residual foci of transmission, including asymptomatic reservoirs of infection and prevention of malaria infection importation from areas where transmission remains endemic[4]. Once national elimination has been achieved, prevention of resurgence should be a priority for elimination programmes through prevention of importation of infection and sustained vector control. Reasons for resurgence post-elimination are multi-factorial and can include cross-border or international importation of infection, funding shortages, failure to implement control programme technical strategy well and relaxing of vector control interventions[5]. In countries at all levels of malaria endemicity, emergence of insecticide and anti-malarial drug resistant parasites will pose a substantial threat to the success of control and elimination programmes[5,6]. Other obstacles to achieving elimination which are more difficult to address also exist such as continuing surveillance and control efforts in conflict settings[7].

As global malaria transmission decreases, and these new associated challenges arise, new, effective epidemiological tools are needed to evaluate progress against national targets, examine spatial patterns of malaria transmission in response to control efforts, and attempt to predict how malaria transmission may be impacted by changing demographics and risk factors. The research presented in this thesis describes challenges which may be faced by malaria control and elimination programmes, and novel epidemiological tools for addressing some of these challenges. These tools can be broadly applied to the three different control settings of malaria elimination – control, elimination and prevention of reintroduction.

7.2 Key research findings

Chapter 2 provides an overview of the current global malaria situation and strategies for malaria control and elimination, describes epidemiological tools that can be used for targeting these interventions and discusses how these tools can be utilised to effectively deal with current and forecasted challenges in malaria elimination. One of the main priorities and biggest challenges for malaria elimination is the impact that human mobility may have on malaria transmission and resurgence post elimination. Several methods have been evaluated for effectively measuring human mobility and assessing the impact that this will have on malaria transmission; however, the most effective and efficient way of doing so being yet to be determined[8-10].

A further obstacle to global malaria elimination, discussed in chapter 2, is the emergence of insecticide and anti-malarial drug resistant parasites, associated with socio-economic disparities and behavioural practices around adhering to anti-malarial treatment regimens[11], amongst other factors. The independent emergence of anti-malarial drug resistance in several geographic locations in the Greater Mekong subregion, as well as the African continent[12], is a threat to global elimination of malaria, making elimination of *P. falciparum* while ACTs are still effective an urgent priority[3], requiring sustained intervention and political will[13].

Appropriate allocation and evaluation of control interventions, including quantification and examination of the spatial distribution of delivered interventions, is essential for improving and maintaining programmatic performance. Effective, sustainable surveillance for continual assessment of malaria incidence, updated in real time, and collated and analysed for feedback to control programmes, is integral to the achievement of global malaria elimination. Chapter 2 discussed how such approaches can be enhanced by epidemiological methods that help make better use of routinely collected operational and surveillance data in control, elimination and prevention of reintroduction settings. Developing simple yet effective technical solutions of this type should be a current focus in the field of operational research.

Chapter 3 describes the current epidemiology and strategies for national malaria control in PNG. Chapter 3 also includes a summary of the previous malaria control programmes and reasons for programme failure or instances of resurgence post elimination[5]. A review of literature identified key challenges to the current malaria control programme in PNG, as well as challenges faced by historical control programmes. Barriers to distribution of control interventions exist, including poor infrastructure and isolation of populations, as well disparities in access to health care, which limits access to proper diagnosis and treatment.

Previous control programmes suffered from unsustained commitment, political will, financing and resources for maintaining interruption of transmission. Critically, a lack of programmatic cohesion and inadequacies in planning were consistently cited as an impediment to the success of all three major previous attempts at malaria control in PNG. This highlights the importance of high-quality evidence to support programmatic decision making including epidemiological tools which allow, for example, visualisation of the variation in risk across the country.

Chapter 4 highlights the importance of assessing the predictive accuracy of statistical models used to predict the spatial distribution of infectious disease. Evidence from this work suggests

that where transmission may be associated with ecological drivers that vary across the country, with complex interactions (including collinearity), a traditional GLM modelling approach may be inappropriate for spatial predictions on a national scale. There are several reasons why graphical models such as BDNs are more accurate in informing predicted probability distributions. Firstly, within a Bayesian framework, flexible structuring of the model enables the integration of expert knowledge about the ecology of the disease with observational data to improve model performance. BDN models quantify associations between covariates in the model, as well as with the defined outcome of interest, allowing collinear variables to be included in the same model.

In countries such as PNG, where the environmental drivers of malaria vary between different areas, and consequent heterogeneity in transmission exists, incorporating explanatory variables which may otherwise have to be excluded due to problems of collinearity may also aid in improvement of model prediction accuracy. An additional benefit of using models of this type is in the ability to demonstrate uncertainty in predictions. From a malaria control point of view, information such as this is important for informing control programmes particularly on a national scale where resources need to be targeted in the most efficient and economical manner possible. Where the degree of uncertainty is high due to uncertainty in measurements, or insufficient data, additional surveillance operations or surveys to obtain new information, or alternative evidence gathering strategies may be employed for improvement in risk estimations.

Chapter 5 demonstrates the application of integrating parasite genotype data with spatial models to infer connectivity between distinct geographic locations in terms of malaria transmission. Model results showed a spatial predominance of specific *P. falciparum* genotypes in distinct areas of PNG. This strategy of examining the parasite population structure may be beneficial in deducing how distinct geographic locations are connected, and identifying sources of infection where imported cases are likely to have originated from. Gaining insight

into transmission dynamics by examining the parasite population structure in this way may also help in assessing risk of resurgence in areas where elimination has been achieved, but remain connected to areas where transmission remains endemic. Consequently this may aid in informing decisions regarding sustained vector control and surveillance operations in areas where local transmission has ceased. Genetic surveillance strategies are financially and logistically demanding however, and may not be feasible in may elimination settings.

The research findings from chapter 6 demonstrate the importance of sustained interventions in areas vulnerable to resurgence if elimination is achieved while connectivity to areas where malaria remains endemic is sustained by human mobility. Where vectorial capacity remains and where the location is connected to an endemic area by human migration, there is potential for resurgence and reestablishment of local transmission. This provides evidence for use of sustained control interventions following elimination and the need for regional collaboration to reduce transmission concurrently across many connected areas[4]. Chapter 6 also demonstrates the application of mathematical models in simulating malaria transmission under hypothetical scenarios such as increased migration between two distinct geographic areas, under sustained or waning vector control and under conditions of increased or decreased abundance of vector populations relative to the human population. For a model to be considered robust, it must reflect the transmission dynamics of the disease accurately enough for predictions to be valid, while not being so specific that predictions are heavily reliant on parameters of the model. A simplified Ross-Macdonald model was applied in this case to the context of preventing malaria resurgence in areas in Solomon Islands to examine the effect of varying human biting, vector abundance and migration rates on resurgent infection.

Results of mathematical modelling work presented in chapter 6 highlighted the importance of vector control through use of LLINs in areas of Solomon Islands aiming for elimination or prevention of reintroduction when connected to areas where malaria remains endemic. The

results obtained from this work is supported by other research findings suggesting that vector control through LLIN use may have the greatest impact on preventing resurgent malaria compared with IRS alone[14,15]. Mathematical modelling work suggests that even while sustaining malaria intervention coverage at current levels, a moderate increase in malaria incidence may occur as a result of partial loss of immunity among populations where interventions have been directed. In order to avoid this, vector control intervention coverage needs to be higher than 80% of at-risk populations and to achieve this level of coverage, innovative epidemiological tools to appropriately direct vector control interventions are needed[3].

Mathematical models can help in planning surveillance operations and allocation of vector control interventions by providing practical tools to assess the feasibility of malaria elimination, identify cost-effective strategies to shorten elimination timelines (if coupled with economic analysis approaches), and provide national control programmes with tools for assessing and tailoring malaria control and elimination strategies to specific settings[16,17]. Mathematical models will also be useful tools in guiding interventions evaluating which control tools will have greatest impact on prevention of resurgence post elimination and in deciding for how long control interventions should be sustained post elimination given the likelihood of resurgence.

7.3 Limitations

There were some limitations to the analytical research contained in this thesis, which were also outlined in each research chapter. One of the primary limitations of this thesis is that data used for the analytical work were collected between 2010 and 2011 in the case of data used for the spatial prediction of malaria in PNG in chapter 4, and between 2008 and 2009, in the case of data used for examination of spatial predominance for parasite genotype data in chapter 5 and

to parameterise the mathematical model in chapter 6. Predictions of the spatial distribution of malaria made using 2010/2011 data, may not correspond to the current prevalence rates in PNG as control interventions which have been ongoing in the intervening time period may have altered transmission dynamics.

Likewise, data used to examine the spatial predominance of *P. falciparum* genotypes were collected ten years ago, and changes in transmission and mobility dynamics of the population may have occurred in the intervening time, which may not be reflected by results obtained here. This work is however, likely to be a good indication of how populations in PNG have historically been connected in terms of malaria transmission and therefore still has relevance for guiding control programmes. Also it is difficult given the lack of other indicators of migration including mobile phone records data and travel history, to explore the validity of these maps as inference for human movement patterns within PNG.

Migration rates used for parameterisation of the mathematical model may not be reflective of daily migration rates of infectious individuals in Solomon Islands. Cell phone data, GPS data or travel surveys may give better estimations of migration rates and improved precision in model predictions on magnitude and time to resurgence in the event of reintroduction of infection post elimination.

7.4 Future research

Crucial to making progress in global elimination will be continuing to improve our understanding of how human mobility connects distinct geographic areas and facilitates transmission of malaria parasites. Future work should focus on integrating parasite genotype data with human mobility data, such as mobile phone records, to examine how these data sources compare in mobility models for the prediction of connectivity. Combining parasite genotype data with movement data such as Google location, census or mobile phone records

data, may give better insights into how human movement impact malaria transmission dynamics. Various different strategies and data sources for measuring human mobility and connectivity of populations in terms of malaria transmission are currently being explored to improve the understanding of spatial transmission of malaria[18,19].

The global malaria elimination strategy has focused on 'shrinking the malaria map' by eliminating on a sub-national or national level and then scaling up to regional level[20]. However, a more nuanced approach might be warranted, whereby areas where similar parasite genotyopes circulate, or areas which are connected in terms of malaria transmission through human movement and circulation of similar parasite genotypes, are identified and targeting with coordinated interventions[21]. Effective regional collaborations are also required to prevent resurgence in countries that have achieved elimination that are connected to endemic areas by human migration[3].

Future research to expand on work carried out within the context of this thesis specifically should focus on:

- Generating maps estimating the predictive risk of *P. falciparum* and *P. vivax* infection across PNG updated with data collected from surveys carried out subsequent to 2010/2011 to examine current predicted malaria prevalence in PNG, and explore how prevalence may be changing in response to control interventions.
- Integrating *Plasmodium* parasite genotype data with call record or google location data to examine how well models using these different data sources compare at identifying routes of transmission as a result of human mobility.
- Examining risk of resurgence in eliminating areas in Solomon Islands using more detailed measures of human migration data such as travel surveys or call record data and parasite prevalence among migratory populations.

- Examining risk of resurgence in eliminating areas using mathematical models which incorporate parameters on likelihood of travelling with infection, or chance of onward transmission given travel of people with asymptomatic infection.

7.5 Conclusions

This thesis provided a review of the current progress being made toward malaria elimination, and some of the challenges inherent in achieving this goal. Strategies used in the control and elimination of malaria and epidemiological tools which can be used in the planning of control and elimination programmers and targeting of resources and interventions were reviewed. In PNG, key future and previous challenges to achieving elimination were identified and recommendations were provided for future directions of control and elimination. The application of several epidemiological methods for examining the spatial distribution of malaria were also examined.

Novel statistical methods (BDNs) were compared with conventional models (GLMs) and the predictive accuracy of the spatial distribution of malaria using both approaches were compared. Both models examined associations of *P. falciparum* and *P. vivax* prevalence from observational data collected in a national malaria survey in 2010/2011 with ecological drivers of transmission derived from remote sensing image data. Cross validation of model results determined that novel BDN models had improved performance, compared with GLM models, for predicting the spatial distribution of malaria in PNG. The geographic niches of *P. falciparum* genotypes were examined using the results of a Dirichlet regression model examining associations of eight distinct *Plasmodium* parasite genotypes with ecological covariates including elevation, latitude and longitude coordinates, population density and distance of survey villages from the coastline. The predicted spatial distribution of these

genotypes based on model results gave insight into the transmission of malaria parasites in PNG and how distinct populations are connected in terms of malaria transmission.

Finally, using a Ross-Macondald metapopulation model, we were able to estimate the likelihood of resurgence in Nggela, an island of low malaria transmission in Solomon Islands, if local transmission has ceased but importation of infection is occurring through connectivity with Guadalcanal, an islands where malaria transmission remains endemic. From model simulations we determined that malaria resurgence is likely to occur within one year, given sustained human mobility between both islands and relaxed vector control interventions in Nggela. We also determined that human biting rate is the most important parameter in malaria resurgence, suggesting that malaria resurgence may be preventable through sustained LLIN use.

This thesis descried epidemiological methods which can be applied in examining malaria transmission dynamics in control, elimination and prevention of reintroduction settings and the findings of this thesis may help in guiding national control and elimination programmes in PNG, Solomon Islands and further afield.

References

1 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. *Malaria journal* **18**, 32 (2019).

2 Wangdi, K., Gatton, M. L., Kelly, G. C. & Clements, A. C. in *Advances in parasitology* Vol. 89 79-107 (Elsevier, 2015).

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

4 Gerardin, J. *et al.* Effectiveness of reactive case detection for malaria elimination in three archetypical transmission settings: a modelling study. *Malaria journal* **16**, 248 (2017).

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

6 World Health Organization. World Malaria Report 2018. (Geneva, 2018).

7 Abeyasinghe, R. R., Galappaththy, G. N., Gueye, C. S., Kahn, J. G. & Feachem, R. G. Malaria control and elimination in Sri Lanka: documenting progress and success factors in a conflict setting. *PLoS One* **7**, e43162 (2012).

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

9 Wesolowski, A., Buckee, C. O., Engø-Monsen, K. & Metcalf, C. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *The Journal of infectious diseases* **214**, S414-S420 (2016).

10 Tatem, A. J. *et al.* Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria journal* **13**, 52 (2014).

11 Anyanwu, P. E., Fulton, J., Evans, E. & Paget, T. Exploring the role of socioeconomic factors in the development and spread of anti-malarial drug resistance: a qualitative study. *Malaria journal* **16**, 203 (2017).

12 Bwire, G. M., Ngasala, B., Mikomangwa, W. P., Kilonzi, M. & Kamuhabwa, A. A. Detection of mutations associated with artemisinin resistance at k13-propeller gene and a near complete return of chloroquine susceptible falciparum malaria in Southeast of Tanzania. *Scientific Reports* **10**, 1-7 (2020).

13 Maude, R. J. *et al.* The last man standing is the most resistant: eliminating artemisinin-resistant malaria in Cambodia. *Malaria journal* **8**, 31 (2009).

14 Corbel, V. *et al.* Combination of malaria vector control interventions in pyrethroid resistance area in Benin: a cluster randomised controlled trial. *The Lancet infectious diseases* **12**, 617-626 (2012).

15 Yukich, J. O. & Chitnis, N. Modelling the implications of stopping vector control for malaria control and elimination. *Malaria journal* **16**, 411 (2017).

16 Stuckey, E. M., Smith, T. A. & Chitnis, N. Estimating malaria transmission through mathematical models. *Trends in parasitology* **29**, 477-482 (2013).

17 Modeling, m. C. G. o. A research agenda for malaria eradication: modeling. *PLoS medicine* **8**, e1000403 (2011).

18 Chang, H.-H. *et al.* The geography of malaria elimination in Bangladesh: combining data layers to estimate the spatial spread of parasites. *BioRxiv*, 421578 (2018).

19 Wesolowski, A. *et al.* Mapping malaria by combining parasite genomic and epidemiologic data. *BMC medicine* **16**, 190 (2018).

20 Feachem, R. G. *et al.* Shrinking the malaria map: progress and prospects. *The Lancet* **376**, 1566-1578 (2010).

21 Tatem, A. J. & Smith, D. L. International population movements and regional Plasmodium falciparum malaria elimination strategies. *Proceedings of the National Academy of Sciences* **107**, 12222-12227 (2010).