Modelling antibody responses to malaria blood stage infections:

A novel method to estimate malaria transmission intensity from serological data

Emilie Pothin

Thesis submitted for PhD examination

Department of Infectious Disease Epidemiology

Imperial College London

2013

Declaration of originality

I declare that the work presented in this thesis is my own work, completed under the supervision of Prof. Azra Ghani and Prof. Neil Ferguson. All individuals who shared their data and commented on the analyses or on the manuscript have been acknowledged. Any other studies mentioned in the manuscript have been appropriately referenced.

Emilie Pothin August 2013

Copyright Declaration

Acknowledgements

Firstly, I would like to thank the Medical Research Council for funding this research and making this work possible.

More importantly, this work would not have been possible without the guidance from my supervisors Professor Azra Ghani and Professor Neil Ferguson. I greatly appreciate all their advice both mathematical and non-mathematical and their suggested directions for the research. Their experience and intellectual stimulation made my research a challenging but enjoyable experience.

I am indebted to Chris Drakeley for sharing his time, expertise, enthusiasm throughout the PhD and from whom I learnt a huge amount. The work in the thesis was made possible as a result of many lively and interesting discussions.

As part of the Malaria Modelling group at Imperial College, I greatly benefited from the dynamic and stimulating environment there. I owe much gratitude to Jamie Griffin, Michael White, Patrick Walker, Hannah Slater, John Marshall, Lucy Okell, Tom Churcher, Michael Bretscher, Bhargavi Rao and Ndukwe Ukoha for their helpful discussion, insightful suggestions, as well as their time and patience.

I am extremely grateful to Lulla Opatowski and Anne Cori for their generosity with time, their many knowledgeable suggestions for my research, their comments on drafts of this thesis, their continuous encouragement and invaluable advice, their inspiring discussions and above all their friendship.

I am indebted to a large number of people within the department who made valuable contributions to my research. In particular, I would like to thank Simon Cauchemez for introducing me to Bayesian methods and for providing advice and modelling expertise throughout the PhD; Deirdre Hollingsworth for introducing me to epidemiological models; Jeff Eaton for sharing his C code to help my programs run much faster; Wes Hinley for his generous help in various programming issues; Thibaut Jombart for all his many great tips in R; Pierre Nouvellet for his numerous advice and Lorenzo Pellis and James Truscott for their interesting and helpful discussions about mathematical modelling. Also, I would like to thank Christl Donnelly for her insightful suggestions on my work during my upgrade and Paul Parham for kindly offering to read my manuscript in spite of his busy schedule.

I owe a great deal to all of those who have provided data for use in this thesis. The work in the thesis was made possible thanks to Chris Drakeley, Jackie Cook, Patrick Corran, Teun Bousema, Alison Rand, Geoff Targett and Carla Proietti, who shared their data, kindly responding to my many questions and provided me

invaluable insight on seroepidemiology of malaria. I would also like to thank all members of the Joint Malaria Program community study teams and the laboratory staff at Kilimanjaro Christian Medical College and the National Institute for Medical Research who contributed to the data collection used in Chapter 4.

I am grateful to Christophe Rogier, Johan Vekemans, Bich-Tram Huynh and Myriam Gharbi for their useful advice and the enjoyable discussions. I benefited greatly from them sharing their enthusiasm and passion about their work on malaria.

I would like to thank all of those with whom I shared my office during the PhD, for their interesting and supportive scientific and less scientific discussions: Marcus Shepherd, Diane Pople, Hannah Clapham, Natsuko Imai, Nick Beckley, Peter Winskill, George Shirreff and Jocelyn Elmes.

Special thanks go to Susan Edwards and Annick Borquez for providing me with unfailing positive encouragement since the beginning of my PhD. I cannot thank them enough for their friendship, trust and support.

Finally, I would like to thank all my friends and family who, from a close or a longer distance, always provided me with support and uplifting discussions when I needed it the most. In particular, I would like to thank my parents, Jean and Marie Thérèse, and my brother Stéphane for always being here for me and unconditionally supporting me through all the decisions that led to this thesis.

Abstract

Infection with the *Plasmodium falciparum* malaria parasite results in an immune response which includes the production of antibodies against the blood-stage of infection. In recent years there has been an increase in the use of serological data to monitor malaria transmission intensity. Traditionally, EIR and parasite prevalence were the preferred tools for measuring malaria transmission intensity. Serology has been shown to be particularly useful in areas of low endemicity where traditional measures (EIR and parasite prevalence) are problematic. Transmission intensity in this case is usually described by the seroconversion rate obtained from fitting a catalytic model to age-stratified serological responses.

The aim of my thesis was to better utilise the continuous measurements of antibody responses provided by serology studies to obtain improved estimates of transmission intensity. To do this, I developed a series of biologically motivated models to mimic the acquisition and decay in blood-stage antibody responses.

In the first part of the thesis, I developed a discrete model as a direct extension of the catalytic model and fitted this to cross-sectional data from several sites in Cambodia to obtain an estimate of the exposure rate. In the second part of the thesis a series of continuous density models were developed to mimic antibody acquisition and loss for *P. falciparum* infections. These models were fitted to both the Cambodian data and separately validated by fitting to data from Tanzanian villages at a wider range of transmission intensities. In the final section I applied and extended the model to encompass a wider range of endemic transmission in Somalia, Bioko Island, Gambia and Uganda in order to assess the robustness of the method.

My results show that estimates of the exposure rate obtained by fitting the density model are highly correlated with classic malariometric indices and that a key advantage of this approach is the increased precision in the estimates compared to estimates of the seroconversion rate, especially in areas of low transmission. This method could therefore be a useful alternative framework for quantifying transmission intensity which makes more complete use of serological data and shows potentials for detecting heterogeneity in malaria exposure.

Contents

List of figures

Chapter 3

Chapter 4

Chapter 6

Appendices

List of tables

Chapter 5

Abbreviations

Ab Antibody

ACT Artemisinin-based Combination Therapy

AMA-1 Apical Membrane Antigen 1

API Annual Parasite Index

BIMCP Bioko Island Malaria Control Program

CMBS Cross-sectional Malaria Baseline Survey

CNM National Center of Malaria Control, Parasitology and Entomology

CrI Credible Interval

CSP CircumSporozoite Protein

DIC Deviance Information Criterion

DDT Dichloro-diphenyl-trichloroethane

EIR Entomological Inoculation Rate

ELISA Enzyme Linked ImmunoSorbent Assay

FOI Force Of Infection

G6PD Glucose-6-Dhosphate dehydrogenase

GMEP Global Malaria Elimination Program

GMP Global Malaria Programme

ibppy infectious bites per person per year

IFA Immuno Fluorescent Antibody

Ig Immunoglobulin

IPT Intermittent Presumptive Treatment

IQR Inter Quartile Range

IRS Indoor Residual Spraying

ITN Insecticide Treated Nets

LLIN Long-Lasting Insecticidal Net

LSHTM London School of Hygiene and Tropical Medicine

MCMC Markov Chain Monte Carlo

MDA Mass Drug Administration

MIS Malaria Indicator Survey

MLE Maximum Likelihood Estimator

$_{mol}$**FOI** Molecular Force Of Infection

MSAT Mass Screen And Treat

MSP-1 Merozoite Surface Protein 1

NMCP National Malaria Control Programme

OD Optical Density

PCR Polymerase Chain Reaction

PR Parasite rate

Pf *Plasmodium falciparum*

Pv *Plamsodium vivax*

RBM Roll Back Malaria

RBC Red Blood Cells

RDT Rapid Diagnostic Test

SCR SeroConversion rate

WHA World Health Assembly

WHO World Health Organisation

Chapter 1: Introduction

1.1 Malaria History

The origin of malaria dates back to 50,000 years ago, as confirmed by phylogenetic studies. Humans may have been infected by malaria since the origin of the species [1, 2]. The malaria parasite, *Plasmodium* has evolved with humans, migrated out of Africa and adapted itself to new environments [3]. The term malaria originates from the Italian mala aria, signifying "bad air" while the French terminology "paludisme" indicating the association with swamps was later introduced.

In 1880, a French army surgeon Laveran observed for the first time the malaria parasite in the blood of one of his patients suffering from malaria. But it is only in 1898 that the British Sir Ronald Ross and the Italian Giovanni Batista Grassi showed that malaria was transmitted by mosquitoes and established the complete cycle of malaria transmission [4]. Ross & McDonald introduced mathematical modelling to describe transmission dynamics and were pioneers in malaria control, giving hope for eradication of malaria [5].

A Global Malaria Eradication Program (GMEP) was launched in 1950 by the World Health Organisation (WHO) with the objective of eliminating malaria from endemic countries. Measures for malaria control consisted of indoor residual spray using the newly discovered insecticide DDT and treatment with chloroquine. The results of the campaign, whilst successful in Western areas including Europe, USA and Australia, were disappointing in tropical countries where malaria transmission was highly intense and no real gains were made [6]. In 1969, it was generally accepted that malaria eradication would not be achieved and the new objective shifted to malaria control.

Foreign soldiers infected with malaria during Korean and Vietnamese wars provided useful samples for the investigation of antibodies from non-immune populations. In the 1960s, serology was shown to be a useful tool for measuring local transmission and assessing the impact of elimination and control interventions. A large number of seroepidemiological studies were subsequently undertaken to assess the potential for elimination [7] and confirm the success of elimination [8, 9]. As the funding for malaria eradication decreased, the use of serology also did. But with the recent renewed interest in the elimination of malaria resurrected by Bill and Melinda Gates and sequentially endorsed by WHO [10], the use of seroepidemiological studies is back on the agenda, with improved and standardised methods [11].

1.2 Epidemiology and Biology of Malaria

1.2.1 Malaria epidemiology

1.2.1.1 Malaria burden

Malaria is widely distributed throughout the tropical and subtropical regions of the world [12–14] and is estimated to be accountable for 219 million clinical cases annually. Malaria remains an important cause of morbidity causing an estimated 660 000 deaths annually worldwide [15], with children bearing the greatest burden [16]. Indeed about 85% of the deaths globally were estimated to be in children under the age of 5 years. According to the WHO, out of the 99 countries with on-going malaria, half of them were on track to meet the World Health Assembly (WHA) and Roll Back Malaria (RBM) target: to achieve a 75% reduction in malaria cases by 2015, compared to levels in 2000.

Malaria is caused by the protozoan *Plasmodium.* Four species are responsible for human malaria: *Plasmodium falciparum, P. vivax, P. ovale and P. malariae.* Among these species, *P. falciparum* accounts for most mortality and *P. vivax* is responsible for most of the malaria infection outside Africa. In 2007, a total of 2.37 billion people were estimated to be at risk from *P. falciparum* [17] geographically wide spread (Figure 2.1).

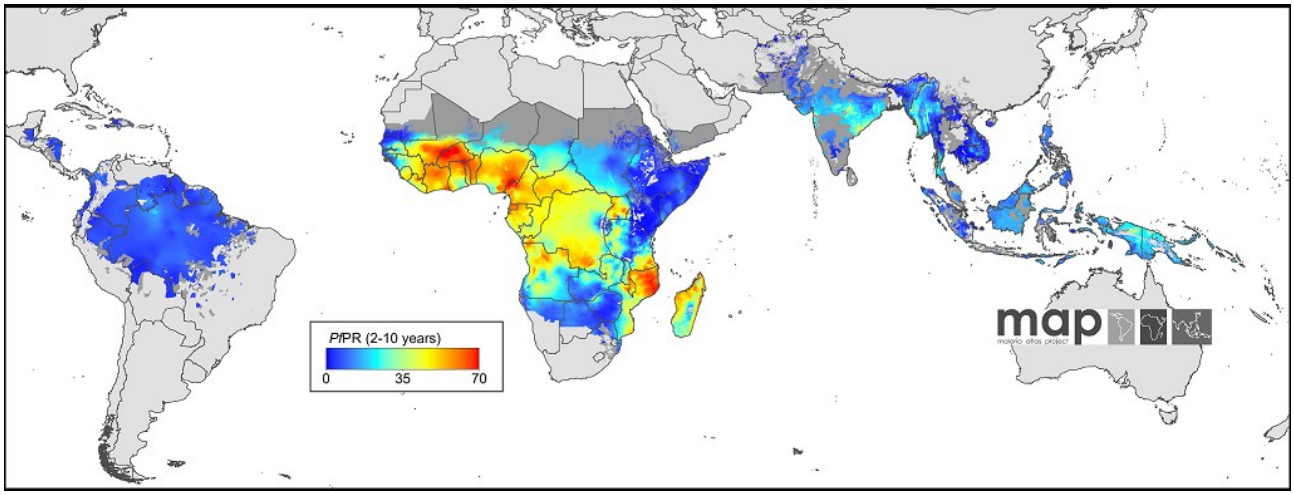


Figure 2.1: The spatial distribution of P. falciparum malaria endemicity in 2010. The land area was defined as no risk (light grey), unstable risk (medium grey) and annual mean PfPR for children between the age of 2 and 10 is presented as a continuum of blue to red from 0%-100% (see map legend). Reproduced from Gething et al [14]**.**

Due to its high associated mortality, its resistance to some anti-malaria drugs and its widespread prevalence in Africa, *P.falciparum* has been regarded as the greatest threat [18, 19]. However, *P. vivax* has a wider geographical range and exposes more people at risk of transmission worldwide, with 2.85 billion people estimated to be at risk [20].The vast majority of people affected by *P. vivax* (91%) live in the Central and South East Asia region [20]. The endemic areas of *P. vivax* often overlap with those affected by *P.*

falciparum with the exception of temperate zones such as sub-Saharan Africa, where populations lack Duffy glycoprotein expression of the red-blood cells which prevents invasion of the *P. vivax* merozoite [20, 21]. There are an estimated 80 to 300 million clinical cases of *P. vivax* every year [21].

1.2.1.2 Malaria control and elimination

Malaria affects essentially poor countries and represents an enormous health as well as a huge economic burden on these populations. . Indeed, if elimination or even control could be achieved and sustained in malaria endemic countries, not only millions of malaria cases would be averted but economic benefits would be observed at different levels including public health, households and industry, and school absenteeism would certainly be reduced [22].

The recent resurgence of interest in malaria elimination [23, 24] along with the call from the Bill and Melinda Gates Foundation [10] for worldwide malaria eradication has proactively engaged the attention of the international health community concerning the global fight against malaria. The target is no longer to control malaria – i.e. to treat disease but rather to eradicate malaria (See Table 2.1 for definitions) or as stated “to reach a day when no human being has malaria, and no mosquito on earth is carrying it". Funds have become available not only to prevent disease and death but to move towards eradication by reducing transmission [25, 26]. Lessons learned from the past have shown that if malaria control activities are reduced after aggressive scale-up, it can have a catastrophic impact [22]. In order to achieve and maintain malaria elimination/control, existing interventions need not only to be scaled up but new interventions and monitoring systems need to be developed [27, 28]. As a result, as transmission intensity decreases, the need for efficient surveillance tools becomes crucial [29], particularly for areas with low levels of transmission. Serological methods are therefore increasingly used for this purpose [30].

Table 2.1: Definition of Control, Elimination and Eradication [31]

Term	**Definition**
Control	Reducing the disease burden to a level at which it is no longer a public health problem
Elimination	Interrupting local mosquito-borne malaria transmission in a defined geographical area, i.e zero incidence of locally contracted cases, although imported cases will continue to occur. Continued intervention measures are required
Eradication	Permanent reduction to zero of the worldwide incidence of malaria infection

1.2.1.3 Malaria interventions

The unsuccessful Global Malaria Eradication Program (GMEP), discontinued in 1969, has shown that malaria elimination cannot be tackled with a global single intervention. Combinations of interventions would be required to be sustainable but firstly, in addition to successful impact of the interventions on

transmission, this program would need long term commitments and the development of efficient surveillance systems [6].

Financial support for malaria research has helped to introduce or scale-up specific interventions to decrease the malaria morbidity and mortality [17, 32]. Clinical malaria can be treated with anti-malarial drugs that target blood-stage infection and clear parasitaemia. Historically, Quinine was used to treat malaria patients. Subsequently Chloroquine was considered first-line therapy until parasite resistance developed against the drug. Today, artemisinin combination therapies (ACT) are currently the recommended first-line treatment in many countries [15]. Malaria burden can also be reduced with the use of chemoprophylaxis and intermittent preventive treatment (IPT) targeting high risk groups, i.e. young children and pregnant women. Additionally, Mass Drug Administration (MDA), in which involves repeatedly treating the entire population and Mass Screening and Treating (MSAT), in which only those diagnosed as parasite positive are treated, can be used to support malaria elimination [33]. In addition to treating or preventing clinical malaria, antimalarials can also reduce transmission. Artemisinin can have some effect in reducing the number of circulating gametocytes. Primaquine and Tafenoquine are also efficient drugs against mature gametocytes [34, 35] but can cause haemolysis in individuals who are glucose-6-phosphate dehydrogenaze (G6PD) deficient [26].

Above all, the most efficient interventions for reducing transmission intensity remain vector control interventions, either by preventing contact between human and vector or by killing the mosquitoes. Use of insecticide treated nets (ITN) or long lasting insecticide-treated nets (LLIN) within households has increased in many African countries [32, 36–38]. Additional vector-control measures have been deployed in some locations, including indoor residual sprays (IRS) and larval control [39]. These interventions have already been associated with encouraging progress in some areas [15]. For instance, in Bioko, following the implementation of a control program with IRS, a successful decrease of 42% of the number of infection was recorded between 2002 and 2004. Also, Langeler conducted a review examining 22 studies and concluded that ITNs had a 17% protection efficacy against mortality in children [40]. Additionally, in southern Africa, Somalia, Eritrea, Rwanda as well as coastal Kenya, The Gambia and on the islands of Zanzibar, and Sao Tome and Principe, declining trends of malaria have been observed following the implementation of vector control interventions [39, 41–48]. Planning, implementing and measuring the impact of malaria interventions, however, depends on the transmission settings.

As the number of new diagnostics, drugs, vaccine and insecticides is increasing, the number of challenges to achieve malaria elimination is unfortunately also increasing [28]. With a call for full coverage with insecticide-treated nets and the roll-out of artemisinin combination therapy (ACT), resistance to insecticide and tolerance/resistance to drugs is rapidly spreading [49, 50], threatening the effectiveness of

interventions. Additionally, *P. vivax*, which has been under less attention, represents an important challenge that needs to be tackled for successful elimination of malaria.

1.2.2 Malaria biology

1.2.2.1 Life cycle

An infected mosquito injects sporozoites into the host (Figure 2.2); these reach the bloodstream and make their way to the liver. Sporozoites invade hepatocytes and differentiate into trophozoites. Trophozoites subsequently undergo schizonic development and mitotic replications. After thousands of replications, hepatocytes release merozoites into the bloodstream. For *P. vivax*, some of these differentiate into the dormant stage, termed hypnozoites, that will stay in the liver and can cause relapse weeks, months or years later. Merozoites are released in the bloodstream and quickly invade red blood cells (erythrocytes); further asexual reproduction takes place in 48 hour cycles for *P. falciparum*, releasing more merozoites by rupture of red blood cells. This stage gives rise to the majority of malaria symptoms. In *P. vivax* infection, merozoites can differentiate into mature gametocytes before any clinical infection and illness develops. Some merozoites develop into male or female gametocytes which initially mature while sequestered in the vasculature. Gametocytes are then taken in a mosquito's blood meal and sexual reproduction occurs in the mosquito midgut, forming diploid zygotes (ookinetes). These penetrate the midgut walls and form oocysts. Meiosis occurs and produces haploid sporozoites which migrate to the salivary glands to form sporozoites and complete the cycle.

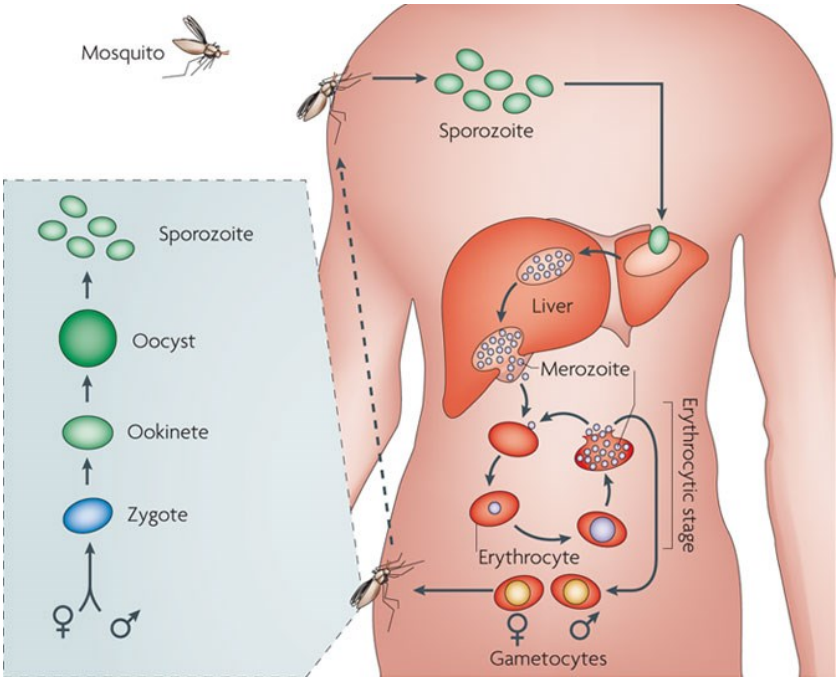


Figure 2.2: Life cycle of Plasmodium falciparum. Reproduced from Dondorp et al [51]

1.2.2.2 Clinical presentation

Infections with *P.falciparum* or *P. vivax* can be asymptomatic or develop into malaria disease which can be uncomplicated or severe. Individuals have asymptomatic malaria when they carry parasites but do not develop any clinical symptoms of infection. Uncomplicated malaria is the most common presentation of the disease. The symptoms are usually fever, headache, weakness and nausea. Severe malaria is often characterised by three main syndromes, i.e. anaemia, cerebral and respiratory distress. Other disease syndromes include seizures and hypoglycaemia. The acute symptoms of *P. vivax* including fever, body ache and headache have resulted in *P. vivax* malaria being classified as a benign disease due to its inability to adhere to the vascular endothelium [20, 21, 52]. However, recent studies have revealed a significant risk of severe disease and death caused by *P. vivax* [18, 19, 53, 54]. Another clinical characteristics of *P. vivax* that distinguishes it from P.*falciparum* is "clinical relapse" - the re-emergence of parasitaemia from liver-stage hypnozoites - which can happen weeks or months after exposure. In endemic settings this is difficult to distinguish from reinfection from biting mosquitoes or re-occurrence (clinical symptoms from a previously subclinical parasitaemia) and hence is to date best characterised in travellers returning from endemic countries [18, 53].

1.2.2.3 Immunity to malaria

Populations living in malaria endemic areas develop immunity to malaria. However, protection against malaria is rarely life-long [55] and is often only partial. Indeed, infections tend to be controlled and tolerated rather than eliminated or prevented [56]. There are different levels of immunity, acquired at different rates (Figure 2.3). Protection against severe malaria is acquired most rapidly, followed by immunity to febrile/uncomplicated clinical disease and finally immunity to asymptomatic blood-stage parasitaemia [57]. The acquisition of immunity depends on exposure [58] and age [59–61]. Infants are first protected by maternal immunity but then become vulnerable to malaria episodes [62]. As age increases, there is a decreasing probability of children experiencing a malaria attack, presented in studies in areas of high endemicity to be independent of cumulative exposure [63]. Also, it was shown that for non-immune individuals migrating to endemic areas that adults acquire a faster protection from clinical attacks than children [64], demonstrating an age-dependency of the acquisition of immunity. However, adults were also more susceptible to severe malaria than children, inferring an exposure-dependency of the acquisition of immunity against severe malaria.

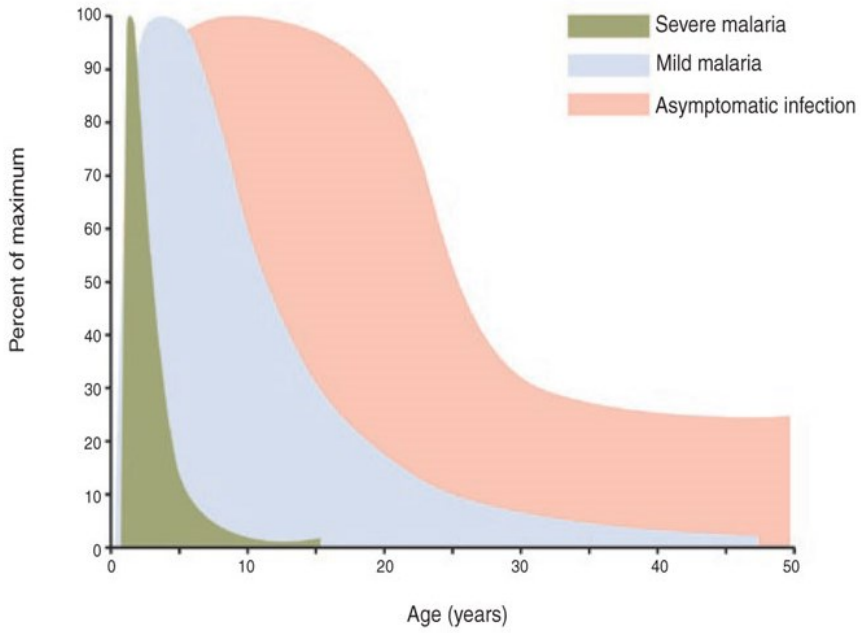


Figure 2.3: Indices of immunity to P. falciparum for severe, mild and asymptomatic malaria. Reproduced from Langhorne et al. [65]**.**

The effect of age on the acquisition of immunity is often difficult to distinguish from the effect of exposure [66] as age represents both an indicator for the duration of exposure but also determines the maturity of the immune system. Additionally, adults, with larger body surface area than children, tend to receive more infectious bites and develop a higher immunity [67].

However, there is an age pattern in immunity to malaria that changes with malaria transmission intensity [58]. The risk of severe malaria in childhood is higher for areas of high intensity and a comparable trend is observed for clinical episodes of malaria (Figure 2.4). Indeed, in high transmission settings, morbidity and mortality mainly affect children at a very young age, while adults who have acquired immunity to severe and mild malaria tend to be asymptomatic. On the contrary, in areas of low endemicity individuals are, independently of their age, at risk of malaria disease (severe or mild). Immunity in these populations is relatively poor and it appears that continuous exposure to the parasite is required to sustain high level of immunity. As a result, reducing transmission intensity will interfere with the natural acquisition of immunity to malaria and might not provide the expected reduction of malaria burden [58, 68, 69].

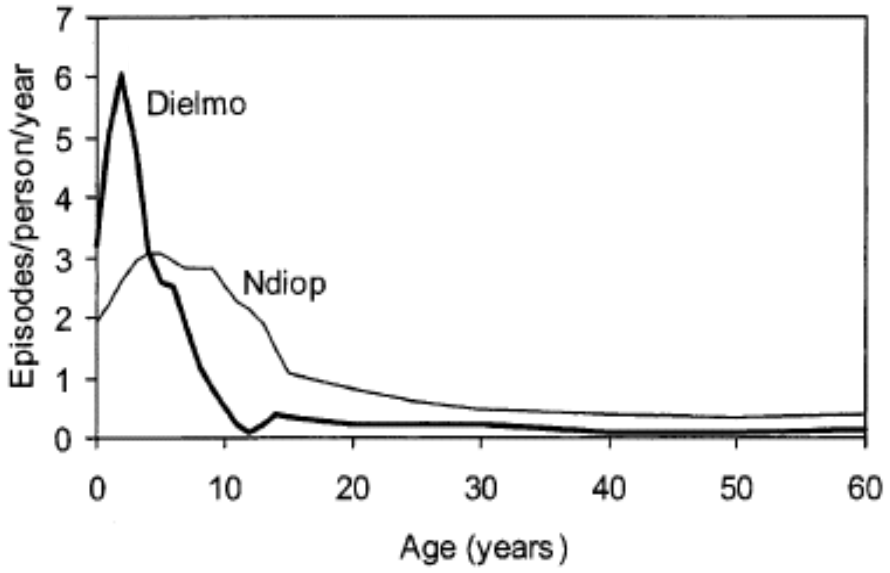


Figure 2.4**: Incidence of clinical malaria in Dielmo (high transmission) and Ndiop (low transmission). Reproduced from T. Smith et al.** [69]**.**

1.3 Malaria transmission

1.3.1 Malaria vector: *Anopheles* mosquito

Anopheles is the genus responsible for the transmission of human malaria. Amongst the 430 *Anopheles* species, 70 of these are malaria vectors and only 40 are of major importance [70]. *Anopheles gambiae s.s., An. Arabiensis* and *An. Funestus* are the most prominent malaria vectors in Africa. The development of the mosquito consists of four stages: egg, larva, pupa and adults and initially takes place in water. Mosquitoes are usually not found more than 2-3 km from their breeding site. Only female *Anopheles* can transmit malaria parasites. Some characteristics of the mosquito might influence the spread of malaria, such as its preference to feed on animals (zoophilic) or on humans (anthropophilic), indoors (endophilic) or outdoors (exophilic), or the longevity of the mosquitoes. Indeed, for mosquitoes lifespans less than 10 days, the parasite is unlikely to have enough time to complete its cycle within the mosquito [71]. Also, temperature and humidity might influence the development and survival of certain species of mosquitoes [57].

1.3.2 Heterogeneity of transmission

Levels of malaria endemicity vary widely between settings, with prevalence of infection ranging close to 0% in low endemicity areas to levels higher than 40% in high endemicity areas for children between 2 and 10 years old [13]. This variation occurs across a range of scales, from differences between and within continents down to variation at the village level and so-called “hotspots” of transmission [72]. Indeed heterogeneity in endemicity of malaria is mainly driven by heterogeneity of malaria transmission. A number of determinants are responsible for malaria transmission [57] and therefore any heterogeneity due to

those factors might result in heterogeneity of transmission. On the one hand, local ecological and climate-based factors are important determinants of the intensity of malaria transmission. For example, temperature affects both vector and parasite development [73], whilst rainfall affects the availability of breeding sites (which vary by species). Both these environmental factors in part determine the seasonal patterns of malaria transmission observed in certain regions of the world [74]. Indeed, many endemic areas are characterised by seasonality of transmission, with low transmission during the dry season and peaks of transmission during the rainy season [75]. Altitude, despite a high correlation with temperature, has also been shown to have an effect on transmission. In addition, many human factors, including antimalarial interventions, socio-economic and behavioural factors, agriculture, land use, urbanization and population movements contribute to variation in the transmission dynamics of malaria.

Additionally, factors related to the efficiency of an infectious mosquito bite are highly related to efficiency of transmission. Indeed, not all infectious mosquito bites progress to blood-stage infection. Therefore heterogeneity in factors related to the efficiency of an infectious mosquito bite can also contribute to the heterogeneity in transmission. Immunity and heterogeneity in mosquito biting have been suggested to be potential causes of this inefficiency of transmission [76]. And, some studies have shown that transmission efficiency decreases with increasing number of infectious mosquito bites per person. Therefore, the hypothesis of heterogeneous biting might be the most plausible to explain the inefficiency of transmission [76] and therefore heterogeneity in transmission. Indeed, Smith has shown that contact between vectors and humans are not random [77] and Carnevale et al. have demonstrated some heterogeneity in mosquito biting with age-related biting patterns [78].

1.3.3 Measuring malaria transmission intensity

An understanding of the relationship between transmission intensity, prevalence of infection and clinical incidence is key for malaria epidemiology. Therefore, a number of methods that measure malaria endemicity have been considered as measurements for malaria transmission intensity.

1.3.3.1 The reproduction number R_0

The basic reproduction number R_0 determines the endemic level of a disease [79]. It is defined as the average number of secondary infections produced from one infected individual introduced in a non-immune and fully susceptible host population [80, 81]. If R_0 is greater than one, the number of infected people increases and if R_0 is less than one then it decreases. For malaria, R_0 is mathematically defined as:

$$R_0 = ma^2bcp^n / (-\log p) \qquad (1.1)$$

where *m* is the ratio of vectors to humans, *a* is the biting rate, *p* the probability that the mosquito survives one day, *n* the extrinsic incubation period, *b* the infectivity of mosquitoes to human and *c* the infectivity of

human to mosquitoes. R_0 provides an index for transmission intensity and is used as a threshold criterion, but its estimation relies on direct measurements of endemicity in the area [82]. A selection of these is presented in Table 1.2.

1.3.3.2 Entomological inoculation Rate (EIR)

The main method for measuring malaria transmission intensity is the entomological inoculation rate (EIR) which is defined as the number of infective mosquitoes bites received per person per unit of time. This is expressed mathematically as:

$$EIR = mas \tag{1.2}$$

where *m* is the number of mosquitoes per person, *a* the biting rate of mosquitoes and *s* the proportion of mosquitoes with sporozoites detectable in their salivary glands [73]. EIR represents a direct measurement of transmission intensity [83]. Indeed, the human biting rate (*ma*) can be directly measured with light-traps or the human bait catch, which consists of collecting mosquitoes trying to feed on exposed individuals, dissecting them and looking for the presence of sporozoites in the salivary glands [73]. EIR data are expensive, difficult to replicate, and are also affected by seasonal variation and geographic over-dispersion of vectors. Furthermore, in areas of low transmission, sampling methods become insensitive as mosquito numbers are low and sometimes below the detection level [73, 84]. In these settings, only a very small proportion of mosquitoes are infectious and it therefore becomes challenging to measure malaria transmission intensity with EIR.

1.3.3.3 Parasite Prevalence

Alternatively, an examination of the peripheral blood for asexual malaria parasite by microscopy would provide more specificity for malaria infection than EIR. The parasite rate (PR) (strictly prevalence not a rate) is defined as the proportion of individuals in the population carrying parasites in their blood and widely used as a measure of endemicity. The PR is typically measured in cross-sectional surveys in communities with the different parasite species distinguished. However, parasite surveys are relatively invasive for participants (requiring blood samples) and logistically demanding. In seasonal settings, interpretation is further complicated as few studies undertake surveys repeatedly throughout the year. Furthermore, it is increasingly being recognised that microscopic methods may miss a substantial proportion of "sub-microscopic" low density infections [85]. Therefore in low transmission settings, parasite prevalence might not be appropriate as it provides unreliable measures of transmission intensity [82, 84], by either missing low density infections or by overestimating levels of transmission due to a lack of sensitivity of the methods [83]. More recently, Polymerase Chain Reaction (PCR) techniques have been developed to improve parasite

detection [73]. Rapid Diagnostic Test (RDT) remains as cheap, easy-to-use and a quick method for detecting malaria antigen from small amount of blood [86].

A particular feature of malaria is that the prevalence of infection in a population attains saturation very quickly. Indeed, there is a non-linear relationship between prevalence and the incidence of infections. Prevalence can change little across a wide range of incidences. It appears therefore difficult to detect incidence from cross-sectional parasite rate surveys across the whole population. The parasite conversion rate for malaria, first introduced by MacDonald, is a method to estimate the force of infection using the infant parasite rate and is defined as the per capita rate at which susceptible individuals contract the infection [87]. The force of infection is defined as the rate of infectious bites successfully causing a blood stage infection. Several methods were suggested for its measurement [88–90]. There is a relationship between EIR and force of infection, although this is not always linear due to heterogeneous biting (not always considered in the estimates of force of infection) [78, 91] and acquisition of pre-erythrocytic immunity (faster in adults than children exposed to high EIR) [60, 61, 67]. Several studies have explored the relationship between EIR, parasite prevalence and the force of infection [67, 83, 92, 93]. Beier *et al.* found a linear relationship for most of the studied sites (31 sites throughout Africa) between parasite prevalence and the logarithm of annual EIR [83] while Smith *et al.* developed a model to predict the force of infection from EIR [67].

1.3.3.4 Serological markers

An alternative method to estimate the force of infection is to distinguish between individuals positive to antimalarial antibodies and those who are not. The proportion of the population who are seropositive can be used to measure exposure [89, 94], but the rate of seroconversion, defined as the rate at which individuals become seropositive, is the preferred method to determine the force of infection from serological data. Models for age-specific antibody prevalence are used to estimate this rate. Serological methods have recently been widely applied in order to estimate endemicity [95]. Exposed individuals may remain seropositive many years after they have been infected [96] and thus the seroprevalence rate represents a tool for assessing malaria exposure over time, as it smoothes the effect of seasonal variation due to the persistence of the antibodies. Serological methods have been proposed as a technique to rapidly assess malaria exposure [95, 97, 98]. One advantage of these methods is that they are simple and cheap. Serological methods are easily reproducible and easily interpretable, which therefore represents a perfectly adequate tool to analyse data in the field. They may also provide more accurate results than parasite rates in areas of low endemicity [99] and therefore help to distinguish differences in exposure when malaria parasites are not detected [99]. Serological markers can identify asymptomatic infections, i.e. individuals at risk of transmitting malaria despite a lack of clinical symptoms.

1.3.3.5 Other methods

Historically the first method of malariometry was introduced in India in 1848 and involved determining the spleen rate (the proportion of the population with an enlarged spleen) [82]. Determination of malaria transmission intensity using spleen rate might not provide great specificity and suggest misleading results. The Global Malaria Eradication Programme (GMEP), coordinated by WHO, tried in 1950 to establish a common nomenclature for measuring malaria. A consensus was finally reached for characterising the population based on spleen rate into 4 categories: hypoendemic (<10%), mesoendemic (11-50%), hyperendemic (51-75%) and holoendemic (>75%) when measured in children between 2-9 years old [100]. Later, the classification was used in conjunction with parasite rates [101].

Measuring clinical incidence (i.e. the number of clinical malaria episodes in a defined population over time) has also been considered as an option for determining transmission intensity. However, it requires a comprehensive surveillance system as well as active case detection and is only considered valid if the proportion of the target population examined is greater than 10% [15]. Moreover, estimation of disease prevalence based on hospital data has been shown to be unreliable due to over-diagnosis of malaria in some patients and under-diagnosis in other communities who do not have easy access to health facilities.

The distribution and transmission of malaria is strongly influenced by climatic factors. Climate-based methods have thus been explored to provide a proxy for malaria transmission intensity [102]. These methods are based on the combination of temperature and rainfall to provide a continuous scale for the probability of malaria infection in a particular area. These methods have been shown to provide estimates that corroborate with field data at regional and country level. However, their ability to estimate malaria transmission intensity at the level of individual communities is limited [103].

Recently, molecular techniques have demonstrated striking features for measuring malaria transmission intensity at both an individual and a population level [104–106]. The molecular force of infection ($_{mol}$FOI) can be estimated with high sensitivity by genotyping *Plasmodium* parasites in longitudinal studies. $_{mol}$FOI can be used to measure malaria transmission intensity in particular in areas of low endemicity where individual and small geographic differences might be important [104]. Malaria transmission intensity can be highly variable even at a small spatial scale [107, 108] and is rarely detected at an individual scale with traditional measures of transmission. One of the great advantages of molecular techniques is therefore that they can capture the microvariability in exposure. $_{mol}$FOI is also useful for measuring the effect of seasonality and age on transmission intensity. However, as natural immunity might prevent infections, $_{mol}$FOI will better estimate differences in transmission with individual with limited levels of naturally acquired immunity [104].

Table 2.2: Summary of measurements of malaria transmission

Measurements of transmission	Advantages	Caveats
Entomological Inoculation Rate (EIR)	+ Direct measurement of transmission through the mosquito	- Expensive, difficult to replicate, affected by seasonal variation and geographic over-dispersion, time consuming, lack precision (low endemicity)
Parasitological markers	+ Direct measurement of malaria infections in humans	- Invasive method affected by seasonal variation, drug use, acquired immunity; underestimate prevalence of infection in low endemic areas (miss parasitaemia of low densities); logistically demanding
Serological markers	+ Simple, rapid, easily reproducible and interpretable (quick & easy application for field data)	- Cannot distinguish between past and present infection
Spleen rate	+ First method suggested to measure malaria	- Non-specific test, overestimate the number of infections
Clinical incidence	+ Data directly available from health facilities	- Overestimation of number of cases in health facilities (overdiagnosis) and underestimation in communities with no access to health facilities
Climate-based markers	+ Good predictor at regional and country level	- No direct measurements in communities
Molecular markers	+ Good predictor at individual and population levels, great sensitivity	- Restricted to individuals with limited acquired immunity, requires highly skilled staff and material

1.4 Seroepidemiology of malaria

1.4.1 Immunology of malaria

1.4.1.1 Humoral mechanisms

When the malaria parasite invades the human host, a series of diverse immunological processes are triggered to protect the host from an infection [109]. The adaptive immune response is mounted as a result of antigen interaction with lymphocyte cells, following the activation of the initial, innate immune response. The T cells generate the cell-mediated immunity and B cells the humoral immunity. B cells differentiate into plasma cells with the help of T cells and secrete a range of immunoglobulin antibodies [110]. There are four main classes of immunoglobulin antibodies, called isotypes: IgM, IgA, IgG and IgE. IgM antibodies are the first antibodies produced by B cells upon infection. Subsequent maturation of the antibody response, often requiring multiple exposure to the same antigen, will trigger the production of a mature antibody response with expression of other isotypes (IgG in particular, being the most abundant), with high affinity and immunological memory. Antibodies play a crucial role in immunity to malaria by preventing the invasion of the parasite with various mechanisms, including opsonisation of merozoites in the case of malaria, to facilitate their phagocytosis by macrophage [111].

An immune response can be triggered at any point in the life cycle of the parasite in the human. As sporozoites, liver stage parasites and gametocytes do not trigger any symptoms, it has been argued that these stages are poorly immunogenic [56]. Studies have shown that pre-erythrocyte stages present limited naturally-acquired immunity [8, 56] as they are substantially short lived. For blood stages, the potential targets for an immune response are the free merozoite or the intra-erythrocytic parasites[65] when the parasite is directly exposed to host immune system. Indeed, intracellular forms (in the liver and in the red blood cells) tend to hide from antibodies. However, blood stage antigens display huge antigen diversity that challenges the immune response.

1.4.1.2 Blood stage antibodies

Anti-malarial antibodies have been known to be associated with malaria clinical immunity [112]. However, the correlation between antibodies and the protection against clinical episodes or death remains unclear. For epidemiological studies, only detectable antibodies used as markers of infection are of interest, and not their role in protection. Therefore, antibodies to the asexual blood stage merozoite antigen, which tend to be more abundant, are the most relevant. Merozoite Surface Protein 1 (MSP-1) is a protein synthesized during schizogony. MSP-1 represents a target of immune responses with antibodies against MSP-1 neutralising the parasite by agglutinating merozoites and preventing red blood cell invasion. Apical

merozoite surface antigen 1 (AMA-1) is a membrane protein located in the apical organelles of developing and free merozoites [113]. Antibodies against AMA-1 prevent the malaria parasite from infecting red blood cells. MSP-1 and AMA-1 are typically found on merozoites, however, AMA-1 can also be found on sporozoites. AMA-1, known to be highly immunogenic [114], is more appropriate to analyse in order to increase the sensitivity of the methods in areas of low endemicity [97]. It is a bigger protein than MSP-1 and has a higher merozoites surface expression rate. As a result, the antibody response is typically higher for AMA-1 compared to MSP-1. However, despite MSP-1 being less immunogenic, antibodies against MSP-1 have been established as useful tools that allow differentiation between short term variations (seasonal) and long term pattern (year to year) of malaria transmission [95]. Additionally, it was shown that once acquired, blood stage antibodies persist for many years [115].

1.4.1.3 Longevity of antibodies

Antibodies appear in the blood not long after parasites invade the bloodstream and may remain there between a few months to a few years. The debate on longevity of antibodies remains controversial [96, 116–120]. Studies have shown that antibody response in vaccinated children tend to be long lasting [121], however a better knowledge of the persistence of antibodies in absence of vaccination would help to differentiate individuals recently infected from those infected months or years before.

On one hand, studies have shown that antibody titres rapidly decline in the absence of re-infection following the time of acute infection, suggesting that naturally acquired immunity against merozoite antigens is short lived following an acute infection [117]**.** A survey conducted in Kenya showed indeed that for children living in endemic areas, antibody responses against merozoites are often very short-lived [116]. It is widely held that immunity to malaria can be lost after a period of time in the absence of re-infection or when an immune person moves away from an endemic area. The underlying hypothesis put forward for explaining this phenomenon is that persistent or frequent exposure to malaria is required to maintain immunity [56]. In endemic areas, blood stage parasites are maintained at low levels by the immune response due to frequent exposure. Under these conditions and in absence of treatment, seropositivity is likely to be maintained [56, 122].

On the other hand, other studies have shown that antibody responses could persist for a very long time [119, 123]**.** A recent study focusing on antibody responses including MSP-1 and AMA-1 has classified antibodies against those antigens as long-lived antibody response once acquired in asymptomatic infections. In such circumstances, antibody responses to those defined malaria antigens can persist for several years for *P. falciparum* and *P. vivax* [96, 117]. Infrequent malaria infections could therefore induce long-lived antibody response [96]. Surveys carried out in Madagascar after almost 30 years of absence of malaria demonstrated that both immune response and protection (from clinical malaria) were observed in

adults previously exposed to malaria and not in young children [124]. Additionally, studies have shown that the mortality rate from malaria is higher in naïve travellers compared to immune people revisiting endemic areas [56].

Another study in a highly endemic area also showed that the antibody response in children presented was short-lived [117] suggesting that long lived immune responses might be acquired after a high number of infections [96]. Therefore the longevity of the immune response might be determined by numerous factors including exposure and age. Indeed, antibodies to merozoite antigens have been shown to decline more rapidly in younger children than in older children. The inability of young children to produce sustained antibody response suggests that age might be a determinant for sustaining antibody levels. The longevity of the antibody response might also differ between antigens and *Plasmodium* strains [117]. For example, in *P.vivax* infections, antibodies can wane between relapses but they rarely completely disappear [125].

1.4.2 Detection of antibodies

Two established immunodiagnostic techniques, introduced by Voller in the late 1970's, are typically used for the detection of malaria antibodies [125]. Traditionally, the indirect method of immune fluorescent antibody (IFA test) was the reference for malarial antibody determination. However, the Enzyme-linked immunosorbent assay method (ELISA) has higher sensitivity and has become the most common test. IFA involves reaction of diluted test serum samples with drops of infected blood dried on microscope slides; the slides are then reacted with fluorescent antiglobulin and finally examined under a fluorescent microscope. Fluorescent antiglobulin only fixes itself to the antigen-antibody complex, indicating that the sera contain antibodies. The antibody level is determined by the last dilution that reacts and is given as a 'titre' [125]. The indirect ELISA method is similar to the IFA test. Antibodies to specific antigens are detected when they bind to a micro-titre plate using a colorimetric enzyme reaction. The amount of antibody present in the serum is determined as being proportional to the amount of colour produced by the test. Optical density is very often used as a proxy of antibody density. Both methods have the advantages of being low technology, easy to reproduce and antigens can be stored for long periods. However, IFA can be laborious for large samples and can potentially be subjective as it relies on the technician's expertise to visually read the results. In contrast, ELISA results are easily interpretable for large epidemiological studies and results can be read with more accuracy as it is better standardised [125] with the wider use of recombinant antigens. Despite a remaining need for broader Quality Control and assay validation, findings can be compared from different laboratories when methods are standardised [126].

1.4.3 Seroepidemiological studies

The presence of anti-malarial antibodies in an individual infers that this person has been exposed to malaria. As a result, serology becomes a valuable tool for epidemiological studies. However, due to remaining maternal antibodies, cross reactivity with other pathogens or other factors, some individuals have detectable levels on antibodies while they have never been exposed to malaria. Defining seropositivity is thus essential, i.e. defining an antibody level above which individuals who have detectable levels of malaria antibodies are considered having been exposed to malaria. The seronegative population is composed of individuals who have never been exposed and those who have previously been exposed and have lost their antibodies.

Conventional serology has provided useful epidemiological information in malaria control programmes and has contributed to define malaria transmission areas and monitor the impact of control interventions. The outcomes of seroepidemiological studies are typically the seroprevalence of infection, a measure of the intensity of transmission and the impact of control measures if the survey was carried out for this purpose. Serological methods are based on detection of circulating antibodies. The amount of antibodies can vary due to relapse, superinfection or reinfection and total exposure history. Serological surveys can be cross-sectional surveys (most common), repeated cross sectional survey or longitudinal surveys. Results are typically recorded by age as the acquisition of antibodies is age dependent [60, 127, 128] and presented with age-seroprevalence curves (Figure 2.5).

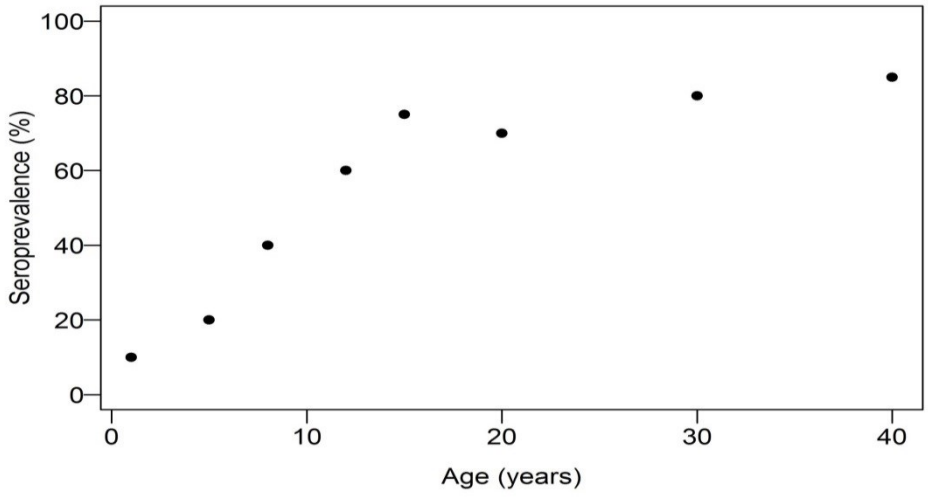


Figure 2.5: Age specific seroprevalence curve. The y-axis represents the proportion of seropositive individuals (seroprevalence) in each age group and the x-axis shows the midpoint of each age group. Fictitious data are represented by the black dots.

The traditional indices of malaria endemicity such as EIR or parasite rate allow an assessment of current transmission and consequently might not reflect malaria transmission intensity over a period of time. Indeed, transmission intensity might be underestimated or overestimated due to false negative or false

positive results attributable to the fluctuating nature or low levels of parasitaemia or even caused by the geographic over dispersion of vectors. Serology can give an improved picture of the intensity of transmission by providing period prevalence [125] and can be used to assess changes in transmission in areas where transmission is considered stable. Consequently, when serology was first established for epidemiological studies, many serological surveys were carried out in different parts of the world. Indeed seroepidemiological studies were performed in Mauritius and Greece to confirm malaria eradication [8], in Tunisia to assess the presence of residual transmission foci after interventions [129], in West Africa [130] and Nigeria during the Garki project [131, 132] to investigate the epidemiology of malaria and measure the impact of house-spraying alone and in combination with mass drug administration. Surveys carried out to confirm elimination of malaria [8] have been performed on children, since long persisting antibodies produced before interventions can confuse the results.

Today an increasing number of studies use serological methods to assess malaria endemicity and risk [95, 99, 133–137] or as a tool for surveillance for malaria elimination [138]. Serological surveys are an established tool to measure malaria transmission intensity [95, 97, 98, 137, 139, 140]. Serological methods can also be useful tools for measuring variation in malaria transmission over time. Studies have shown that such methods are capable of distinguishing between long term patterns of malaria transmission from short-term variations [95, 97]. Similarly, a cross-sectional survey, performed in Vanuatu where malaria transmission had been widely reduced over the past years [135], highlighted the potential for serological methods to monitor changes in malaria transmission. As a result, serology has become an established tool to measure changes in transmission and was used for instance in Cambodia to show changes in transmission pattern during the rainy season [134] and in Bioko to demonstrate the heterogeneity of the effectiveness of the interventions [133]. Additionally, a study conducted in Somalia, showed that serological markers can be used to determine heterogeneity of malaria transmission in areas of low endemicity, where parasitaemia is undetectable [99].

Serological methods also have some drawbacks. Saturation of prevalence at high transmission intensity or very long lived antibodies can be problematic for detecting recent variations from historical trends [97, 123]. As a result, they might not be appropriate to assess malaria exposure at the individual level.

1.5 Mathematical models

> *"Models are a useful means of collating knowledge and experience to determine whether success is possible under a given set of constraints and conditions, and, if not what changes are required."* [141]

1.5.1 Why model malaria?

Mathematical models are routinely used in epidemiology to simulate disease transmission, identify the key factors of transmission and inform the underlying process that drives the transmission dynamics. A model is a simplified representation of the complex reality but provides a good qualitative description of the system. Additionally, mathematical modelling also represents a compromise between "purely applied" and "purely theoretical" approaches that allows extrapolation of the findings when the collection of data is expensive, large, ethically challenging or time-consuming.

One of the first model for infectious disease originates with the malaria model developed by Ross during a trip for malaria control in Mauritius [5] and extended by MacDonald [142] during the Global Malaria Eradication Program (GMEP). These models were applied to guide malaria interventions and highlighted the importance of vector control by spraying with DDT [143]. Later, Dietz further developed the simple models to include immunity and superinfection during the Garki project in Nigeria [144]. This model played a key role in the design and analysis of the interventions. New indices to measure transmission were then introduced, notably vectorial capacity and the human blood index [145, 146]. Subsequent extensions of the basic models were developed to consider heterogeneity [147, 148], immunity [149], within host dynamics of the parasite [150], interventions [151], strain theory [152] and other phenomenon. Today, mathematical models are still developed with different levels of complexity in order to evaluate the impact of interventions to inform policy and guide research for control and elimination [93, 148, 153–155].

There is an extensive literature on modelling the dynamics of malaria transmission. Most of these models are based on parasitological, entomological, clinical and epidemiological data. However, a much lower number of models are developed for serological data.

1.5.2 Assessing transmission using models for serological data

1.5.2.1 Catalytic models

A variety of mathematical models for serological data have been developed [127, 156, 157], of which the catalytic model is most widely used in the context of seroepidemiology. This was first introduced for malaria by Draper, Voller and Carpenter in 1972 and termed the 'constant infection rate model' as it can be considered a catalytic model with a constant force of infection [158] . This early model was applied in east Africa [159], Mauritius [8] and Guyana [160] to estimate past infection rates. In these models human hosts were assumed to move from seronegative to seropositive with a constant seroconversion rate λ. Initially, the decline of antibodies was not considered and hence the proportion seropositive at age t is given by:

$$y(t) = 1 - e^{-\lambda t} \qquad (1.3)$$

VanDruten [161] subsequently proposed a model incorporating a decay of antibodies with a reversion from seropositive to seronegative occurring at rate ρ (Figure 2.6). The proportion seropositive at age t is then:

$$y(t) = \frac{\lambda}{\lambda + \rho}(1 - e^{-(\lambda+\rho)t}) \qquad (1.4)$$

This reversible catalytic model corresponds to Ross's original malaria model for transmission [142].The seroconversion rate λ and the rate of seroreversion ρ can be obtained by fitting the model to empirical age-prevalence curves [97].

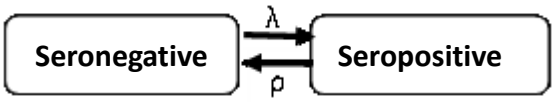


Figure 2.6: Flow diagram for catalytic model

Titre data can be converted into prevalence data using control sera from Europeans who represent truly unexposed individuals to define a threshold. However, different genetic make-up or exposure to other pathogens in the local area may mean that these do not represent appropriate controls. An alternative method is to use the data from endemic settings and define the cut-off using mixture models [162]. For a mixture model, positivity is defined as a measurement more than three standard deviation above the mean measurement of a panel of non-exposed individuals [163, 164]. However, in high prevalence settings there may be insufficient truly negative responders to appropriately define the cut-off threshold.

In catalytic models it is assumed that in a given interval of time, all individuals have the same probability of seroconverting (at rate λ) and this probability is a function of the immunogenicity of the antigen and the likelihood of being infected [97]. The outputs from such a model provide a proxy for malaria transmission. Indeed the seroconversion rate λ is closely related to the force of infection of malaria [165]. The seroconversion rate has also been shown to correlate with EIR [95]. However, note that the seroconversion rate is assumed to be independent of the antibody density. Indeed, the number of time an individual gets infected is considered unrelated to its current antibody level.

As a result, these models fitted to seroprevalence data allow a rapid and local assessment of malaria transmission intensity [95, 97–99, 135]. However, even though this approach is simple and practical the information contained in the continuous serological titre data is only partly considered. The use of catalytic models to evaluate transmission intensity based on serological data requires that the population is discretised between seropositive and seronegative individuals. The force of infection corresponds to the incidence of seroconverting. However, this discretisation and therefore the force of infection consequently highly rely on the choice of the threshold between seropositive and seronegative individuals, which lacks standardisation. Under these models any seropositive individuals that get infected (and therefore produce

higher levels of antibodies) would not be captured in the force of infection as such discretisation does not account for dynamics of individuals once they are seropositive. This simplistic model is indeed a partial representation of the reality as it does not address the complexity of biological mechanisms including antibody priming and boosting.

1.5.2.2 Other models

Many other models have been derived from the simple catalytic models and modified to allow the seroconversion rate to potentially vary with time or be a function of age and exposure [127].

Gatton presents a model [139] similar to the reversible catalytic model used in the Garki project [166] to estimate the impact of interventions targeted for elimination in a study area in Africa. However, unlike Van Druten's model, Gatton fits antibody decay to serological data collected in longitudinal surveys and uses it to derive malaria transmission rates from cross-sectional surveys. The decay probability is related to the time the antibody persisted at a high level in the individual. It is assumed that malaria prevalence is not age-dependant. The output from this model is the number of people on each day with high and low antibody levels. Survival curves are then produced and compared with the curves from the actual data. Once the decay rate has been estimated, a modified catalytic model is then used to derive the transmission rate of malaria.

Burattini [140, 167] developed a stochastic model to estimate malaria transmission rates from serological data, based on cross-sectional data. The compartmental structure of the model takes into consideration parasite and serological prevalence data. This model assumes an age-dependent force of infection and includes acquisition and exponential decay of antibodies (with mean persistence of 10 years). The model is applied to data from Brazilian Indians over a period of 25 years. They suggest that changes in malaria transmission indices are due to age rather than changes in malaria transmission in time.

1.5.3 Modelling antibody levels

1.5.3.1 Malaria models

All the models described so far have a compartmental structure and rely on fixing a cut off value for distinguishing seropositive from seronegative, or modelling the probability of belonging to each compartment accounting for misclassifications [168]. However, antibodies can persist for many years and individuals might stay seropositive for a while. To that extent, a binary structure does not represent the actual fluctuations in antibody levels. Indeed, the magnitude of the antibody response might reflect changes in transmission that seroprevalence might miss. Despite some occasional use of the frequency distribution of antibody titre to characterise malaria endemicity [169–171] or to identify risk factors for

malaria infections [128] , there has generally been a lack of methods developed to make full use of the distribution of antibody responses. Models fitted to titre data have however been used to assess the duration of antibody response using longitudinal data (M. White – personal communication).

Recently, Bretscher and colleagues have further developed mixture models [168] that considers antibody titres to derive the seroconversion rate, without any arbitrary values in order to estimate malaria transmission intensity [137]. The method uses finite mixture models with a decomposition of the range of antibody titres. Hidden Markov Chains are applied to estimate the rates of seroconversion and seroreversion from individual-level longitudinal data. This method is more robust to noise in titre measurements than a threshold-based method and makes better use of the information in the data as more weight is given to large titre changes than small ones. However, as longitudinal data is required, as transmission intensity decreases, the number and duration of follow-ups needs increase to ensure good precision, and this is unlikely to be achievable.

To my knowledge, no other studies have directly modelled antimalarial antibody titres using a mechanistic model which consider the underlying biological processes.

1.5.3.2 Other pathogens

Most mathematical models which mimic the antibody response generally assess the persistence of antibody response after vaccination [172, 173]. However, Wilson and colleagues used mathematical modelling for the generation of immune response and antibody titre in response to hepatitis B vaccines [174]. Their model quantifies the rate of antibody responses and the development of immunity. However, their model also lacks an explicit understanding of the mechanisms underlying vaccine-induced immune response.

Modelling antibody titre data is commonly used to inform epidemiological parameters. For instance, the kinetics of pertussis antibody response was modelled to assess the incidence of pertussis and infer the distribution of times from infection [175]. An age structured dynamic model was also constructed to assess the dependence of age on the rate of decline of antibodies to pertussis [176].

Additionally, age-stratified serological data has been modelled for Dengue to determine incidence of infection and its variation over time. A study carried out by Ferguson et al. [177] provides an appropriate method for obtaining a good estimate of the force of infection from cross-sectional serological data and will provide a starting point for the development of density models in this thesis.

1.6 The scope of the thesis

The overall aim of this PhD is to develop a density model for antibody dynamics that reproduces the age-structured distribution of antibodies from cross-sectional data and to establish its assessment of exposure as a valid tool for measuring malaria transmission intensity using serological data. This model will take into consideration the level of exposure in the area and the kinetics of the antibodies, accounting for antibody priming, boosting and decay to describe the full information contained in antibody levels. The specific objectives and the approach taken are outlined below.

- Chapter 2 presents, as a preliminary method, an extension of the catalytic model as a "proof of concept" to assess whether a model that takes into consideration multiple arbitrary levels of antibodies (rather than a single cut-off value) can provides estimates of the exposure rate that correlate with other measures of transmission.
- Chapter 3 presents a variety of continuous density models explored for different hypotheses for antibody boosting and for the acquisition of antibodies for individuals who have undetected levels of antibodies. A unique model, which will be further used in the subsequent chapters, is selected.
- Chapter 4 establishes the exposure rate estimated using the density model as a valid measure of transmission intensity by validating the method against currently used indices and additional seroconversion-based metrics. Methods are validated using well known data from Tanzania that has previously been used to demonstrate strong correlations between the seroconversion rate and malaria exposure [95].
- Chapter 5 applies the density model to a wider range of endemic settings to assess the robustness of the method. The density model is also further extended to account for additional complexities such as heterogeneity in exposure, spatial or temporal changes in transmission and an age effect on exposure.
- Chapter 6 summarises the key findings of the thesis, indicates the implications of the research, highlights the limitations of the methods and finally points out future directions.

Chapter 2: Development of an antibody density model to assess malaria transmission intensity – a discrete framework.

2.1 Introduction

Better estimates of malaria transmission intensity are invaluable for planning and monitoring malaria control interventions. In areas of low endemicity, the limited number of infected mosquitoes and the low density of parasites in the human host render entomological and parasitological methods inadequate to measure malaria exposure. In such settings, serology has shown to be more sensitive at detecting exposure [97, 99], due to the longevity of the antibodies [96, 117, 124]. Serological data, reflecting past exposure [178], have been used in various contexts as a tool for epidemiological monitoring [7, 125].

In Cambodia, malaria burden is relatively low [15] but most of the malaria cases reported are through passive case detection, when individuals consult health facilities [179]. This does not reflect the true picture of malaria intensity among remote populations. More importantly, the asymptomatic carriage of parasite that significantly contributes to malaria transmission is poorly documented in Cambodia. Malariometric indices are available from a few studies that have investigated malaria exposure [179]. The results have shown that transmission intensity in Cambodia is heterogeneous and characterised by forest malaria that represents a major problem for implementing effective control interventions. The distance to the forest has been identified as a risk factor for malaria exposure suggesting high transmission in the deep forests and the male human population representing a group at high risk of malaria infection [180]. Individuals who move into forests and forest-fringe areas are an important at-risk group. This includes forest-fringe inhabitants, temporary migrants, traditional forest inhabitant or new settlers who have been relocated to forest area [180]. Malaria interventions need to target these particular populations to successfully eliminate malaria.

In low transmission areas, malaria in one host very often consists of more than a single species. Both *Plasmodium falciparum* and *Plasmodium vivax* are present in Cambodia with some spatial heterogeneity, presumably due to a difference in ecological differences between areas [134]. Although little is known

about the interaction between species, *P. vivax* has been reported to become predominant over *P. falciparum* as malaria intensity decreases [179, 181, 182]. With the progress of *P. falciparum* specific interventions, such as vaccines, this raises some questions about the implication of reducing the prevalence of *P. falciparum* as it could result in an increase of infection with *P. vivax*. Nevertheless, in a seroepidemiological context, antibodies to both species have been used to estimate malaria exposure [95, 134, 136].

The outcome of serological data is often viewed as a binary event in which the host either does or does not present circulating antibodies. However, antibody titre is a continuous measurement. Seroprevalence, commonly used to summarise serological data, requires the definition of a threshold for seropositivity. The force of infection is typically determined using a catalytic model for the seroprevalence data [95, 97]. Summary statistics of antibody titres are used to describe current settings [132] but the full information contained in antibody titres has rarely been used to infer the force of infection [137]. Here I develop mathematical models to characterise the relationship between antibody levels and malaria exposure. This relationship is used to estimate malaria transmission intensity based on the magnitude of the antibody response. Models were fitted to Immunoglobulin G antibody titres to *P. falciparum* antigens collected in cross sectional study in Cambodia.

2.2 Setting

2.2.1 Data source

The Mekong region and Cambodia in particular are considered to be areas of low endemicity with substantial heterogeneity in exposure mainly due to forest-related malaria (See Figure 2.1) and heterogeneity of vectors population [179, 180, 183]. Also, the presence of both *P. falciparum* and *P. vivax* in this region can hinder the assessment of transmission intensity. In 2004, the Cambodia Malaria Baseline Survey (CBMS) was conducted in more than 8,000 individuals to measure the population “Knowledge, Attitude, Behaviour and Practice” (KABP) towards malaria and obtain baseline estimates of transmission intensity across the country. This country-wide cross-sectional randomised survey was carried out by the National Centre for Malaria Control, Parasitology and Entomology (CNM) in Cambodia. Serological measurements were analysed by the London School of Hygiene and Tropical Medicine (LSHTM). Other covariates including age, gender of the participant and the distance to the forest were identified as risk factors for malaria exposure.

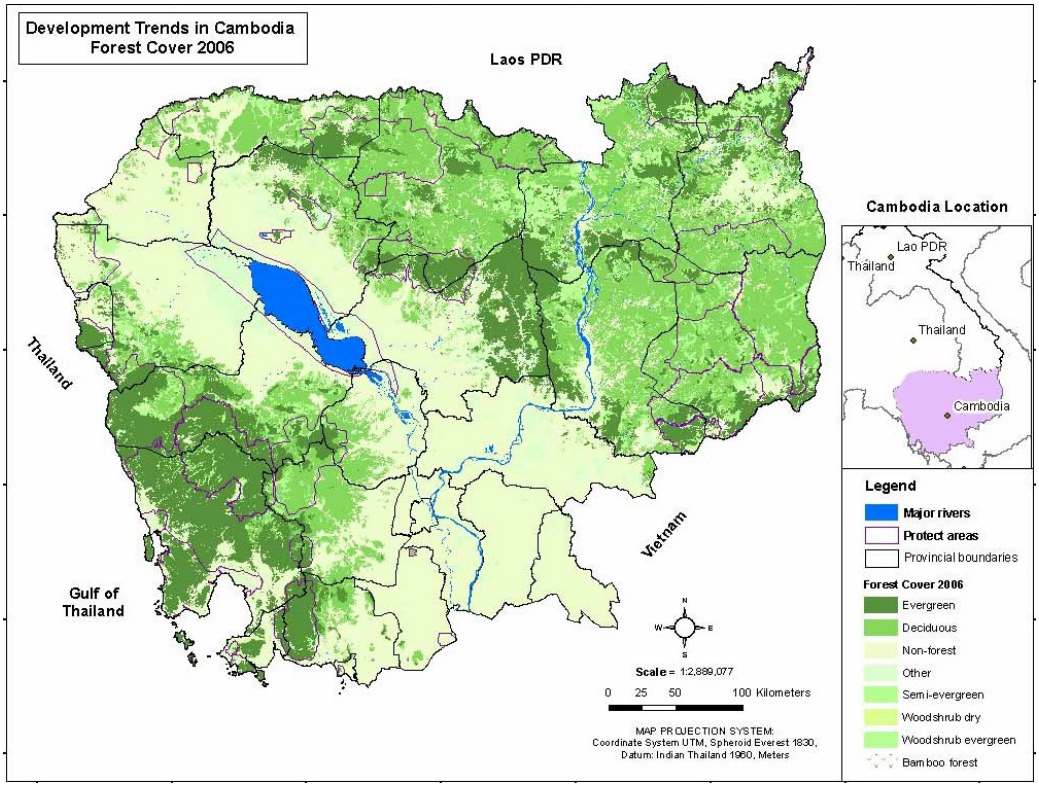


Figure 2.1: **Forest cover in Cambodia Source: sithi.org** [184]

Blood spots were collected onto filter paper in order to measure exposure to both *P. falciparum* and *P. vivax.* Serological analyses were performed using ELISA and Immunoglobulin G antibodies to the asexual stage merozoite antigens determined. Samples were tested against two antigens, Merozoite Surface Protein (MSP-1) and Apical Membrane Antigen (AMA-1). The ELISA technique produces measurements as optical densities which are subsequently translated into estimated titres [97]. Titres were log-transformed prior to analysis with zero (and negative) measurements arbitrarily assigned a low value of zero on the log10 scale. Also, measurements greater than three on a log10 scale were set to three as higher values were considered to be unreliable. The age range of the studied population is wide with a range from zero to 89 years. However older people (those >60 years) represent a very small proportion of the population and the majority of these did not have any antibodies. As infants may present with maternal antibodies only individuals between 1 and 60 years were included in the analysis.

2.2.2 Descriptive analysis

The data collected in Cambodia included antibody titres against AMA-1 and MSP-1 antigens for both *P. falciparum* (*Pf*) and *P. vivax (Pv)*. The summary results of the collated information are presented in Table 2.1. Titres were reported on a log10 scale and the median and inter quartile range (IQR) for seropositive individuals only are presented together with the overall prevalence of antibodies for each antigen and plasmodium species. Titre values for antibodies against antigens AMA-1 are generally higher than for antibodies against MSP-1.

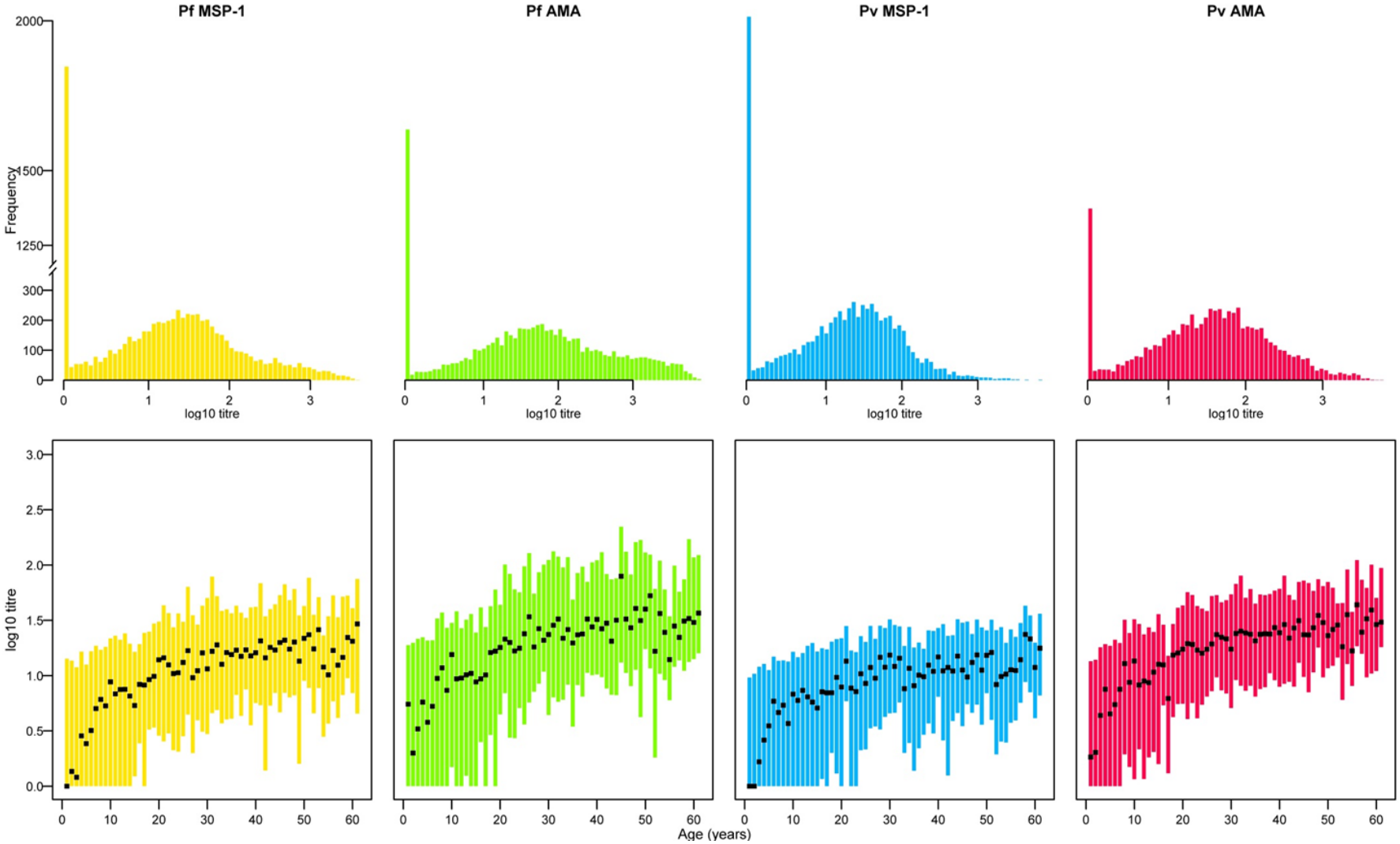


Figure 2.2: Antibody titre frequency distribution (above) and age structured antibody titre distribution (below) for Pf MSP-1, Pf AMA-1, Pv MSP1, Pv AMA-1 (left to right). Median of log10 titre presented by black dot and each coloured box represent the interquartile range (25th-75th percentile).

Table 2.1: Summary statistics for MSP-1 and AMA-1 antigens for both *P. falciparum* and *P. vivax*.

	Log10 Titre for seropositive individuals Median (IQR*)	**Seroprevalence (no. positive/no. tested)**
***Pf* MSP-1**	1.2 (0.8-1.6)	16 (1218/7577)
***Pf* AMA-1**	1.5 (1.1-2.1)	24 (1730/7315)
***Pv* MSP-1**	1.2 (0.8-1.5)	8 (626/7722)
***Pv* AMA-1**	1.4 (1.0-1.8)	16 (1177/7583)

***IQR, interquartile range (25th-75th percentile)**

Figure 2.2 shows the overall distribution of the antibodies for each antibody type and each plasmodium species. Seronegative individuals are represented by the peak at log titre=0. The age-structured distribution of the antibodies is also presented in Figure 2.2, showing how low and variable the antibody titres are for younger ages whilst they become higher and less variable for older individuals.

As malaria transmission intensity was known to vary with forest malaria, the distribution of antibodies was categorised by the distance to the forest (see Figure 2.3). People who live in the forest tend to have higher antibody levels for both *P. falciparum* and *P. vivax* and both antibody types.

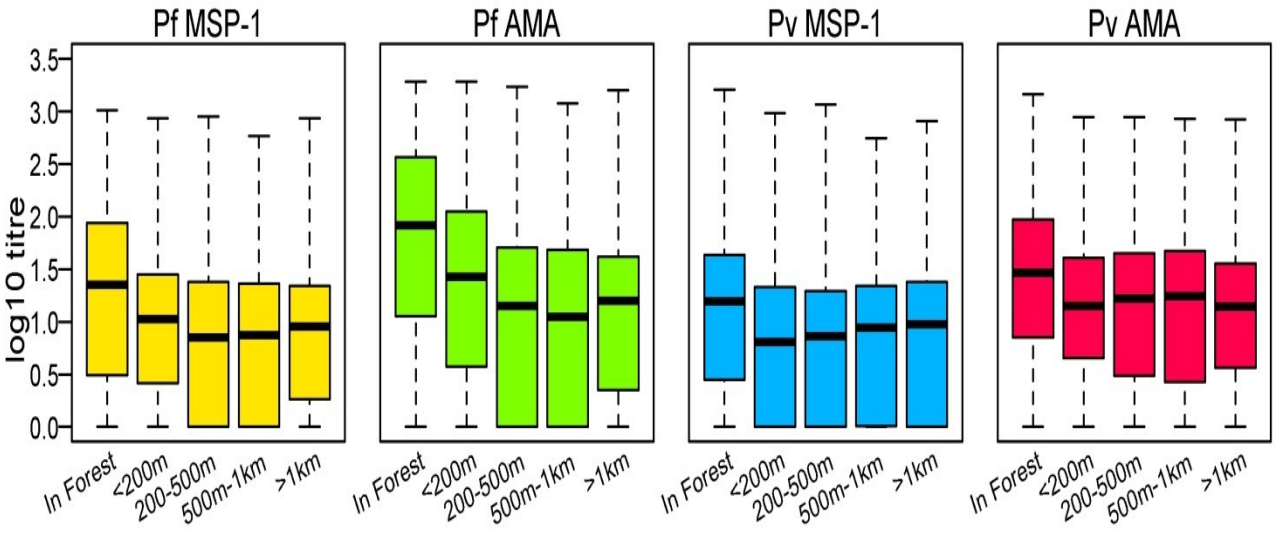


Figure 2.3: Antibody titre distribution according to the distance to the forest for antibodies against MSP-1 and AMA-1 antigens for both *Plasmodium falciparum* (*Pf*) and *Plasmodium vivax* (*Pv*).

2.3 Material & Methods

2.3.1 Mathematical models

The objective was to develop and fit a model to titre data and extract information about exposure to infection, the boosting of the antibody-mediated immune response and potentially the decay of antibodies. The aim is to extend the original catalytic model, typically used to analyse serological data, to one with multiple compartments to represent the full dynamics of acquisition and loss of antibodies in the population.

2.3.1.1 Categorisation of the population

The population was stratified into different groups according to an individual's antibody level on a log10 scale at the time of the survey. An arbitrary number of categories (here six) were chosen and the range of titre, varying from zero to three was split into equally sized intervals of 0.6 logs. An individual *i* is classified in category *k* if their log10 antibody titre $x_i \in (X_{k-1}, X_k]$ where $k \in \{2,3,4,5,6\}$ or in category 1 if $x_i = 0$. The population stratification is represented schematically in Figure 2.4.

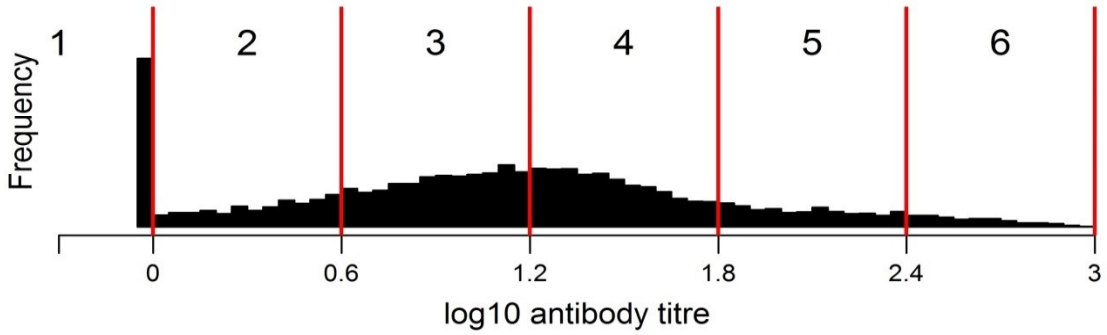


Figure 2.4: Schematic representation of the discretisation of the antibody level of individuals. For each antibody class, its index *k* is presented at the top and its boundaries *(X_{k-1},X_k)* on the x-axis. Frequency of the population in each class is presented in black.

2.3.1.2 Model specifications

A compartmental model was used to model the dynamics of the acquisition and loss of antibodies. This density model is an extended version of the reversible catalytic model [158] as presented in Chapter 1. The rate of movement from titre class *j* to titre class *i* is λk_{ij}, where λ represents the 'force of infection' as a proxy for measure of exposure and k_{ij} is the probability that once infected an individual with a titre in class *j* will be boosted to a titre in class *i*. $K = (k_{ij})$ is termed the transition matrix (2.1). The structure of this transition matrix can take different forms depending on the assumed biology. In its most general form, transitions can occur from any state to any other higher state (Figure 2.5).

$$K = \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & \\ k_{21} & 0 & & & & \vdots \\ k_{31} & k_{32} & \ddots & & & \vdots \\ k_{41} & k_{42} & k_{43} & \ddots & & \vdots \\ k_{51} & k_{52} & k_{53} & k_{54} & \ddots & \vdots \\ k_{61} & k_{62} & k_{63} & k_{64} & k_{65} & 0 \end{bmatrix} \quad (2.1) \qquad G = \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & \\ \rho & 0 & & & & \vdots \\ 0 & \rho & \ddots & & & \vdots \\ \vdots & \ddots & & \ddots & & \vdots \\ & & & \rho & \ddots & \vdots \\ 0 & \cdots & & 0 & \rho & 0 \end{bmatrix} \quad (2.2)$$

The rate of decay of antibodies is assumed to be constant over time, resulting in exponential decay ρ. The matrix G, representing the decay of antibodies has the form presented in (2.2). Estimates for antibody decay and exposure rate might be separately identified with difficulty from cross-sectional data. Therefore, the antibody decay was fixed to an estimate obtained by fitting the catalytic model to the data of $\rho = \frac{0.03}{\Delta} = 0.05$, with $\Delta = 0.6$ the size of the interval [134].

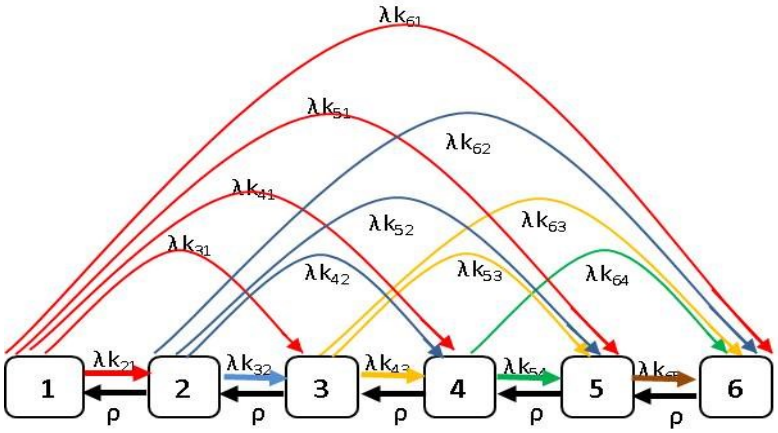


Figure 2.5: Flow diagram illustrating the dynamics of the population building immunity upon exposure. Population stratified into 6 classes. Individuals acquire antibodies at rate λk_{ij} and lose their antibodies at rate ρ.

Let $y_i(t)$ be the proportion of individuals with categorised titre i∈{1,…,6} at age t∈{1,…,60}. Age is considered in months as continuous time and averaged over a year. The matrix form of the model can be written as:

$$\begin{bmatrix} dy_1/dt \\ dy_2/dt \\ dy_3/dt \\ dy_4/dt \\ dy_5/dt \\ dy_6/dt \end{bmatrix} = \lambda \cdot \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & y_1 \\ k_{21} & 0 & & & & \vdots \\ k_{31} & k_{32} & \ddots & & & \vdots \\ k_{41} & k_{42} & k_{43} & \ddots & & \vdots \\ k_{51} & k_{52} & k_{53} & k_{54} & \ddots & \vdots \\ k_{61} & k_{62} & k_{63} & k_{64} & k_{65} & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} \cdot \begin{bmatrix} 0 & \rho & \cdots & \cdots & \cdots & \\ 0 & 0 & \rho & & & \vdots \\ & 0 & \ddots & & & \vdots \\ \vdots & \ddots & & \ddots & & \vdots \\ & & & 0 & \ddots & \\ 0 & \cdots & \cdots & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} - \lambda \begin{bmatrix} 0 & k_{21} & k_{31} & k_{41} & k_{51} & k_{61} \\ \vdots & & & & & \\ \vdots & & \ddots & & & \\ \vdots & & & \ddots & & \\ \vdots & & & & \ddots & \\ 0 & \cdots & \cdots & \cdots & & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & \\ & & & & & \vdots \\ & & \ddots & & & \vdots \\ & & & \ddots & & \vdots \\ & & & & \ddots & \vdots \\ & & \cdots & & & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix}$$

or more generally as:

$$\frac{dy_i}{dt} = \lambda \sum_j k_{ij} y_j + \rho y_{i+1} - \lambda y_i \sum_j k_{ji} - \rho y_i \qquad (2.3)$$

2.3.1.3 Parameters

The exposure rate λ is assumed to be constant over time. The focus of the work presented here is to explore different parameterisations for the acquisition of antibodies which are represented by transition matrix *K*. I want to characterise the boost in antibody levels that occurs following infection. Therefore an understanding of the underlying mechanisms for the boost distribution is essential for an accurate model able to separate the effect of exposure and immunogenicity of antigens. One would ideally be able to directly measure the boost of antibodies upon infection. However this information cannot be captured in cross-sectional surveys. I therefore explored model structures with different levels of complexity that were biologically motivated.

The full model as presented in Figure 2.5 represents the least constrained scenario where, following infection, antibody levels can be boosted to any higher level. In this 6-compartments model, this requires estimating 15 parameters for the boost and 1 for exposure (λ). To avoid over-parameterisation one of the coefficients in the transition matrix *K* is set to a fixed value and therefore 15 parameters are estimated with the following constraints.

1. Each parameter $k_{ij} \in [0,1]$;
2. The sum of each column of the transition matrix is ≤ 1 as it represents the total probability of the transition from state j.

A simplified version of this model is also considered and represents the case where individuals once exposed can only boost their antibody levels to the next level of antibodies. This model is illustrated in Figure 2.6 with its associated matrix *K*. The total number of parameters to estimate in this instance is strongly reduced (four for the boost distribution and one for the exposure). Here the following constraint remains $k_{ij} \in [0,1]$. In addition, to avoid over-parameterisaton k_{43} was set to a fixed value, chosen to be 0.1, therefore assuming a low probability of boosting between antibody classes 3 and 4, which includes most of the population.

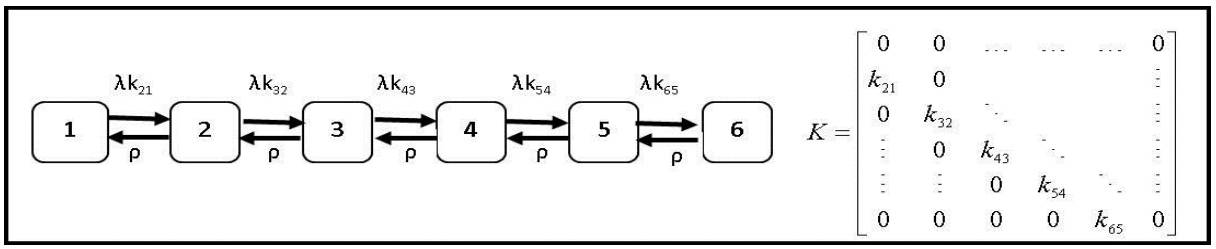


Figure 2.6: Flow diagram illustrating population dynamics during acquisition of antibodies for the simplified model and its associated transition matrix K. Population stratified into 6 classes. Individuals acquire antibodies at rate λk_{ij} and lose their antibodies at rate ρ.

2.3.1.4 Application to different patterns of endemicity

The simplified model was further developed to take into consideration different patterns for endemicity. In this case, based on biological grounds, the parameters determining the dynamics of antibodies (k_{ij} and ρ) were considered to be identical for each individual regardless of the area in which they live. Only the force of infection λ varies for the different areas. The distance to the forest is used as a proxy measurement for the transmission risk and was categorised into five groups. The resulting model has five parameters for the transition matrix, one for the decay of antibodies and five forces of infection parameters representing different levels of endemicity. The decay, as well as one of the parameter from the boosting matrix (k_{43}) was fixed, resulting in a total of 9 parameters to estimate.

2.3.2 Model Fitting and Model Selection

2.3.2.1 Bayesian Model

The model parameters were estimated by fitting the models to the data using a Bayesian approach [185]. Let $D = \{(x_j, t_j)\}$ denote the observed data constituted of the log10 antibody level x_j and age t_j of individual j and $\theta = \{\lambda, K\}$ denote the model parameters. The joint density of observed data and parameters is:

$$P(D, \theta) = P(D / \theta)P(\theta) \quad (2.4)$$

with $P(D / \theta) = \Pi_t P(D_t / \theta)$ and $P(\theta)$ respectively the likelihood and prior of the model parameters and D_t the data for age range *t*.

2.3.2.2 Likelihood

Let $D_{i,t} = \frac{n_{i,t}}{\sum_i n_{i,t}}$ denote the observed proportion of individuals at age t in antibody titre class i and let $n_{i,t}$ denote the number of individuals of age *t* in this class. The model predicted proportion of individuals at age *t* in titre class *i* is denoted $y_{i,t}$ solution of (2.3). The likelihood is therefore defined by:

$$P(D_{i,t} / \theta) = {y_{i,t}}^{n_{i,t}} \quad (2.5)$$

Consequently, assuming observations are independent, the multinomial likelihood for the data is:

$$P(D / \theta) = \prod_i \prod_{t \neq 0} {y_{i,t}}^{n_{i,t}} \quad (2.6)$$

and the log-likelihood given by : $$l = \log(P(D \mid \theta)) = \sum_{t \neq 0} \sum_{i} n_{i,t} \log(y_{i,t}) \quad (2.7)$$

The differential equations were numerically integrated in C using the Runge-Kutta method [186]. The number of individuals in each compartment was derived for each time step (every second day) and the mean over a year was used as the predicted values. Results were validated with the R solver [187] using the lsoda function in the deSolve package [188] . The starting values for the algorithm were taken from the data when individuals were at age 0. However, for model fitting, only individuals aged above 1 year were included to remove the confounding effect of maternally-derived antibodies.

When estimating multiple exposures the log-likelihood becomes:

$$l(D/\theta) = \sum_{v} \sum_{i} \sum_{t} n_{i,t,v} \log(y_{i,t,v}) \quad (2.8)$$

where $y_{i,t,v}$ *and* $n_{i,t,v}$ are respectively the predicted proportion and the observed number of people in antibody titre class i and age t in area v.

2.3.2.3 Prior distribution

Uniform priors were chosen for the different parameters. The coefficients from the boosting matrix *K* were given a prior that is uniform $U[0,1]$. In addition, they were subject to the constraint $\sum_i k_{ij} \leq 1$. The exposure rate also had an uninformative prior and was drawn from a uniform distribution $U[0,5]$.

2.3.2.4 MCMC Sampling

A Monte Carlo Markov Chain (MCMC) method was used for the parameter estimation. Parameters were sampled using a standard random-walk Metropolis-Hasting algorithm [185, 189]. To reduce correlations in the chain, model parameters were updated together. At each iteration, all model parameters were resampled; if θ was the current value of the parameter, a candidate point θ^* is sampled from a proposal distribution so that $\log(\theta^*) = \log(\theta) + \sigma\mu$, where μ is drawn from a normal distribution $N(0,1)$ and the random walk rate σ was tuned to obtain optimal mixing (See Section *2.3.2.5*). The candidate point is then accepted with probability $\alpha(\theta, \theta^*)$ where:

$$\alpha(\theta,\theta^*) = \min\left(1, \frac{P(\theta^*)P(\theta/D)}{P(\theta)P(\theta^*/D)}\right) \quad (2.9)$$

If the candidate point is accepted, the current point for the next iteration becomes $\theta = \theta^*$.

A total of 100,000 iterations were performed for each run of the MCMC algorithm. The first 1,000 were discarded as the burn-in period. The output was then sampled every 250 iterations to constitute a sample from the posterior distribution. Multiple chains were run with different parameter starting values and combined to obtain an overall posterior sample of 3,600 iterations. For each parameter the *posterior* median and a 95% credible interval were computed.

2.3.2.5 Random walk tuning

The standard deviation of the proposal distribution, also called the random walk rate σ, was tuned in order to achieve appropriate mixing of the chains and an acceptance rate close to 20% [190]. During the burn-in period, at each iteration m the random walk rate was updated as below (Jamie Griffin, personal communication March 2011):

$$\sigma_{m+1} = \sigma_m \times \exp\left(\frac{0.4 \times (\alpha - \alpha_0)}{35 \times \frac{m}{M+1}}\right) \text{ and } \begin{cases} \sigma_{m+1} = 0.001 \text{ if } \sigma_{m+1} \leq 0.001 \\ \sigma_{m+1} = 10 \text{ if } \sigma_{m+1} \geq 10 \end{cases} \quad (2.10)$$

With α the acceptance probability defined in (2.9), α_0 the optimal acceptance rate equal to 23% and M the total number of iteration.

2.3.3 Validation of measures of exposure

A classic catalytic model was fitted to the data with the purpose of comparing estimated measures of exposure. A cut-off value above which individuals were considered seropositive was defined using a mixture model [98, 162]. Cut-off values were generated separately for each antigen and each plasmodium species. The proportion of seropositive individuals who are seropositive at age *t* is given by:

$$y(t) = \frac{\lambda_c}{\lambda_c + \rho_c}(1 - e^{-(\lambda_c + \rho_c)t}) \quad (2.11)$$

with λ_c is the seroconversion rate and ρ_c the reversion rate for the catalytic model.

The reversion rate was fixed to a constant value of 0.05, resulting for *Pf* MSP-1 from another study also carried out in Cambodia [134]. However, we assume that loss of antibodies is not antigen- or species-specific. Models were fitted with standard maximum likelihood estimation techniques [191], using *optim* function in R [188]. Denoting the maximum likelihood estimate of the force of infection λ*, a 95% confidence interval for the parameter λ is then determined by $\lambda^* \pm 1.96.se$ with $se = \sqrt{diag(H^{-1})}$ and *H* is the Hessian matrix provided by R, corresponding to the negative of the inverse of the covariance matrix of the estimates.

2.4 Results

2.4.1 Estimating malaria exposure and acquisition of antibodies rate

The simplified model was fitted to data available from Cambodia. Parameters were estimated using a Bayesian approach. The algorithm converged and a good model fit was obtained, as demonstrated by the results below.

Posterior densities of the parameters and MCMC trace are presented in Figure 2.7, demonstrating good convergence of the MCMC chain with smooth posterior distributions obtained for each parameter. The auto-correlations were small enough (between 0.2 and 0.3) to consider the observations in each parameter posterior sample to be independent from each other. The lowest effective sample size is around 233.

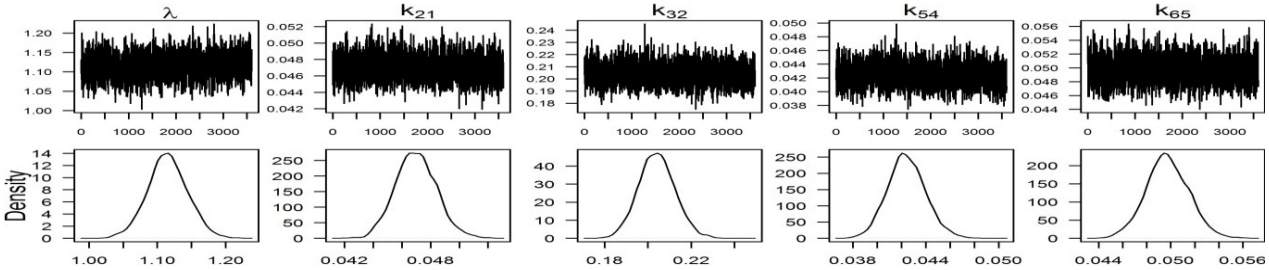


Figure 2.7: MCMC trace (top row) and *posterior* distribution (bottom row) for measure of exposure (λ) and coefficients for the boosting matrix

Correlations between parameters were examined to understand the relationship between parameters, as presented in Figure 2.8. A high degree of correlation was observed between parameters, as cross-sectional data provide limited information to distinguish between exposure and biological mechanisms for acquisition of antibodies. Indeed, estimates of the force of infection are negatively correlated with estimates of all the parameters from the boosting matrix.

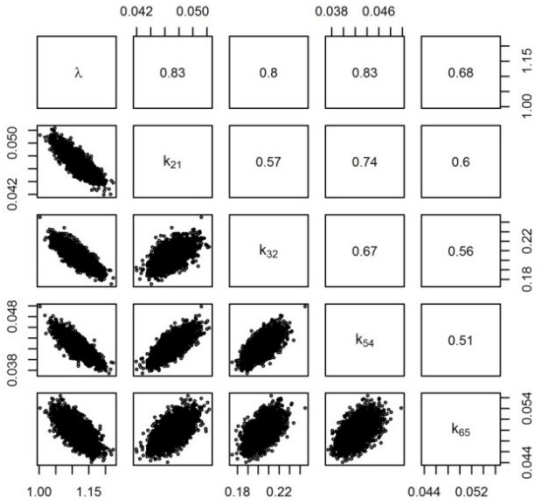


Figure 2.8: Bivariate plots of the marginal posterior distributions of all the model parameters. Pearson correlation coefficients are presented in the upper panels.

Figure 2.9 and Figure 2.10 show the predicted output from model fit, illustrating a qualitatively good fit and a very narrow credible interval around the median fit. The model fit is constructed through simulation using sets of parameters, here 100, from the posterior distribution. Age specific median fit and 95% credible intervals of the resulting simulations are then computed and presented on the figures. Age-structured seroprevalence curves, categorised by antibody titre, are presented in Figure 2.9.

At a very young age, individuals with no antibodies against Pf MSP-1 represent a high proportion of the population. However, with increasing age, an antibody response is first initiated and then boosted and people acquire higher antibody titres.

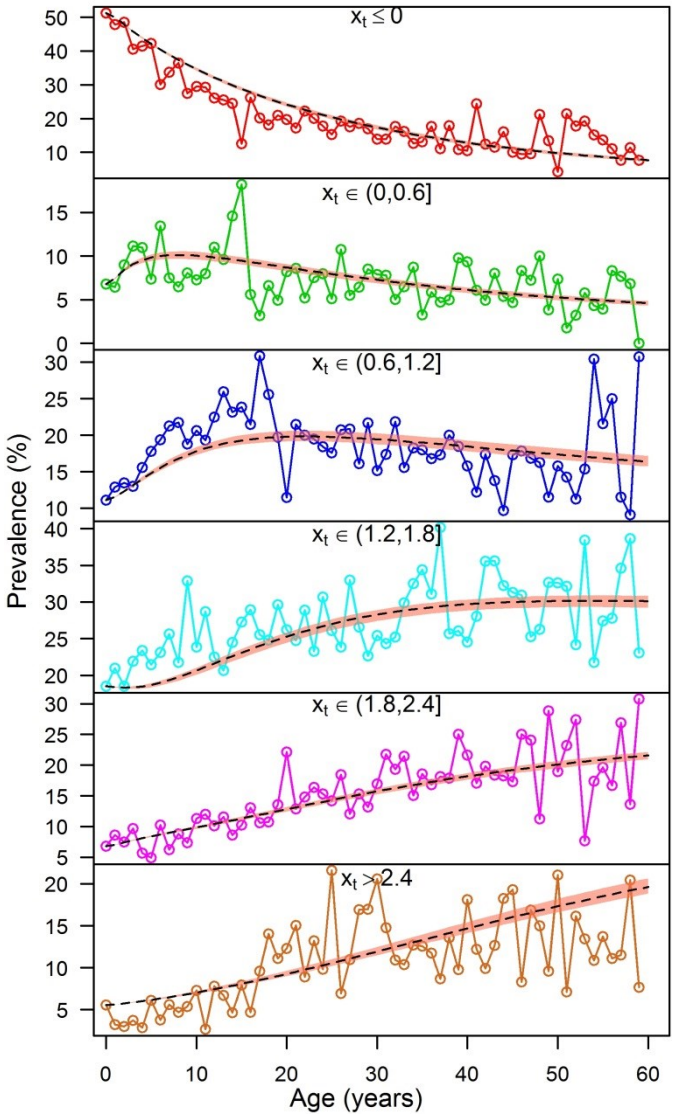


Figure 2.9: Seroprevalence curves as a function of age categorised by titre values for Pf MSP-1; the coloured lines correspond to the data seroprevalence and the black line the model fit median and the pink shaded area represents the 95% credible interval. x_t represents individual's antibody titre on log10 scale.

In each antibody class individuals are assumed to have, on average, the median value of the class, except for class 1 where individual's log10 titre is zero. The mean titre for each age group is therefore defined as the mean of the proportion of individuals in each antibody class multiplied by the median value of the class (0 for class 1). The mean antibody titre by age is presented in Figure 2.10 for *Pf* MSP-1. Individuals acquire antibodies through their life seemingly faster at younger age than at older age. Analogous results are presented in Figure 2.11 for MSP-1 and AMA-1 for both *P. falciparum* and *P. vivax*.

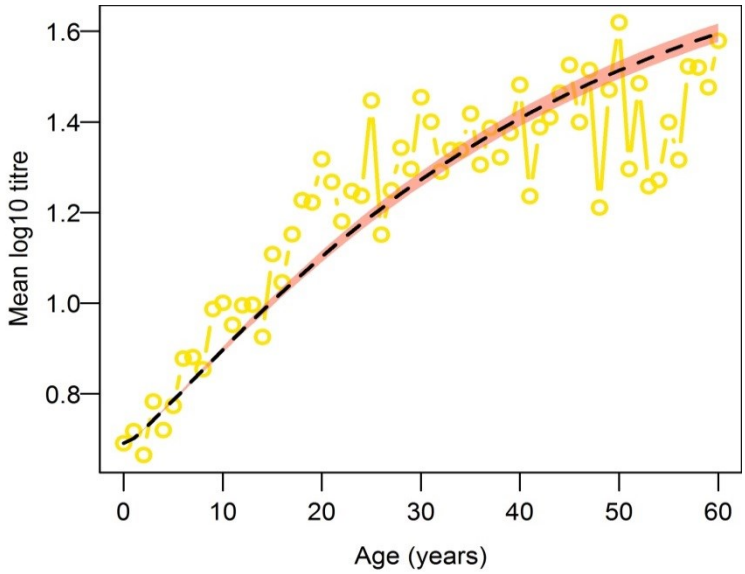


Figure 2.10: Mean antibody titre again Pf-MSP1 antigen; the coloured line represents the data and the black line the model. Here, median (black line) and 95% credible interval (pink shaded area) of 100 simulations of the mean log10 titre are computed.

Figure 2.11 compares the average antibody titre by age for both antibody types and both *Plasmodium* species. The model fit consistently reproduces the data qualitatively well. We observe higher titres for antibodies against AMA-1 compared to those against MSP-1. However, *P. falciparum* and *P. vivax* appear to have similar trend in mean antibody levels across age ranges.

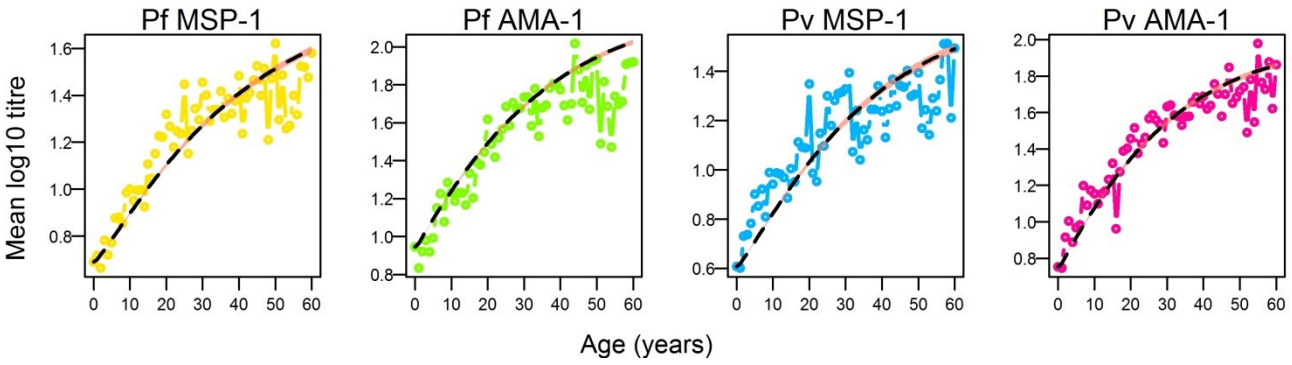


Figure 2.11: Individual mean antibody titre for MSP-1 and AMA-1 for *P. falciparum* and *P. vivax*. The black lines correspond to the median fit for a sample of the MCMC simulation results and the colours dots correspond to the data. The pink shaded areas represent the narrow credible interval around the model fit. Note that the y-axes are on different scales.

Similar to the results found with the mean antibody titre, Figure 2.12 indicates that the proportion of individuals with high antibody levels (>2.4) appears to be higher for antibodies against AMA-1 antigens, consistently for both *P. falciparum* and *P. vivax*. This could be reflecting the higher immunogenicity of AMA-1. In addition, across all ages, a higher proportion of individuals present with *P. falciparum* antibodies compared to *P. vivax* antibodies (the model predicts up to 20 to 40% of the population for respectively MSP-1 and AMA-1 for old individuals compared to 8 to 25% for *P. vivax*).

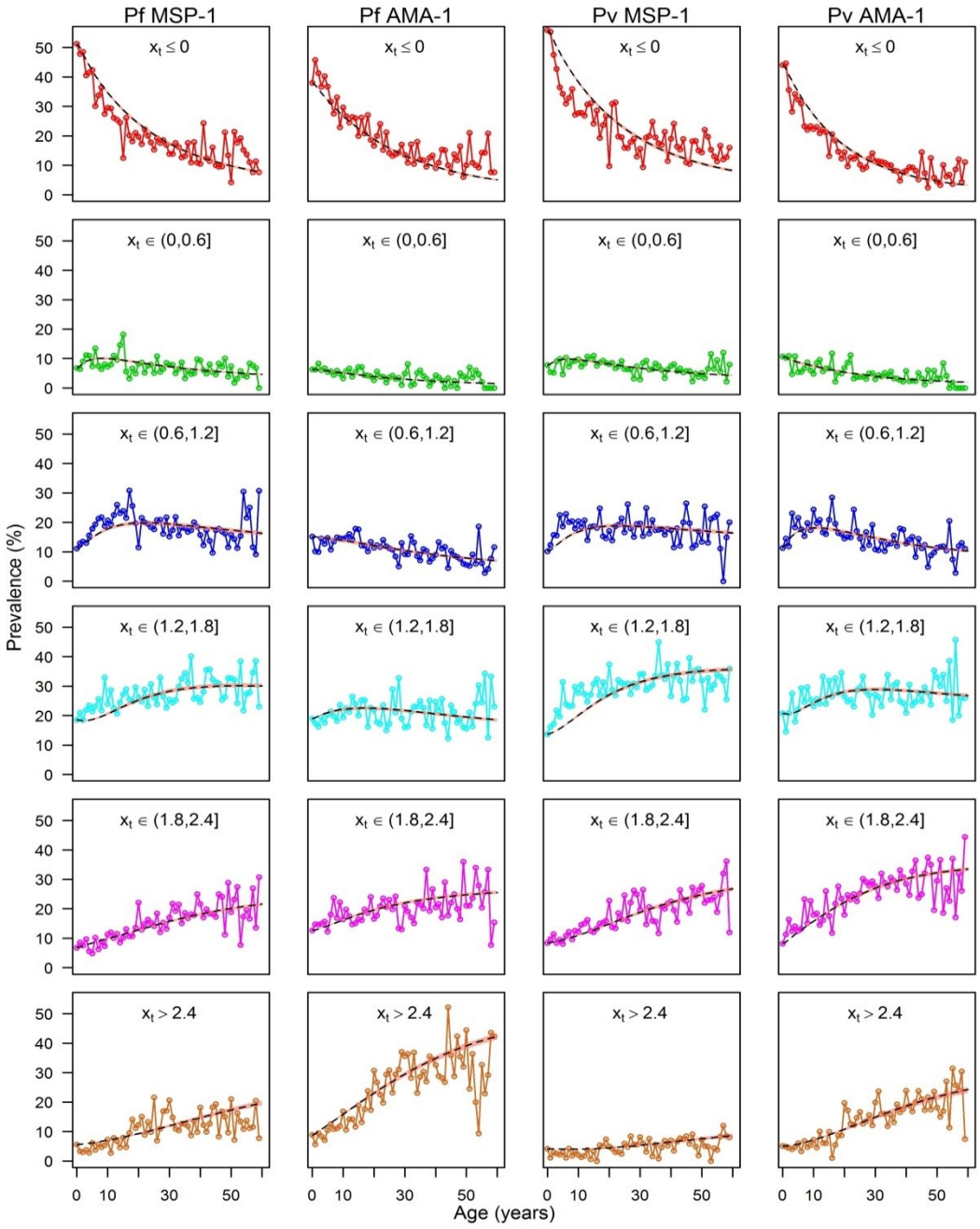
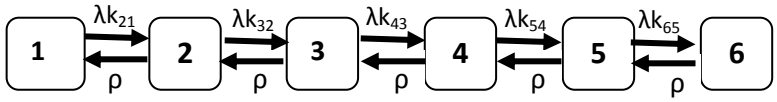


Figure 2.12: Seroprevalence curves categorised by 6 titre ranges on log10 scale (x_t) for antibodies against MSP-1 and AMA-1 antigens for *P. falciparum* and *P. vivax.* The black lines correspond to the median fit of a sample of the MCMC simulation results and the coloured dots correspond to the data (colours are used for visual clarity).

Table 2.2 presents posterior median (95% equal-tailed credible interval) for model parameters for antibodies against MSP-1 and AMA-1 for both *P. falciparum* and *P. vivax*.

Table 2.2: Posterior median (95% credible interval) for model parameters and transition rates λk

Model Parameter	*Pf* MSP-1 Median (95% credible interval)	*Pf* AMA-1 Median (95% credible interval)	*Pv* MSP-1 Median (95% credible interval)	*Pv* AMA-1 Median (95% credible interval)
λ	**1.11** (1.06-1.17)	**1.71** (1.61-1.81)	**1.27** (1.2-1.33)	**1.53** (1.45-1.61)
k_{21}	**0.047** (0.0443-0.0499)	**0.0266** (0.0248-0.0284)	**0.0394** (0.0372-0.0418)	**0.0415** (0.0391-0.044)
k_{32}	**0.204** (0.188-0.221)	**0.204** (0.185-0.225)	**0.196** (0.181-0.212)	0.218 (0.201-0.237)
k_{54}	**0.0424** (0.0395-0.0457)	**0.0511** (0.0471-0.0552)	**0.0357** (0.0332-0.0383)	**0.1** (0.1-0.1)
k_{65}	**0.0497** (0.0463-0.0533)	**0.0552** (0.0513-0.0591)	**0.0155** (0.0143-0.017)	**0.0485** (0.0452-0.0521)
λk_{21}	**0.0523** (0.0507-0.0541)	**0.0454** (0.0439-0.0469)	**0.05** (0.0484-0.0515)	**0.0634** (0.0615-0.0655)
λk_{32}	**0.227** (0.216-0.239)	**0.349** (0.328-0.372)	**0.249** (0.236-0.261)	**0.333** (0.316-0.352)
λk_{43}	**0.111** (0.106-0.117)	**0.171** (0.161-0.181)	**0.127** (0.12-0.133)	**0.153** (0.145-0.161)
λk_{54}	**0.0473** (0.0454-0.0492)	**0.0872** (0.084-0.0905)	**0.0452** (0.0434-0.047)	**0.0742** (0.0716-0.0769)
λk_{65}	**0.0554** (0.0526-0.0584)	**0.0942** (0.0906-0.0978)	**0.0197** (0.0183-0.0211)	**0.0423** (0.0404-0.0443)



The results suggest that the initiation of immunity, corresponding to the transition from no antibodies, (antibody class 1), to some antibodies (antibody class 2), is occurring at a slower rate compared with the acquisition of immunity during subsequent infections, consistently for both antigens and both *Plasmodium* species. The rate is around fivefold higher for second infections (from class 2 to class 3) and two to three fold higher for third infections (from class 3 to class 4) compared to first infection (assuming infection corresponds to a boost of immunity).

Applying the full model

Parameter estimation was also conducted using the complex full model presented in Figure 2.5. The MCMC algorithm was run for *Pf*MSP-1 for 1 million iterations. Although the trace for the likelihood was converging, and both the seroprevalence curves and the mean antibody titre indicated a visually good fit to the data, the acceptance rate was still very low (around 2%) and it is clear from the MCMC trace plots for the parameters (not presented here) that the fitting has not converged. This is very likely due to model over-

parameterisation, where multiple combinations of parameter values provide the same likelihood and are thus correlated in the sampling process.

As algorithms were not converging for this more complex model, the simplified model was considered for the remainder of this chapter. The generalizability of the model was subsequently addressed by switching to a continuous model (see Chapter 3).

2.4.2 Comparison of estimates with those from a catalytic model

One of the objectives in developing this model was to be able to separate the effect of exposure and immunogenicity of antibodies. If it is to be used as a measure of transmission intensity, the 'force of infection' estimated from these models should rank different settings in the same way as other methods. A first step was therefore to compare the estimates from the density models with those obtained using a catalytic model. The data available are stratified by risk of exposure, considering the distance to the forest as a proxy for malaria transmission intensity.

Figure 2.13 presents the seroprevalence curves and the fitted catalytic model when the seroreversion rate ρ_c is fixed to 0.05. The estimates of the seroconversion rates λ_c (Maximum Likelihood (ML) estimates and 95% confidence interval) are presented in Table 2.3.

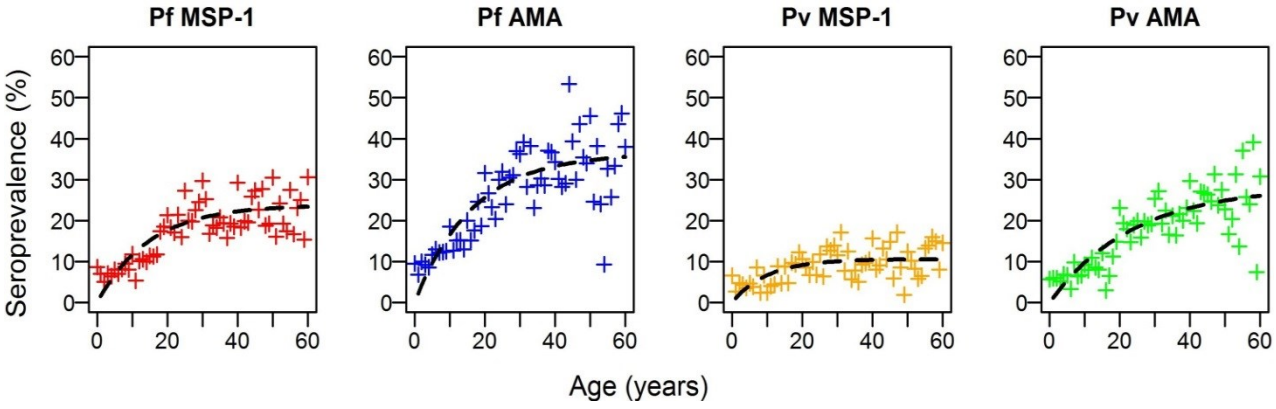


Figure 2.13: Age-structured seroprevalence curves for MSP-1 and AMA-1 for both *P. falciparum* and *P. vivax*.

The results suggest that fixing the seroreversion rate ρ_c to 0.05 is a reasonable assumption as the estimated values from both models are consistent, with rates for *P. falciparum* higher than for *P. vivax* consistently for both antigens. Similarly the conversion to seropositive is higher with AMA-1 antigen rather than MSP-1, consistently for both *Plasmodium* species (Table 2.3).

Table 2.3 : Parameter estimation for seroconversion rate for a catalytic model for antibodies against MSP-1 and AMA-1 for both *P. falciparum* (*Pf*) and *P. vivax (Pv)*.

	Model Parameter	*Pf* MSP-1 ML Estimates (95% conf. interval)	*Pf* AMA-1 ML Estimates (95% conf. interval)	*Pv* MSP-1 ML Estimates (95% conf. interval)	*Pv* AMA-1 ML Estimates (95% conf. interval)
Estimating ρ_c	λc	**0.016** (0.014-0.019)	**0.022** (0.019-1.025)	**0.010** (0.008-0.013)	**0.012** (0.010-0.014)
	ρc	**0.05** (0.04-0.06)	**0.04** (0.03-0.05)	**0.09** (0.06-0.12)	**0.03** (0.02-0.04)
Fixing ρ_c	λc	**0.016** (0.015-0.017)	**0.025** (0.024-0.027)	**0.007** (0.006-0.008)	**0.015** (0.014-0.016)

2.4.3 Estimating exposure rate in multiple endemicity settings

As heterogeneity in malaria transmission is often believed to be related to the forest in Cambodia, the distribution of the antibody titre was presented in Figure 2.3 according to the distance between the individual's house and the forest and showed higher antibody levels for individuals living in the forest.

The density model applied to different transmission settings fitted the data well with convergence of the MCMC algorithm. *Posterior* distributions for the coefficients of the boosting matrix are summarized in Table 2.4. The parameter estimates show the same patterns between antigens and between species as the model fitted assuming no variation in endemicity.

Table 2.4: Parameter estimation for boosting matrix for MSP-1 and AMA-1 for both *P. falciparum* and *P. vivax*

Model Parameter*	*Pf* MSP-1 Median (95% credible interval)	*Pf* AMA-1 Median (95% credible interval)	*Pv* MSP-1 Median (95% credible interval)	*Pv* AMA-1 Median (95% credible interval)
k_{21}	**0.076** (0.073-0.078)	**0.032** (0.031-0.033)	**0.069** (0.067-0.071)	**0.06** (0.058-0.062)
k_{32}	**0.3** (0.28-0.31)	**0.23** (0.21-0.24)	**0.31** (0.3-0.32)	**0.27** (0.26-0.28)
k_{54}	**0.04** (0.039-0.042)	**0.043** (0.042-0.045)	**0.02** (0.019-0.021)	**0.038** (0.036-0.039)
k_{65}	**0.037** (0.035-0.039)	**0.038** (0.037-0.04)	**0.015** (0.013-0.016)	**0.016** (0.015-0.016)

***k_{43} was fixed to 0.1**

Table 2.5 shows the parameter estimates from this simple density model (median and 95% credible interval) alongside those from the catalytic model (MLE and 95% confidence interval) for antibodies against AMA-1 and MSP-1 for both *P. falciparum* and *P. vivax*. The results for the density model show a decreasing exposure rate with an increasing distance to the forest consistently for both antigens and both *Plasmodium* species (Figure 2.14). The same conclusions hold for the catalytic model.

Table 2.5: Parameter estimation measuring 5 different exposure rates λ based on the distance to the forest (1 : In forest, 2 : <200m, 3 : 200-500m, 4 : 500m-1km, 5 : >1km). Rates estimated by simple density model and catalytic model for MSP-1 and AMA-1 for both *P. falciparum* and *P. vivax.*

Risk area	Pf MSP-1		Pf AMA-1		Pv MSP-1		Pv AMA-1	
	Density Median (95% CrI)	Catalytic ML Estimates (95% conf. int)	Density Median (95% CrI)	Catalytic ML Estimates (95% conf. int)	Density Median (95% CrI)	Catalytic ML Estimates (95% conf. int)	Density Median (95% CrI)	Catalytic ML Estimates (95% conf. int)
1	**1** (0.97-1.03)	**0.038** (0.031-0.044)	**2.32** (2.27-2.39)	**0.055** (0.048-0.063)	**0.99** (0.96-1.02)	**0.021** (0.016-0.027)	**1.51** (1.47-1.54)	**0.023** (0.019-0.027)
2	**0.84** (0.81-0.87)	**0.014** (0.011-0.017)	**1.35** (1.31-1.41)	**0.022** (0.018-0.026)	**0.74** (0.71-0.76)	**0.0063** (0.0039-0.0087)	**1.13** (1.09-1.16)	**0.0091** (0.007-0.011)
3	**0.81** (0.78-0.83)	**0.012** (0.0093-0.014)	**1.26** (1.22-1.3)	**0.012** (0.0099-0.014)	**0.74** (0.71-0.76)	**0.0062** (0.0042-0.0083)	**1.15** (1.12-1.18)	**0.009** (0.0072-0.011)
4	**0.74** (0.71-0.76)	**0.0071** (0.0052-0.009)	**1.22** (1.18-1.26)	**0.0091** (0.0071-0.011)	**0.77** (0.75-0.8)	**0.0071** (0.045-0.0096)	**1.21** (1.17-1.25)	**0.0098** (0.0076-0.012)
5	**0.65** (0.63-0.66)	**0.0056** (0.0044-0.0067)	**0.89** (0.86-0.91)	**0.007** (0.0059-0.0081)	**0.78** (0.76-0.8)	**0.0072** (0.0051-0.0093)	**0.83** (0.81-0.86)	**0.0061** (0.005-0.0072)

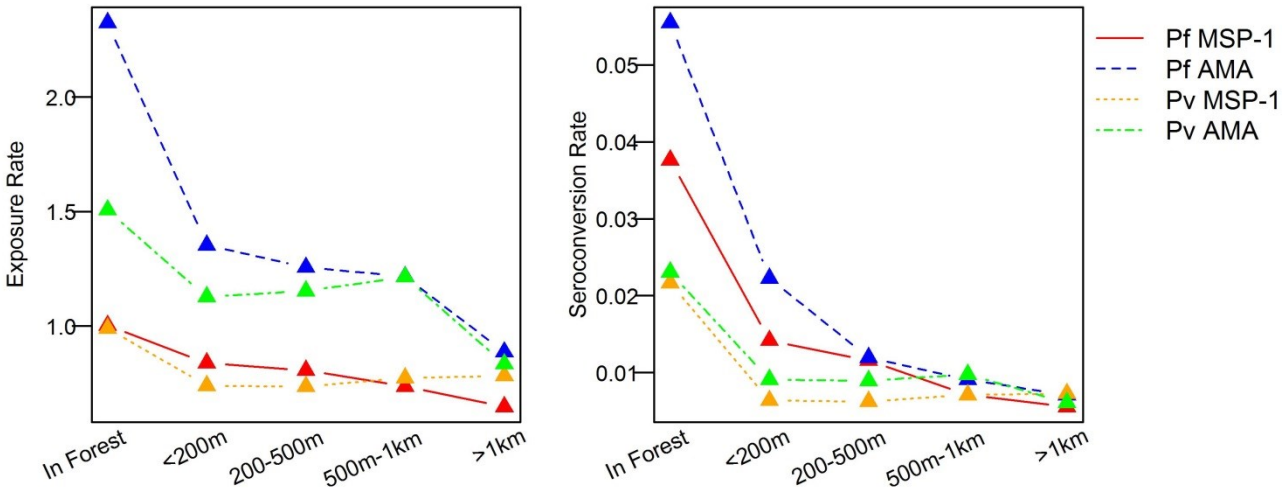


Figure 2.14: Median force of infection estimated by density model (left) and catalytic model (right) for Pf MSP-1, Pf AMA-1, Pv MSP1 and Pv AMA-1. Note that the scale is expected to be different between the two models.

Figure 2.15 illustrates the relationship between the exposure parameter for the two different models (catalytic and density). A Pearson correlation of 0.78 and a Spearman correlation of 0.76 were found, indicating that both exposure rate from the density model and seroconversion rate from catalytic model would rank endemicity settings in an equivalent order.

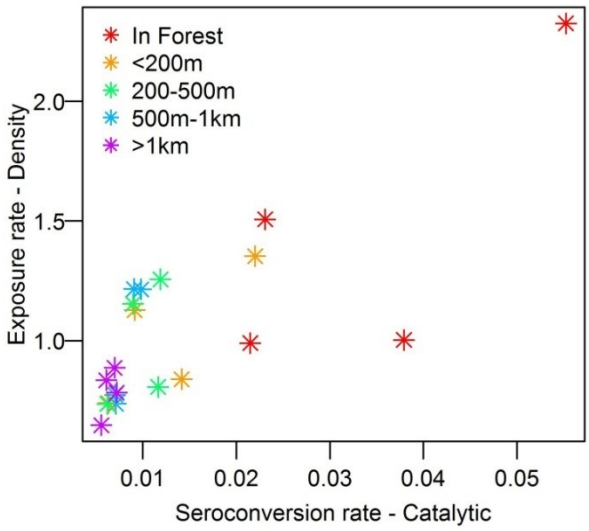


Figure 2.15: Correlation between estimates from the catalytic model and the density model, categorised by distance to the forest, irrespective of *Plasmodium* species and antibody types.

2.5 Discussion

In this chapter, I developed a compartmental model to mimic the dynamics of the acquisition and loss of blood-stage antibodies in the population using a discrete model. The model separates the effect of exposure, antibody production and decay of antibodies as part of the wider immune response by using the full information contained in serological measurements. The results demonstrate that using a simple discretised version of a density model not only provides measurements of exposure consistent with those obtained with classical methods, but also gives insight into the dynamics of the acquisition of antibodies. The use of more complex models would ideally better inform the biological mechanisms of the acquisition of antibodies and help in understanding the immune responses against *P. falciparum* and *P.vivax* and their potential interactions.

2.5.1 Boost of antibodies

A key result is that the change in the rate of acquisition of antibodies is estimated to depend on current titre. The simplistic model showed that at lower antibody densities, rates of boosting conditional on exposure are slow, and then increase before slowing down again. This suggests a density dependency in the acquisition of antibodies for both anti-MSP-1 and anti-AMA-1 antibodies, whereby circulating antibodies are necessary to facilitate a faster boost of antibodies and as saturation begins, the acquisition of antibodies happens at a slower rate. An explanation of this mechanism could be a saturation of antibodies such that antibodies in subsequent infections occurring with existing high titres are boosted at a slow rate and eventually stop production if sufficient levels already exist. Alternatively, the existence of a threshold above which antibody production is triggered could explain the minimum level of circulating antibodies required to activate a faster boost. This assumption would therefore give weight to the choice of a cut-off value for the assessment of seroprevalence using the catalytic model, although what a catalytic model would not capture is the first infections leading to low levels of antibodies. However, such a finding could also simply be an artefact of the heavily constrained model. Indeed, the simple model representing a gradual acquisition of antibodies forces individuals to have an arbitrary identical fixed boost upon infection, irrespective of the current titre or first or subsequent infections. Only the probability of developing an antibody boost varies according to the current level of antibodies. Additionally, I have assumed that the probability of boosting is independent of individual's age. In the available dataset, age and antibody levels are collinear, it would be impossible to distinguish whether the boost in antibodies depend on age or on current antibody levels.

The full model allows for greater flexibility in the assumptions about the acquisition of antibodies, implying that individuals can experience antibody boosts of different magnitudes irrespective of their current level

of antibodies. The interesting question this model is trying to answer is whether there is a saturated titre above which antibodies do not get boosted anymore or whether there is a constant or density-dependent boost constantly occurring. Even though conclusions cannot be made from this model, with the current discretisation of the population, it would still be difficult to differentiate between both hypotheses. Indeed, a discretised model might mask the effect of underlying threshold values and generate heterogeneity within discretised population. Therefore a discrete model with very small intervals (i.e. large number of states) or ideally a continuous model would overcome this issue and will be developed in the following chapters. When a higher degree of complexity was added to the model to attempt to recreate the biological mechanisms allowing antibodies to be boosted to any level, it was not possible to properly estimate the parameters due to over parameterisation.

The lack of convergence of the MCMC algorithm limits the validity of the complex models. Increasing the number of states to overcome the issue of creating heterogeneous population would only increase the number of parameters of a model already over-parameterised. Setting a few parameters to constant values could help fixing the issue, but this would be tedious and would only revert back to a constrained model leading to a model such as the simple one developed here. Consequently, fitting a continuous distribution for the acquisition of antibodies would represent a trade-off for decreasing the number of parameters and minimizing the constraints and better mimic the biological mechanisms underlying the acquisition of antibodies.

2.5.2 Measure of transmission intensity

Serology is currently used as an indicator of malaria transmission and estimates of the rate of seroconversion made using catalytic models [97]. Studies have shown that these estimates of the seroconversion rate correlate well with EIR [95]. However it was not clear whether density models that utilise the full information contained in the serological measurement, i.e. titres, would provide similar estimates as the parameter here is meant to represent the force of infection rather than a rate of seroconversion. The results show that there is a positive correlation between the parameters estimated from both traditional catalytic and the simple density model fitted here. This suggests that the 'force of infection' estimated in the density model can represent a measure of transmission intensity and such models could therefore be used to measure changes in transmission intensity in a similar way to the use of catalytic models [99, 135]. However, an application of the model to other datasets across a wider range of transmission is first needed to validate the method as a tool for measuring malaria transmission intensity.

Also, the assessment of transmission intensity based on serological data relies on the residual antibodies sustained by individuals after infection. The maintenance of long-term immunity is highly dependent on the

rate of decline of these antibodies. The model has assumed an exponential decay of antibodies by assuming that the number of individuals becoming seronegative is exponentially decreasing with increasing duration of exposure (age). The value of the rate of fading antibodies was fixed in the model. However, the underlying antibody response relies on antibody secreting cells (plasma cells) and memory B cells [56, 65]. The duration of the antibody response for long term immunity remains poorly understood. The current literature provides, indeed, substantially variable estimates for its duration [96, 115, 117, 119, 124, 192]. The rate chosen for this analysis, 0.03 yrs^{-1} represents a half-life for antibodies around 23 years. This estimate was chosen based on the catalytic model which assumes individuals would remain seropositive for this duration. However, despite the fact that the longevity of the antibody response can vary due to a number of factors such as age [96, 117] or treatment [118], this rate might be unrealistic and therefore a better estimate based on longitudinal data would be more appropriate.

The model outlined here is subject to a number of other assumptions. In particular it assumes that the parameter measuring exposure is not age dependent. Both adults and children are equally exposed to infectious mosquitoes. However, a study conducted by Carnevale [78] has shown that the mosquitoes biting rate is age-dependent, with a preference three fold higher for adult. Therefore a constant force of infection across all age groups is unlikely to occur. The developed model would benefit from an age-specific force of infection based on similar concept used by Smith *et al.*[67] or by Griffin *et al.* [93] for the transmission model where an age-specific biting rate was fitted to the data.

I have fitted the model to data on antibodies against MSP-1 and AMA-1 antigens, assuming they were similarly responding to the infection. However, the immunogenicity for these antigens is relatively different [97]. Antibodies to MSP-1 antigens might bind faster while antibodies to AMA-1 might produce a higher antibody response [114]. These differences in antibody acquisition might explain the discrepancy observed between the predictions of antibody levels of MSP-1 and AMA-1 antibodies, already observed in the data with higher prevalence and higher average titre for antibodies against AMA-1 antigens.

The model has assumed that infections with *P. vivax* and *P. falciparum* were undergoing the same biological process. However, the main characteristic which differs from *P. falciparum* is the presence of dormant hypnozoites in the liver. The exposure to blood stage antigen might therefore be due to a relapse of hypnozoites [181]. Therefore assuming a systematic boost of blood stage antibodies following exposure to *P. vivax* might not be an adequate representation of reality. Results might underestimate antibody levels and consequently exposure, if hypnozoites remain in the liver upon a mosquito bite. This might therefore explain the lower predicted prevalence of *P. vivax* antibodies in comparison with *P. falciparum* antibodies.

Additionally, as demonstrated in this chapter, the model has replicated conclusions made by a classic catalytic model showing seroprevalence for *P. falciparum* was consistently higher than for *P. vivax.* However, Cook *et al.* published an evaluation of transmission in Cambodia using serological data [133] and showed that there was spatial heterogeneity of the distribution of the *Plasmodium* species with *P. vivax* dominating in the west and *P. falciparum* in the east, probably due to ecological differences. The country-wide data used for the model in the chapter might therefore experience some heterogeneity in the distribution of plasmodium species that would explain differences in antibody levels. Therefore in order to differentiate between biological and environment hypotheses to explain differences in antibody level between species, it would be interesting to adjust for environmental factors (e.g. test the model where levels of *P. falciparum* and *P. vivax* are known to be the same). The results would help to understand the extent to which the model would be valid for *P. vivax*.

2.6 Conclusion

This chapter has summarised the development and application of a discretised density model allowing for separation of the effect of exposure and immunogenicity of antigens on the immune response. Correlation of the measures of exposure obtained with the density model and current methods was demonstrated. On the other hand, given the simplistic form of the model, only qualitative conclusions indicating a density-dependency of the rate of acquisition of antibodies were established. Complex models providing a more realistic insight of the biological processes could not be assessed. A continuous model is needed to better capture the acquisition of antibodies.

Chapter 3: Development of a density model for antibody dynamics

In the previous chapter, I investigated the acquisition of immunity in a discrete framework assuming independence of the parameters responsible for the boost of antibodies. In this chapter, I extend this model and use a continuous distribution to model the boost of antibodies upon infection. A continuous framework will be more representative of the actual process, while decreasing the total number of parameters and providing more easily interpretable information.

3.1 Introduction

As presented in Chapter 1, a blood stage antibody response is initiated by the presence of the *Plasmodium* parasite at the erythrocytic stage. These antibodies have eligibly an important role in malaria immunity. However, the protective function of some induced antibody responses remains unclear. Indeed, some studies have shown a lack of correlation with lower prevalence of parasitaemia or with lower rate of disease [65, 110]. Also, protective immunity presents some important challenges with antibodies which might not last long enough to protect from subsequent infections [115, 116, 118]. Additionally, antigenic variation might render naturally acquired immunity obsolete [115, 193].

In addition to their role for protection against malaria, antibodies have a role of marker of exposure, useful in a seroepidemiological context. The presence of antibody titre to blood stage antigens give some indication on the interaction, past or present, between malaria parasite and individuals [194]. Antibody titres have been associated with malaria exposure and high concentrations of antibodies have been reported in areas of high EIR [111, 195]. Also, in areas where transmission was reduced by insecticide treated nets, a decrease of concentrations of antibodies was recorded [196].

However, many other factors are also responsible for variation in antibody response. Despite being less susceptible than parasite density to seasonal variation, antibody levels have been recorded to be higher during peaks of malaria transmission [195, 197]. This is likely to be a consequence that concentration of antibodies is higher during an acute malaria infection [198, 199]. Antibody concentration peaks around one week after a malaria episode and decay to low levels in 6 to 8 weeks [200]. The rapid decay of antibodies, widely discussed [128], the unmeasured heterogeneity of exposure and the unmeasured difference in individual's susceptibility to mount antibody response [200] are all important determinants for the levels of

antibody response and might infer some heterogeneity in the seropositive and seronegative subpopulations . For instance, individuals can lack antibodies because they are not under exposure or they have been exposed and loss their antibodies. Similarly, the seropositive population consists in a mixture of non-immune individuals recently treated for an episode of malaria and parasite positive or negative individuals with some degree of immunity. It might therefore be difficult to accurately distinguish between immune and susceptible and between exposed and unexposed individuals and this might lead to misclassification issues.

A better understanding of the processes underlying the acquisition and loss of antibodies would greatly benefit the interpretation of seroepidemiological studies. In the previous chapter, a discrete framework was used to characterise the acquisition of antibodies. However, due to the limited number of the classes of antibody level, only qualitative result could be drawn on the dynamics of the antibodies. To better characterize the kinetics of antibodies, a much higher number of classes is required to approximate a continuous model. In this Chapter, I explore a number of continuous models to mimic the biological mechanisms responsible for antibody production. In particular, I investigate whether a model in which the amount of antibodies produced following exposure depends on the current individual's antibody levels provides a more parsimonious fit to cross-sectional serological data than one in which the response is uniform. The structures of the models are based on the model presented in Chapter 2 with a higher number of compartments.

3.2 Methods

3.2.1 Data

A cross-sectional Malaria Baseline Survey (CMBS), described in detail in Chapter 2, was conducted across Cambodia in 2004. Serological samples were collected and tested against blood stage antigens. In this chapter, only antibodies against Merozoite Surface Protein (MSP-1) antigens were analysed. Also, only individuals older than 1 year old (disregarding maternal antibodies) were included in the analyses. Antibody titres were reported on a $\log_{10}$ scale. The laboratory methods employed provided reliable measurement between -2 and 4 (Patrick Corran, personal communication), therefore values outside this range were set to these limits. The data for individuals at age 0 were used as the initial state to solve the equations of the model.

3.2.2 Framework for a continuous model

The compartmental model developed here reproduces the dynamics of antibodies when individuals are exposed to malaria infection. It is based on the previously described model and focuses on the continuous aspect of the acquisition and loss of antibodies.

3.2.2.1 Individual interpretation of antibody dynamics

I developed a mathematical model to describe the dynamics of acquisition and loss of antibodies in the population. The model assumes that, following exposure to an infectious bite which occurs at rate λ, an individual's antibody level is boosted by $\delta(x_t)$ where x_t is the base-10 logarithm of antibody density. In the absence of exposure I assume antibodies decay exponentially at a constant rate ρ. A schematic representation of an individual's dynamics of naturally acquired antibodies in presented in Figure 3.1.

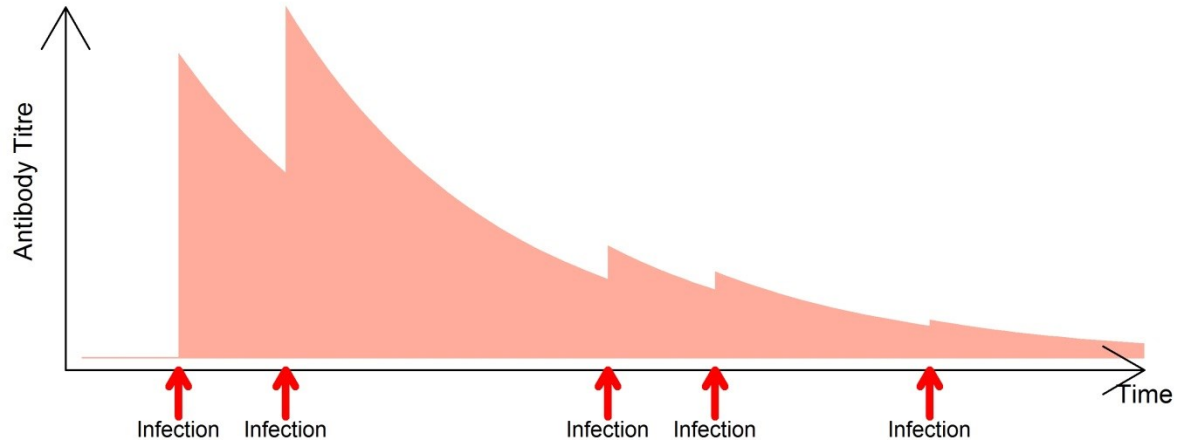


Figure 3.1: Schematic kinetics of antibodies in response to successive malaria infection

3.2.2.2 Model description

Let $y(x)$ denote the proportion of the population with $\log_{10}$ antibody level x at time t and $K(x^*, x)$ denote the probability that individuals with $\log_{10}$ antibody level x are boosted to level x^* $(x^* > x)$ on exposure to an infectious bite. Then the distribution of antibody levels in the population is given by:

$$\frac{\partial y(x)}{\partial t} = \int \lambda \left[K(x, x^*) y(x^*) - K(x^*, x) y(x) \right] dx^* + \rho \frac{\partial y(x)}{\partial x} \qquad (3.1)$$

The model was numerically approximated by a version in which the $\log_{10}$ antibody density variable, x, was discretised by dividing the range of the variable into N compartments each of width Δ, with x_i denoting

the value of ($\log_{10}$) antibody density at the mid-point of antibody class i . The first class represents measurements below the limit of detection, $x_{\min}$. I used N=51, with Δ=0.12 and $x_{min} = -2$. I also investigated the impact of the number of states on the results by analysing the model with *N=41* and *N=61* states. The resulting discrete model describes the dynamics of the proportion of the population in each antibody density category *i,* denoted y_i, and is defined by the following set of ordinary differential equations:

$$\frac{dy_i}{dt} = \lambda \sum_{j \le i} k_{ij} y_j + \frac{\rho}{\Delta} y_{j+1} - \lambda \sum_{h \ge i} k_{hi} y_i - \frac{\rho}{\Delta} y_i \qquad 1 \le i \le N \qquad (3.2)$$

where h, i, j index the N antibody level classes. The rates of exposure and decay of antibodies, λ and ρ, are assumed to be independent of antibody density and age. Multiple functions are explored to describe the probability that, following exposure, antibody levels are boosted to class *i* from class *j*, k_{ij} .

3.2.2.3 Modelling acquisition of antibodies

There is limited knowledge of the dynamics of acquisition of antibodies and even less about the density-dependency of the boost of antibodies following malaria infection. I explored a few hypotheses for the underlying mechanisms of acquisition of antibodies by testing different parameterisations of the boost of antibodies that individuals experience upon exposure.

A first assumption is what occurs in individuals with no circulating antibodies. These individuals can be "seronegative" either because they have never been exposed or because any previous antibody response has decayed to below detectable levels. Such individuals are assumed to have an antibody response that differs from those who already have circulating antibodies. I further investigate this underlying assumption about the composition of the "seronegative" population in section 3.2.3.

In order to assess the density dependency of the antibody boost size I considered the following scenarios (see Figure 3.2):

- The size of the boost of antibodies upon infection is independent of the individual's current antibody level (constant boost size)
- The size of the boost of antibodies upon infection only depends on a threshold value for the current antibody level; above this level the antibody boost is lower than below it (step boost size)
- The size of the boost of antibodies upon infection decreases continuously with increasing current antibody levels with a given parametric form (logistic, exponential or linear boost size)

This is referred to as functions for the "boost size **mean"**.

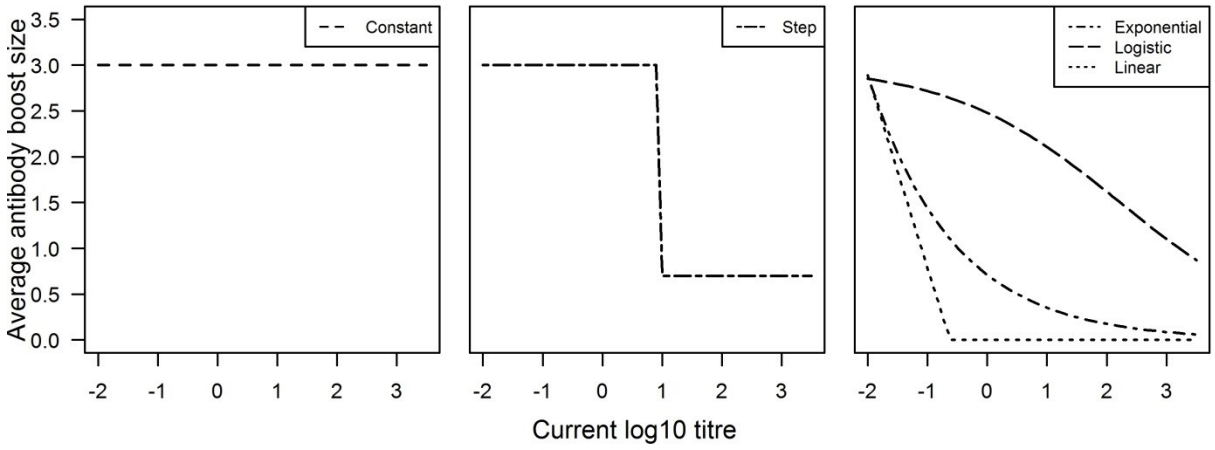


Figure 3.2: Representation of functional forms for average boost size using the parameters in Table 3.1.

In addition, there is likely to be variation between individuals in the magnitude of the boost even if they have the same current antibody level on exposure to a new infection. I therefore also compared models in which this between-individual variation was incorporated with either a Normal or a Lognormal distribution to one in which there was no between-individual variation. This is referred to as the "boost size **distribution**". Therefore, a total of 15 models were considered for comparison encompassing 5 functions for the "boost size mean" and 3 different distributions for the between-individual variation in boost size ("boost size distribution"). These are summarized in Table 3.1.

Mathematically, the between-individual variation in the boost size is calculated in the following way. The probability that following exposure, antibody levels are boosted to class *i* from class *j*, k_{ij}, is distributed according to a discretised distribution:

$$k_{ij} = \begin{cases} 0 & \text{if } i < j \\ F(x_i + \Delta/2 - x_j; \delta(x_j), S) - F(x_i - \Delta/2 - x_j; \delta(x_j), S) & \text{if } j \le i < N \\ 1 - F\left(x_{N-1} + \Delta/2 - x_j; \delta(x_j), S\right) & \text{if } i = N \end{cases} \quad (3.3)$$

where $F\left(z, \delta(x), S\right)$ is the cumulative density function at point *z* of the boost distribution with mean $\delta(x)$ and standard deviation S (see Table 3.1). Here $\delta(x)$ (the boost size mean) is a function of the current $\log_{10}$ antibody level, x, assumed to be given by: $\delta(x) = \begin{cases} d(x) & \text{if } x > x_{\min} \\ \eta & \text{otherwise} \end{cases}$

with $d(x)$ representing the function for "boost size mean" dependent on current antibody levels as shown in Table 3.1 and Figure 3.2 and η the boost size mean for individuals with no circulating antibodies.

Table 3.1: Summary of mean and distribution of boost size for each of the models

Models	"boost size distribution"[1]	Function for "boost size mean"[2]	"Boost size mean" parameterisation	"Boost size distribution " parameterisation
1	None	Constant	$d(x) = a$	$F = 1\{\delta(x) \in (X_{i-1} - x, X_i - x]\}$
2	None	Exponential	$d(x) = a\exp(-bx)$	$F = 1\{\delta(x) \in (X_{i-1} - x, X_i - x]\}$
3	None	Linear	$d(x) = ax + b$	$F = 1\{\delta(x) \in (X_{i-1} - x, X_i - x]\}$
4	None	Logistic	$d(x) = a / (1 + A * \exp(bx))$	$F = 1\{\delta(x) \in (X_{i-1} - x, X_i - x]\}$
5	None	Step	$d(x) = \begin{cases} a \text{ if } x < x_o \\ 0 \text{ otherwise} \end{cases}$	$F = 1\{\delta(x) \in (X_{i-1} - x, X_i - x]\}$
6	Normal	Constant	$d(x) = a$	$F = Norm(mean = \delta(x), sd = S)$
7	Normal	Exponential	$d(x) = a\exp(-bx)$	$F = Norm(mean = \delta(x), sd = S)$
8	Normal	Linear	$d(x) = ax + b$	$F = Norm(mean = \delta(x), sd = S)$
9	Normal	Logistic	$d(x) = a / (1 + A * \exp(bx))$	$F = Norm(mean = \delta(x), sd = S)$
10	Normal	Step	$d(x) = \begin{cases} a \text{ if } x < x_o \\ 0 \text{ otherwise} \end{cases}$	$F = Norm(mean = \delta(x), sd = S)$
11	Log Normal	Constant	$d(x) = a$	$F = \log Norm(mean = \delta(x), sd = S)$
12	Log Normal	Exponential	$d(x) = a\exp(-bx)$	$F = \log Norm(mean = \delta(x), sd = S)$
13	Log Normal	Linear	$d(x) = ax + b$	$F = \log Norm(mean = \delta(x), sd = S)$
14	Log Normal	Logistic	$d(x) = a / (1 + A * \exp(bx))$	$F = \log Norm(mean = \delta(x), sd = S)$
15	Log Normal	Step	$d(x) = \begin{cases} a \text{ if } x < x_o \\ 0 \text{ otherwise} \end{cases}$	$F = \log Norm(mean = \delta(x), sd = S)$

[1] Between-individual variation in size of the boost

[2] Dependent on current antibody level

3.2.2.4 Model parameters

The 15 models presented above are all characterised by the same following input variables: exposure rate, rate of decay of antibodies, mean boost for acquisition of antibodies for "seronegative" individuals. Only the acquisition of antibodies for those individuals with existing circulating antibodies differs. A detailed description of the parameters is presented in Table 3.2 and their use in each model is summarised in Table 3.3. Literature provides a wide range of estimates of duration of immune response but based on White's estimate [192], I have assumed throughout that the half-life for circulating antibodies is around 1 year, fixing the rate of loss of antibodies to 0.7 yr^{-1} using $e^{-\rho t_{1/2}} = 1/2$ where $t_{1/2}$ represents the antibody half-life.

Table 3.2: Parameters description & associated uninformative priors

	Parameter descriptions	**Prior distribution**
λ	Rate of exposure	U[0,100]
ρ	Rate of loss of antibodies	-
η	Mean boost for individuals with no current circulating antibody	U[0,10]
a	Maximum antibody boost size on exposure	U[0,10]
b	Slope of dependence of antibody boost on current $\log_{10}$ antibody titre	U[0,10]
xo	Maximum antibody titre above which individual's do not get a boost of antibodies	U[0,10]
A	Parameter associated with the slope in the logistic function	U[0,10]
S	Standard deviation for boost size distribution	U[0,100]

Table 3.3: Model parameters

Models	**"Boost size mean"**	**"Boost size distribution"**	Model parameters								**Total # of parameters**
						a	b	x_o	A	S	
1	Constant	None	☒	☒	☒	☒					**4**
2	Exponential	None	☒	☒	☒	☒	☒				**5**
3	Linear	None	☒	☒	☒	☒	☒				**5**
4	Logistic	None	☒	☒	☒	☒	☒		☒		**6**
5	Step	None	☒	☒	☒	☒		☒			**5**
6	Constant	Normal	☒	☒	☒	☒				☒	**5**
7	Exponential	Normal	☒	☒	☒	☒	☒			☒	**6**
8	Linear	Normal	☒	☒	☒	☒	☒			☒	**6**
9	Logistic	Normal	☒	☒	☒	☒	☒		☒	☒	**7**
10	Step	Normal	☒	☒	☒	☒		☒		☒	**6**
11	Constant	Log Normal	☒	☒	☒	☒				☒	**5**
12	Exponential	Log Normal	☒	☒	☒	☒	☒			☒	**6**
13	Linear	Log Normal	☒	☒	☒	☒	☒			☒	**6**
14	Logistic	Log Normal	☒	☒	☒	☒	☒		☒	☒	**7**
15	Step	Log Normal	☒	☒	☒	☒		☒		☒	**6**

3.2.3 Heterogeneity in the « seronegative » population

The models presented so far considered a specific antibody response for individuals with no circulating antibodies, corresponding to an antibody titre equal to the limit of detection $x_{\min}$ and termed "seronegative" population in this context. However, this assumption might not be valid as there is inherent heterogeneity between individuals presenting with no circulating antibodies. Indeed, individuals classified as "seronegative" could either be individuals who have never experienced an infection or individuals who had an infection and whose circulating antibodies have been lost or fallen below the limit of detection. As a result, I additionally considered parameterising the « seronegative » individuals as a mixture of sub-populations representing these different populations. Five hypotheses about the antibody response for "seronegative" individuals were assessed. A schematic representation of these is presented in Figure 3.3.

- Hypothesis H1: "Seronegative" individuals do not develop any specific antibody response and their mean antibody boost η is constrained to equal the maximum boost size seropositive individuals can experience, $\eta = a$
- Hypothesis H2: "Seronegative" individuals have a specific antibody response and so the model is not constrained, i.e $\eta \neq a$
- Hypothesis H3: A proportion $(1-\gamma)$ of the "seronegative" individuals have never been exposed and will never be exposed and therefore won't produce any antibody response whilst a proportion γ of the population is exposed and develop a specific antibody response η_1, therefore the overall mean antibody boost size for seronegative individuals is $\eta = \gamma\eta_1$.
- Hypothesis H4: "Seronegative" individuals represent a mixed population of $w\%$ of never infected individuals (and are now susceptible) and $(1-w)\%$ of previously infected individuals with no detectable antibodies. The latter group are assumed to have a mean antibody boost η_2 and the other group a different mean antibody boost η_1. Thus the overall mean antibody boost is $\eta = w\eta_1 + (1-w)\eta_2$.
- Hypothesis H5: "Seronegative" individuals represent a mixed population of never exposed individuals and previously infected individuals with no detectable antibodies but only 80% of them will ever get exposed, i.e $\eta = 0.8*(w\eta_1 + (1-w)\eta_2)$, and $\eta_2 = \eta_1 * \beta, \quad \text{with } \beta \in [0,1]$. This hypothesis was based on the observation in the cross-sectional data from Cambodia with at least 20% of individuals being seronegative (with antibody level below the level of detection) for all age classes counfounded at the time of the survey.

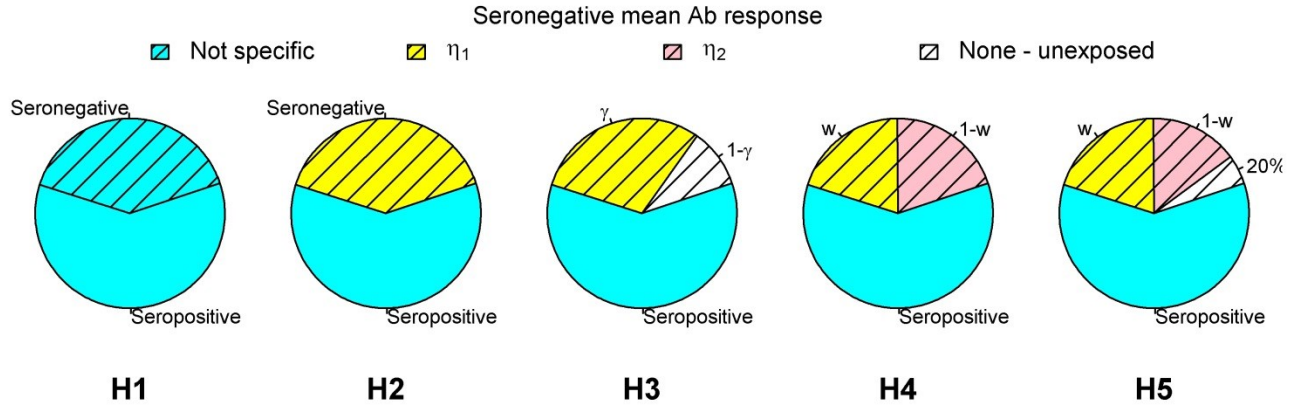


Figure 3.3: Schematic representation of the five different hypotheses to explain the composition and the antibody response of the seronegative population.

For each of these 5 hypotheses, the model used has the same structure as *Model 12* from above, with a Lognormal distribution for the boost size and an exponentially decreasing mean boost size with increasing current antibody level for modelling the density dependent acquisition of antibodies. The four parameters common to all five model variants were the exposure rate λ, the decay of antibodies rate ρ, and the parameters associated with antibodies kinetics (a, b and S). The extra parameters associated with each hypothesis are presented in Table 3.4.

Table 3.4: Model parameters for seronegative population and associated uninformative priors.

Parameters \ Hypotheses	H1	H2	H3	H4	H5	Prior distribution
Mean boost size η **(or** η_1**)**		✗	✗	✗	✗	**U[0,10]**
Mean boost size η_2				✗		**U[0,10]**
Proportion of exposed individuals γ			✗			**U[0,1]**
Proportion of previously infected individuals w				✗	✗	**U[0,1]**
Scaling factor β					✗	**U[0,1]**
Total number of parameters.	**4**	**5**	**6**	**7**	**7**	

3.2.4 Bayesian framework for inference of model parameters

I used a Bayesian approach to estimate the model parameters by fitting the model to the data. For anti-MSP-1 antibodies, the rate of decay of antibodies was fixed. Using θ to denote the estimated parameter vector and $D = \{(x_i, t_i)\}$ the data including the observed antibody level and age of individual *i*, the multinomial log-likelihood is given by:

$$l = \log(P(D \mid \theta)) = \sum_{t \neq 0} \sum_{i} n_{i,t} \log(y_{i,t}) \qquad (3.4)$$

Here $n_{i,t}$ and $y_{i,t}$ are, respectively, the observed number and predicted proportion of individuals in antibody category *i* at age *t*.

MCMC methods were used to calculate the posterior distribution of the parameters (details in Appendix I). As all parameters were positive-definite I used a Log-normal random walk proposal density and assumed uninformative uniform priors (see Table 3.4). After a burn-in period of variable length, I performed 500,000 iterations for each run of the MCMC algorithm. Chain convergence was checked visually. The output was then recorded every 100 iterations to generate a sample from the posterior distribution of a minimum size of 4,500. The standard deviation of the proposal distribution was tuned in order to achieve appropriate mixing of the chains and an acceptance rate close to 20% [190].

3.2.5 Model selection

To discriminate between models one should ideally use a criterion based on a trade-off between complexity of the model and the fit to the data. The Deviance Information Criterion (DIC) [201] provides a measure of the goodness of fit weighted against the number of parameters being estimated and is defined as: *DIC = goodness of fit + complexity of the model*

The goodness of fit is evaluated through the deviance defined as:

$$Dv(\theta) = -2\log(P(D/\theta)) \quad (3.4)$$

for the likelihood P(D/θ). The complexity of the model is assessed with the estimation of the 'effective number of parameters', defined as "posterior means deviance - deviance of posterior means":

$$p_{Dv} = E_{\theta/D}[Dv(\theta)] - Dv\left(E_{\theta/D}[\theta]\right) = \overline{Dv(\theta)} - Dv(\bar{\theta}) \quad (3.4)$$

The DIC is then defined as:

$$DIC = Dv(\bar{\theta}) + 2p_{Dv}$$

$$= \overline{Dv(\theta)} + p_{Dv} \quad (3.4)$$

Models with smaller values better support the data. It is imperative to note that only the differences in DIC are important and its absolute size is irrelevant.

In total, this Chapter presents 75 different models; 15 for the boost size (presented in Section 3.2.2) combined with 5 hypotheses for the heterogeneity in the "seronegative" population (presented in section 3.2.3). However, in practice, a model selection was first performed to select the best model for the boost size (assuming a specific antibody response for the "seronegative" population: H2). Subsequently, based on this selected model, the 5 hypotheses for the heterogeneity in the "seronegative" population were compared to define the best model.

3.2.6 Fitting the model to simulated datasets

To assess the performance of my method to reproduce estimates of transmission intensity, I designed simulations based on the following scenario. A population was exposed at a constant exposure rate λ=5.5 yrs^{-1} and constant rate of decay of antibodies ρ=0.7 yrs^{-1}. I used *Model 12* from the model presented above, assuming a Log normal distribution for the boost distribution with an exponentially decreasing mean boost size with increasing current antibody levels with the following values; the maximum size of the boost of antibodies is $a = 0.5$, while the decrease of boost size $b = 0.4$ and the standard deviation $S = 0.1$. "Seronegative" individuals had a mean boost size of $\eta = 0.015$. The values used for simulation were chosen so the model outcomes were similar to those observed in field datasets.

I assumed that a survey sampled individuals with the same age structure that was found in the dataset from Cambodia. The number of individuals in each antibody level category was drawn from a multinomial distribution with its associated probability corresponding to $y_{i,t}$ representing the solution of the model described above. Individuals were assigned a same antibody level if they were from the same discretised interval, corresponding to the median antibody of that interval. I simulated 100 datasets and I performed parameter estimation for each of them by fitting the model to simulated data using MCMC methods as described above.

3.2.7 Fitting the model to Cambodian datasets

In order to determine the best model that reproduces age-specific antibody levels from a cross-sectional survey, I fitted the different models to data from Cambodia using MCMC methods as described above. For the selection of the best model, I assumed the exposure level was constant over the country. In a second part, I used the best selected model to assess exposure from different areas, using the distance to the forest as a proxy for transmission intensity. Therefore, I fitted a single model that estimates simultaneously exposure levels from different areas.

When estimating multiple exposures, the log-likelihood becomes:

$$l(D/\theta) = \sum_v \sum_i \sum_t n_{i,t,v} \log(y_{i,t,v}) \qquad (3.4)$$

where $y_{i,t,v}$ and $n_{i,t,v}$ are respectively the predicted proportion and the observed number of individuals with titre i and age t in area v.

3.3 Results

3.3.1 Simulation study

The simulation study assessed the ability of the fitting algorithm to estimate the parameters. The posterior median and credible intervals of each parameter were computed for each simulated dataset and are presented in Figure 3.4. As a result, the original parameter fell within the estimated *posterior* 95% credible interval on average for 82.8% of the simulations. I observed a correlation between the rate of exposure (λ) and the maximum boost size (a). This is not surprising since individuals at a given antibody level can be there either because they were recently at a slightly lower level, were exposed and as a result had a small boost or because they were at a lower level (having been exposed less) and recently received a large boost in antibodies. The proportion of times either the rate of exposure or the maximum boost size is correctly estimated is relatively high (74%), which suggests that, despite the underlying correlation between these two parameters, the model is capable of identifying the correct exposure rate (and therefore maximum boost size). Note that the product of the exposure rate by the maximum boost size ($\lambda * a$) is correctly estimated in 96% of the cases. The estimated antibody levels are shown in Figure 3.5 (see Appendix II for details of how these are calculated). As can be seen on the plot, the predicted values from the model closely match the values from the simulated datasets.

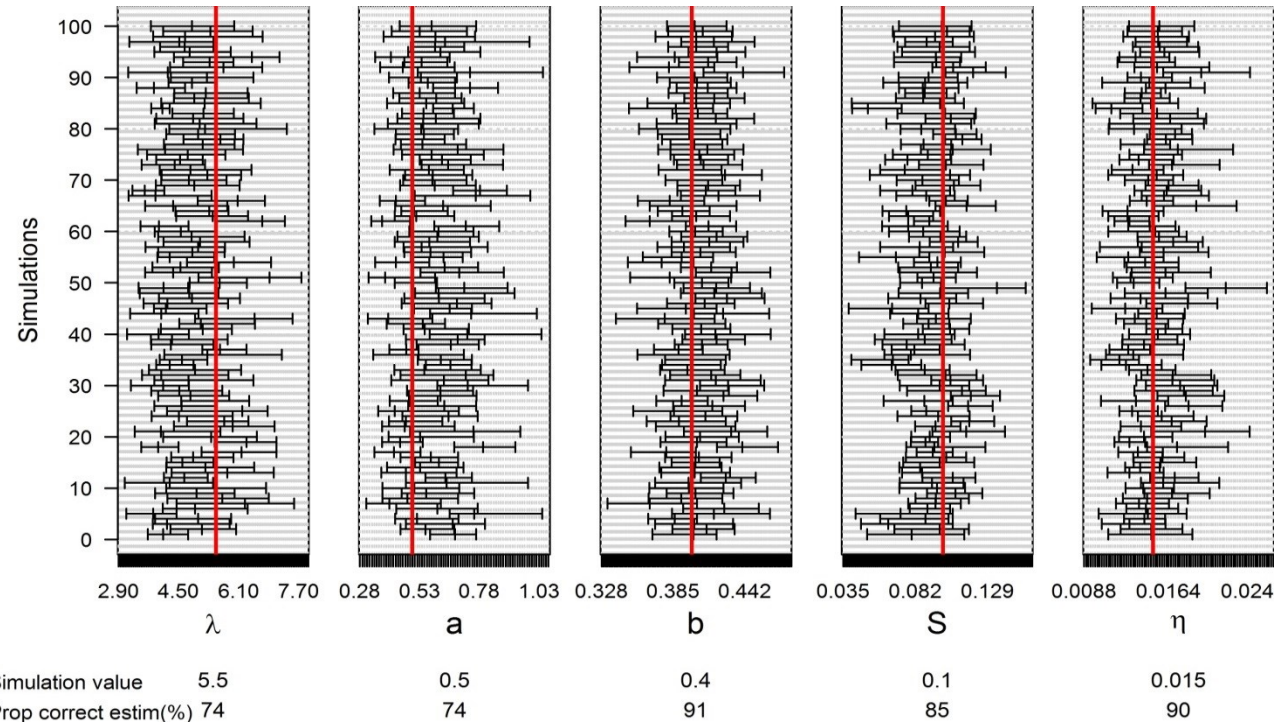


Figure 3.4 : Posterior 95% credible interval for each parameter (each panel) estimated for each of the 100 simulated datasets. The values used for the simulation are shown by the red line and the proportion of the intervals that contains the simulated value is noted below by the proportion correctly estimated. Note that when λ and a are fixed, unbiased estimates of their values were obtained from additional simulation studies (not presented).

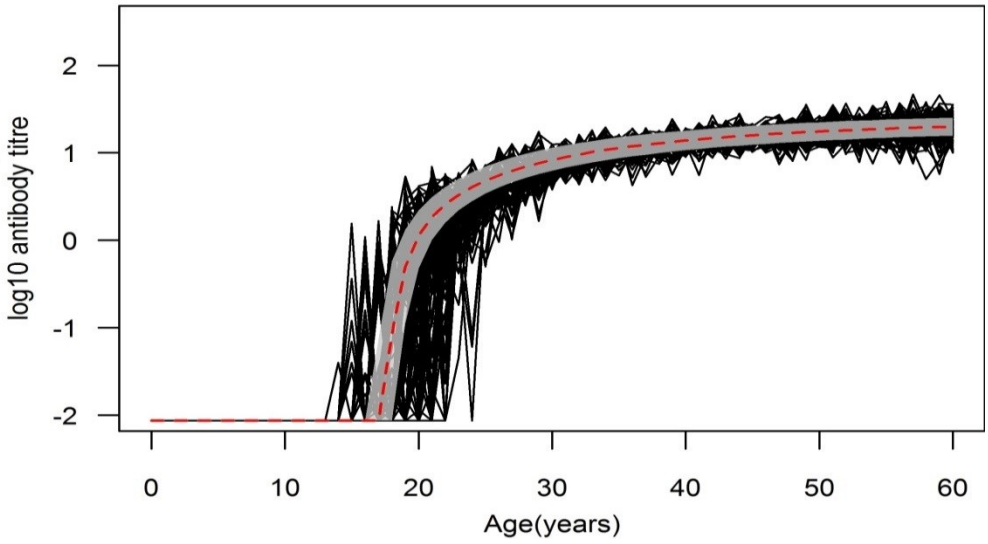


Figure 3.5 : Predicted median antibody titre. The black lines represent the antibody titres for each of the 100 simulated dataset. The grey shaded area correspond to the 95% credible interval for the model fit using the median parameters sets resulting from the estimation on each of the 100 simulated datasets and the red line corresponds to the associated median antibody titre.

Note here that I made the assumption that at birth all individuals have no antibody and the size of the boost for seronegative individuals is chosen to be very low. This explains why individuals before 15 years old do not present any antibody in the simulated dataset. However, in field data, this theoretical assumption might not be valid.

3.3.2 Model selection by fitting to Cambodia data

A series of 15 models that suggest various scenarios for the acquisition of antibodies were investigated in order to best describe this mechanism despite the limited within-host information captured in cross-sectional data. Each model was successfully fitted to the serological data from Cambodia and summary results are presented in Table 3.5. The simplest model in which there is no variation in the antibody boost size between individuals had a substantially poorer fit as judged by the DIC than other models (*Models 1* to *5*). Allowing there to be variation between individuals (*Models 6 to 15*) improves the fit of the models, with lower values of DIC obtained assuming a Lognormal boost distribution (*Models 11 to 15*). Overall, with the exception of the constant boost size distribution model (*Models 1, 6* and *11*), all models provided a reasonably good fit to the data (not presented here). However, the estimates of exposure rate were noticeably sensitive to the choice of the model. Thus relative values of exposure rate would provide better information rather than their absolute values. In addition, the exponential boost size distribution consistently provided the lowest values of DIC. As a result, the model with a Lognormal boost distribution and an exponential boost size distribution was retained as the best model as it provided a more parsimonious fit than the model with a logistic boost size distribution, which had the second lowest DIC.

Table 3.5 : Model comparisons & estimates of exposure

Models	Boost size	Boost distribution	Exposure estimate Median (95% CrI)	# param.	LogLik Median	ΔDIC*
1	**Constant**	**None**	0.4 (0.39-0.41)	3	-11996.1	5091.1
2	**Exponential**	**None**	0.9 (0.89-0.91)	4	-9844.96	772.7
3	**Linear**	**None**	0.88 (0.87-0.9)	4	-10029	1158.8
4	**Logistic**	**None**	0.9 (0.88-0.91)	5	-9844.1	784.0
5	**Step**	**None**	0.96 (0.95-1)	5	-9983.43	1065.8
6	**Constant**	**Normal**	0.4 (0.39-0.41)	4	-11996.4	5092.2
7	**Exponential**	**Normal**	5.3 (4.9-5.4)	5	-9488.46	78.5
8	**Linear**	**Normal**	1.3 (1.2-1.3)	5	-9585.31	272.4
9	**Logistic**	**Normal**	5.3 (5.2-5.4)	6	-9487.9	77.8
10	**Step**	**Normal**	0.45 (0.44-0.46)	6	-10859	2814.1
11	**Constant**	**Log Normal**	0.4 (0.39-0.41)	4	-11996.4	5091.8
12	**Exponential**	**Log Normal**	5.6 (4.1-6.9)	5	-9495.66	0.0
13	**Linear**	**Log Normal**	1.8 (1.6-2.1)	5	-9506.79	114.7
14	**Logistic**	**Log Normal**	5.4 (3.8-6.8)	6	-9496.08	53.2
15	**Step**	**Log Normal**	2.2 (2-2.3)	6	-9516.77	132.3

* $\Delta DIC_i = DIC_i - DIC_{12}$

3.3.3 Best model

The most parsimonious model (*Model 12*) assumes that, following a new infection, the size of the boost in antibody levels decreases exponentially for increasing current antibody levels. As can be seen in Figure 3.6, the data and the fitted antibody density curves for anti MSP-1 antibodies show an increase in antibody titre with age. There is a good fit to both the age-specific antibody titre and overall antibody distribution for individuals with circulating antibodies, despite the model over-estimating the number of individuals without circulating antibodies.

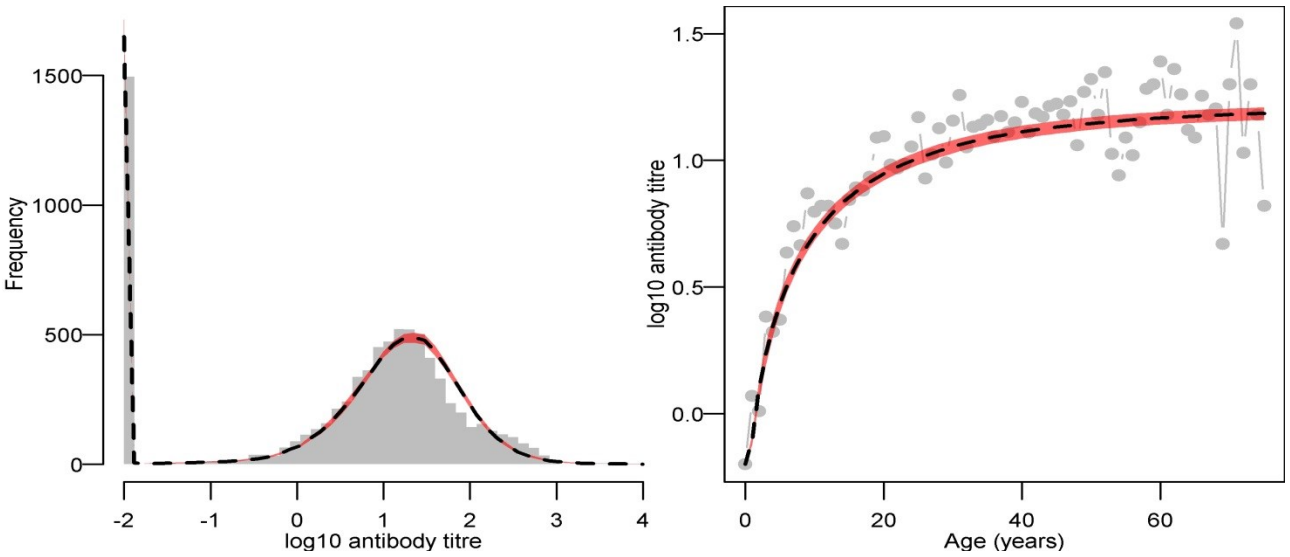


Figure 3.6 : Model fit for the overall antibody distribution (left) and age specific antibody distribution (right). Actual data are represented in grey while the dashed line represent the median fit and the pink shaded area the associated 95% credible interval.

Estimates of the different parameters are shown in Table 3.6. The key result here was the measure of exposure rate, estimated to 5.6 yr^{-1} (95% credible interval: (4.1-6.9)) over the whole country in Cambodia. This estimate might however be overestimated due to local heterogeneity in transmission intensity [202], not accounted for in the model. A secondary aim was to gain insight into the parameters used to explain the antibody kinetics. The small antibody boost size for individuals with no current antibodies indicate that these "seronegative" individuals tend to produce very small amount of antibodies and multiple infections are therefore required to acquire higher levels of antibodies. Indeed, upon exposure, I estimated that only 1% of the "seronegative" individuals will acquire antibodies, which would explain why the model overestimated the number of "seronegative" individuals. In addition, the significant difference between the size of antibody boost for "seronegative" individuals and the maximum boost for "seropositive" individuals might indicate that individuals need to be primed to achieve a high antibody response. The density-dependent antibody boost size is illustrated in Figure 3.7. It can be seen that the uncertainty around the size of the boost decreases as individual's current antibody titre decreases.

Table 3.6 : Parameter estimates for the most parsimonious model (*Model 12*)

Param.	Description	Estimates
λ	Exposure Rate	5.6 (4.1-6.9)
α	Maximum antibody boost size on exposure	0.44 (0.31-0.75)
b	Slope of the decreasing boost size	0.39 (0.33-0.46)
S	Standard deviation for boost size distribution	0.11 (0.09-0.13)
η	Mean boost for individuals with no current antibody	0.016 (0.013-0.022)

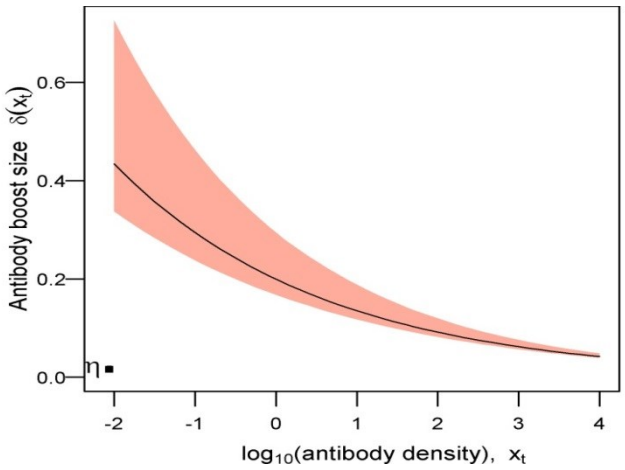


Figure 3.7 : Antibody boost size (δ) depends on individual's current antibody level when individual get exposed (x_t) Black line (and the pink shaded area) represents median boost size for individuals with circulating antibodies (95% credible interval). Black point indicates boost size mean for "seronegative" individuals.

MCMC Diagnostics:

Four MCMC chains were run in parallel, sampled every 1,000 iterations and combined to obtain a sample of size 1,800. The lowest effective sample size was 100. *Posterior* densities of the parameters and the MCMC trace are presented in Figure 3.8, demonstrating relatively good convergence of the MCMC chain with smooth posterior distributions obtained for each parameter.

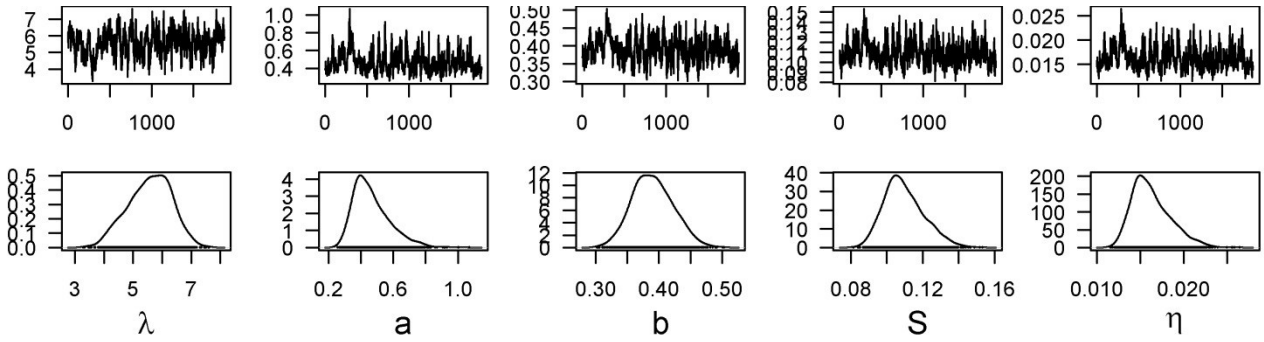


Figure 3.8 : MCMC trace (above) and posterior distributions (below) for the exposure rate (*λ*), the maximum boost size (*a*), the slope of the decreasing boost size (*b*), the standard deviation of the size of antibodies (*S*) and the mean of antibodies for individuals with no circulating antibodies (*η*).

There was some *posterior* correlation between the model parameters. In particular, as expected, the exposure rate was negatively correlated with the size of the antibody boost (average for "seronegative" and maximum for "seropositive" individuals) with a correlation of 0.96 with *η* and 0.97 with *a*. Also, the variability amongst individuals, modelled by *S*, had unexpected associations, i.e. a negative correlation with maximum boost size for "seropositive" individuals (0.89) and a positive correlation with mean boost size for "seronegative" individuals (0.97).

Additional analyses show that, for this specific model, changing the number of antibody categories does not affect the estimates of the exposure rate (See Annex III).

3.3.4 Heterogeneity in the "seronegative" population

In addition to investigating the dynamics of acquisition of antibodies when individuals already had circulating antibodies, I was interested in better understanding the acquisition of antibodies for individuals that do not have any circulating antibodies. A total of 5 hypotheses were investigated. Models corresponding to each of these hypotheses were fitted to the data and the MCMC algorithms all converged reasonably well. Posterior median and 95% credible intervals are presented in Table 3.7. With the exception of H3 and H4 that present very similar estimates for most of the parameters, the estimates of the exposure rate, as well as the estimates for the biological parameters (a, b and S), differed substantially from each other between the different hypotheses. The estimate of w for hypothesis H5 was uncertain (wide credible interval), giving limited information about the exact proportion of the population that gets their antibody level boosted upon infection. Similarly, the estimate of the variability between individuals was imprecise for models H1 and H5. Note that the results with *H4* are consistent with those found not only in *H3* but also in *H2*. Indeed, when the model assumes a mixture in the seronegative population, it predicts that 96% of the seronegative have an antibody response with a mean boost size around 0.016 (similar to results found in *H2* when the mixture in the seronegative population is ignored) and 4% have an antibody response with a mean boost size around 2.3 (similar to the results found with *H3*).

Table 3.7 : Parameter estimates for of models for heterogeneity of the "seronegative" population.

Hyp	λ	η (or η_1)	η_2 (or β)	w (or γ)	a	b	S
H1	**0.43** (0.4-0.45)				**3.9** (3.8-4.1)	**0.45** (0.43-0.47)	**0.067** (0.02-0.21)
H2	**5.6** (4.1-6.9)	**0.016** (0.013-0.022)			**0.44** (0.31-0.75)	**0.39** (0.33-0.46)	**0.11** (0.09-0.13)
H3	**1.5** (1.3-1.7)	**2.3** (1.8-2.6)		**0.036** (0.029-0.043)	**4.6** (3.3-7)	**0.75** (0.67-0.87)	**0.26** (0.23-0.29)
H4	**1.5** (1.3-1.7)	**0.016** (0.007-0.042)	**2.3** (1.7-2.6)	**0.96** (0.96-0.98)	**4.7** (3.3-7.2)	**0.75** (0.67-0.88)	**0.26** (0.20-0.29)
H5	**0.8** (0.78-0.81)	**0.0054** (0.001-0.029)	*(*β*=)* **0.49** (0.11-0.96)	**0.48** (0.04-0.97)	**5.5** (5.3-5.8)	**0.71** (0.69-0.73)	**0.15** (0.07-0.21)

The different models were compared using the DIC (see Table 3.8) and the model *H2* had the lowest DIC, suggesting it might represent the best model.

Table 3.8 : Model comparison for the various hypotheses about heterogeneity of the « seronegative » population.

Hypotheses	# parms	LogLik	ΔDIC*
H1	4	-10713	2498.0
H2	5	-9495.66	0
H3	6	-9496.65	91.1
H4	7	-9497.23	91.7
H5	7	-10109.2	1319.9

* $\Delta DIC_i = DIC_i - DIC_{H2}$

When looking at the model outputs (see Figure 3.9), model *H1* predicts that all "seronegative" individuals will systematically increase their antibody level when they are exposed. However, this model also predicts a lower number of seronegative individuals than observed in the data. In contrast, model *H2* predicts the lowest proportion of individuals that will acquire antibodies upon an infection and also predicts the highest number of "seronegative" individuals. As seen above, the estimated parameters for models *H3* and *H4* were very similar and hence the predicted numbers of "seropositive" and "seronegative" individuals were identical. Additionally, the predicted number of "seropositive" individuals for *H2* was similar to the results found for *H3* and *H4*, while the number of seronegative was different. All models provided a reasonably good fit to the data. However, models *H1* and *H5* gave the worst overall fit to the distribution of antibodies by underestimating the levels of antibodies for adults and overestimating it for children. Indeed, the better fitting models as judged by the DIC (*H2*, *H3* and *H4*) visually gave the best fit for individuals and overestimate the number of "seronegative" individuals, in particular for *H2*.

Overall, all models provide a reasonably good fit. I chose to retain model *H2* as it presents the best DIC and represents the most parsimonious model despite marginally overestimating the number of seronegative individuals.

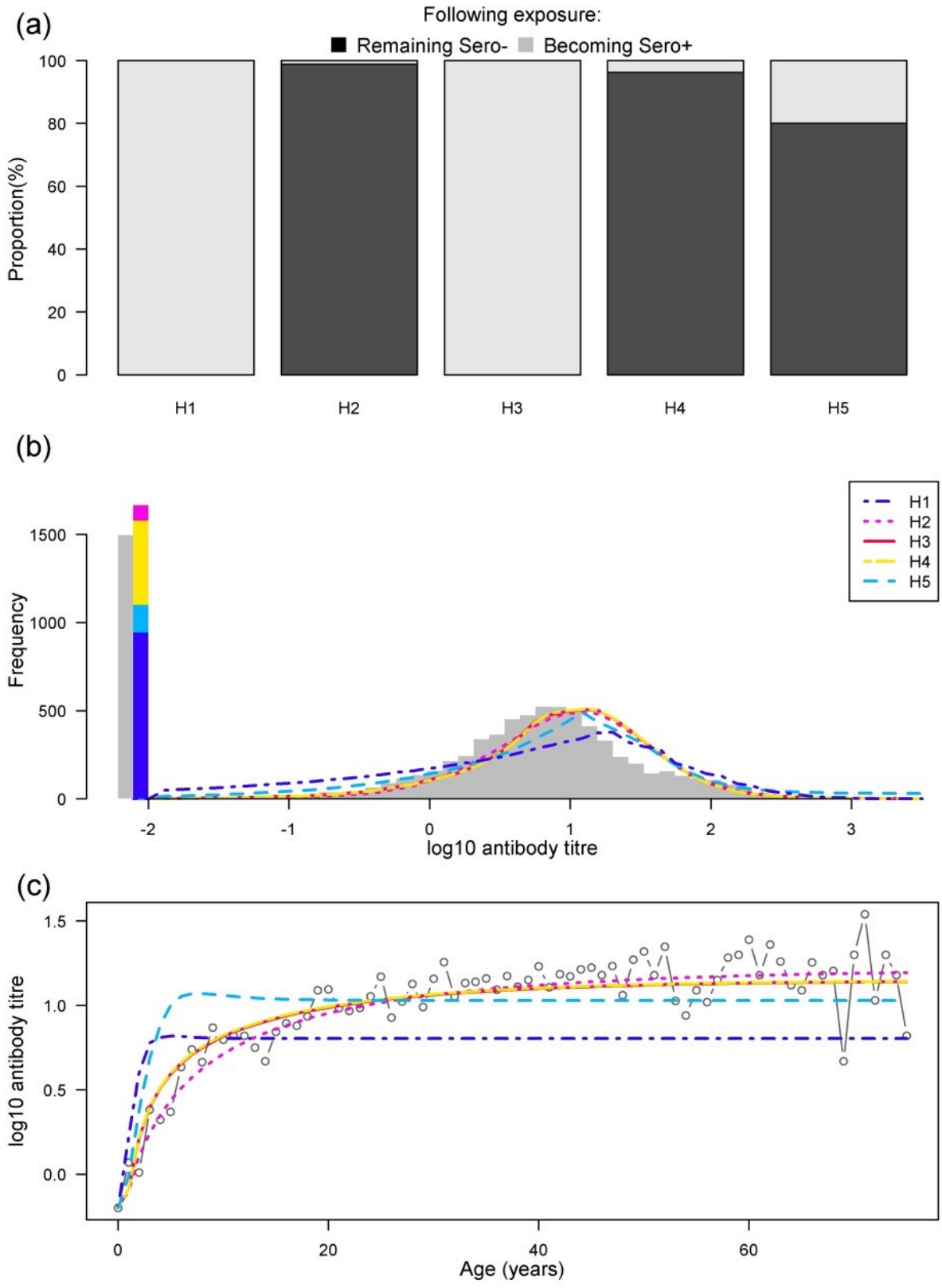


Figure 3.9 : Model predictions for (a) the proportion of individuals that remain seronegative/become seropositive upon infection, (b) the overall antibody distribution, (c) age specific antibody distribution categorised by model. The 5 models for heterogeneity of the seronegative population are represented by the different coloured lines and the data are represented in grey.

3.3.5 Multiple regions

In Cambodia, the distance to the forest is likely to be a proxy for exposure levels [179]. Therefore *Model 12* was extended, as presented in section 2.4, to simultaneously estimate the exposure rate in 5 regions categorised according to their distance to the forest. The MCMC algorithm converged well and the parameter estimates are presented in Table 3.9. As can be seen in the table, there is a trend where exposure rate, represented by λ, decreases with increasing distance to the forest. This extended model provided a better fit to the data according to the DIC (equal to 18311), lower than the DIC estimated from the homogeneous model in section 3.3.2 (DIC=18902). The estimated exposure rates were the same order of magnitude as those from the simpler model, with a much higher level of exposure for individuals within the forest. There were no significant difference in the estimate of the maximum boost size *a* between this extended and the original models which did not consider multiple exposure levels.

Table 3.9 : Parameter estimation for model extended to account for multiple exposure

Parameters		Estimates Median [95% Credible Interval]
λ	In forest	5.9 (5.8-6)
	<200m	5.3 (5.2-5.4)
	200m-500m	5.2 (5.1-5.3)
	500m-1km	5.1 (5-5.2)
	>1km	5 (4.9-5.1)
a		0.65 (0.62-0.69)
b		0.5 (0.48-0.51)
S		0.015 (0.013-0.023)
η		0.006 (0.005-0.020)

The fit of the model to the data is presented separately for each region in Figure 3.10. The model fit for the "seropositive" population fit the data well for the five regions, only the predicted number of "seronegative" individuals appear to present some lack of fit, in particular for individuals in the forest. Overall, the model fits the age structured data relatively well. However a few discrepancies between regions are worth noting. One of the main differences between the age structured output of the model and the data lies in the presence of extremely small values for children below the age of 1 (used as a starting point for the resolution of the equation of the model). Indeed, I had assumed that individuals start their life with the same antibody levels than children at age 0 (taken from the data) but in regions between 200m and 1km away from the forest, the median antibody titre for individuals below 1 year of age is equal to -2, causing the model to underestimate the antibody level between 0 and 20 years old. However, by ignoring the decay of maternal immunity, the model underestimates by around 3 fold the number of "seronegative" individuals living in the forest. Also, in areas where individuals tend to be less exposed (>1km), the antibody level is underestimated for individuals older than 5.

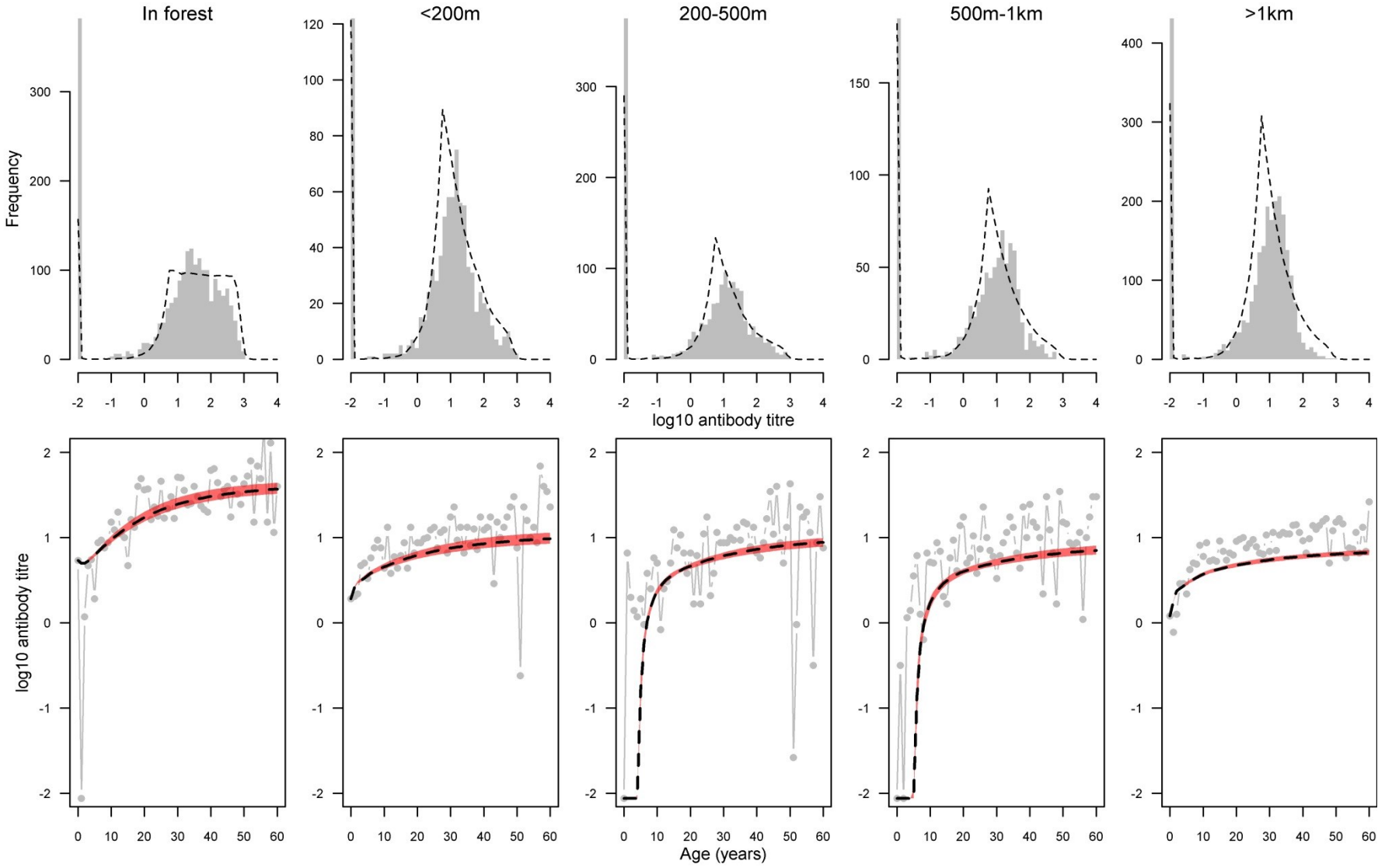


Figure 3.10 : Model fit for the overall antibody distribution (above) and age specific antibody distribution (below) across regions categorised by their increasing distance from the forest (left to right). Data are represented in grey while the dashed line represents the median fit and the pink shaded area the associated 95% predicted interval.

3.4 Discussion

The objective of this chapter was to develop a model that incorporates an exposure rate and antibody kinetics to determine continuous levels of antibody density in a cross sectional survey. The rationale for developing a density model was to better estimate malaria transmission intensity by taking into consideration the full information contained in serological data, i.e. antibody levels rather than just seropositivity as has typically been done. The model exploits the continuous characteristic of the antibody quantity measurements (discretised for implementation reasons) and, using a simple model of the acquisition and loss of antibodies following exposure, replicates the observed antibody levels measured in individuals during a population-wide survey in Cambodia. Given the limited understanding of the precise role of specific blood-stage antibodies in protection from infection and disease, I developed and tested a series of models assessing different hypotheses for the acquisition of antibodies. The most parsimonious model was one which assumes the quantity of antibodies produced is Lognormally distributed between individuals with a mean that depends on the antibody level individuals have when they get exposed. The boost is assumed to be decreasing exponentially in individuals with increasing levels of antibodies. I have also assumed that individuals who do not have any circulating antibodies, or are below the detection limit equal to x_{min}=-2 on a log10 scale, would have a constant specific antibody response. The validation of the method through simulation study demonstrates that the MCMC method used to estimate the parameters performs correctly. By fixing the rate of decay of antibodies, the fitting could identify all parameters with good chain convergence.

In the simulation study, the model was fitted to 100 simulated datasets, produced with the same model and same set of parameters. This resulted in the *posterior* 95% credible interval for the parameters to include the values used for the simulation 83% of the time. There is a high level of correlation between the parameters. Therefore, while the model tends to slightly underestimate the exposure rate, it also tends to simultaneously overestimate the maximum size of the boost of antibodies. However, the proportion of individuals being boosted to a fixed level of antibodies (i.e. frequency x size of boost) remains constant and in agreement with the simulation values. The size of the boost for seronegative individuals is however systematically well estimated despite the correlation with other parameters.

It is broadly agreed that antibodies to malaria infection are produced relatively quickly during an infection [203] with antibodies to blood stage antigens appearing within days of infection [56, 116, 128]. However, to my knowledge, there is no detailed description relating the amount of antibodies produced to the current antibody levels. The model explored here is a very simplistic representation of the true picture by ignoring the underlying processes responsible for the production of antibodies. Indeed, parasite antigens are first recognised by B cells that will proliferate. B cells will then differentiate into plasma cells (antibody secreting

cells) which will secrete circulating antibodies that will be measured in surveys [204]. In reality the model did not allow to distinguish with the available data whether individuals will produce fewer antibodies due to a large amount of circulating antibodies or due to any other feature about memory B cells and/or plasma cells [205]. Also, some studies suggest that levels of antibody are maintained due to the continuous production of antibodies by long- lived plasma cells that migrate to the marrow bones [96, 206]. This is not an assumption that I considered in the model but the fit of the model to the data could indicate that the model represents an approximation, averaged over a year, of the real mechanism responsible for the antibody response.

The persistence of antibodies in individuals previously exposed to malaria infection is debated [56, 124, 207]. In the model developed in this chapter, I made the simple assumption that, in the absence of exposure, individuals lose their antibodies at a constant rate 0.7 yr^{-1}, corresponding to a half-life of 360 days. This estimate was based on the results from a study on the duration of antibody response to *Plasmodium* infection. Indeed, White estimated the half-life of MSP-1 long-lived B cells to be 376 days (233-600) days for children in Gambia [192]. These estimates were obtained from longitudinal data assuming that there are short- and long- lived B cells producing circulating antibodies. Therefore, I have also made the underlying assumption that the measured antibodies were produced from long- lived cells. However, if the measured antibodies are in reality short lived, the model might in consequence over-predict the antibody levels and consequently also overestimate the exposure level. Further analyses of the antibodies kinetics using longitudinal data might be necessary to better understand their duration and better inform the model.

By investigating multiple models for the "seronegative" population I wanted to assess the heterogeneity in the immune response for the acquisition of antibodies during the first infection. There is inherent heterogeneity between seronegative individuals [200]. Indeed, individuals with no circulating antibodies are people who either have never experienced an infection or had an infection and lost their circulating antibodies. Thus, despite their lack of circulating antibodies, some apparently "seronegative" individuals may have been exposed in the past and therefore have low levels of memory cells. This immune memory would thus facilitate a better immune response once re-infected. Although the model that assessed heterogeneity in the "seronegative population" (Hypothesis 4) did not have the lowest DIC, the parameter estimates for this model suggest that the observed antibody levels for "seronegatives" in these data could be captured by a model in which those that have never previously been infected experience a small boost on infection while those previously-infected are already primed and experience a larger boost in antibody levels. If this assumption is valid, this form of a density model would better estimate the acquisition of antibodies for individuals who never had infection or antibodies. However with available data collected

from cross-sectional surveys, the model might be over-parameterised and hence any inference might be limited. To better estimate this I would need longitudinal data that would allow me to calibrate my model with the parameters that account for the antibody kinetics. I might also expect heterogeneity in the seronegative population to be due to the fact that some individuals might have been treated and therefore develop slower immune responses [200].

Additionally, asymptomatic individuals who have not been treated are most likely to have been exposed before and are very likely to already have an immune memory. The age of individuals has not been considered as a determinant for the acquisition of antibodies but might be responsible for the heterogeneity in the seronegatives, as children might experience smaller boosts in antibodies than adults [60, 208]. Finally, measurement error or undetectable antibody titre [85] could also mislead the results by misclassifying individuals who actually had antibodies, as individuals with no circulating antibodies.

The aim of this work was to develop a model that could better estimate malaria intensity levels from serological data. My results demonstrate that it is possible to infer differences in exposure rate. Here I used data from Cambodia to assess exposure in 5 different regions categorised by their distance to the forest, considered to be a surrogate for malaria exposure and intensity [179]. I observed a trend between increasing distance to the forest and decreasing exposure rate. In addition, in comparison with the results found for the catalytic model in chapter 2, I observed a correlation between the exposure rate estimated with this density model and the seroconversion rate estimated using a classic catalytic model. This indicates that this density model could be a novel method used to measure malaria transmission intensity.

It is worth noting that the models considered here make numerous inferences on within host parameters that could not be directly measured using cross-sectional surveys. I am therefore making assumptions about unobservable data while, with good quality longitudinal data I could estimate parameters associated with the acquisition of antibodies and therefore better inform the measure of interest, i.e. transmission intensity. Despite limited information on the loss of antibodies, I could calibrate the model with values similar to the available estimates [192]. So if I could calibrate models using longitudinal data to inform the other biological parameters, I would also be able to consider other models that were disregarded because of their level of complexity.

3.5 Conclusion

In summary, my findings show that we can use a density model that mimics individual's antibody kinetics and takes into consideration exposure rate to predict age specific antibody titres collected in a cross sectional survey. The model assumes a constant exposure rate and loss of antibodies with a density dependent acquisition of antibodies. The most parsimonious model assumes that individuals will produce decreasing (exponentially) amounts of antibodies as their current antibody titre increases and the size of this boost varies between individuals which is best captured by a Lognormally distribution. The fitted model was able to reproduce data from a survey carried out in Cambodia and the resulting estimates of malaria intensity correlate with the distance to the forest used as a proxy for exposure. However, levels of endemicity in Cambodia are very low and some of the results might be artefacts from such data. The methods developed here therefore need to be tested on data with from a wider range of endemicity levels and where malariometric indices are available for comparison.

Chapter 4: Estimating transmission intensity across a range of transmission settings in Tanzania

In the previous chapters I developed models that reproduce antibody levels based on malaria exposure and assumptions made on the kinetics of antibody acquisition and decay. Here, I further explore the best model obtained and test its ability to measure malaria transmission intensity simultaneously in multiple regions where exposure levels differ significantly. Additionally, I compare the results obtained by fitting my density model to data from a seroepidemiological study in Tanzania with estimates of the seroconversion rate and traditional measures of transmission intensity (EIR) from the same region.

4.1 Introduction

Malaria remains a leading cause of morbidity and mortality worldwide [15], with heterogeneous levels of endemicity across the globe [14]. Measuring malaria transmission intensity is a key element of monitoring changes in transmission and assessing the impact of anti-malaria interventions. The standard reference historically used for reporting malaria transmission intensity is the entomological inoculation rate (EIR), defined as the number of infectious bites per person per year (ibppy), estimated by catching mosquitoes in order to estimate the sporozoite (infectious) rate. Despite its usefulness in providing a direct estimate of the force of infection, this method is time consuming, expensive and can lack precision, especially in low endemicity areas. The prevalence of individuals carrying the parasite (parasite prevalence) is an alternative measurement of malaria transmission intensity. This can be estimated in cross-sectional surveys using microscopy, rapid diagnostic tests (RDTs), or increasingly using PCR methods to detect infection in individuals. This approach is in widespread use, with population-level representative samples now undertaken in many malaria endemic areas as part of national Malaria Indicator Surveys [209]. However, although parasite prevalence can be estimated rapidly in populations, it is subject to seasonal variation (with up to 30% variation between the peak and trough in highly seasonal areas in West Africa), is affected by anti-malarial treatment levels, requires highly skilled staff (for microscopy and PCR), and lacks precision in low endemicity settings.

Serological data, which measure antibody responses to one or more blood-stage antigens, offer an alternative means to estimate past exposure to malaria [125]. Serological data are typically analysed using

catalytic models to estimate the antibody seroconversion rate (SCR) - the rate at which seronegative individuals become seropositive – as a proxy for the force of infection [8,9]. However, one limitation of this method is that it is necessary to distinguish seropositives from seronegatives using continuous measures of antibody levels. For malaria this is typically achieved using sera from European (i.e. unexposed) volunteers to define a cut-off. However, there may be underlying differences between the immune responses in these unexposed volunteers and those living in endemic countries, as for instance, genetic differences is already observed between ethnic groups from a same country [210, 211]. An alternative approach is thus to fit mixture models to the bi-modal distribution of antibody levels [137, 163, 168]. This approach has the advantage of using only data from the study and assumes similar characteristics throughout the whole population, with the exception of seropositivity status. Two Gaussian distributions can be fitted for seropositive and seronegative sub-populations. However, this approach can be problematic in highly endemic areas where a large proportion of the population are seropositive. Additionally, another caveat of such method is that the population is likely to be made of many distributions, rather than only two.

A study conducted by Drakeley and colleagues in Tanzania [95] has highlighted the importance of the use of serological markers for estimating transmission intensity. In this study, altitude was used as a proxy for transmission intensity. Indeed, as transmission intensity varies with temperature affecting the development of the vector and of the parasite, transmission intensity is therefore affected by altitude associated with changes in temperature. Mosquitoes' density and parasite prevalence are influenced by topography and tend to decrease with increasing altitude, causing variation in malaria transmission intensity. Drakeley et al showed that malaria transmission intensity can be estimated using the seroconversion rate (SCR) across multiple endemicity settings.

Here I use the continuous model of the acquisition and loss of antibodies developed in the previous chapter and fit this to individual-level data on measured antibody levels from cross sectional surveys. An advantage of this approach is that it takes into consideration the full information contained in measurement of antibodies levels rather than reducing the data to seropositive / seronegative status. Estimates of malaria transmission intensity are thus derived without the use of cut-offs and with better precision than estimates obtained with binary seroprevalence data. I illustrate the utility of the approach by fitting the model to data from a cross-sectional survey in Tanzania across areas with differing endemicity and compare the results to traditional measures of malaria transmission intensity as well as to estimates obtained from traditional catalytic models.

4.2 Setting

4.2.1 Data source: Tanzania cross-sectional survey

A program investigating malaria disease burden and transmission intensity was conducted in Tanzania during short rainy seasons in 2001. The objective of the study was to use altitude and rainfall to predict malaria transmission intensity. The full study surveyed individuals from 24 villages for two age stratified cross sectional surveys, which were pooled for the analysis [212]. In his seroepidemiological study, Drakeley include a subset of 12 of these villages from West Usambara, North Pare and South Pare regions [95]. The studied area has mountains ranging between 300 and 1870m high. In each region, also referred to as transect, 4 villages were selected with different altitude: 1 at low (<600m), 2 at intermediate (600-1200m), and 1 at high altitude (>1200m). A map of the region is presented in Figure 4.1. The exposure in this area ranged from less than 1 infected bite per person per year (ibppy) for the highest villages up to ≈100 ibppy for low altitude villages close to the coast [95]. In each village, around 250 individuals were recruited (~32% between 0 and 4 years of age, ~32% between 5 and 14 years of age, ~36% between 15 and 45 years of age). Individuals were recruited on a first come first served basis over a period of three days until the required sample size was achieved. Previous publications have already described the study design and epidemiology of malaria in this area [212]. Ethical approval was obtained from the institutional review boards of the National Institute of Medical Research of Tanzania, Kilimanjaro Christian Medical Centre, and the London School of Hygiene and Tropical Medicine. Individuals' sera were collected and antibodies to the asexual stage merozoite antigens, Merozoite Surface Protein (MSP-1) and Apical Membrane Protein (AMA-1) were determined using ELISA [194].

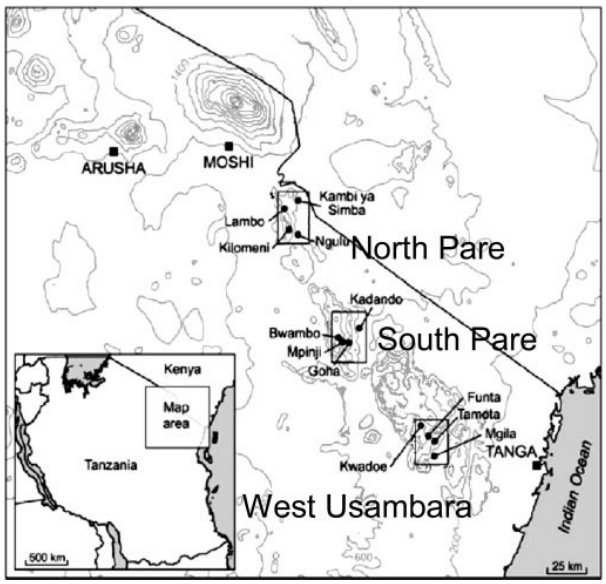


Figure 4.1: Map of the studied area showing the 3 regions and 12 villages. Figure reproduced from *Drakeley et al* [95].

During the study, parasitaemia data were also collected for each individual and parasite rate (PR) is reported in Table 4.1. Direct measurements of EIR were not available for the studied area, however EIR could be predicted for this area based either on altitude *h*, according to $\mathrm{EIR} = 331.5\ \mathrm{e}^{-0.0057h}$ [213] or based on the relationship between EIR and parasite prevalence demonstrated by Griffin *et al* [93] (See Table 4.1).

Table 4.1: Summary of study data. PR denotes parasite prevalence measured in children 0-4 years of age. EIR[1] is estimated from altitude [214]**. EIR[2] was estimated from the parasite prevalence data using a previously published relationship** [93] **.**

Transect	Village	Altitude (m)	N	PR (%)	EIR[1] (ibppy)	EIR[2] (ibppy)
North Pare	Kilomeni (Ki)	1 556	411	1	0.047	0.08
	Lambo (La)	1 188	355	10	0.38	1.11
	Ngulu (Ng)	832	486	8	2.9	0.84
	Kambi ya Simba (KyS)	746	494	10	4.7	1.11
South Pare	Bwanbo (Bw)	1 598	485	3	0.037	0.26
	Mpinji (Mp)	1 445	461	2	0.088	0.17
	Goha (Go)	1 163	453	13	0.44	1.57
	Kadando (Ka)	528	382	25	16	4.45
West Usambara	Kwadoe (Kw)	1 564	357	4	0.045	0.37
	Funta (Fu)	1 240	429	17	0.28	2.38
	Tamota (Ta)	1 055	449	19	0.81	2.74
	Mgila (Mg)	375	465	34	39	8.31

In the first part of my analyses, I selected the same 12 villages to fit the model to the data and compare measurements of exposure. Additionally, for validation purposes, 8 villages from Rombo and West Usambara regions were considered. As infants may present with maternal antibodies, only individuals between 1 and 45 years were included in my analyses. Data on individuals at age 0 were however used as initial state to solve the equations in the model. Only anti-MSP-1 antibodies were analysed. Originally, measurements were recorded as optical densities and I log-transformed them prior to analysis due to the high number of individuals with low levels of antibodies. Measurements below the limit of detection were assigned to approximate limit of detection (LoD) of 0.01 (i.e. -2 on a log10 scale). Also, assay results above a value of 4 were considered to lack accuracy and were therefore set to an upper limit of 4.

4.2.2 Descriptive analysis

The distribution of the antibody level, measured in optical density unit (OD) is presented for each studied village in Figure 4.2. The altitude of villages decreases from left to right on the graph. Hence, for higher altitude as in Kilomeni, Bwambo and Kwadoe, most of the population have very low antibody level. In contrast, in villages with lower altitude, such as Kadando and Mgila, the number of individuals with low antibody level is much lower while the number of individuals with a medium or high antibody level is

higher. Consistently for each transect, corresponding to rows on the figure, the distribution of the antibody density seems to shift towards higher levels of antibodies as the altitude decreases. It is worth noting that given their altitude, villages such as Ngulu and Funta present an unexpected distribution of antibodies (low number of individuals with low antibody levels and high number of individuals with higher levels) suggesting that factors other than altitude alone may influence exposure to malaria in this region.

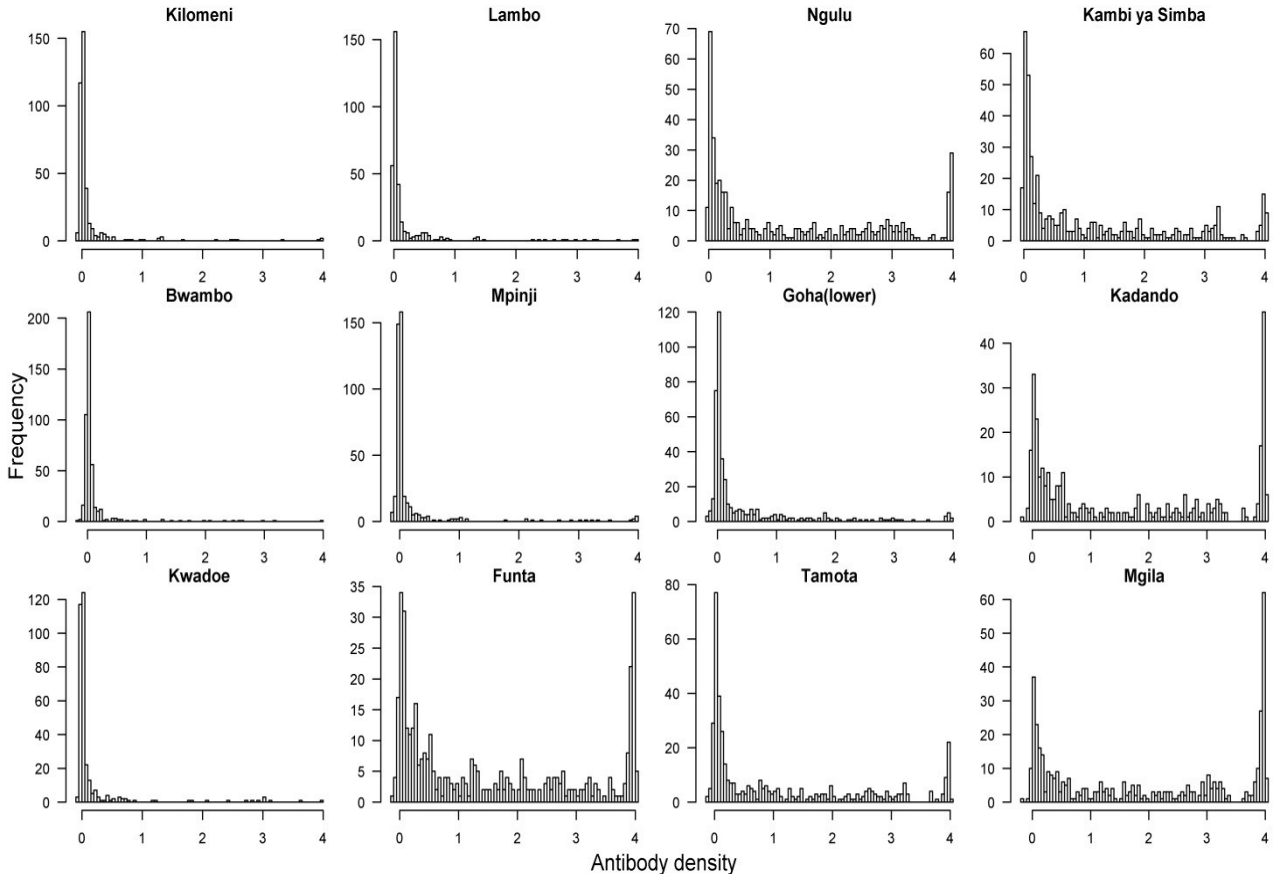


Figure 4.2: Distribution of optical density per village in North Pare (first row), South Pare (second row) and West Usambara (third row). In each transect (each row), villages are presented with decreasing altitude (left to right). Note y-axes are on different scales.

Figure 4.3 show the antibody level distribution on a log10 scale for each village. The median antibody density, presented in black in each box, tends to increase with decreasing altitude, with the exception of Ngulu and Funta, as seen in Figure 4.2. The mean antibody density, presented in red, appears to be greater than the median antibody density for higher altitude and lower for lower altitude. This reflects a very high number of individuals with low antibody levels in villages at high altitude and a high number of individuals with high antibody level in villages at low altitude, i.e. the distributions are over-dispersed.

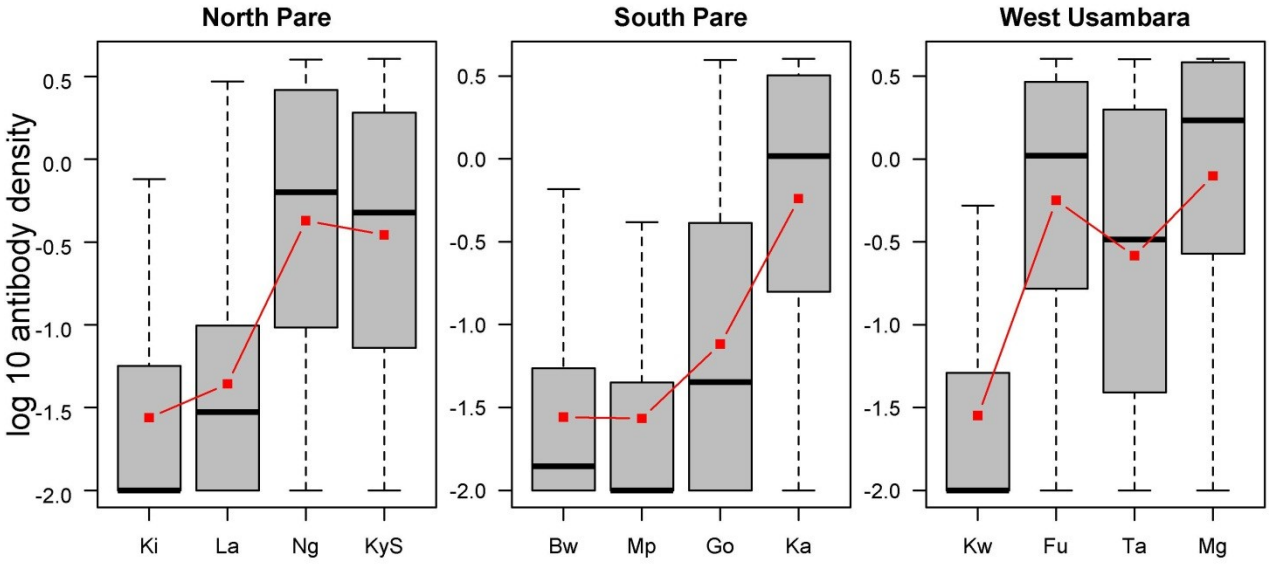


Figure 4.3: Summary of optical density distribution on a log10 scale, per village and categorised per region (North Pare, South Pare and West Usambara). The grey boxes characterize the interquartile range (25%-75%) of the optical density distribution while the thick black line represents the median. Error bars include all values 1.5 times the interquartile range. The red squares show the mean optical density.

These observations suggest that as altitude decreases, the density of antibodies increases, consistently for each region. However, the acquisition of antibodies is also dependent on age. Indeed, as shown in Figure 4.4, the level of antibodies increases with age, consistently for all villages. However, those in villages at lower altitude have levels of antibody that appear to saturate at high levels by 15 years of age, while for those residing villages at higher altitude, most individuals tend to have low levels of immunity throughout their life. As altitude is considered to be a proxy for exposure in this study, individuals highly exposed to malaria tend to build immunity at a young age and this immunity is long lasting. However, those residing in villages at higher altitude never reach this immunity level in their lifetime (up to 45 years). In such settings, it might be difficult to maintain immunity due to limited exposure, antibody levels tending to decay in the absence of reinfection. As noted earlier, the villages of Ngulu and Funta appear to have patterns of antibody density distribution that are more consistent with those residing in villages at low altitude.

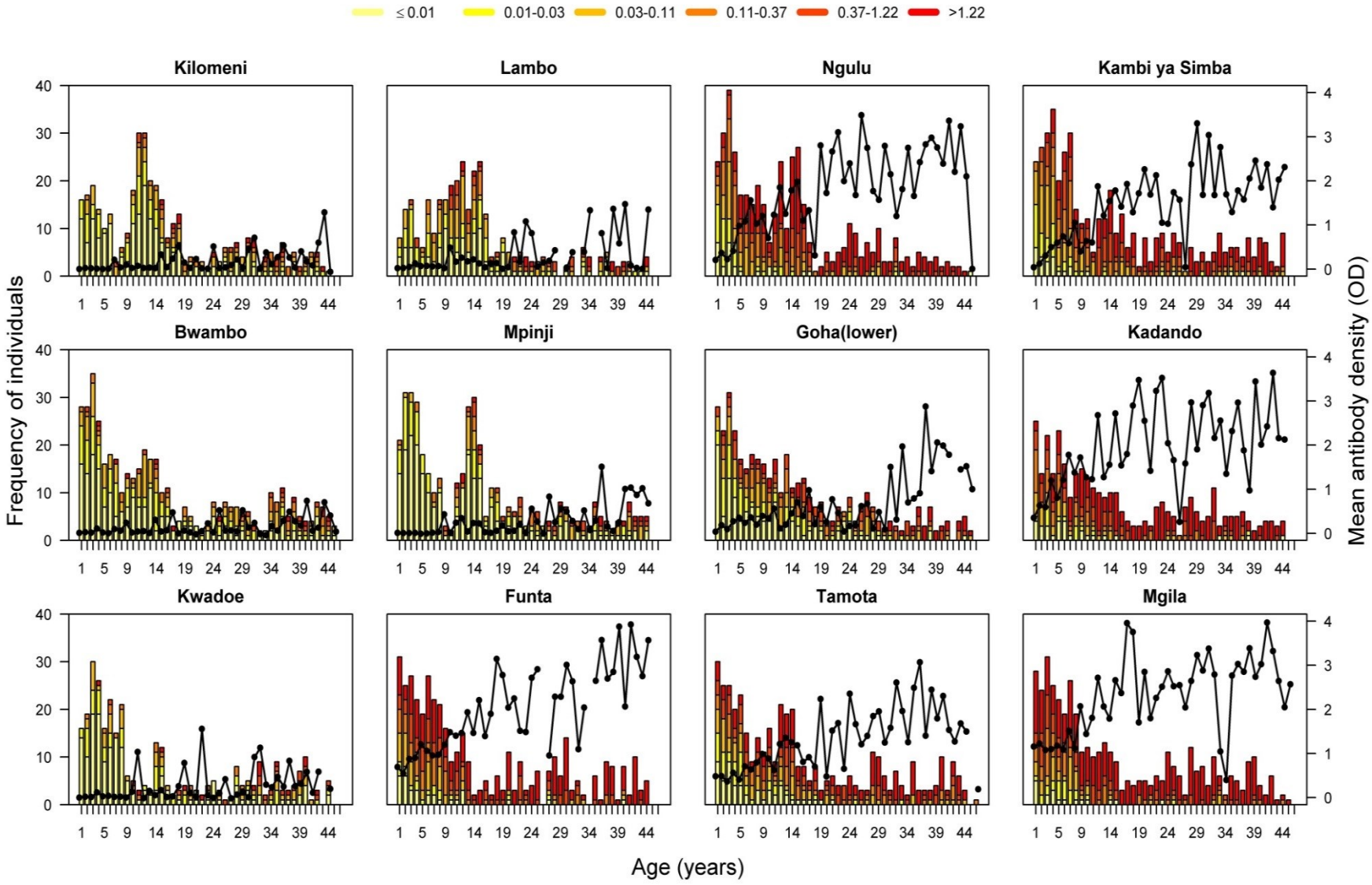


Figure 4.4: Age specific distribution of antibody density (categorised into OD intervals: ≤0.01, 0.01-0.03, 0.03-0.11, 0.11-0.37, 0.37-1.22, >1.22 in OD unit) for each village. The black curve represents the mean age-dependent antibody level.

In addition to antibody levels (OD here), the prevalence of seropositive individuals is commonly used to characterise serological data. Seropositivity for these individuals was previously defined as individuals whose antibody level is greater than 0.5 OD unit (Chris Drakeley – personal communication). Based on this definition, the number of seropositive and seronegative individuals is presented for each village in Figure 4.5. The proportion of seronegative individuals is substantial in high altitude villages and the proportion of seropositive individuals in the population increases with decreasing altitude but is never markedly higher than the number of seronegative individuals. Therefore, in each altitude transect the proportion of seropositive individuals in the population increases as the exposure increases, as expected. Here also, Ngulu and Funta do not follow the trend and show high proportions of seropositive individuals despite their medium altitude. For those residing in areas of higher exposure, there appear to be a saturation in the proportion of seropositive individuals at less than the whole population, suggesting that some individuals might never get exposed. Another possible explanation would be that in those settings, individuals become seropositive at the same rate than they reverse to seronegative. The rate of acquisition of immunity for the exposed individuals would be similar to the rate of losing immunity if this assumption is valid.

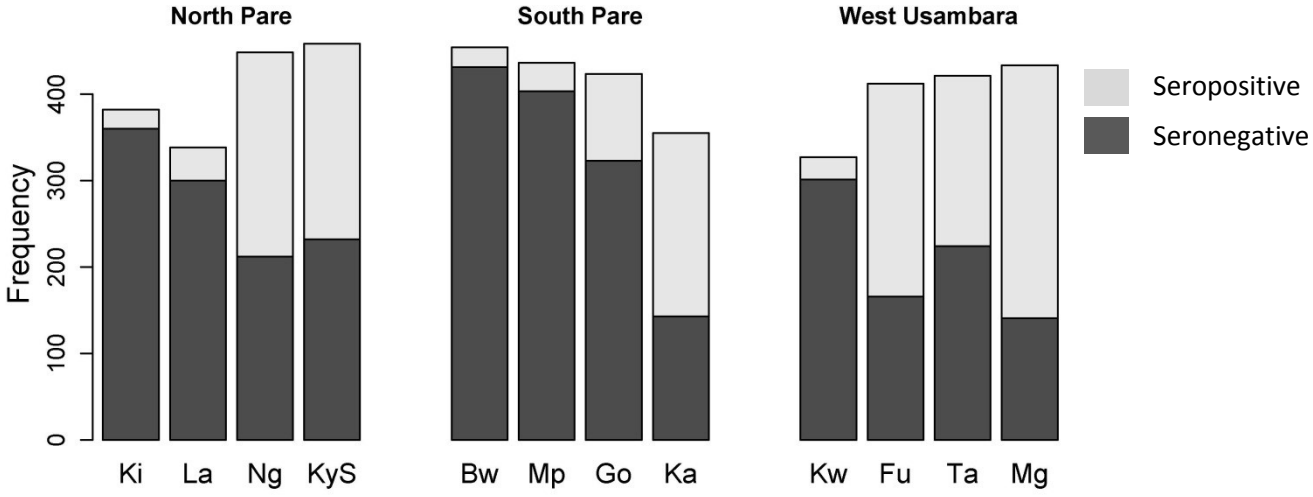


Figure 4.5: Number of seropositive/seronegative individuals for each village and categorised per region. The dark shaded area illustrates the number of seronegative individuals while the light shaded area represents the number of seropositive individuals.

4.3 Methods

4.3.1 Density model to assess transmission from serological data

4.3.1.1 Model formulation

In Chapter 3, I developed a mathematical model to describe the dynamics of acquisition and loss of antibodies in the population. In this chapter I use the same model, i.e a discretised approximation of the density model, that mimics the proportion of the population in each antibody density category i, denoted y_i. The following equation is the same as the equation presented in *Section 3.2.2.2*:

$$\frac{dy_i}{dt} = \lambda \sum_{j \le i} k_{ij} y_j + \frac{\rho}{\Delta} y_{i+1} - \lambda \sum_{h \ge i} k_{hi} y_i - \frac{\rho}{\Delta} y_i \qquad 1 \le i \le N \tag{4.1}$$

where *h, i, j =1,…,N* index the N antibody level classes. The rates of exposure and decay of antibodies, λ and ρ, are assumed to be independent of antibody density and age. The numerical approximation of the continuous model was achieve by categorising the log10 antibody optical density variable, x, into *N=51* compartments each of width $\Delta = 0.052$ with the first class represents measurements below the limit of detection, $x_{\min} = -2$.

Let x_i be the value of (log10) antibody optical density at the mid-point of antibody class i. The probability that, following exposure, antibody levels are boosted to class *i* from class *j*, k_{ij}, is distributed according to a discretised lognormal distribution, as it was already presented in *Section 3.2.2.3*:

$$k_{ij} = \begin{cases} 0 & \text{if } i < j \\ F(x_i + \Delta/2 - x_j; \delta(x_j), S) - F(x_i - \Delta/2 - x_j; \delta(x_j), S) & \text{if } j \le i < N \\ 1 - F\left(x_{N-1} + \Delta/2 - x_j; \delta(x_j), S\right) & \text{if } i = N \end{cases} \tag{4.2}$$

where $F\left(z; \delta(x), S\right)$ is the cumulative density function at point z of the lognormal distribution with mean $\delta(x)$ and standard deviation S. $\delta(x)$ is the mean boost size, a function of the current log10 antibody level, x, assumed to be given by:

$$\delta(x) = \begin{cases} ae^{-bx} & \text{if} \quad x > x_{\min} \\ \eta & \text{otherwise} \end{cases} \tag{4.3}$$

where a, b and η are parameters. This model assumes that exposure increases the log of antibody density by an exponentially decreasing amount as current density increases.

4.3.1.2 Parameter estimation

A Bayesian approach was used to estimate the model parameters, summarised in Table 4.2, by fitting the model to the optical density data from the 12 villages simultaneously, allowing only the exposure rate, λ, to vary by village, with local value for village v denoted λ_v. As in Chapter 3, the rate of decay of antibodies was fixed to ρ=0.7 yr^{-1}. Using $\theta = \{\lambda_v, a, b, S, \eta\}$ to denote the estimated parameter vector and $D = \{(x_i, t_i)\}$ the data including the observed antibody level and age of individual *i*, the multinomial log-likelihood is given by:

$$l = \log(P(D \mid \theta)) = \sum_v \sum_{t \neq 0} \sum_i n_{i,t,v} \log(y_{i,t,v}) \qquad (4.4)$$

where $n_{i,t,v}$ and $y_{i,t,v}$ are, respectively, the observed number and predicted proportion of individuals in antibody category i in village v at age t. MCMC methods were used to sample from the posterior distribution of the parameters. As all parameters were positive-definite I assumed uninformative uniform priors on [0, max], where max was the maximum permitted value of each parameter as listed in Table 4.2 and I used a log-normal random walk proposal density. I performed two runs of 500,000 iterations for the MCMC algorithm with a burn-in period of 50,000 iterations. The output was then recorded every 200 iterations to generate a sample of size 5,000 from the posterior distribution.

Table 4.2: Summary of model parameter values

Related to	Param.	Description	*Prior* Distribution
Exposure	λ_v	Annual antibody acquisition rate for village v	Uniform on [0;100]
Boost size	*a*	Maximum antibody boost size in exposure	Uniform on [0;10]
	b	Slope of the dependence of antibody boost on current log10 Optical density	Uniform on [0;10]
	s	Standard deviation for boost size distribution	Uniform on [0;10]
Seronegatives	*η*	Mean boost for individuals with no current antibody	Uniform on [0;100]

4.3.1.3 Model validation

The density model was then validated using data from 8 additional villages in the Rombo and West Usambara regions [212]. The *posterior* distribution of the boost parameters obtained from the first 12 villages was used to inform the model in the parameter estimation using data from 8 nearby different villages. The *priors* were assumed to be normally distributed with the previous *posterior* mean and a coefficient of variation of 1% as parameters. Priors for exposure rate were uninformative (Uniform on [0,100]).

4.3.2 Use of a mixture model for seroprevalence data

Mixture models can be used in a number of different forms for defining seroprevalence and deriving the seroconversion rate for the measurement of malaria exposure. Here, I present three methods which were further investigated. First, a mixture model was fitted to the optical density measurements in the population as a whole and a cut-off chosen based on this fitted mixture model used to define seropositivity. In a second method, I fitted a mixture model to the data but additionally accounted for potential misclassification while modelling the dynamics between seropositive and seronegative individuals. Finally, in the third method, I developed a Bayesian approach that provides estimates of seroconversion rates using augmented data for unobserved seropositivity status, based on the probability of being seropositive.

4.3.2.1 Defining seropositivity

Commonly, seropositivity is determined using European control populations as the seronegative population. Using these data, a cut-off value for seropositivity is defined as the mean plus 3 standard deviations of the distribution of the negative antibody density (optical density or titre) in that population. This cut off value was suggested to be 0.5 OD unit in Drakeley's study (Chris Drakeley - personal communication).

Here, I explore alternative methods to derive a cut-off value without resorting to external data. A mixture model was fitted to the normalised optical density distribution for each village to determine appropriate cut-offs for defining seropositivity. This method assumes that the population is composed of two subpopulations with proportions *p* and *(1-p)* respectively denoting seropositive and seronegative individuals. The optical density distribution for each sub-population is assumed to be a Normal distribution with parameters (μ_1, σ_1) for seronegative and (μ_2, σ_2) for seropositives. A cut off value that differentiates between seronegative and seropositive individuals is defined as $\mu_1 + 3\sigma_1$.

The parameters of the mixture model were estimated using MCMC method (as previously described) with the following likelihood considering the parameter vector $\theta = \{\mu_1, \sigma_1, \mu_2, \sigma_2, p\}$ and the data $D = \{(x_i, t_i)\}$

$$\begin{aligned} L(D;\theta) &= P(D/\theta) \\ &= \prod_{i=1}^{n} P(X_i = x_i/\theta) \\ &= \prod_{i=1}^{n} \{P(S_i = 0)P(X_i = x_i/\theta, S_i = 0) + P(S_i = 1)P(X_i = x_i/\theta, S_i = 1)\} \\ &= \prod_{i=1}^{n} \{(1-p)\phi(x_i;\mu_1,\sigma_1) + p\phi(x_i;\mu_2,\sigma_2)\} \end{aligned} \quad (4.5)$$

where $\phi(x_i, \mu, \sigma)$ represents the probability density function for an observed antibody titre x_i that is Normally distributed with mean μ and standard deviation σ and S_i denotes the (unobservable) individual's

seropositivity state of individual *i* equal to 1 for seropositive and 0 for seronegative. Parameter estimation is performed for all villages simultaneously to determine the cut-off value.

Once individuals have been classified based on the derived cut-off value, the standard approach using a catalytic model, as seen in the previous chapters, is applied to estimate seroconversion rates for each village. The proportion of individuals who are seropositive at age *t* is given by:

$$y_C(t) = \frac{\lambda_C}{\lambda_C + \rho_C}(1 - e^{-(\lambda_C + \rho_C)t}) \quad (4.6)$$

where λ_C is the annual mean rate of conversion (SCR) from seronegative to seropositive in a village and ρ_C the annual mean rate of reversion from seropositive to seronegative. A Bayesian MCMC approach as described above was used for parameter estimation with the following log-likelihood:

$$l = \sum_v \sum_t n_{t,v} \log(y_C^v(t)) \quad (4.7)$$

where $n_{t,v}$ and $y_C^v(t)$ are, respectively, the observed number and predicted proportion of individuals in village v at age t .The model was fitted to all villages simultaneously allowing λ_C to vary between villages but with the constraint of a single value for the reversion rate ρ_C , which was also estimated.

4.3.2.2 Mixture model for estimation of SCR

This method incorporates both the mixture model fitted to optical density as well as the catalytic model in a single model and therefore captures misclassification issues. As before, the population is assumed to be composed of two subpopulations with proportions $y_C(t)$ and *(1-* $y_C(t)$ *)* respectively for seropositive and seronegative individuals for each age group *t* where $y_c(t)$ is defined in (4.6). As in the previous method, the antibody level in each sub-population is assumed to have a Normal distribution with parameters (μ_1, σ_1) for seronegative and (μ_2, σ_2) for seropositive individuals. The log-likelihood for the model is then given by:

$$\begin{aligned} l &= \log L(D, \theta) \\ &= \log(\textstyle\prod_{i=1}^{n} P(D_i / \theta)) \\ &= \sum_{i=1}^{n} \log\{(1 - y_C(t_i))\phi(x_i; \mu_1, \sigma_1) + y_C(t_i)\phi(x_i; \mu_2, \sigma_2)\} \end{aligned} \quad (4.8)$$

with n the total number of observations, $\theta = \{\lambda_C, \rho_C, \mu_1, \sigma_1, \mu_2, \sigma_2\}$ the model parameters and $D_i = \{x_i, t_i\}$ the data including the observed antibody level and age of individual *i*. The model was fitted to all villages simultaneously; parameter estimation uses MCMC methods as described above with (4.8) as the likelihood.

4.3.2.3 Use of augmented data

In theory seronegative individuals should be those who have categorically no antibodies contained in their blood serum. However, due to potential cross-reactivity of the antibodies, residual maternal antibodies, analytical method errors and other reasons, in practice measured low levels of antibodies might be detected in seronegative individuals. The potential for misclassification was addressed in the previous method by jointly estimating the cut-off value and the model fit.

An alternative method is to impute the unobserved "true" sero-status of each individual. To do this, I used MCMC sampling to explore different possibilities. As before, I define an unobserved variable S_i corresponding to the sero-status for an individual i ($S_i=1$ if individual is seropositive, else *0*). This is imputed based on the probability of being seropositive. The level of antibodies measured for an individual *i* is denoted by the continuous variable x_i .

Let $\theta = \{\lambda_C, \rho_C, \mu_1, \sigma_1, \mu_2, \sigma_2\}$ denote the model parameters and $S = (S_i)$ the vector of unobserved individual's sero-status. The joint density of parameters, observed and unobserved data is given by:

$$P(X, S, \theta) = P(X / S, \theta)P(S / \theta)P(\theta) \quad (4.9)$$

with,
$$P(X / S, \theta) = \prod_{i=1}^{n} P(X_i / S_i, \theta) \quad (4.10)$$

and,
$$P(S / \theta) = \prod_{i=1}^{n} P(S_i / \theta) \quad (4.11)$$

and $P(\theta)$ *prior* distribution for θ, *n* the total number of individuals. The following calculations are presented for a single village.

The first part of the equation (4.9) corresponds to the observation level and ensures that augmented and observed data are compatible. The seropositivity status is a binary variable; individuals can either be seropositive (individuals have been exposed in the past) or seronegative (unexposed individuals or exposed individuals who have lost their immunity). There is no direct relationship between antibody level and seropositivity status but I assumed that antibody density distribution for each sub-population is a Normal distribution with parameters (μ_1, σ_1) for seronegative and (μ_2, σ_2) for seropositive subpopulations.

I write,
$$P(X_i = x_i / S_i, \theta) = \phi(x_i; \mu_1, \sigma_1)1_{S_i=0} + \phi(x_i; \mu_2, \sigma_2)1_{S_i=1} \quad (4.12)$$

with $\phi(x_i; \mu, \sigma)$ the probability density function of a normal distribution for an observed antibody titre x_i with parameters μ and σ and $1_{\{Z\}}$ the indicator function equal to 1 if *Z* is true, else 0.

The second part of the equation (4.9) refers to the dynamics between seropositive and seronegative individuals (as allocated by the augmented data) given the parameters. The model here simply is a reversible catalytic model where individuals become seropositive at a rate λ and revert to seronegative

status at a rate ρ. The probability that individual *i* at age *t* is seropositive is $y_C(t)$ defined in (4.6). Hence, conditional on the seropositivity status:

$$P(S/\theta)=\prod_{i=1}^{n}P(S_i/\theta)=\prod_{t=1}^{n_t}\prod_{i/t_i=t}\{(1-y(t_i))1_{S_i=0}(S_i)+y(t_i)1_{S_i=1}(S_i)\}=\prod_{t=1}^{n_t}\{(1-y(t))^{n^+(t)}y(t)^{n^-(t)}\} \quad (4.13)$$

where $n^+(t), n^-(t)$ represent the number of seropositive and seronegative individuals at age *t* and n_t the number of age groups. Finally, the joint distribution is given by:

$$\begin{aligned} l &= \log L(X,S,\theta) \\ &= \log(P(X_i/S,\theta)P(S/\theta)) \\ &= \sum_{i=1}^{n}\log\left\{\phi(X_i;\mu_1,\sigma_1)1_{S_i=0}+\phi(X_i;\mu_2,\sigma_2)1_{S_i=1}\right\}+\sum_{t=1}^{n_t}\{n^+(t)\log(1-y_C(t))+n^-(t)\log(y_C(t))\} \end{aligned} \quad (4.14)$$

MCMC sampling was used for parameter estimation so that the stationary distribution of the chain is the joint distribution of the parameters and the augmented data $P(X,S,\theta)$.

A random walk Metropolis Hasting sampling was performed. At each iteration, the following algorithm was performed:

a. Independently sample model parameters
b. Sample seropositivity status with the proposal for the candidate status $S_i^*: Q(S_i^*/S_i)=1-S_i$

Positive model parameters were sampled in (a) with a log Normal proposal. Only μ_1 and μ_2 were sampled with a Normal proposal. The random walk for all parameters was tuned for better mixing. The algorithm was run for 50,000 iterations and I fixed a burn-in period of 1,000 steps. The output was recorded every 10 iterations to constitute a sample for the posterior distribution of size 4,900.

In summary, in this chapter I used not only the density model developed in the previous chapters but I also developed methods to improve estimation of seroconversion rate**.** All of the developed methods were compared between themselves and against EIR, derived from altitude [214] and from parasite prevalence [93]. Except when specified otherwise, Pearson correlations were used to assess the association between results. A summary of the different methods is presented in Table 4.3.

Table 4.3: Summary of the methods applied to Tanzania dataset to measure malaria transmission intensity

Density Model	**Catalytic model**				**EIR**	
Exposure rate	SCR	SCR	SCR	SCR	EIR	EIR
Newly developed	European control	Mixture model defines seropositivity	Mixture model to assess SCR	Use of augmented data to assess SCR	Derived from altitude	Derived from parasite prevalence

4.4 Results

4.4.1 Malaria transmission assessed using a density model

4.4.1.1 Antibodies density

Data from 5,227 individuals in the 12 villages were included in the analysis. The antibody levels of individuals in each village stratified by age are shown in Figure 4.6. The overall trend shows an increase in mean antibody level in adults in the village with decreasing altitude and hence increasing transmission intensity in this setting [214–216]. As previously noted [95], two villages - Ngulu and Funta - have higher than expected antibody densities, suggesting higher transmission despite being at medium altitude. As expected, the trend is for antibody density to increase with age in each village, representing cumulative exposure to infection.

I fitted the previously developed age-structured density model to 12 villages from different altitude in Tanzania to estimate exposure rate based on antibody levels. The algorithm converged well (See Appendix IV). The fitted density model is able to capture antibody density patterns across most of the villages (Figure 4.6).

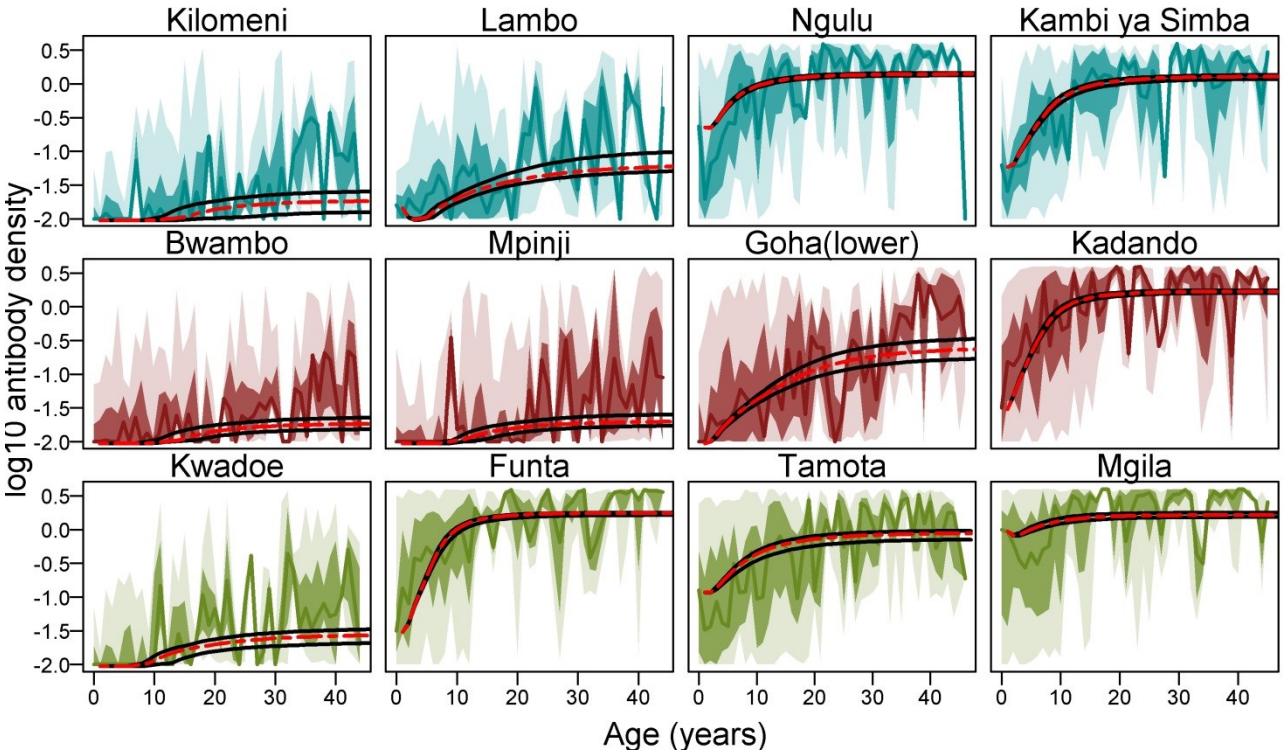


Figure 4.6: Antibody levels associated with age and village altitude. Median fits (95% credible interval) for median antibody levels are represeted by the red (black) lines. Dark and light shaded area represent respectively 25th/75th IQR and 2.5th/97.5th IQR for the data. Individuals surveyed were between 0 and 45 years old.

4.4.1.2 Estimates of exposure

In each transect, North Pare, South Pare and West Usambara, the estimate of the exposure rate increased with increasing transmission intensity (as indicated by decreasing altitude), with the exception, as expected, of Ngulu and Funta villages (See Figure 4.7 and Table 4.4).

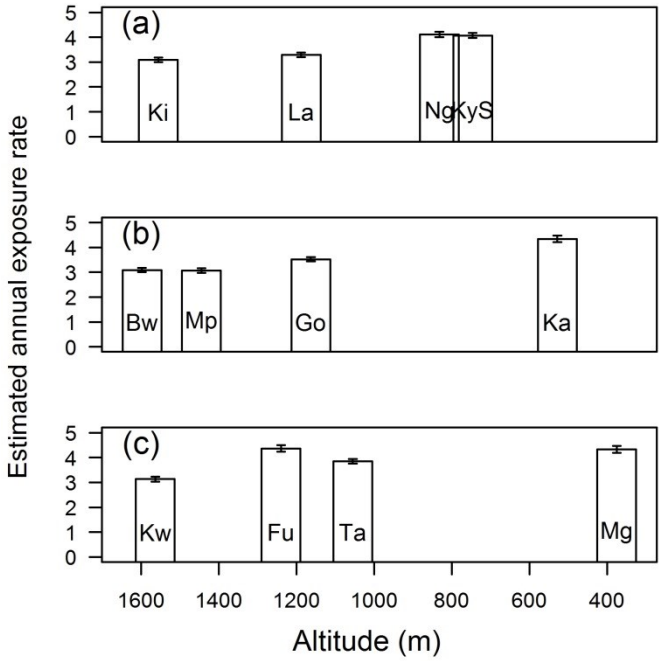


Figure 4.7: Measure of antibody acquisition rate by village and by region (a) North Pare (b) South Pare and (c) West Usambara. Estimates represented here are *posterior* median ± 95% credible intervals.

Table 4.4: *Posterior* median and credible intervals for each estimated model parameter

	Posterior Median(95% CrI)
Exposure parameters (λ)	
Kilomeni	3.09 (3 - 3.18)
Lambo	3.29 (3.2 - 3.38)
Ngulu	4.11 (4 - 4.22)
Kambi ya Simba	4.07 (3.97 - 4.17)
Bwanbo	3.09 (3 - 3.17)
Mpinji	3.07 (2.98 - 3.16)
Goha	3.52 (3.44 - 3.61)
Kadando	4.34 (4.21 - 4.48)
Kwadoe	3.14 (3.04 - 3.24)
Funta	4.36 (4.24 - 4.49)
Tamota	3.86 (3.76 - 3.95)
Mgila	4.33 (4.2 - 4.46)
Antibody boost size parameters	
a	0.24 (0.24 - 0.25)
B	0.09 (0.04 - 0.12)
S	0.02 (0.01 - 0.03)
H	0.026 (0.025 - 0.029))

4.4.1.3 Antibody boost size

I estimated the maximum boost size, a, to be 0.24 (95% credible interval 0.24-0.25) for individuals with an antibody level prior to exposure above the detection threshold. The rate of decline of the boost size with increasing antibody level prior to exposure was b=0.09 (95% CrI 0.04-0.12). As illustrated in Figure 4.8, the low value of b means that the estimated mean boost size declines approximately linearly with the $\log_{10}$ antibody level prior to exposure. For the previously unexposed population, I estimated the mean boost size as η=0.026 (95% CrI 0.025-0.029). The lack of overlapping credible intervals for the estimates of a and η suggests that there is a difference in the antibody responses of individuals who have never been exposed or have antibody levels below the detection threshold compared to those whose antibody levels are above the detection threshold at the time of exposure. With these estimates, the probability of becoming seropositive is negligble meaning that most of the seronegative individuals remain seronegative upon exposure and it therefore takes a number of exposures for individuals to become seropositive.

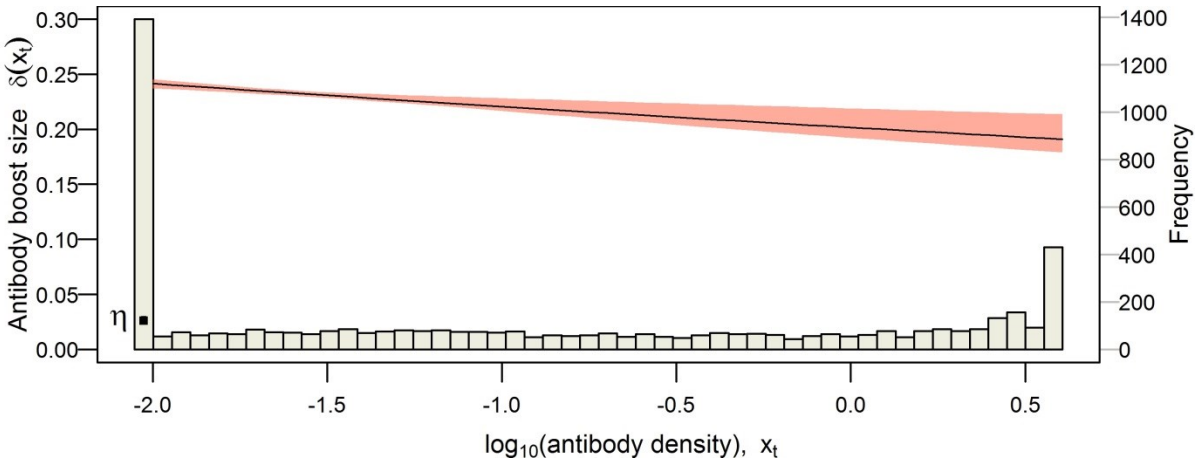


Figure 4.8: Antibody boost size (δ) depends on individual's current antibody level when an individual get exposed (x_t). Black line (and the pink shaded area) represents median boost size for individuals with circulating antibodies (95% credible interval). Black point indicates boost size mean for "seronegative" individuals. A histogram of the distribution of anti-MSP1 antibody OD is shown in grey.

4.4.1.4 Association between measures of exposure from density and catalytic models

A strong correlation was observed between the estimates of exposure rates obtained using the density model, and those estimated by fitting a catalytic model to the data, using European controls to derive the cut-off (see Figure 4.9a). As anticipated, villages at high altitude (Bwambo, Kilomeni, Mpinji and Kwadoe) had lower estimates of exposure while villages at lower altitude (Kadando and Mgila) had higher estimates of exposure. This was consistent for estimates obtained with both the density and catalytic models.

One of the advantages of estimating the exposure rate using the density model rather than estimating the seroconversion rate using a catalytic model is that it makes fuller use of the continuous nature of the data, thus potentially increasing inferential power. The coefficient of variation of the estimated exposure rate (standard deviation of posterior/mean of posterior), which measures the precision of the estimates, is consistently smaller for the density model than for the catalytic model (Figure 4.9b). This result is more marked for villages with overall levels of lower transmission (i.e. those at higher altitude).

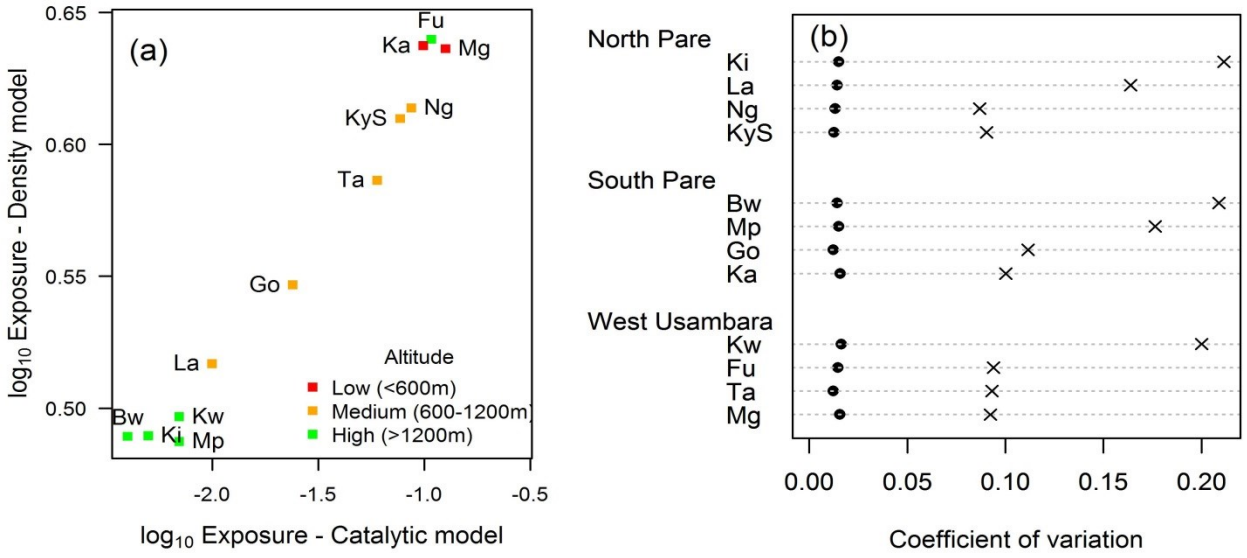


Figure 4.9: (a) Association between median estimates of exposure obtained with density model (y-axis) and catalytic model using European control (x-axis) for each village. (b) Coefficient of variation for both the density model (●) and the catalytic model (x).

4.4.1.5 Correlation between exposure rate and derived EIRs

Figure 4.10 shows that the estimates of exposure rate were also highly correlated with the two different estimates of the EIR available for the study villages, whether derived from altitude (Figure 4.10a) or derived from parasite prevalence (Figure 4.10b). It is worth noting that despite there being a strong correlation between the estimate of the force of infection and the two estimates of EIR (Figure 4.10c), the estimated values of EIR are substantially different in particular in lower endemicity settings. The range of EIR varies between ~0.03 and ~32 ibppy when derived from altitude, whereas its variation is between ~3 and ~20 ibppy when derived from parasite prevalence. However, despite this, both methods rank the villages in a similar order of increasing endemicity.

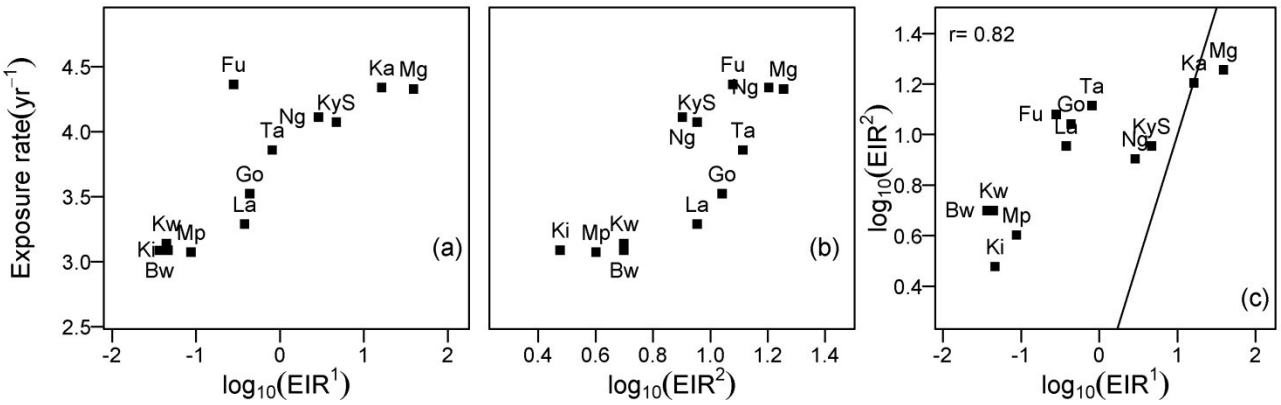


Figure 4.10: Association between exposure rate obtained from density model and EIR^1 when derived from altitude (a) or derived from Parasite Rate EIR^2 (b) and the correlation between both calculations (c). Coefficient of correlation, using Pearson method are denoted with r. If estimates were equal they would fall on black line in (c).

4.4.2 Estimates of the seroconversion rate

Classical methods for estimating seroconversion rate as a measure of the force of infection consist in fitting a simple reversible catalytic model to the data with European sera used to specify seronegatives. This method was applied to the current data using a cut off value of 0.5 OD unit and the estimates are reported in Table 4.5. In a first development of this method, I used a mixture model to define seropositivity and distinguish between seropositive and seronegative individuals. A resulting cut-off value equal to 0.75 OD unit was obtained. In a second development of the method, I jointly estimated the cut-off using a mixture model and estimated the seroconversion for each village. Finally, in a third development of the method, I iteratively simulated seropositivity status as a data augmentation step whilst fitting the model to estimate the seroconversion rate. The estimates of the seroconversion rate for each village for each of these four methods are presented in Table 4.5.

Table 4.5: Estimates of seroconversion rate SCR (λ_C) and rate of decay of antibodies (ρ_C) using 4 different methods. *Posterior* median and 95% credible intervals are presented for each village.

Median(95%CrI)	***Seropositivity defined using European control***	***Seropositivity defined by mixture model***	***Estimation of SCR using mixture model***	***Estimation of SCR using augmented data***
Exposure (λ_C)				
Kilomeni	**0.0049** (0.0032-0.0072)	**0.0041** (0.0026-0.0062)	**0.011** (0.0081-0.014)	**0.011** (0.0073-0.016)
Lambo	**0.011** (0.0076-0.015)	**0.0071** (0.0046-0.01)	**0.019** (0.014-0.025)	**0.019** (0.014-0.027)
Ngulu	**0.095** (0.08-0.11)	**0.079** (0.067-0.094)	**0.2** (0.17-0.24)	**0.2** (0.16-0.26)
Kambia ya Simba	**0.082** (0.069-0.096)	**0.062** (0.052-0.073)	**0.16** (0.13-0.19)	**0.15** (0.12-0.21)
Bwanbo	**0.0045** (0.0028-0.0065)	**0.0033** (0.002-0.0052)	**0.0092** (0.0066-0.013)	**0.0092** (0.0064-0.013)
Mpinji	**0.0069** (0.0049-0.0096)	**0.0065** (0.0044-0.0091)	**0.014** (0.011-0.019)	**0.014** (0.01-0.021)
Goha	**0.026** (0.021-0.033)	**0.019** (0.015-0.024)	**0.046** (0.037-0.057)	**0.045** (0.035-0.062)
Kadando	**0.11** (0.092-0.13)	**0.09** (0.075-0.11)	**0.23** (0.19-0.29)	**0.23** (0.18-0.33)
Kwadoe	**0.0076** (0.005-0.011)	**0.0049** (0.003-0.0076)	**0.014** (0.0096-0.019)	**0.014** (0.0096-0.021)
Funta	**0.12** (0.1-0.15)	**0.097** (0.081-0.12)	**0.28** (0.23-0.35)	**0.28** (0.22-0.38)
Tamota	**0.068** (0.057-0.082)	**0.057** (0.047-0.068)	**0.11** (0.09-0.13)	**0.11** (0.084-0.15)
Mgila	**0.16** (0.14-0.19)	**0.13** (0.11-0.16)	**0.31** (0.25-0.39)	**0.31** (0.24-0.44)
Seroreversion rate (ρ_C)	**0.02** (0.015-0.026)	**0.02** (0.015-0.027)	**0.017** (0.011-0.024)	**0.017** (0.0077-0.029)

A comparison of these estimates is shown in Figure 4.11. The method using European controls (method 0), with a lower cut-off value, provides estimates of SCR that are higher than estimates obtained with the method using a mixture model to define seropositivity (method 1). For villages that are at low levels of endemicity, where the antibody levels are lower, the difference in the estimate of the seroconversion rate is not as substantial as in villages at higher endemicity levels where a large proportion of the population have antibody levels between the cut-off values defined in the two methods. Estimates derived from methods using cut-off values (method 0 & 1) appear to give lower estimates than the other methods. Jointly fitting a mixture model and estimating the SCR (method 2) or using the augmented data technique

(method 3) allows misclassification issues to be taken into consideration. Both methods (2 & 3) provide similar median estimates of the seroconversion rate, but with lower precision for the augmented data technique, reflecting the variability between individuals. By accounting for misclassification of the individuals, these methods appear more realistic than methods that define overall cut-off value as the models are also adjusted per age and per village. In addition, all methods resulted in high correlation between the estimated SCR, the derived EIR and the exposure rate estimated using the density model (see Figure 4.12).

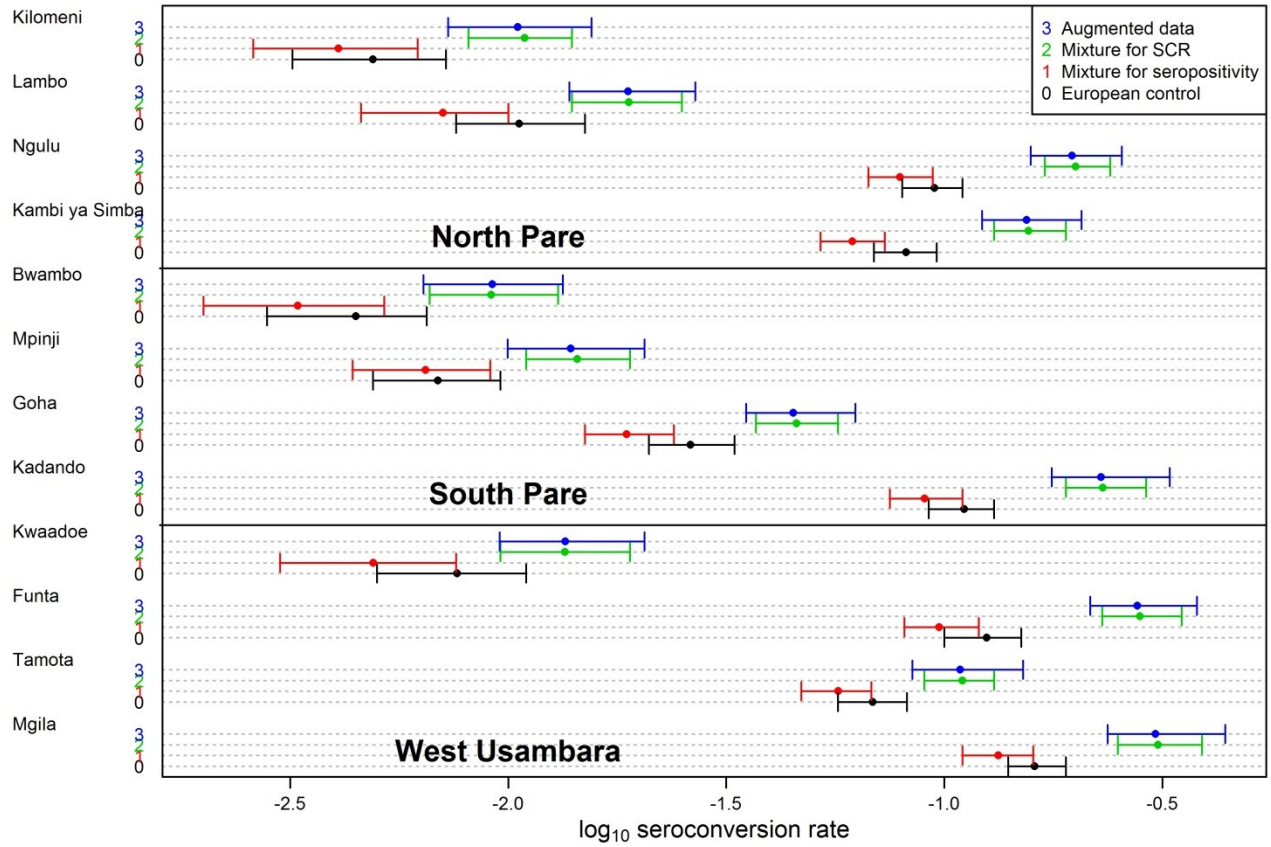


Figure 4.11: Seroconversion rate and associated 95% credible interval presented for each village and categorised by method used to derive SCR.

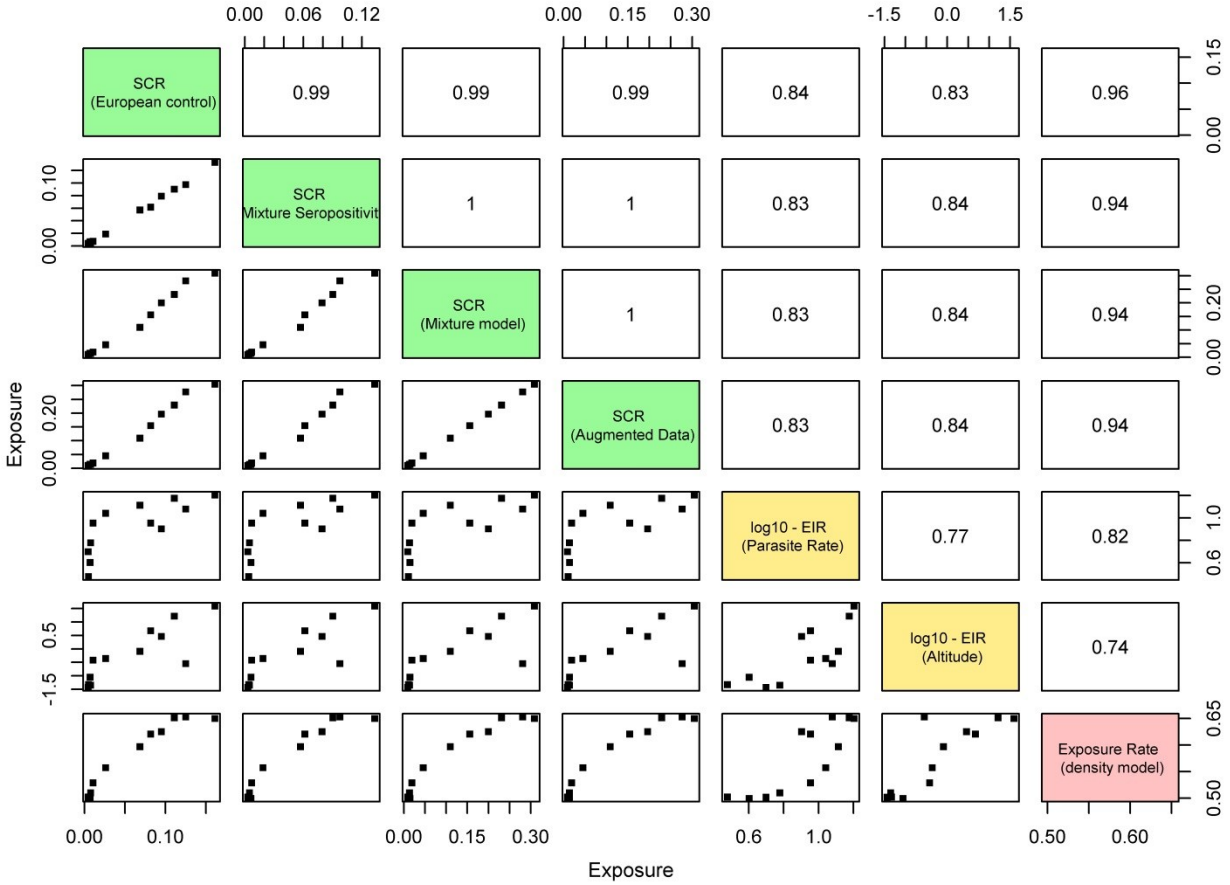


Figure 4.12: Correlation between multiple measures of exposure of each study village: SCR (yr^{-1}) using 4 different methods described previously, EIR (ibppy) derived from Parasite Rate and from altitude and exposure rate, obtained from the density model (yr^{-1}). Spearman correlations are presented in the upper panels.

4.4.3 Validation of the density model using data from different villages in Tanzania.

A validation study was performed for the density model by using dataset from 8 villages studied during the same survey in Tanzania but not used in the original fit. I tested whether the model developed provided consistent estimates of exposure and comparable biological parameters. The *posterior* distribution of the boost parameters obtained from the first 12 villages was used as a *prior* distribution to fit the model to data from these 8 different villages in Rombo and West Usambara regions. In each region, data from 4 villages with different altitude (exposure levels) were analysed. The distribution of the optical densities for anti-MSP-1 antibodies in these 8 villages are represented in Figure 4.13.

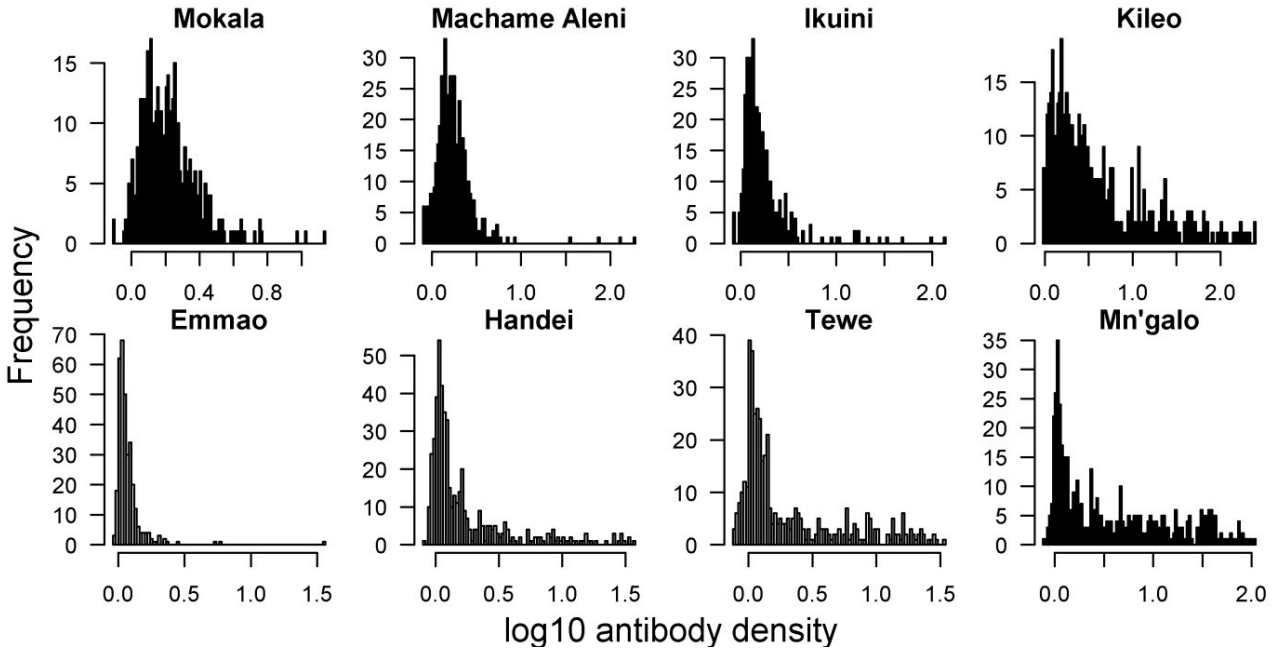


Figure 4.13: Anti-MSP-1 antibody distribution for 8 villages from Rombo (first row) and West Usambara (second row) regions. Antibody density measured in Optical density and presented on log10 scale. Altitude decreasing from left to right.

Seroprevalence was defined using the cut-off value obtained from European control (0.5 OD unit). A catalytic model was used to estimate seroconversion rates. The simple reversible catalytic model reproduces the age-stratified seroprevalence (Figure 4.14). The parameters for the 8 villages were estimated simultaneously with the rate of sero-reversion, estimated to be 0.039 years^{-1} (95% credible interval 0.025-0.056).

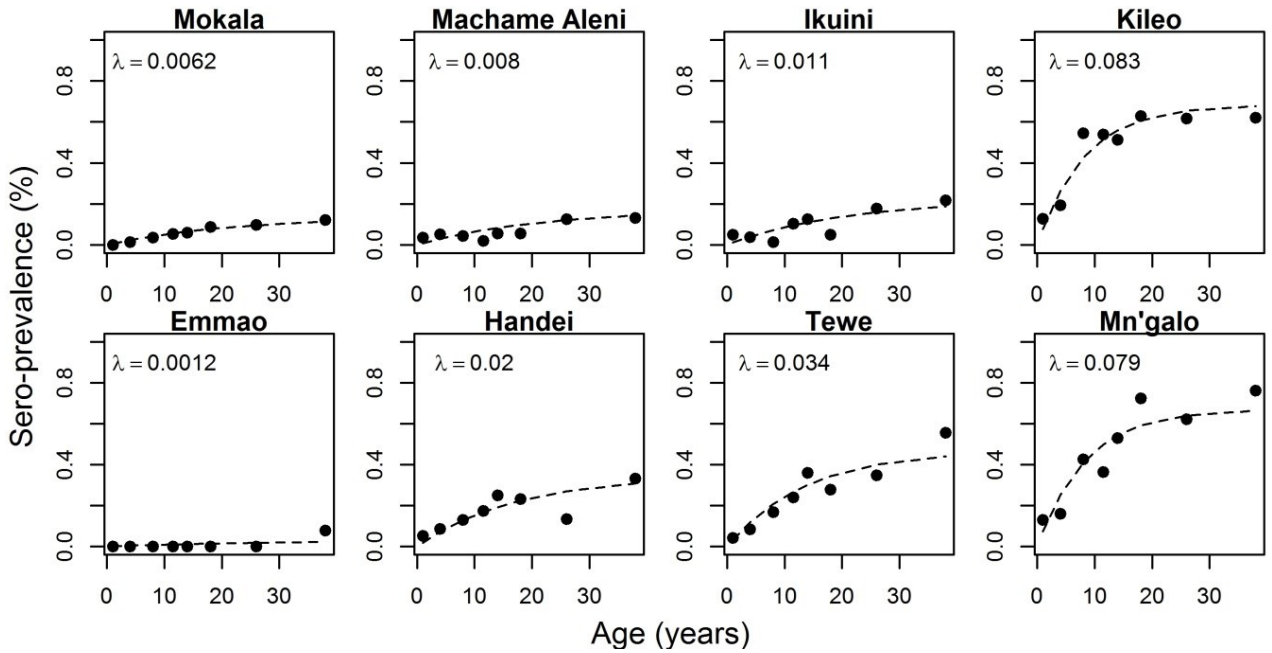


Figure 4.14: Age specific seroprevalence for 8 villages. Data points are represented in black while median model fit is represented by the dashed line. λ: village specific annual rate of converting from MSP-1 seronegative to MSP-1 seropositive, assuming a cut-off value for the density equal to 0.5 OD unit.

The antibody density model was used on the optical density for anti-MSP-1 antibodies from these 8 villages. Figure 4.15 shows represented the optical densities by age for all the villages for both the data and the model fit. Despite some variability in the data, the median fit for optical density reproduces well the median from the data. As seen previously, antibody levels, measured with optical density increase with age and appear to be higher for villages with low altitude / high exposure such as Kileo and Mn'galo. However, some villages, despite being at higher altitude (e.g., Mokala: 1702m) than villages studied previously (e.g., Kilomeni: 1556m) have higher levels of antibodies from the youngest age with values around $10^{-0.75}$ OD unit and their levels do not increase much with age.

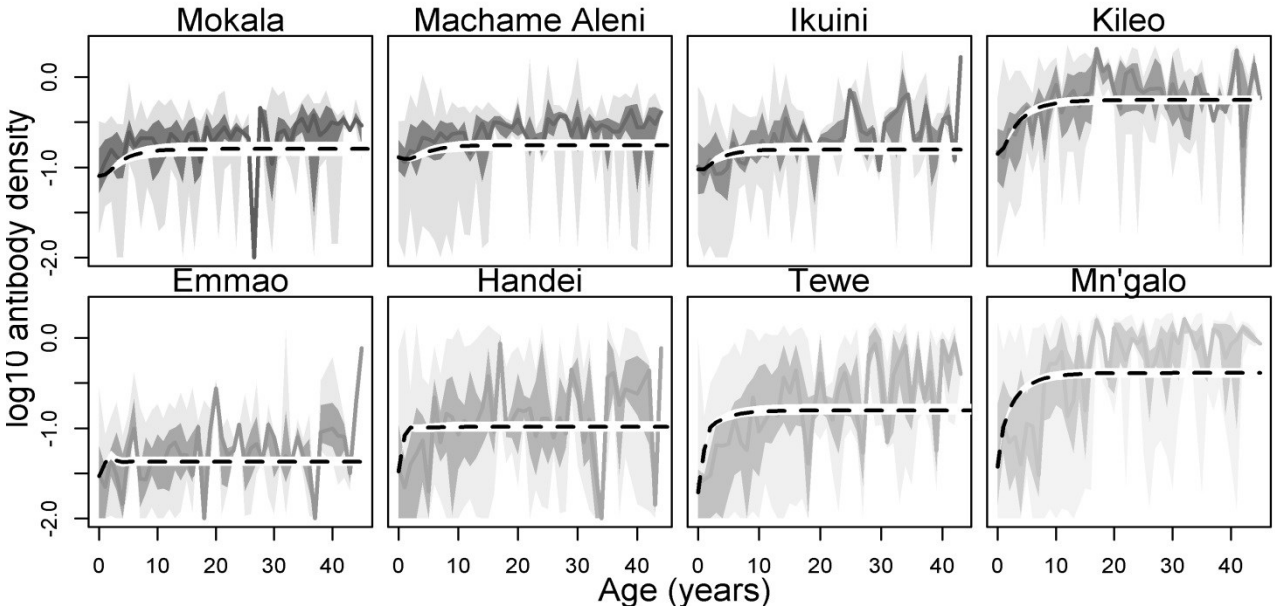


Figure 4.15: Antibody levels associated with age and village altitude. Median fits (95% credible interval) for median antibody levels are represeted by the black dashed line (white area around median). Dark and light shaded area represent respectively 25th/75th IQR and 2.5th/97.5th IQR for the data.

Parameter estimation for the model with this set of data for 8 villages was informed by the results from the model with 12 villages. Indeed an informative *prior* was used in the MCMC algorithm for the biological boost parameters (*a, b ,S* and *η*). *Posterior* distributions for those parameters were used to inform the *prior* while the exposure parameter had non-informative priors. The results from parameter estimation are presented in Table 4.6.

Table 4.6: *Posterior* median and credible intervals for each estimated model parameter

	***Posterior* Median** (95% Credible Interval)	***Prior*** Mean (SD)
Exposure (λ)		
Mokala	3.3 (3.1 - 3.5)	-
Machame Aleni	3.4 (3.2 - 3.5)	-
Ikuini	3.3 (3.1 - 3.5)	-
Kileo	4.1 (4 - 4.3)	-
Emmao	1.8 (1.6 - 2.1)	-
Handei	3 (2.8 - 3.1)	-
Tewe	3.3 (3.1 - 3.4)	-
Mn'galo	3.9 (3.8 - 4.1)	-
a	0.24 (0.23 - 0.24)	0.24 (0.007)
b	0.16 (0.15 - 0.17)	0.09 (0.068)
S	0.018 (0.018 - 0.019)	0.02 (0.02)
η	1.2 (1.1 - 1.3)	0.03 (0.003)

In order to assess the validity of this method, it is interesting to compare not only the estimates of the exposure rate with other metrics, such as seroconversion rate, parasite prevalence (PfPR) or EIR (derived from altitude) but also to visualize how the relationship between metrics compare with what was found previously with 12 villages.

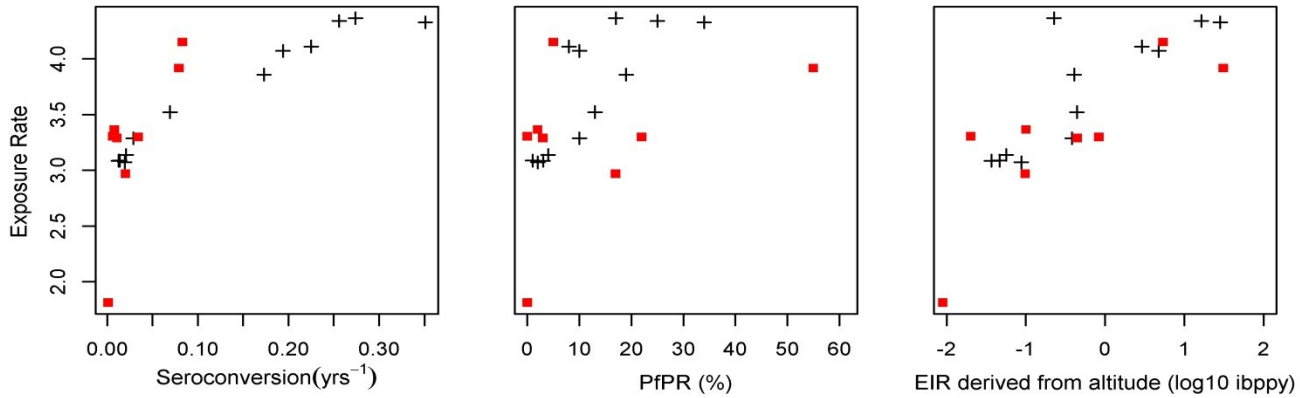


Figure 4.16: Association between exposure rates and seroconversion rate (a), parasite prevalence (b) and EIR derived from altitude (c) for the estimations of 8 (▪) and 12 villages (+).

As shown in Figure 4.16a, a high correlation between exposure rate and seroconversion rate can be observed for the 8 villages (r=0.62). Similarly, there is also a high correlation between seroconversion rate and exposure rate for the 12 original villages (r=0.96). However, if I assume a linear relationship between both metrics, the slope of regression is different for the fit to 8 villages (α=15.8) and for the fit of 12 villages (α=4.3). Nevertheless, the relationship between exposure rate and parasite prevalence (Figure 4.16b) is also strong for these 8 villages (r=0.31) as for the original 12 villages (r=0.82) and the slope of the regression line appears to be more similar when I compare estimates from 8 villages (α=0.01) with the original 12 villages (α=0.04). If I compare the estimated exposure rate with EIR derived from altitude (Figure 4.16c) I also observe a high correlation between variables (r=0.71 for 8 villages vs. r=0.74 for 12 villages). Also, the slope of the regression line is comparable (α=0.47 for 8 villages and α=0.45 for 12 villages).

4.5 Discussion

The utility of serological data to measure malaria transmission intensity has gained recognition in recent years [95, 97, 98] and is increasingly being incorporated in cross-sectional and longitudinal studies to monitor changes in transmission [26, 41, 48, 82, 217], identify "hotspots" of transmission [72, 218, 219] and to identify high risk groups [11]. One of the key advantages of the methods is their ease of use in the field, with new laboratory techniques enabling serological responses to multiple antigens to be made from dried blood spots that can be stored and transported without the need for refrigeration [220]. Classically the approach to analysing such data has been to distinguish seropositives from seronegatives using a cut off value informed by unexposed European control populations [95, 97]. This has recognised limitations as the European control population may differ genetically in their immunological response to infection from the populations being analysed [211]. To avoid the need to incorporate a cut-off independent of the antibody background level, I developed and fitted a density model to serological data from a malaria endemic setting in Northern Tanzania. For comparison purposes, I also further developed methods based on catalytic model that uses seroconversion rates to measure transmission intensity.

The results demonstrate that estimates of the exposure rate obtained from fitting the density model correlate highly both with previous estimates of the seroconversion rate obtained from the catalytic model as well as with traditional measures of transmission intensity (EIR) derived from altitude or from parasite prevalence data [93, 95]. The model therefore provides an alternative method to estimate transmission intensity from serological data that avoids the need to determine a cut-off between seropositivity and seronegativity. One alternative approach to using European controls to define a cut-off has been to using a mixture model, already broadly used in epidemiological studies in countries including Tanzania [98], Uganda [220], Somalia [99], Bioko Island [133] and Vanuatu [135]. The use of mixture models had already been explored by Irion and colleagues [168] who adopted an approach with latent class models to estimate prevalence of positivity. In this Chapter I have also further developed the use of mixture model to estimate the seroconversion rate without defining a cut-off value, either based on average proportion of responders or using each individual's response (using augmented data techniques). All methods provided high correlation with indices of transmission. These mixture methods can also take into consideration the potential for misclassification of seropositive and seronegative individuals and do not require any external data for standardisation. Therefore, they provide alternative appealing methods for analysing serological data. However, one limitation of the mixture method approach is that its application in high transmission settings has limited validity as the antibody distribution of seropositive and seronegative individuals becomes impossible to distinguish. In contrast, the density model presented here performed equally well across all transmission settings.

An additional advantage of the density model demonstrated here is the improvement in the precision of the estimate of transmission intensity obtained by fitting a density model in comparison to those obtained by fitting a catalytic model. In particular, this was notably better in lower transmission settings. This result is not surprising, as with a density model a greater degree of information in the data was intrinsically used. However, such methods are also relevant from a practical perspective, as serological measures are likely to be of greatest use in areas of low transmission intensity where other commonly used measures (in particular parasite prevalence) lack precision. Thus by utilising the full data set, this increased precision will improve the ability to distinguish temporal and spatial trends in settings in which malaria has recently fallen to low levels [221, 222] as well as to monitor low-level transmission in countries working towards local elimination of the parasite [223, 224].

However, the application of the method to the different villages makes the assumption that there is no biological difference between the individuals/regions. Data from longitudinal studies from the same areas, if available, would alternatively be useful to calibrate the model for these biological parameters and assess the validity of these assumptions.

The validation of the exposure rate as metrics for transmission intensity was based on values of EIR, either derived from altitude [214] or from parasite prevalence [93]. Derived EIR might not provide as accurate estimates as measured EIR. But, as SCR, based on anti-MSP-1 antibodies, had already shown high correlation with measured EIR in areas including Lower Moshi region in Tanzania [220], I expect similar correlations between measured EIRs and exposure rates. However, it would be of great interest if this approach could be directly validated against measured EIRs to further confirm that the exposure rate estimated by fitting the density model is truly an indicator of malaria exposure.

Whilst the density model clearly has many advantages over the classical catalytic model, there are some limitations worth noting. Firstly the density model outlined here, whilst biologically grounded, is a simplistic representation of the true process of antibody-acquisition and loss: it does not take into consideration more complex immune responses (such as interaction between antibody responses to different blood-stage antigens and/or between antibody- and cell-mediated immune responses [225]) or any ethnic or genetic difference between the different regions [210]. Nevertheless, it would be interesting to check what effect, if any, incorporating these factors would have on the estimates of transmission. Also, while providing a simplistic representation of the complex process, the model does not consider age dependence in the affinity of the response [117, 226]. Whilst such aspects are clearly important, they cannot be estimated from cross-sectional data.

4.6 Conclusion

In summary, the density model used in this Chapter provides a new method for analysing serological data that complements existing widely utilised tools for measuring malaria transmission intensity. These estimates are consistent with estimates of SCR, obtained using European controls or mixture models, and provide better precision. Further development of this method is needed to test it against a wider set of transmission settings, incorporate methods for assessing spatial and temporal variations in exposure and to assess its utility in capturing changes in transmission following scaling up of malaria interventions.

Chapter 5: Application of the antibody density model to assess malaria transmission intensity in diverse settings.

In the previous chapters, I developed a density model capable of reproducing antibody levels and measuring malaria transmission intensity. The data used to measure malaria intensity in Tanzania originated from a study designed to assess the potential of measuring malaria transmission with serological data. If the density model I developed so far is to be re-used in other studies, it is of interest to further understand its range of applicability. In this chapter, I fit the same density model to data on antibody levels from four countries with diverse endemicity settings to estimate malaria transmission intensity. I also apply the density model to different antigens and compare their performance in measuring malaria transmission. Finally, I extend the density model to capture heterogeneity in transmission, the effect of age on transmission and to estimate temporal change in transmission.

5.1 Introduction

Malaria transmission intensity varies worldwide. As a result, the distribution of antibodies in different populations varies substantially. In addition to exposure to malaria, other factors could contribute to the differences in immunological responses that would give rise to different antibody distribution. It is therefore of interest to understand whether using an antibody density model to estimate exposure rate, as presented in the previous chapters, is valid across a range of different transmission settings and for individuals from different countries and hence ethnic backgrounds. As seasonal malaria might represent a target for interventions, it is worth exploring the performance of the model to assess malaria transmission just after the rainy season. Additionally, with the increasing number of control and elimination interventions, it becomes of interest to assess the applicability of this method in places in which interventions have been applied. In this Chapter, I therefore use the density model to estimate the exposure rate from age-stratified antibody data from Somalia, Bioko Island, The Gambia and Uganda.

Somalia presents substantial spatial variation in malaria transmission intensity [221]. In some areas, the levels of transmission is so low that parasitaemia drops below the detection limits of microscopy [85]. The use of serological markers in such settings has already been demonstrated to be useful in predicting malaria exposure [99]. Due to the longevity of antibodies, serology measures exposure to malaria over time

and can therefore indicate the occurrence of malaria transmission in areas that report no parasite positive slides. I will apply the density model to data from Somalia to assess the performance of the method in low transmission settings.

In Bioko Island, Equatorial Guinea, national malaria interventions have been scaled up since 2004. As a result, parasite prevalence has dropped following the application of intensified vector control and improved access to treatment [227]. However, transmission intensity has still not uniformly been reduced to very low levels. Serology has previously been used there to show recent changes in transmission and the heterogeneity in the effectiveness of these interventions [135]. I will use data from Bioko to assess transmission intensity when interventions have had an impact on transmission intensity.

In The Gambia, parasite positive slides show a highly seasonal distribution of malaria infections that corresponds to the rainy seasons [41]. Transmission levels have significantly dropped in the last 20 years, setting The Gambia as an example for successful interventions. Indeed, areas of seasonal malaria might represent great targets to achieve reductions in malaria transmission intensity as levels of transmission drop substantially during the dry season [228]. However, in order to monitor the impact of these interventions, serology represents a useful tool for measuring malaria transmission intensity. I have used data from The Gambia to assess the performance of the density model to measure transmission intensity just after the wet season in areas of seasonal malaria.

Despite on-going intensive interventions in Uganda, essentially bed nets distribution, efforts have been inefficient to reduce malaria transmission and the associated disease burden [229, 230]. Indeed, despite the free distribution of bed nets, the proportion of bed nets ownership is relatively low; when three ITNs per household would represent total coverage in Uganda, only 8% of the households has three or more ITNs [231]. Malaria remains predominantly holoendemic with intense and perennial transmission. Serology has been used to confirm the relative lack of impact of the interventions [136]. However, the use of seroprevalence in high endemicity settings is limited by the small number of unexposed individuals required to derive the cut off value. By using the full information contained in the antibody titre, the density model was applied in Uganda to assess transmission intensity in a high endemicity setting.

The density model was fitted to data on antibodies against both AMA-1 and MSP-1 antigens. These are blood stage antigens and share some characteristics. Levels of antibodies to MSP-1 and AMA-1 antigens are strongly associated with increasing exposure to *Plasmodium falciparum.* However, the quality of these antibodies differs amongst individuals as antibody avidity shows some respectively negative and positive correlation with age for anti- MSP-1 and anti- AMA-1 [232]. By fitting the model separately and

simultaneously to both antibody types, the extent to which the density model can be used for any of these antibodies to measure malaria exposure is investigated.

Malaria transmission is influenced by many different sources of heterogeneity, including genetic, behavioural and spatial factors [80]. For instance, the difference in exposure between individuals can be due to the distance to mosquitoes breeding site, the use of protective nets or antimalarial drug intake. Additionally, the diversity of host characteristics might affect the heterogeneity in biting [78] and the development of acquired and innate immunity that limits the frequency of infection [67], therefore reducing transmission. Thus, it is of interest to explore whether including heterogeneity in transmission in the density model improves the estimate of the exposure rate.

Differences in human behaviour often contribute to heterogeneity of malaria transmission, in particular due to individual's occupations. In South-East Asia, many studies have shown the effect of working in the forest on exposure [233–235]. Children starting to work in the forest become exposed to a much higher degree. This makes it difficult to differentiate the effect of age and cumulative biting, so a further development of the density model to account for a change in behaviour at a particular age would be useful to characterise forest malaria in specific areas.

A consequence of the interventions is that temporal and spatial heterogeneity of malaria transmission might become apparent as transmission declines. Indeed, the impact of interventions can be spatially heterogeneous and happen at different time [133]. A classic method to assess the impact of any temporal change in transmission consists of conducting multiple cross sectional surveys (before and after the change). Some studies have shown that it was possible to detect a change in transmission by estimating two seroconversion rate (SCR) from a single cross-sectional survey [133, 135]. Accounting for a change in transmission in the density model would therefore increase the potential applications of the model.

In the first section of this chapter, I summarise the epidemiology in the different settings used for the analyses. Next, I fit the density model to each of these data sets to obtain estimates of the exposure for each site. In the last section, I extend the density model to allow for an assessment of heterogeneity in transmission in the population, the effect of age on transmission and a temporal change in malaria transmission. Using simulations, I reproduce multiple scenarios to assess the ability of my method to estimate variation in transmission. These extended versions of the density model are then fitted to the data to assess temporal changes in transmission in Bioko Island and the effect of age on transmission in Cambodia.

5.2 Settings

5.2.1 Data collection

5.2.1.1 Somalia

Located in the eastern part of Africa, Somalia is referred to as "the horn" of Africa (Figure 5.1). Somalia has a relatively semi-arid landscape and seasonal rivers. In Somalia, the rainfall pattern is bimodal with peaks in April and August. Somalia is a country with very low malaria intensity. The parasite prevalence is below 5% for most part of the North West Zone [221]. The dominant parasite species are *b*oth *P. falciparum* and *P. vivax*.

A cross-sectional survey was conducted in August-September 2008 in Somaliland, in North Western Somalia. Three villages were selected randomly in the Gebiley district, Xuunshaley, Badahabo and Ceel-Bardaale. The timing of the study corresponds to the end of the wet season. Each household was visited. Demographics characteristics and bednet use data were collected for households that agreed to participate. The presence of parasite was assessed using rapid diagnostic tests (RDT). The presence of Anopheles species were determined by larvae collections. The full description of the study is presented elsewhere [236]. Ethical approval for this study was granted by the Research Ethics Committee of the World Health Organisation and the Ethical Committee of the Ministry of Health and Labour, Republic of Somaliland.

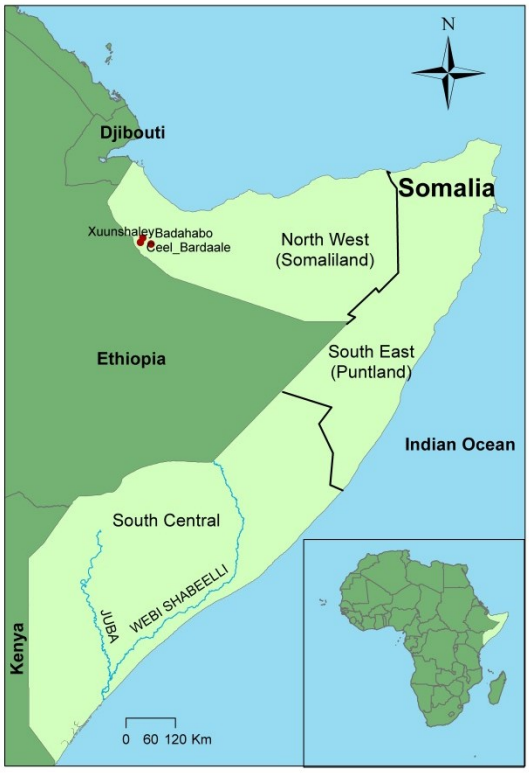


Figure 5.1: Map of Somalia

5.2.1.2 Bioko Island

Bioko is an island in Equatorial Guinea, located around 30 miles west of Cameroon (Figure 5.2). The climate is characterised by peaks of rainfall towards September and October. The total rainfall is on average 2,000 mm/year. Parasite prevalence is around 26% on average over the island but varies substantially between regions, from 5% in high altitude in the South East to 72% in the South West of the Island [237].

A malaria control intervention program named Bioko Island Malaria Control Program (BIMCP), sponsored by Marathon Oil Company [227] was launched in 2004 with the intention of eliminating malaria from the island. Since the launch of this program, Malaria Indicators Surveys (MIS) were conducted annually in 18 randomly selected sentinel sites, grouped in 5 geographical regions and covering most of the island. The surveys collected demographic data as well as data on household spraying, illness history and compliance with interventions. Children under 15 years old were tested for malaria parasitaemia using rapid diagnostic tests (RDT). The research protocol was approved by the Equatorial Guinea Ministry of Health and Social Welfare.

Figure 5.2: Map of Bioko Island

5.2.1.3 The Gambia

The Gambia is a small West African country divided by the river Gambia (Figure 5.3). The country experiences highly seasonal rainfall with an average annual rainfall of 800 mm/year. A short rainy season takes place between July and October and a long dry season from November to June [238]. Malaria transmission in The Gambia is very low during the annual dry season and high during the wet season. The South bank has a longer malaria transmission season than the North bank due to a longer rainy season and its denser vegetation which provides breeding sites for mosquitoes. The three main mosquitoes' species that transmit malaria in The Gambia are *Anopheles Gambiae*, *Anopheles Arabiensis* and *Anopheles Malas*. The biting rate for children is between 3 and 5 infective bites per person per year (ibppy), while an adult receives between 11 and 24 infective bites annually [239]. The predominant plasmodium species in the region was *P. falciparum*.

Between 1990 and 1991, four cross-sectionnal surveys were carried out as an investigation in malaria immunology, epidemiology and control in this region [240]. Data presented here only focus on the survey carried out in November 1990, just after the wet season. Around 1,200 participants were recruited in two sets of hamlets on either side of the river. Each participant had to complete a survey form recording demographic characteristics. The collection of the data stopped when the expected sample size was achieved. Ethical approval for the study was obtained from the Scientific Coordinating Committee of the Medical research Council (The Gambia) and the Joint The Gambia Government Medical Research Council Ethical Committee.

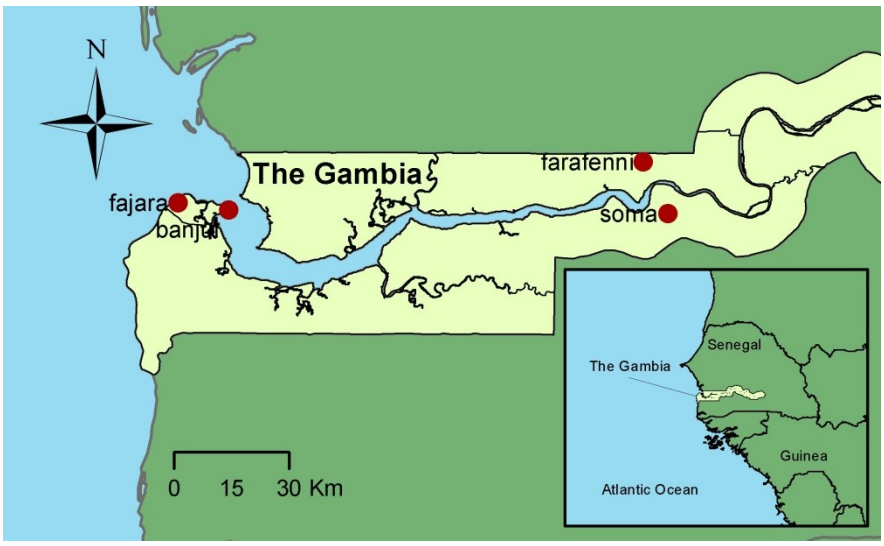


Figure 5.3: Map of The Gambia

5.2.1.4 Uganda

Uganda is a landlocked country in eastern Africa with vast inland water bodies. Uganda experiences a dry season from November to February and two short rainy seasons from April to May and September to October. The Apac Sub-County, a district in Northern Uganda located between Kwania Lake and the Victoria Nile (Figure 5.4) reports the highest estimate of EIR in Africa [230] with EIR ranging from 4 to over 1,500 infectious bites per person per year. As a consequence, parasite prevalence is also extremely high, with around 80% prevalence in children younger than 10 years old. Since 2001, a number of unsuccessful malaria control programs have been launched in order to control malaria and reduce morbidity and mortality [136].

In October 2009, a survey was carried out by Proietti and colleagues to assess potential changes in transmission [136]. The study was conducted in four parishes: Apac District hospital, two health facilities in Abedi and Akere and a school in Atopi. Subjects were recruited per age category until the required sample size was reached (around 200-300 individuals per parish). Demographic characteristics, clinical information, use of antimalarial and prevention against mosquitoes were recorded through questionnaires. Presence of parasite was first detected using rapid diagnostics test (RDT) and confirmed by examination using microscopy and PCR. Ethical approval for the study was obtained from the ethical review committee of the London School of Hygiene and Tropical Medicine, the ethical committee of the Medical Biotech Laboratory, and the national ethical committee of Uganda.

Figure 5.4: Map of Uganda

5.2.2 Laboratory methods

In each of the studies, a fingerprint blood sample was obtained from each participant for malaria parasite examination and humoral assays. This sample was placed on filter paper as described by Corran [97] to measure antibody density. Filter papers were stored with dessicant until processed. Samples were diluted to a serum dilution equivalent of 1/1000 (1/750 in Uganda) for human immunoglobulin G antibodies against *P. falciparum* merozoite surface protein 1_{19} (MSP-1_{19}) and 1/2000 for antibodies against apical membrane antigen 1 (AMA-1). Antibodies were detected by ELISA, as previously described in [220]. Antibody data were collected as optical density and converted to arbitrary titres using a standard curve based on dilutions of hyperimmune serum on each assay plate.

5.2.3 Seropositivity definition

Serological data were reported as antibody titre and will later be used in conjunction with the density model to characterise endemicity levels. However, in a preliminary descriptive analysis, seroprevalence was defined using a method widely used and presented in Chapter 4. Individuals were considered antibody positive when their antibody titre was above a cut-off value derived using a mixture model. In brief, the distribution of antibody titre was fitted as the sum of two Gaussian distributions using maximum likelihood methods. The cut-off value was then defined as the mean of the Gaussian corresponding to the seronegative population added with 3 standard deviation of the same distribution. I have assumed similar characteristics within countries and different characteristics between countries. As a result, cut-off values were derived at a country level but specific for each country. The resulting cut-off values are presented for each antigen in Table 5.1. Each country might present different genetic and ecological backgrounds that might explain country specific definition of seropositivity. However, in Uganda the number of seronegatives might be so low that the threshold obtained with a mixture model might over-estimate the real one.

Table 5.1: Cut off values obtained using mixture model

	MSP-1 (units/mL)	AMA-1 (units/mL)
Somalia	40	86
Bioko	87	233
The Gambia	168	279
Uganda	156	1006

An overall summary of the malaria endemicity in the four countries that will be considered in the following analyses is presented in Table 5.2. Despite a lack of a standard method to compare the endemicity level of each country, Somalia and Uganda represent respectively areas of low and high levels of transmission while Bioko and Gambia had intermediate endemicity levels at the time of the survey (with endemicity in The Gambia expected to be marginally higher).

Table 5.2: Summary of studies and malaria transmission intensity for each studied country

	Malaria transmission intensity	Year of study	Study reference
Somalia	**Parasite prevalence:** <5%	2008	Youssef et al. [236]
Bioko	**Parasite prevalence:** 26% - 72%	2008	Kleinschmidt et al.[227]
The Gambia	**EIR:** - 3 to 5 ibppy (children) - 11 to 24 ibppy (adults)	1990	Lulat [240]
Uganda	**Parasite prevalence:** 80% (children <10 years old) **EIR:** 4 to over 1500 ibppy	2004	Proietti et al.[136]

5.2.4 Descriptive analysis

5.2.4.1 Overall seroprevalence and antibody titres

The endemicity level in each country differs and this is reflected in the distribution of antibody levels in the population. Figure 5.5 demonstrates that with increasing transmission intensity (left to right), the proportion of individuals with high antibody levels increases, consistently for both antibody types. It appears that using a mixture model to determine seropositivity status is appropriate with the exception of Uganda where transmission intensity is high and therefore, it becomes difficult to distinguish between seropositive and seronegative individuals, in particular when considering the AMA-1 antigen. Using a density model which takes into account the full information contained in the antibody titre is therefore likely to be better able to characterise transmission intensity in this setting.

Additionally, note that the anti- MSP-1 antibody distributions for The Gambia and Uganda are very similar (with a marginally higher proportion of seropositive individuals in The Gambia) despite a large differential of exposure between both countries (see Table 5.2 and Figure 5.5). As a result, in Uganda, it appears that the distribution of anti- MSP-1 antibodies is unexpectedly low. A number of plausible hypotheses for these observations could be suggested, including technical and biological explanations.

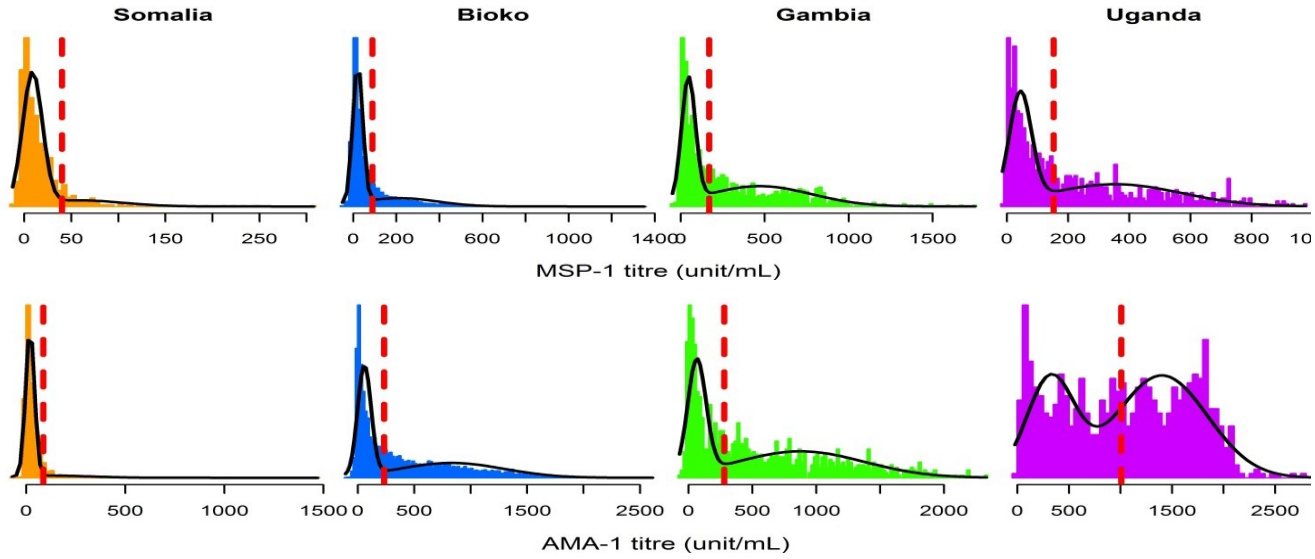


Figure 5.5: Antibody distribution in Somalia, Bioko, Gambia and Uganda (left to right) for anti- MSP1 (first row) and anti- AMA1 (second row). Cutoffs between seronegative and seropositive individuals are presented in red and the black line represents the Gaussian fit.

In all four studied countries, data were collected in different geographical areas; three villages in Somalia, North and South river banks in The Gambia, six regions in Bioko and four parishes in Uganda. Figure 5.6 illustrates the distribution of antibodies against MSP-1 and AMA-1 antigens for the different geographical areas in each country. Details of the summary statistics for antibody titres against MSP-1 and AMA-1 antigens and the associated seroprevalence are provided in Table 5.3.

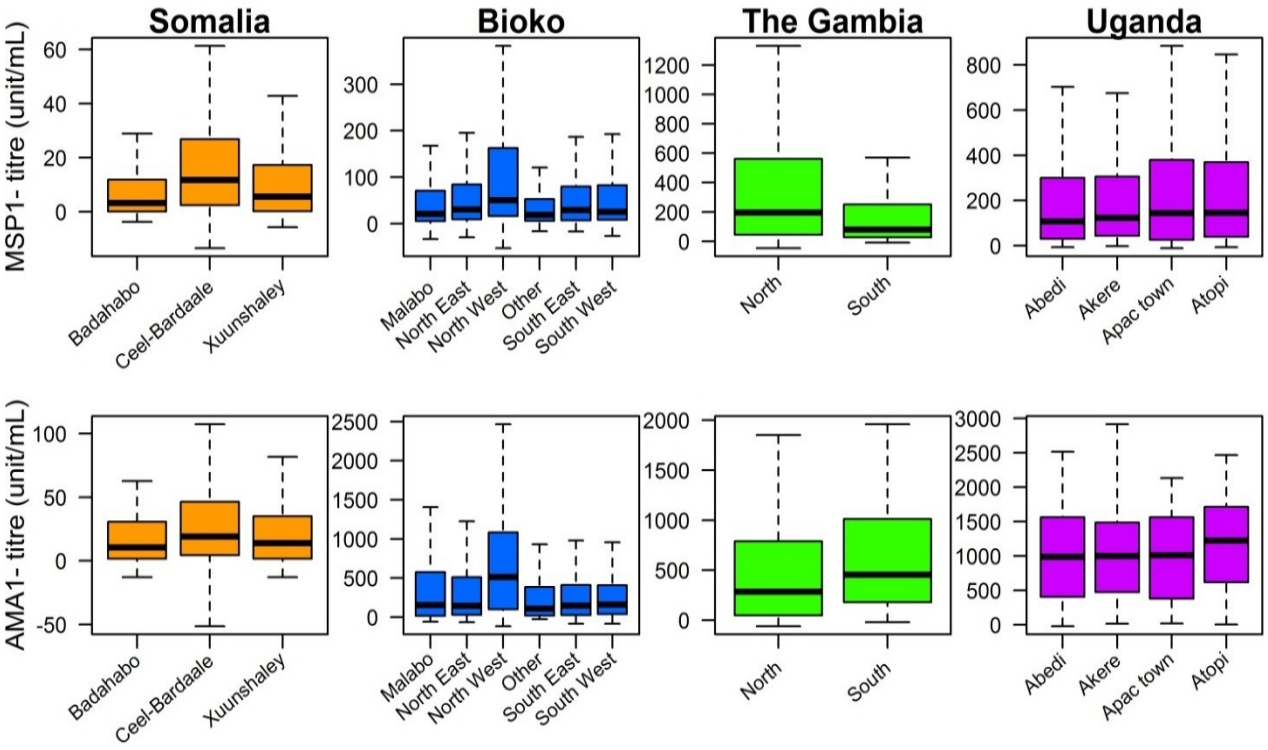


Figure 5.6: Distribution of antibodies against MSP-1 (first row) and AMA-1 (second row) antigens for Somalia, Bioko, Gambia and Uganda (left to right) and presented per region. Note that the y-axes are on different scales.

Table 5.3: Summary statistics for anti- MSP-1 and anti- AMA-1 antibody titres and seroprevalence for regions of Somalia, Bioko, Gambia and Uganda.

	# ind.	MSP-1		AMA-1	
		Seroprevalence %(no. positive/no. tested)	Titre Median (IQR)	Seroprevalence %(no. positive/no. tested)	Titre Median (IQR)
Somalia					
Badahabo	160	6.6 (7/106)	3.3 (0.1 - 11.7)	9.2 (13/141)	10.5 (1.7-30.7)
Ceel-Bardaale	697	15 (95/634)	11.8 (2.5 - 26.8)	9.2 (60/653)	19 (4.5-46.3)
Xuunshaley	271	5.1 (13/254)	5.6 (0.2 - 17.3)	5.6 (14/252)	14 (1.7-35.1)
Total	1128	11.6 (115/994)	8.9 (1.3 – 23.1)	8.3 (87/1046)	17.2 (3.1-40.4)
		Overall Seroprevalence % (no. positive/no. tested)			
		17.7 (171/967)			
Bioko					
Malabo	2328	21.3 (473/2218)	20.9 (5.2-70.5)	44 (960/2181)	156.3 (18.1-574.3)
North East	1323	24 (299/1247)	30.5 (9.1-84.2)	42.3 (495/1171)	146.3 (30-509.8)
North West	1749	37.3 (598/1604)	50.4 (16.1-162.7)	64.8 (1053/1626)	513 (100.7-1082.7)
South East	700	23.8 (144/604)	29.1 (7.2-79.1)	37.3 (245/656)	150.7 (27.6-411.4)
South West	588	23.2 (134/577)	25.5 (7.9-82.7)	40.7 (231/568)	163.1 (39.8-408.4)
Other	699	16.7 (107/641)	19.1 (6-52.6)	34.5 (226/655)	109 (20.9-386.4)
Total	7387	25.5 (1755/6891)	28.7 (8.3-89.9)	46.8 (3210/6857)	196.4 (34.4-639)
		Overall Seroprevalence % (no. positive/no. tested)			
		54 (3663/6788)			
Gambia					
North Bank	841	52.7 (394/748)	195.6 (44.8-558.2)	50.5 (373/739)	285.1 (48.3-788.8)
South Bank	344	33.8 (93/275)	79.6 (26.6-250.7)	64.5 (178/276)	454.7 (179.4-1009.2)
Total	1185	47.6 (487/1023)	151.2 (38.2-448.3)	54.3 (551/1015)	355.1 (64.7-865.7)
		Overall Seroprevalence % (no. positive/no. tested)			
		66.4 (675/1017)			
Uganda					
Abedi	251	41.1 (97/236)	107.7 (30.8 - 298.5)	49.2 (119/242)	989.9 (404.3-1561.1)
Akere	217	42.1 (83/197)	123 (43.6 - 305)	48.8 (98/201)	999.9 (474.1-1485.6)
Apac Town	213	49.4 (76/154)	145.1 (27 - 377.1)	50.6 (78/154)	1010.4 (382.8-1561.1)
Atopi	202	48.7 (93/191)	146.1 (39.4 - 369.4)	59.2 (113/191)	1222.8 (616-1712.3)
Total	883	44.9 (349/778)	125.7 (36.3-337.5)	51.8 (408/788)	1045.8 (453.3-1603.6)
		Overall Seroprevalence % (no. positive/no. tested)			
		70 (545/779)			

*IQR, interquartile range (25^{th}-75^{th} percentile)

Despite some level of variation within each country, the major variability in antibody titre is observed between countries.

In Somalia, the number of individuals taking part in the study differs between the three villages Xuunshaley (n=271), Badahabo (n=160) and Ceel-Bardaale (n=697). The seroprevalence of individuals in each of these villages is low (around 11.6% for MSP-1 and 8.3% for AMA-1) and so is the median titre (8.9 units/mL for MSP-1 and 17.2 units/mL for AMA-1). Overall, only 17.7% of the population is seropositive to either MSP-1 or AMA-1 antigens, as seen in Figure 5.7. Data were collected in three villages from the same district. However, in Ceel-Bardaale individuals appear to have a higher antibody titre on average and the distribution of antibodies against both MSP-1 and AMA-1 antigens is wider than in other villages. This could potentially be explained by the fact that Ceel-Bardaale is a larger and wealthier village where households are located along the seasonal river [236] and therefore individuals might be more exposed to malaria and present with higher antibody levels.

In Bioko, serum samples were collected from 7,387 individuals from the Malabo region (n=2,328), the North East region (n=1,323), the North West region (n=1,749), the South West region (n=700), the South East (n=688) and from other regions (n=700). More than half of the population (54%) were seropositive for antibodies against MSP-1 or AMA-1 antigens. The North West region has seroprevalence and median antibody titre for antibodies against MSP-1 and AMA-1 antigens much higher than the other five regions, and also a much wider distribution. If we assume that high antibody levels are associated with high exposure, then this result corroborates with what was already observed previously, i. e. high transmission in the North West region [241].

In The Gambia, serum samples were collected from South (n=344) and North (n=841) banks of the river Gambia. Overall, a high seroprevalence was recorded for both MSP-1 (47.6%) and AMA-1 (54.3%) antigens, with a seroprevalence to either MSP-1 or AMA-1 antigen of 66.4%. The antibody titre was also relatively high with a median of 151.2 units/mL for MSP-1 and 355.1 units/mL for AMA-1. In The Gambia, the malaria transmission season is expected to be shorter in the North Bank, which would explain some heterogeneity in the distribution of antibodies in this region. Some individuals might get exposed and increase their antibody levels while some individuals might not get exposed at all.

In Uganda, similar numbers of individuals were reported in each parish: Abedi (n=251), Akare (n=217), Apac Town (n=213) and Atopi (n=202). Around 70% of the overall population is seropositive to either MSP-1 (44.9%) or AMA-1 antigen (51.8%). Seroprevalence was comparable for each parish, ranging from 41.1% in Abedi to 49.4% in Apac Town for MSP-1 and from 48.8% in Akere to 59.2% Atopi for AMA-1. The distribution of both antibodies types is very similar between the different parishes. These results infer a lack of local heterogeneity as the data were collected from four parishes geographically relatively close to each other.

Titres for anti- AMA-1 antibodies are much higher than for anti- MSP-1 antibodies in Uganda, The Gambia and Bioko while the difference is less marked in Somalia. Similarly, the prevalence of individuals with antibodies against MSP-1 antigens appears to be consistently lower than against AMA-1 antigens for most of the regions in each country, with the exception of Ceel-Bardaale in Somalia and the North Bank in The Gambia where prevalence of anti-MSP-1 is higher than prevalence of anti- AMA-1. The lowest and highest antibody titres for MSP-1 are associated with, respectively, the lowest and highest antibody titre for AMA-1 within each country, with the exception of The Gambia. While seroprevalence for MSP-1 and AMA-1 ranks regions in the same order for Somalia and Bioko, there are some discrepancies in ranking in The Gambia and Uganda.

In a country of low endemicity such as Somalia, most of the population is seronegative. As intensity increases between the different countries, the prevalence of seronegative individuals decreases (Figure 5.7). In Uganda, the proportion of seronegative individuals is similar to The Gambia when the exposure is exected to be much higher. As transmission increases, it also appears that the proportion of individuals with anti- MSP-1 antibody increases, with the exception of Uganda where seroprevalence to MSP-1 only is lower than in The Gambia. There is no clear pattern for the acquisition of AMA-1 with increasing transmission intensity.

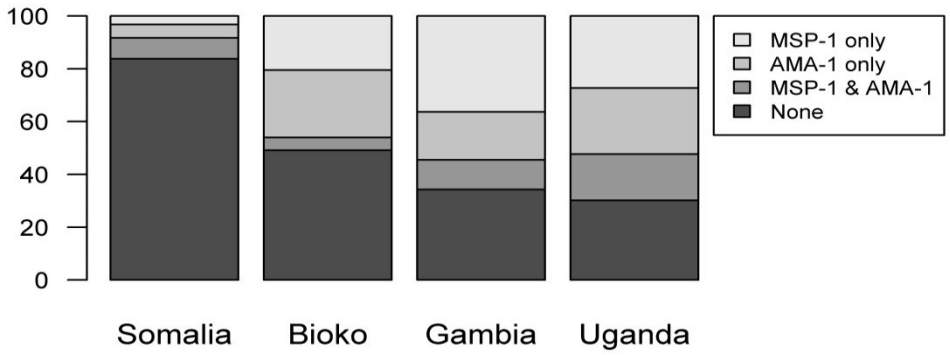


Figure 5.7: Prevalence of antibody types in each country

There is an association between seroprevalence and antibody titres between the different countries. As expected, all villages in Somalia report the lowest seroprevalence/antibody titre, all regions in Bioko records medium seroprevalence/antibody titre and all areas in Uganda and The Gambia present higher seroprevalence/antibody titres. This consistent association between seroprevalence and median antibody titre is also observed within countries. In most of the regions, high seroprevalence is associated with high antibody titre and low seroprevalence is associated with low antibody titre. Figure 5.8 shows a linear relationship between seroprevalence and median log10 antibody titre for both anti- MSP-1 and anti- AMA-1 antibodies. The correlation between seroprevalence and log10 antibody titre is significantly high with Spearman correlation coefficients of 0.87 for MSP-1 and 0.91 for AMA-1, with antibody titre which tends to increase as seroprevalence increases.

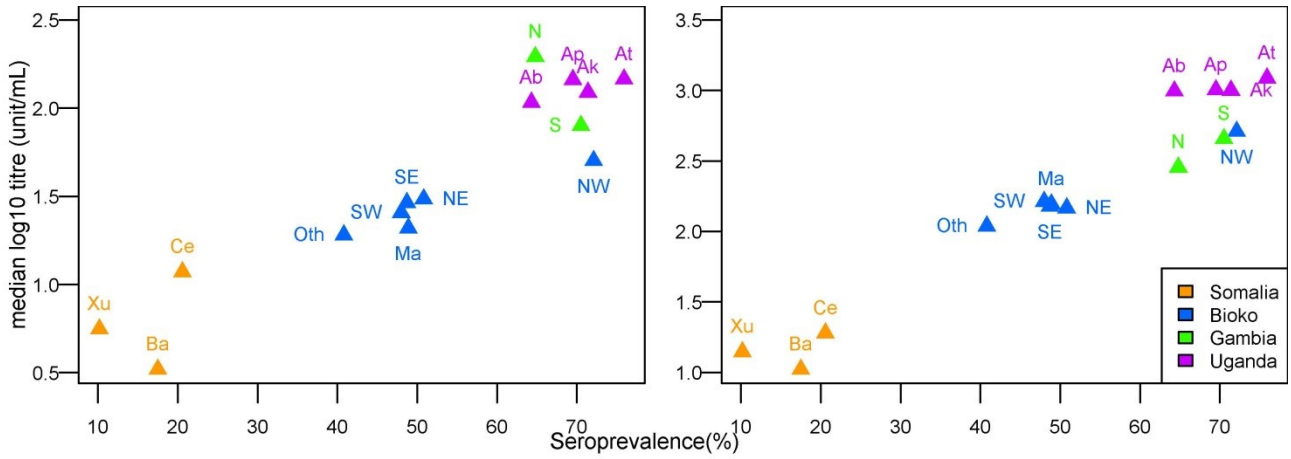


Figure 5.8: Association between seroprevalence and antibody titre for anti- MSP1 (left) and anti- AMA-1 (right) antibodies. Note that the y-axes are not on the same scale. Data are presented per region and categorised per country.

5.2.4.2 Age structured seroprevalence and antibody titres

Age is a significant determinant of malaria exposure and is an important factor when analysing serological data, as antibody levels vary with age. Individuals' ages were recorded during the data collection and the age distribution of individuals included in the study in each country is presented in Figure 5.9. A large proportion of the data are from individuals less than 20 years old. In Somalia, the median age is 15 years old (IQR: 6-37), 14 years old (IQR: 6-30) in The Gambia, 12 years old (IQR: 4-30) in Bioko and 15 years old (IQR: 5-29) in Uganda.

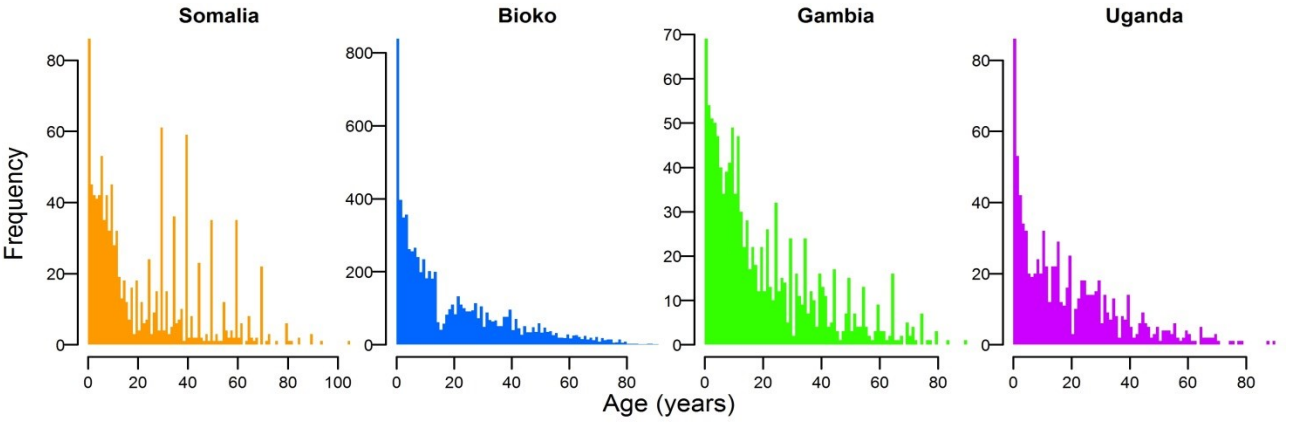


Figure 5.9: Age distribution in the studies in Somalia, Bioko, The Gambia and Uganda (left to right)

Figure 5.10 illustrates age-specific seroprevalence for all countries for both MSP-1 and AMA-1 antigens for the data and model fit using the catalytic model (described in the Chapter 4) for the following 11 age classes (chosen to have comparable number of individuals in each class): 0-1, 1-3, 3-5, 5-8, 8-10, 10-15, 15-25, 25-32, 32-40, 40-55, >55 years old. In all four countries, antibodies tested against both MSP-1 and AMA-1 antigens show a clear increase in seroprevalence with a person's age. In Somalia, The Gambia and Bioko, the catalytic model reproduces the data reasonably well for both antigens. The estimated seroconversion

rates are presented in Table 5.4. In Uganda, the model for antibodies against AMA-1 antigen provides a poor fit to the data with an under-prediction of seropositive individuals aged between 1 and 32 years old and an over-prediction for older individuals. In contrast, for the MSP-1 antigen, the model fit appears to be relatively good for individuals above 10 years old despite some lack of fit for younger children. Poor model fit may suggest some heterogeneity in the population, some changes in transmission intensity in time or some cross-reactivity of the antibodies.

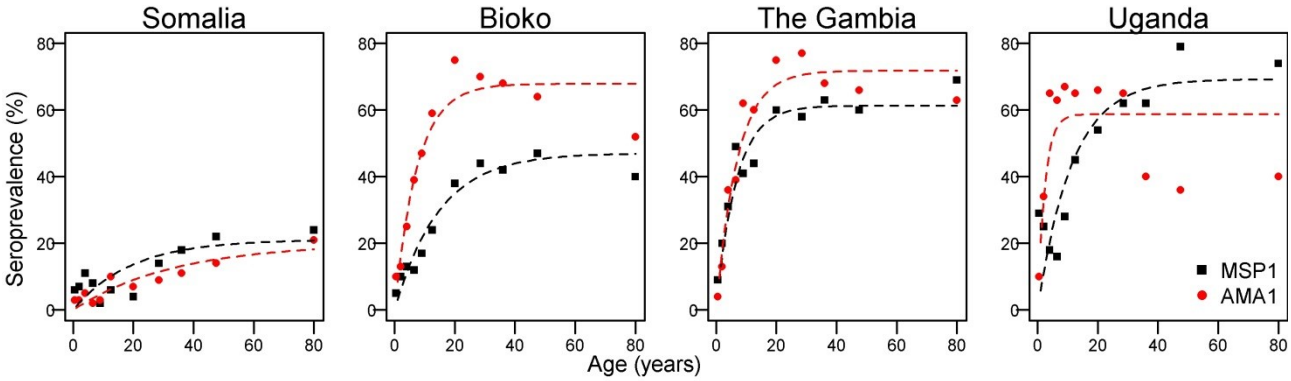


Figure 5.10: Seroprevalence for actual data (dots) and model fits (dashed lines) for Somalia, Bioko, The Gambia and Uganda (left to right) for both anti- MSP1 (black) and anti- AMA1 (red) antibodies.

Table 5.4: Estimates of seroconversion rates (SCR) for the studied countries. Mean and 95% confidence intervals are reported.

	MSP-1 SCR Mean (95% conf. interval)	**AMA-1 SCR** Mean (95% conf. interval)
Somalia	0.011 (0.005-0.016)	0.0059 (0.0034-0.0084)
Bioko	0.032 (0.029-0.035)	0.091 (0.085-0.097)
The Gambia	0.092 (0.07-0.11)	0.10 (0.084-0.12)
Uganda	0.061 (0.045-0.077)	0.25 (0.18-0.32)

The age-specific antibody titre distribution is presented in Figure 5.11. Antibody titre increases with age for all countries. In areas of low endemicity such as Somalia, the median level of antibodies only slowly increases between 0 and 15 units/mL for MSP-1 and up to 30 units/mL for AMA-1. However, for countries where transmission is much higher, a large proportion of the population have antibody titres exceeding values of 600 units/mL for MSP-1 in The Gambia and 1500 units/mL for AMA-1 in Uganda. The wide range around the median titre shows the variation in individual antibody titres in the population and is larger as age and exposure increase. Although in all countries while anti-MSP-1 antibody titres appear to continuously increase with age, anti- AMA-1 antibody titres seem to increase steeply with younger age and around 20 years old, titres start to decrease, with the exception for Somalia where antibody titre increases monotonically with age. The difference in antibody acquisition against MSP-1 and AMA-1 could be explained by a difference in immunogenicity of the antigens.

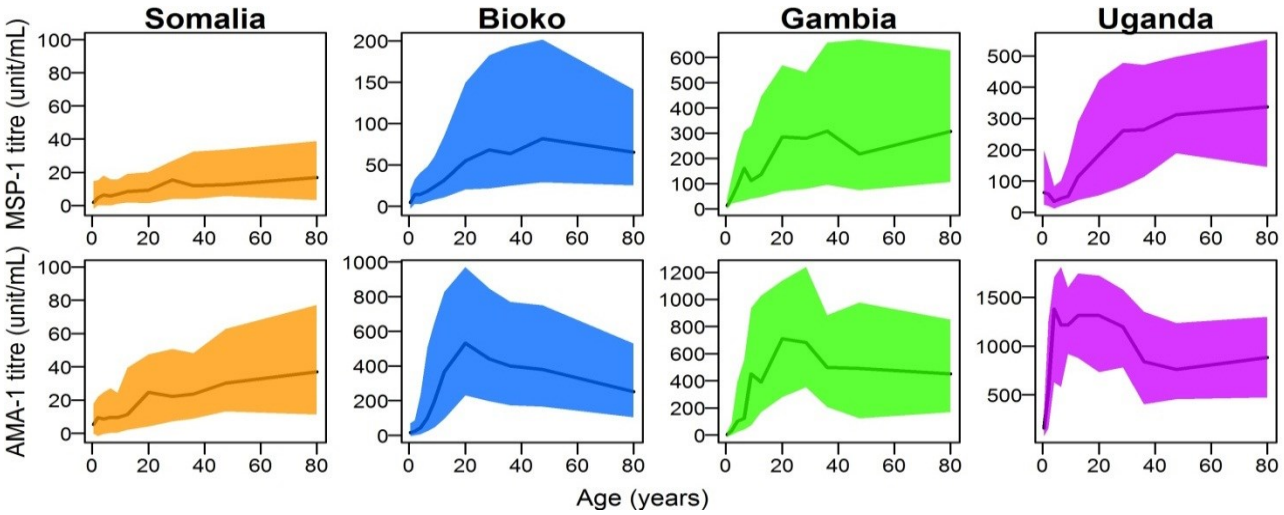


Figure 5.11: Age structured antibody titres against MSP-1 (first row) and AMA-1 (second row) for studied countries. Median titre represented by lines and shaded areas correspond to 25th and 75th percentiles of the antibodies distribution (IQR). Note that the y-axes are on different scales.

The prevalence of the parasite in human blood is indicative of malaria transmission. In areas of low exposure, parasite prevalence tends to be very low or equal to zero as it is the case in Somalia (0 parasite positive individuals / 1128 tested individuals). Parasite prevalence decreases with increasing age as expected, as shown in Figure 5.11. In high transmission settings, such as Uganda, the parasite prevalence is as high as 80% for young children and drops to 10% for older individuals. In The Gambia and Bioko, parasite prevalence appears to have a comparable pattern by age. For countries with medium to high transmission, it appears that the difference in exposure is noticeable in young children but for adults older than 20 years parasite prevalence becomes comparable.

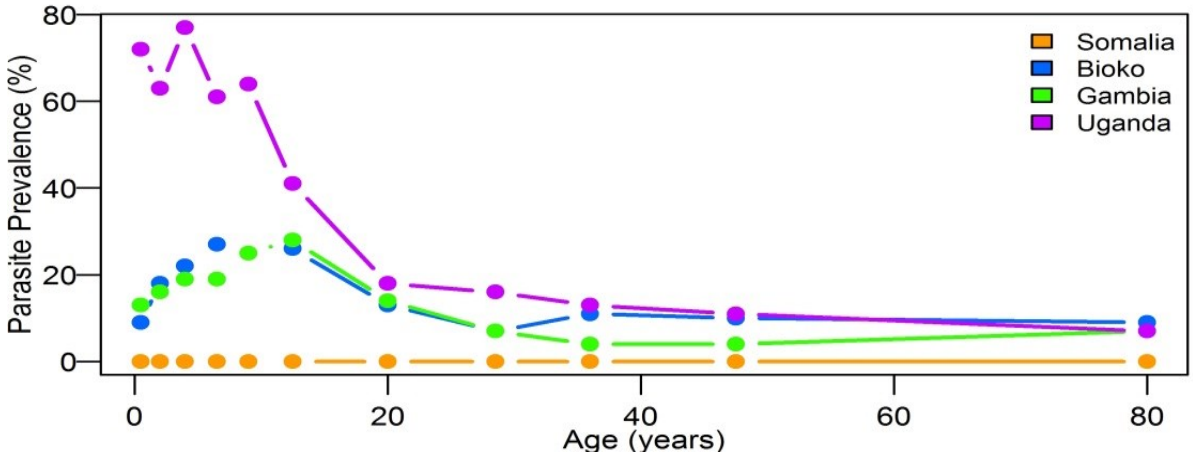


Figure 5.12: Parasite prevalence across age categorised by studied country

5.3 Methods

5.3.1 Measuring malaria transmission intensity using density model

The density model previously developed was fitted to data from Somalia, Bioko, The Gambia and Uganda to estimate malaria transmission intensity at country levels using both anti-MSP-1 and anti-AMA-1 antibodies. The laboratory methods were standardised across the four studies and reported antibody titres used for the analysis. The limit of detection was set to be 0.01 and antibody titres were $\log_{10}$ transformed. For all countries and both antigens, the model included *N=51* compartments and *Δ=0.12*. As before, the first class contained all individuals with log10 antibody titre equal to $x_{\min} = -2$ and the last class all individuals who had antibody titre greater than 3.88 on a log10 scale. Country-specific parameters for the exposure rate were estimated and the model was fitted to all countries simultaneously with the other parameters not varying by country. The rate of decay of antibodies, *ρ*, was fixed to a constant value of 0.7 years^{-1} (half-life: 360 days) for both MSP-1 and AMA-1 antigens. The model was identical to the one presented in Chapters 3 & 4 (*Sections 3.2.2.2 and 4.3.1.1*):

$$\frac{dy_{i,v}(t)}{dt} = \lambda_v \sum_{j \le i} k_{ij} y_j(t) + \frac{\rho}{\Delta} y_{i+1}(t) - \lambda_v \sum_{h \ge i} k_{hi} y_i(t) - \frac{\rho}{\Delta} y_i(\mathrm{t}) \qquad 1 \le i \le N \qquad (5.1)$$

with $y_{i,v}(t)$ the proportion of individuals in country v in antibody category *i* at age *t* and k_{ij} the probability that following exposure antibodies are boosted from antibody category *j* to category *i*, as defined previously. Let $Y(y_0, \lambda_v, t; \zeta, \rho)$ be the solution of (5.1) with y_0 the initial state (i.e. the distribution of the population according to individual's antibody level at the beginning of the exposure time, namely birth), λ_v the exposure rate for country v, t the age class, *ρ* the rate of decay of antibodies and $\zeta = \{a, b, S, \eta\}$ the set of boost parameters required to define k_{ij} .

As in earlier chapters (See *section 4.3.1.2*), the likelihood is given by:

$$l = \log(P(D \mid \theta)) = \sum_{v} \sum_{t \ne 0} \sum_{i} n_{i,t,v} \log(y_{i,v}(t)) \qquad (5.2)$$

with $n_{i,t,v}$ the observed number of individuals in antibody category *i* in country *v* at age *t.*

MCMC methods were used to sample from the posterior distribution of the parameters as described in the previous chapters. Because of the computational time required I ran two parallel chains with different starting points for MSP-1 (and three for AMA-1). I performed 1,000,000 iterations respectively for each of the runs of the MCMC algorithm with a burn-in period of 250,000 steps (50,000 for AMA-1). The output of

each of the chains was then recorded every 500 iterations (1,000 for AMA-1) to generate an aggregated sample from the posterior distribution of size 3,000 (2,850 for AMA-1).

5.3.2 Combining both MSP-1 and AMA-1 antigens to inform on transmission intensity

The model was extended to simultaneously fit data on anti- MSP-1 and anti- AMA-1 antibody levels. I assumed the country-specific rate of exposure, λ_v was independent of the antibody type. The boost parameters i.e $\zeta = \{a, b, S, \eta\}$ and the rate of decay of antibodies ρ were assumed to be antibody-specific (presented in Table 5.5). Parameter estimation was performed fitting simultaneously to data from Somalia, Bioko, The Gambia and Uganda at country-level as described before but with the following likelihood:

$$l = \log(P(D \mid \theta)) = \sum_{v} \sum_{t \neq 0} \sum_{i} (n_{i,t,v}^{MSP-1} \log(y_{i,v}^{MSP-1}(t)) + n_{i,t,v}^{AMA-1} \log(y_{i,v}^{AMA-1}(t))) \quad (5.3)$$

with $n_{i,t,v}^{MSP-1}$, $n_{i,t,v}^{AMA-1}$, $y_{i,v}^{MSP-1}(t)$ and $\mathrm{y}_{i,v}^{AMA-1}(\mathrm{t})$ respectively the observed number and predicted proportion of individuals in anti-MSP-1 and anti-AMA-1 antibodies category *i* in country *v* at age *t* defined as:

$$\begin{cases} y_{i,v}^{MSP-1}(t) = Y(y_0^{MSP-1}, \lambda_v, t; \zeta^{MSP-1}, \rho^{MSP-1}) \\ y_{i,v}^{AMA-1}(t) = Y(y_0^{AMA-1}, \lambda_v, t; \zeta^{AMA-1}, \rho^{AMA-1}) \end{cases} \quad (5.4)$$

with $y_0^{MSP-1} = y^{MSP-1}(0)$ and $y_0^{AMA-1} = y^{AMA-1}(0)$ the antibody distributions observed in the data at age 0. However, in Bioko (where age was recorded in months for individuals younger than 1) and in The Gambia, there was a significant drop of anti- AMA-1 antibodies between the age of 0 and 1. Therefore, in order to discard maternal immunity, in those countries I assumed individuals start their life with the anti- AMA-1 antibody level observed at age 1 and therefore $y_0^{AMA-1} = y^{AMA-1}(1)$.

Here, I performed 150,000 iterations of the MCMC algorithm with a burn-in period of 25,000 steps and the output of the four chains was recorded every 200 iterations to generate an aggregated sample for the posterior distribution of size 2,500.

Table 5.5: List of model parameters

Parameters	Description
λ_v	Exposure rate for country *v*
$a^{MSP\text{-}1}, a^{AMA\text{-}1}$	Maximum antibody boost size on exposure
$b^{MSP\text{-}1}, b^{AMA\text{-}1}$	Slope of dependence of antibody boost on current $\log_{10}$ titre
$S^{MSP\text{-}1}, S^{AMA\text{-}1}$	Standard deviation for boost size distribution
$\eta^{MSP\text{-}1}, \eta^{AMA\text{-}1}$	Mean boost for individuals with no current antibody
$\rho^{MSP\text{-}1}, \rho^{AMA\text{-}1}$	Rate of decay of antibodies

5.3.3 Extension of the density model

The density model can be further developed to address heterogeneity in exposure potentially attributable to spatial or temporal variations in transmission or any other factor. The effect of age on exposure can be accounted for by considering different exposure levels before and after a specific age. Also, any change in transmission that happened in the past can be additionally included in the model. Spatial heterogeneity in transmission can be considered by simultaneously assessing specific exposures in different regions. Here, I present three extensions of the density model that respectively account for heterogeneity of exposure, age-effect in exposure and changes in transmission intensity.

1. *Heterogeneity in exposure*

   Malaria transmission is rarely homogeneous in the population. Some individuals might be less exposed than other for instance due to genetic differences, individual's behaviour, use of intervention or spatial variations. I made the assumption that the population is made of two subpopulations exposed to malaria at different levels (See Figure 5.13). A proportion p of the population is exposed at a rate λ (subpopulation A) and a proportion $(1-p)$ is exposed at a lower rate $\lambda * \beta$ (subpopulation B), where $\beta \in [0,1]$. The model that captures heterogeneity in exposure can be written as:

   $$y(t) = py^A(t) + (1-p)y^B(t) \qquad (5.5)$$

   where $y^A(t) = Y(py_0, \lambda, t; \zeta\ , \rho\ )$ and $y^B(t) = Y((1-p)y_0, \lambda * \beta, t; \zeta\ , \rho\ )$ solutions of (5.1), represent the distribution of the proportions of individuals in each antibody class at age *t* in respectively subpopulations A and B.

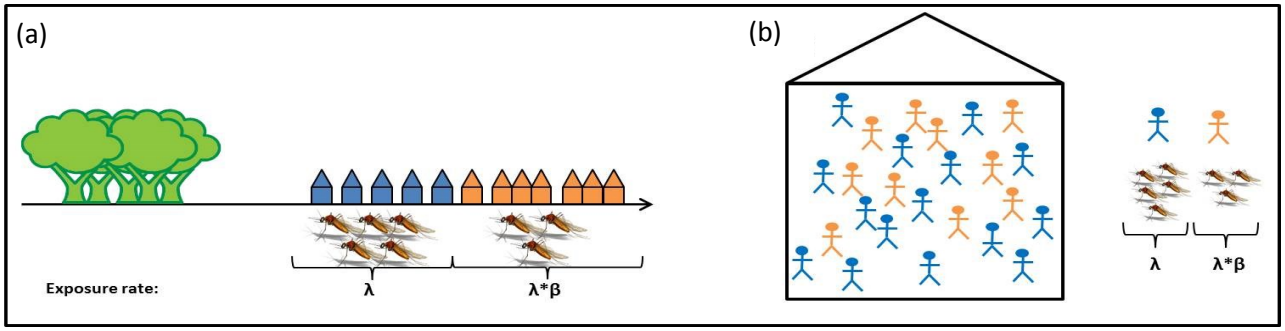


Figure 5.13: Examples of scenarios for heterogeneity of exposure. (a) Houses closer to the forest might get higher exposure (λ) than houses that live further away exposed at λ*β (with 0≤β≤1). (b) Within a same community/house, individuals might be exposed at different rates.

2. *Age-dependent exposure*

In some countries malaria exposure can vary between children and adults due to a number of different reasons including use of bed nets preferred for children only or occupational malaria with increasing exposure when children start to work. Here, I assumed that individuals older than ω years old (referred as adults) are exposed at a rate λ and individuals younger than ω years old (referred as children) to be exposed at a lower rate $\lambda * \beta$, where $\beta \in [0,1]$ (See Figure 5.14A). The distribution of individuals in each antibody class at age *t* can be modelled as:

$$y(t) = \begin{cases} Y(y(0), \lambda * \beta, t; \zeta, \rho) & \text{if } t \leq \omega \\ Y(y(\omega), \lambda, t; \zeta, \rho) & \text{if } t > \omega \end{cases} \qquad (5.6)$$

with the distribution of the population according to an individual's antibody level at the beginning of the exposure time being at birth ($y(0)$) for children and at age ω for adults ($y(\omega)$).

3. *Changes in transmission intensity*

Due to interventions put in place or other environmental factors, in some areas malaria transmission intensity can vary through time. Here I extend the density model to assess whether a decline in transmission can be detected. I assumed a sudden change, reflecting effective control interventions, happened Ω years before the survey and the exposure rate dropped from λ to $\lambda * \beta$, $\beta \in [0,1]$ (See Figure 5.14B). So individuals, aged t years old at the time of the survey, were exposed from birth to $t - \Omega$ at rate λ and at rate $\lambda * \beta$ for the last Ω years between the change and the time of the survey. Individuals born after the change (less than Ω years ago) were exposed at a single rate $\lambda * \beta$. The distribution of individuals in each antibody class at age *t* can be modelled to account for the change in transmission as:

$$y(t) = \begin{cases} Y(y(0), \lambda * \beta, t; \zeta, \rho) & \text{if } t < \Omega \\ Y(y(t-\Omega), \lambda * \beta, t; \zeta, \rho) & \text{if } t \geq \Omega \end{cases} \quad \text{where} \quad y(t-\Omega) = Y(y(0), \lambda, t-\Omega; \zeta, \rho) \qquad (5.7)$$

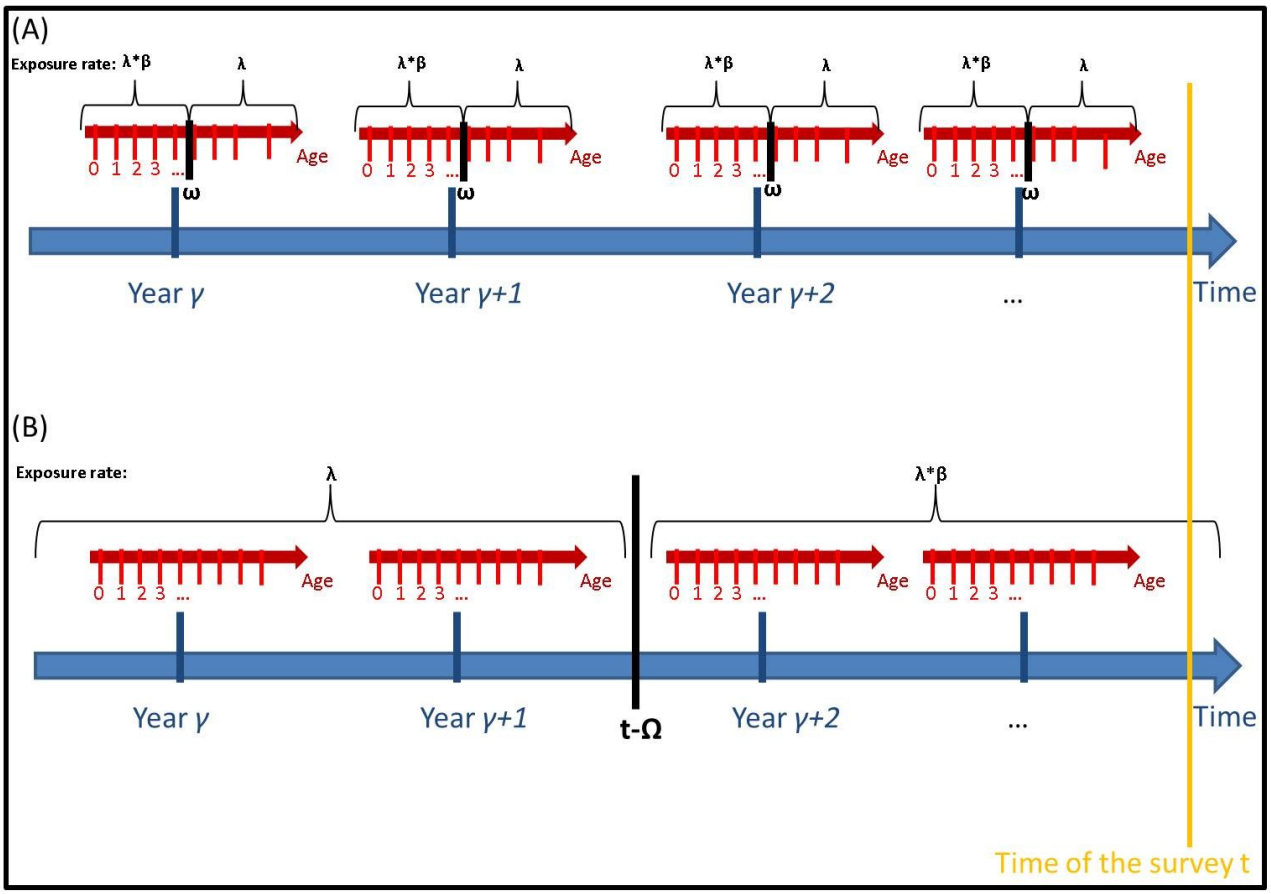


Figure 5.14: Schematic representation of the variability in exposure due to age and time. (A) At age ω, children who were exposed at rate λ*β are now exposed at rate λ, indifferently of the time. (B) A change in transmission, due to successful intervention for instance, happened Ω years ago. Before the change all individuals were exposed at rate λ and since the change individuals are exposed at the reduced rate λ*β.

A summary of the model parameters for the three extensions of the density model is presented in Table 5.6.

Table 5.6: Summary of the model parameters related to exposure for the three extended models

Scenarios	Parameters	Description	Prior Distributions
Heterogeneity	λ	Exposure Rate (years^{-1})	Uniform on [0,100]
	β	Scaling for exposure	Uniform on [0,1]
	p	Proportion of individuals exposed at rate λ	Uniform on [0,1]
Age Effect	λ	Exposure Rate (years^{-1})	Uniform on [0,100]
	β	Scaling for exposure	Uniform on [0,1]
	ω	Age at change in transmission (years)	Uniform on [1,100]
Changes in transmission	λ	Exposure Rate (years^{-1})	Uniform on [0,100]
	β	Scaling for exposure	Uniform on [0,1]
	Ω	Time since change in transmission (years)	Uniform on [1,100]

5.3.4 Fitting extended models to simulated data

To assess the performance of the method to quantify transmission intensity in complex settings, I designed simulations based on the following three scenarios:

1. *Heterogeneity in exposure*

   I assumed that data were collected in a survey where the population had heterogeneous exposure of transmission with a proportion $p = 0.8$ (subpopulation A) exposed at a constant rate $\lambda = 3.5 \text{ yrs}^{-1}$ while the other individuals (subpopulation B) were exposed at a constant rate $\lambda * \beta, \text{ with } \beta = 0.2$.

2. *Age-dependent exposure*

   I made the assumption that data were collected in an area where adults, after the age of $\omega = 15$, were exposed at a rate $\lambda = 3.5 \text{ yrs}^{-1}$, while children younger than $\omega = 15$ years old were exposed at a rate $\lambda * \beta, \text{ with } \beta = 0.8$

3. *Changes in transmission intensity*

   A change in transmission was assumed to have happened $\Omega = 15$ years ago. Before the change, individuals were exposed at a rate $\lambda = 3.5 \text{ yrs}^{-1}$ and since the change individuals are exposed at a rate $\lambda * \beta, \text{ with } \beta = 0.8$.

For these three scenarios, biological parameters modelling the boost of antibodies are assumed to be constant with the maximum size of the boost of antibodies $a = 0.75$, the slope of dependence of the boost size $b = 0.5$ and the standard deviation $S = 0.02$. "Seronegative" individuals had an average boost size of $\eta = 0.01$ and the rate of decay of antibodies was set to 0.7 years^{-1}. I used the same age structure that was used for the simulation in section 3.3.1, based on the Cambodian data set. The number of individuals in each antibody level category was drawn from a multinomial distribution with its associated probability corresponding to $y_i(t)$, where $y(t) = (y_i(t))_i$ is described in Section 5.3.3 for each of the scenarios. I also assumed that at birth no individuals presented with antibodies, thereby ignoring maternal immunity. For each scenario 50 datasets were simulated and parameter estimation was performed for each of them using MCMC approach using the following multinomial log-likelihood:

$$l = \log(P(D \mid \theta)) = \sum_t \sum_i n_{i,t} \log(y_i(t)) \qquad (5.8)$$

with $n_{i,t}$ and $y_i(t)$ respectively the observed number and modelled proportion of individuals in antibodies category i at age t. Only parameters related to exposure (the exposure rate λ, the scaling factor β, the

proportion p, the age of change ω and the time since the change Ω) were estimated and their associated uninformative priors are presented in Table 5.6. The biological parameters (the maximum boost size a, the slope of dependence b, the standard deviation S, the maximum boost size for "seronegative" individuals η and the rate of decay of antibodies ρ) were set to the values used for the simulation. I performed 40,000 iterations for each run of the MCMC algorithm with a burn-in period of 1,000 steps. The output was recorded every 10 iterations to generate a sample from the posterior distribution of size 3,900.

5.3.5 Confidence regions from profile likelihood

So far, Bayesian methods have been used to provide inference on model parameters. However, classic methods such as estimation by maximum likelihood were also used in this chapter, as the use of Bayesian methods becomes computationally intensive when exploring temporal heterogeneity in transmission. Here, I present the use of profile likelihood methods [242] to construct the confidence intervals for the parameters of interest θ. Let ψ be the other parameters of the model, also termed nuisance parameters. Denote by $L(\theta,\psi)$ the likelihood function and by $\hat{\theta}$ the maximum likelihood estimates. The profile likelihood $P_l(\theta)$ is defined by $P_l(\theta)=\max_{\psi} L(\theta,\psi)$. Confidence regions are based on likelihood ratio test. Indeed, the likelihood ratio test statistic of the hypothesis $H_0:\theta=\theta_0$ (where θ_0 is a fixed value), given by

$-2\log\frac{P_l(\theta_0)}{P_l(\hat{\theta})}$ follows a χ^2_p distribution with p the dimension of θ. So the profile likelihood-based confidence interval is given by: $\left\{\theta:\log P_l(\hat{\theta})-\log P_l(\theta)\le\frac{1}{2}\chi^2_{p,1-\alpha}\right\}=\left\{\theta:\log P_l(\theta)\ge\log P_l(\hat{\theta})-\frac{1}{2}\chi^2_{p,1-\alpha}\right\}$

with $\chi^2_{p,1-\alpha}$ the (1-α) percentile of χ^2_p distribution.

5.4 Results

5.4.1 Measuring malaria transmission intensity in Somalia, Bioko, Gambia and Uganda

5.4.1.1 Measuring malaria transmission intensity analysing each antibody separately

The fitted model for each separate antibody is presented in Figure 5.15 and Figure 5.16 respectively for MSP-1 and AMA-1 and the parameter estimates are shown in Table 5.7.

In all countries the model reproduces the data relatively well for both anti- MSP-1 and anti- AMA-1 antibody levels.

Around 15 and 5 years old, respectively for Bioko and Uganda, anti-AMA-1 antibody levels appear to plateau and then decrease with age. The model is not able to capture these trends. Instead, in Bioko, while the model fits the levels of anti- AMA-1 antibodies between birth and 10 years old well, they are under-predicted for individuals between 10 and 50 years old and over-predicted for individuals older than 50 years. In Uganda, the fitted model overestimates anti- AMA-1 antibody levels from the age of 25 onwards. There is also a decrease in anti- AMA-1 antibodies observed in the data but this is less marked than for Bioko. The fitted model tends to over-predict the number of individuals with no circulating antibodies, due to the low estimates of the average boost size for these individuals, η. Nevertheless, the age-specific distribution of antibodies does not appear to show these results. This might be due to the overall number of individuals with high levels of anti- AMA-1 antibodies being over-predicted to balance the under-prediction of lower antibody levels.

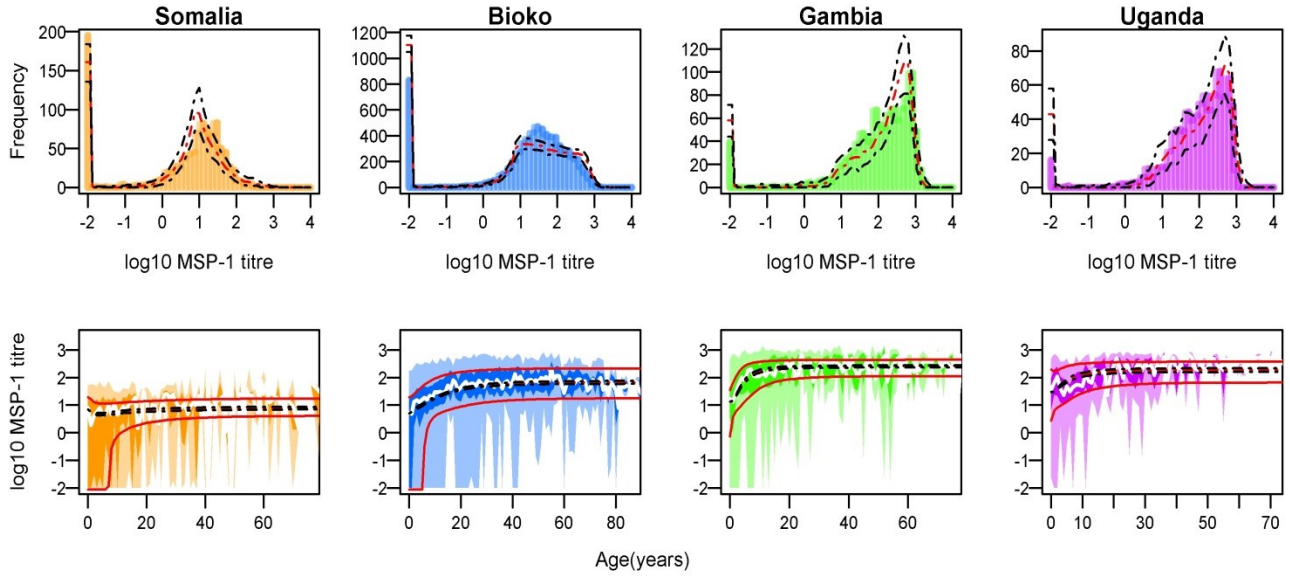


Figure 5.15: Anti- MSP-1 antibody distribution. Overall (first row) and age specific (second row) antibody distributions for Somalia, Bioko, Gambia and Uganda (left to right). Median fit (95% prediction interval) for median antibody levels are represented by the red (black) dashed lines. Dark and light shaded area represent respectively 25th/75th IQR and 2.5th/97.5th IQR for the data. Solid red lines represent 25th/75th IQR for the model fit.

In the age specific antibody distribution (Figure 5.16), the median fit of the median antibody level and its associated 95% credible interval are represented with respectively red and black dashed lines while the median fit of the 25th and 75th percentiles (IQR) are represented by the solid red lines (no credible intervals presented for the IQR model fit).

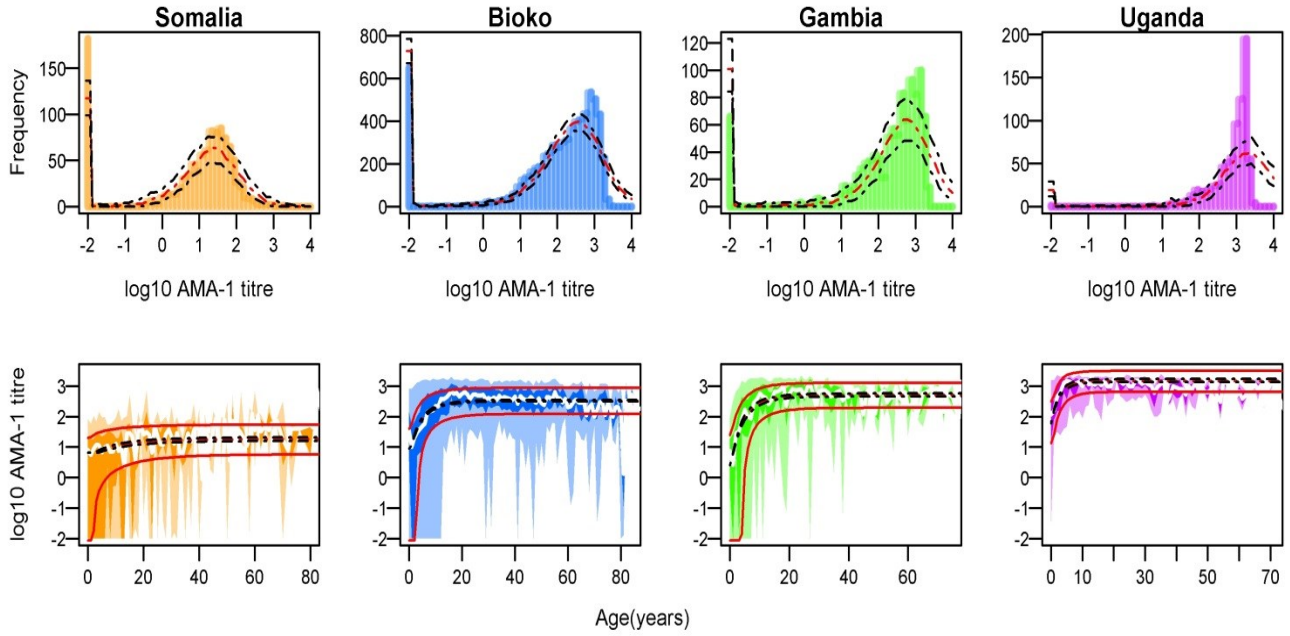


Figure 5.16: Anti- AMA-1 antibody distribution Overall (first row) and age specific (second row) antibody distributions for Somalia, Bioko, The Gambia and Uganda (left to right). Median fit (95% credible interval) for median antibody levels are represeted by the red (black) dashed lines Dark and light shaded area represent respectively 25th/75th IQR and 2.5th/97.5th IQR for the data. Solid red lines represent 25th/75th IQR for the model fit.

It is important to note that, despite a good fit of the model to the data for AMA-1, the *posterior* distributions found for the parameters were bimodal (See Appendix VI), suggesting either a high rate of exposure and a small mean antibody boost or a low rate of exposure with a large mean boost. Given the available data on anti- AMA-1 antibodies the model was not able to distinguish between these hypotheses.

Exposure rate estimates for each country are presented in Figure 5.17A for anti- MSP-1 antibody and the bimodal distribution is presented in Figure 5.17B for anti- AMA-1 antibody. Countries are shown according to their expected level of transmission increasing from left to right. The exposure rate estimates increase with increasing expected level of transmission. However, the estimated exposure rates differ depending on the antibody type fitted to. For example, the estimated rate of exposure in Uganda is lower than that in The Gambia when the model is fitted to anti- MSP-1 antibody data. However, when fitted to the anti- AMA-1 antibody data the exposure rate for Uganda is much higher (considering either of the two modes). Despite the bimodal distribution of exposure for AMA-1, either of the parameter sets corresponding to these two modes remains correlated with increasing level of transmission between countries.

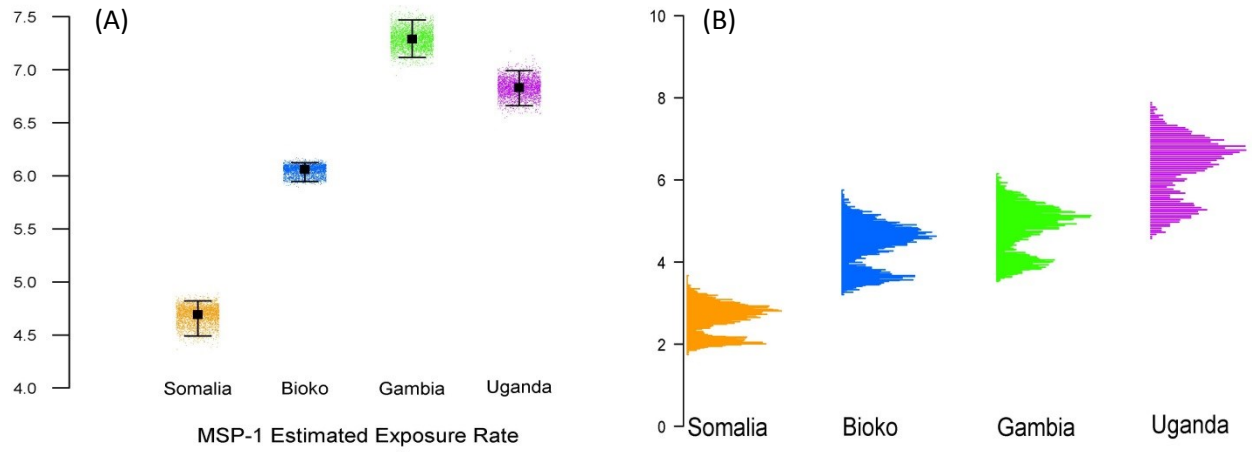


Figure 5.17 Estimates of exposure rate for anti- MSP-1 (left) and anti- AMA-1 (right) antibodies. Median and *posterior* 95% credible intervals are represented by the black square and error bars. Individuals estimates from the MCMC algorithm are presented in colours: Somalia (orange), Bioko (blue), The Gambia (green) and Uganda (purple). Nothe that the y-axes are on different scales.

Despite expected biological differences between MSP-1 and AMA-1 antigens, the estimates of the parameters modelling the antibody boost are somewhat comparable, but with a higher standard deviation for AMA-1 (See Table 5.7). Also, "seronegative" individuals tend to have higher anti- AMA-1 antibody boost size, but for both antigens, following the first exposure, individuals are estimated to remain "seronegative". Indeed given the lognormal distribution of the boost with mean size η and standard deviation S, only a negligible proportion of the "seronegative" individuals will increase their anti-MSP-1 antibody level (2% for MSP-1).

Table 5.7: Parameter estimation for estimation using respectively anti- MSP-1 and anti- AMA-1 antibodies.

Parameter	MSP-1	AMA-1		
	Median (95% CrI)	Median 1	Median 2	(95% CrI)
Exposure rate λ				
Somalia	4.7 (4.5-4.8)	2.1	2.8	(1.9-3.2)
Bioko	6.1 (5.9-6.1)	3.7	4.7	(3.4-5.2)
The Gambia	7.3 (7.1-7.5)	4	5.1	(3.7-5.7)
Uganda	6.8 (6.7-7)	5.3	6.7	(4.9-7.4)
Maximum boost size a	0.75 (0.69-0.86)	1.3	0.86	(0.73-1.4)
Slope of dependence b	0.52 (0.5-0.55)	0.42	0.38	(0.36-0.43)
Standard deviation S	0.02 (0.017-0.033)	0.15	0.14	(0.11-0.18)
Maximum boost size for "seronegative" η	0.021 (0.007-0.045)	0.044	0.034	(0.028-0.049)

Note here that the modes for the parameter distribution for anti- AMA-1 antibodies were clearly distinguishable; therefore the *posterior* distribution was split in two subsets to derive the median for each mode.

5.4.1.2 Measuring malaria transmission intensity analysing simultaneously both antibody types

The model was next extended to simultaneously estimate parameters associated with the acquisition of multiple types of antibodies. The estimates obtained are presented in Table 5.8.

The estimates obtained for the boost parameters are comparable to those obtained when fitting the model separately to the different antigens. The estimates of the exposure rates are similar to those obtained by fitting the model to the MSP-1 antibody data with the exception of Uganda, which is estimated to be 7.2 years^{-1} (95% CI: 7.0-7.4), higher than the exposure rate in The Gambia equal to 6.8 years^{-1} (95% CI: 6.7-6.9).

Table 5.8: Parameter estimation when both antibody types are simultaneously included in the model.

Estimates **Median (95% CrI)**					
Exposure		*MSP-1*		*AMA-1*	
Somalia	4.1 (4-4.2)	***a***	0.93 (0.85-1)	***a***	0.57 (0.53-0.61)
Bioko	5.9 (5.8-6)	***b***	0.57 (0.55-0.59)	***b***	0.35 (0.33-0.36)
The Gambia	6.8 (6.7-6.9)	***S***	0.021 (0.018-0.021)	***S***	0.082 (0.074-0.093)
Uganda	7.2 (7.0-7.4)	***η***	0.025 (0.009-0.042)	***η***	0.020 (0.018-0.021)

The model fits reproduce well the overall and age-specific distributions of both anti-MSP-1 and anti-AMA-1 antibodies (See Figure 5.18). Nevertheless, the model tends to overestimate the number of individuals with no circulating antibodies (as the model predicts “seronegative” individual to stay “seronegative”). The model fits (Figure 5.18) suggest similar predictions to those obtained by fitting the model to the two antibodies independently.

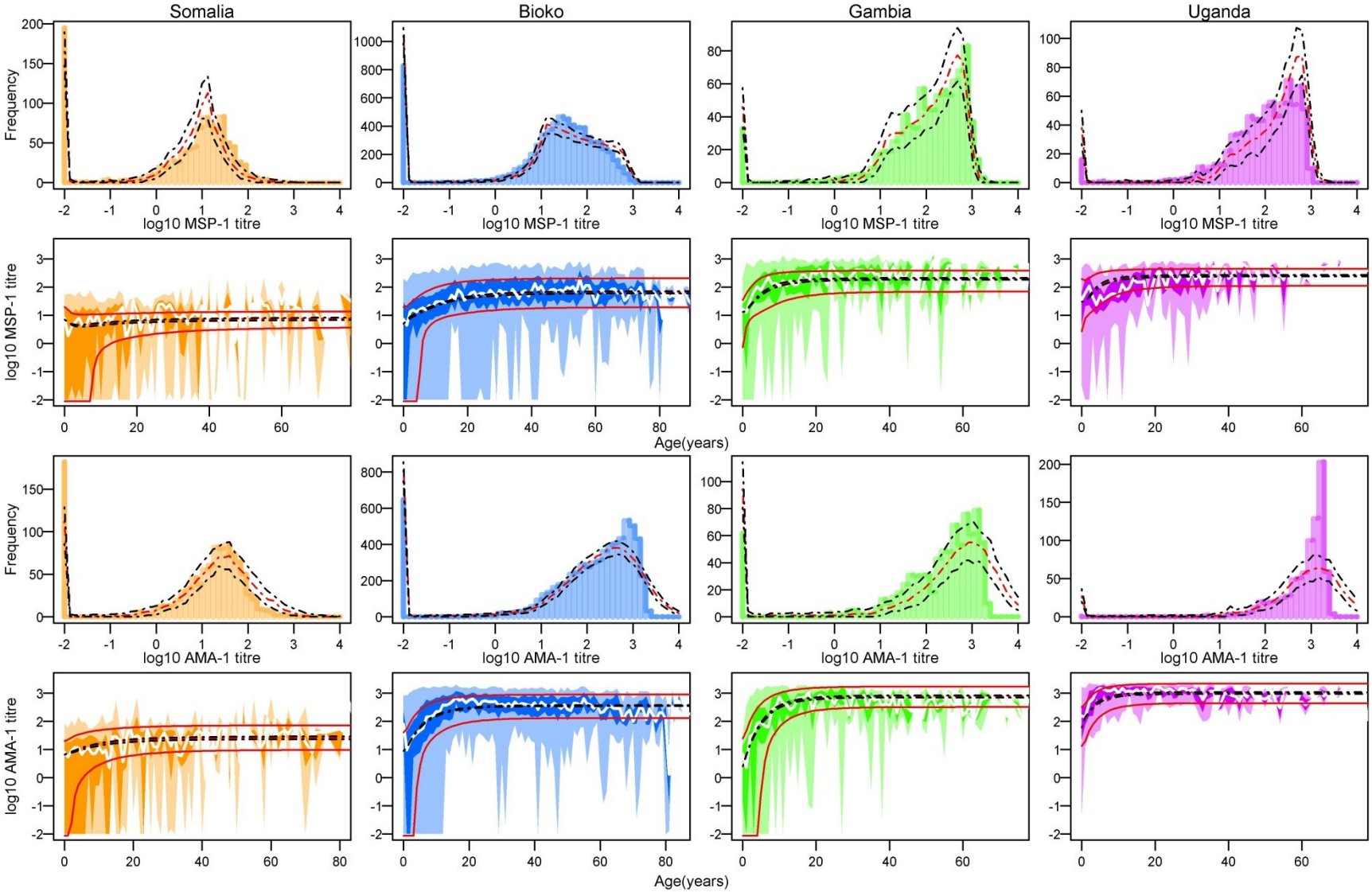


Figure 5.18: **Anti- MSP-1 (top two rows) and anti-AMA-1 (bottom two rows) antibody distributions. Overall (first and third rows) and age specific (second and fourth rows) antibody distributions for Somalia, Bioko, The Gambia and Uganda. Median fit (95% credible interval) for median antibody levels are represeted by the red (black) dashed lines. Dark and light shaded area represent respectively 25th/75th IQR and 2.5th/97.5th IQR for the data. Solid red lines represent 25th/75th IQR for the model fit.**

A comparison of the estimated exposure rates is presented in Figure 5.19A with their association with estimates of seroconversion rates (SCR) obtained using the catalytic model. A high degree of correlation is observed between seroconversion rates derived using a catalytic model and exposure rates obtained fitting a density model (Spearman correlation r=0.94 between density model using both antigens and SCR derived using MSP-1). However, the models fitted to the different antibodies independently rank the exposure from each country in different orders to those fitted to the two antibodies together. It is worth noting here that this difference is also observed when exposure is assessed using seroconversion rates. In addition, when both antibodies are analysed in a single model, the ranking of villages based on exposure correspond to the expected ranking (Table 5.2), also found considering anti-AMA-1 alone (Uganda has higher exposure than The Gambia). As seen in Figure 5.19B, using both antibodies in the evaluation of transmission intensity improves the precision of the estimations. Note that the precision of the estimated exposure rate using density models is relatively consistent for each country, while precision for SCR highly depends on the endemicity level (here country).

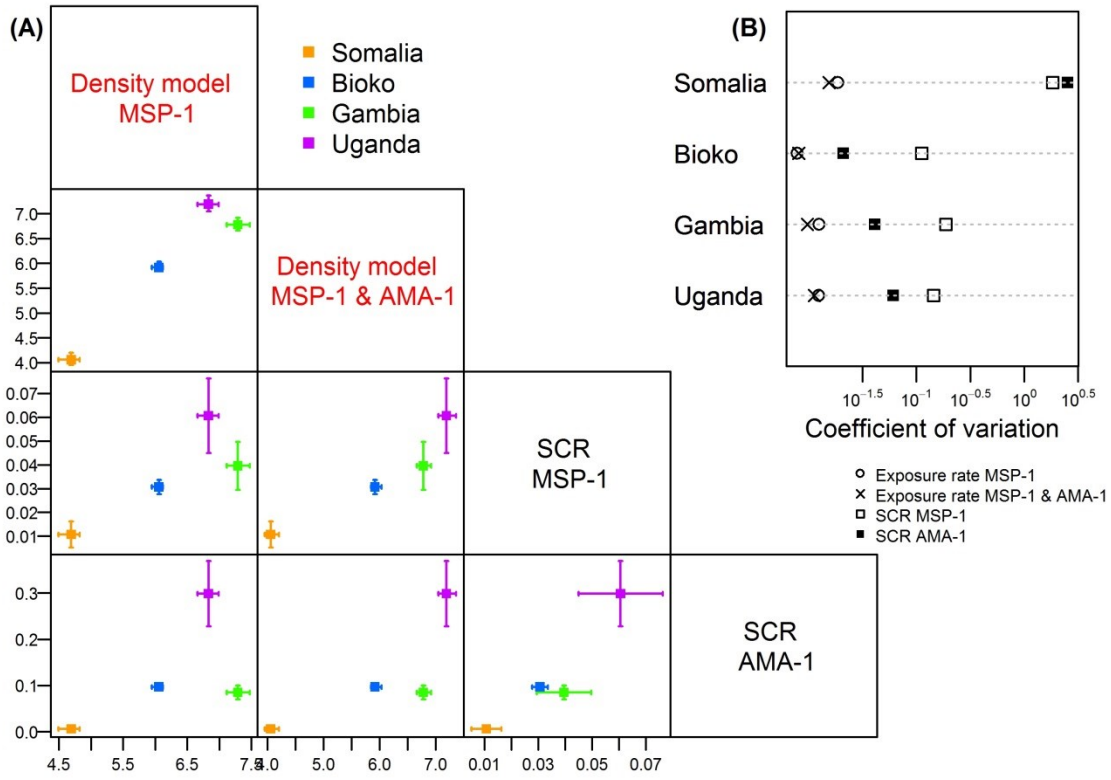


Figure 5.19: (A) Association between exposure rate from the density model and seroconverion rate (SCR). Median estimates of exposure and their associated *posterior* 95% credible interval are presented for each country when density model used simultaneously both anti- MSP-1 and anti- AMA-1 antibodies (first column), separately anti-MSP-1 and anti- AMA-1 antibodies (second and third columns). Estimates of SCR are presented when anti-MSP-1 (fourth colum) and anti- AMA-1 (last colum) antibodies are considered. (B) Coefficient of variation for exposure rate derived for each country categorised by the different methods.

5.4.2 Simulation study

A total of 50 datasets were generated for each of the scenarios described in the methods section using extended versions of the density model. In the first scenario, antibody levels are modelled assuming that the population is exposed to two different levels (scenario A). In the second scenario children and adults are assumed to be exposed at different rates (scenario B). In the third scenario a change in transmission is assumed to have occurred in the past (scenario C). For all the simulations, the "true" parameter value used in the simulation fell within the 95% credible interval obtained from fitting the model to the simulated data in 96.8% of the cases for scenario A, 90.7% for scenario B and 97.3% in scenario C. Details of these probabilities are given for individual parameters in Table 5.9.

Table 5.9: Probability of correctly estimating each of the parameters

Scenarios	Param.	Description	Probability of correctly estimating individually each parameter
(A) Heterogeneity	λ	Exposure Rate	92%
	β	Scaling for exposure	100%
	p	Proportion of individuals exposed at rate λ	98%
(B) Age Effect	λ	Exposure Rate	78%
	β	Scaling for exposure	94%
	ω	Age at change in transmission	98%
(C) Changes in transmission	λ	Exposure Rate	96%
	β	Scaling for exposure	96%
	Ω	Time since change in transmission	100%

Figure 5.20 illustrates the results from the simulation studies. These scenarios do not necessarily provide a clear pattern in the data that would clearly indicate the underlying simulated heterogeneity or change in transmission. When individuals experience a high degree of heterogeneity in exposure in the population (across all age classes) as in simulation A this is noticeable in the wide range of the antibody levels. However, even this would be difficult to determine without the comparison dataset. The parameter estimation was able to recapture the model parameters and therefore the model fits the data extremely well.

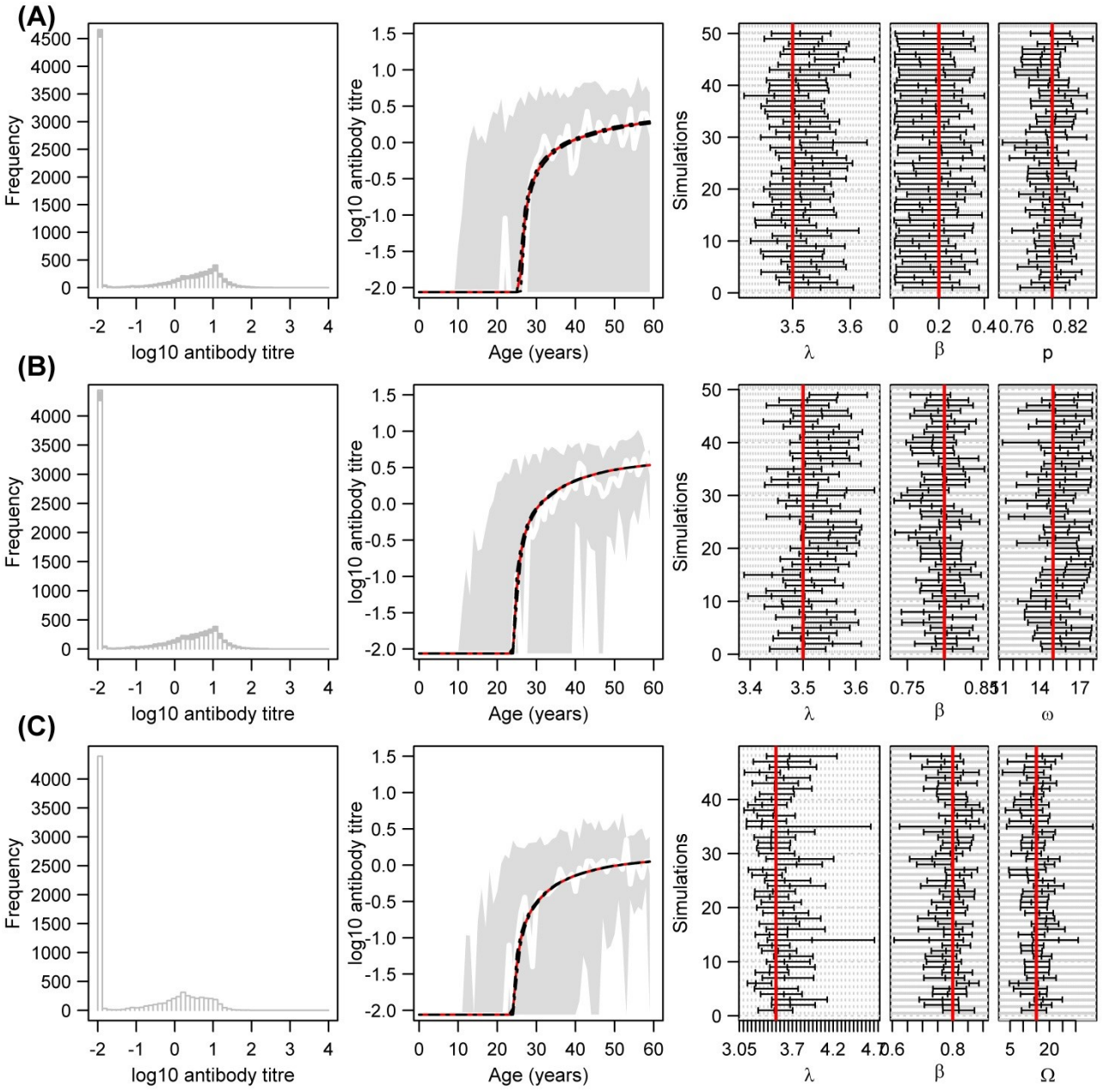


Figure 5.20: Simulation study results for models considering heterogeneity in exposure (A), the effect of age on exposure (B) and a change in transmission intensity (C). For each scenario, the distribution of the antibody levels in each of the simulated dataset is presented in the left panels. The middle panels illustrate the model fit for a set of simulated data (25^{th}/75^{th} IQR in grey and median in white) against the median fit obtained using the *posterior* median of the model parameters (red line) with its associated credible interval (black dashed lines). The right panels represent the *posterior* 95% credible intervals and median for each of the parameters estimates for each simulated dataset. The red line corresponds to the value of the parameter used during the simulations.

Note also that only 50 simulations were performed here, therefore it is likely that the confidence on the proportion of simulations in which the parameters are correctly estimated might increase by increasing the number of simulations.

5.4.3 Applying the extended models to real data

5.4.3.1 Assessing changes in transmission intensity in Bioko Island

Malaria control interventions have been put in place in Bioko Island since 2004. Multiple studies have shown that since the implementation of the interventions, a heterogeneous impact has been recorded around the island [135, 237, 241]. I therefore extended the density model to be able to capture any change in malaria transmission that has happened in the different areas on the island. Due to the computational complexity of the model, I did not attempt to estimate change in transmission. Instead, I used the posterior median of the boost parameters obtained when applying the model to both MSP-1 and AMA-1 antigens separately (See Table 5.10) and I used profile likelihood methods to explore how different combinations of the parameters (the exposure rate λ, the scaling factor β and the time since the change Ω) influence the fit of the model (as measured by the likelihood). By analysing each region separately, I accounted for heterogeneity in exposure between the regions and assumed that the exposure, the impact and the time of the change are specific for each region. To construct the grid, I used the ranges of exposure rate obtained with a regional analysis performed in Bioko (not presented) i.e. between 4 and 8 for MSP-1 and 2.5 and 5 for AMA-1. Additionally, I assumed that a change could have happened between 1 and 20 years ago. As a consequence, the likelihood was computed annually for 10 values of the scaling factor β equally spaced between 0 and 1, and 15 values of exposure rate for each of the antigens, resulting in a grid of 3,000 points. The likelihood, function of the scaling factor β and the time since the change in transmission Ω are illustrated in Figure 5.21. Additional results presenting both of these factors against the exposure rate (presented in the Appendix VI) show that the precision associated with the exposure rate is relatively good for both antigens.

Table 5.10: Parameter values used for nuisance parameter for derivation of likelihood for both MSP-1 and AMA-1 data.

Parameter	MSP-1	AMA-1
Maximum boost size a	1.7	1.3
Slope of dependence b	0.69	0.42
Standard deviation S	0.025	0.017
Maximum boost size for "seronegative" η	0.034	0.055
Rate of antibody decay ρ	0.7	0.7

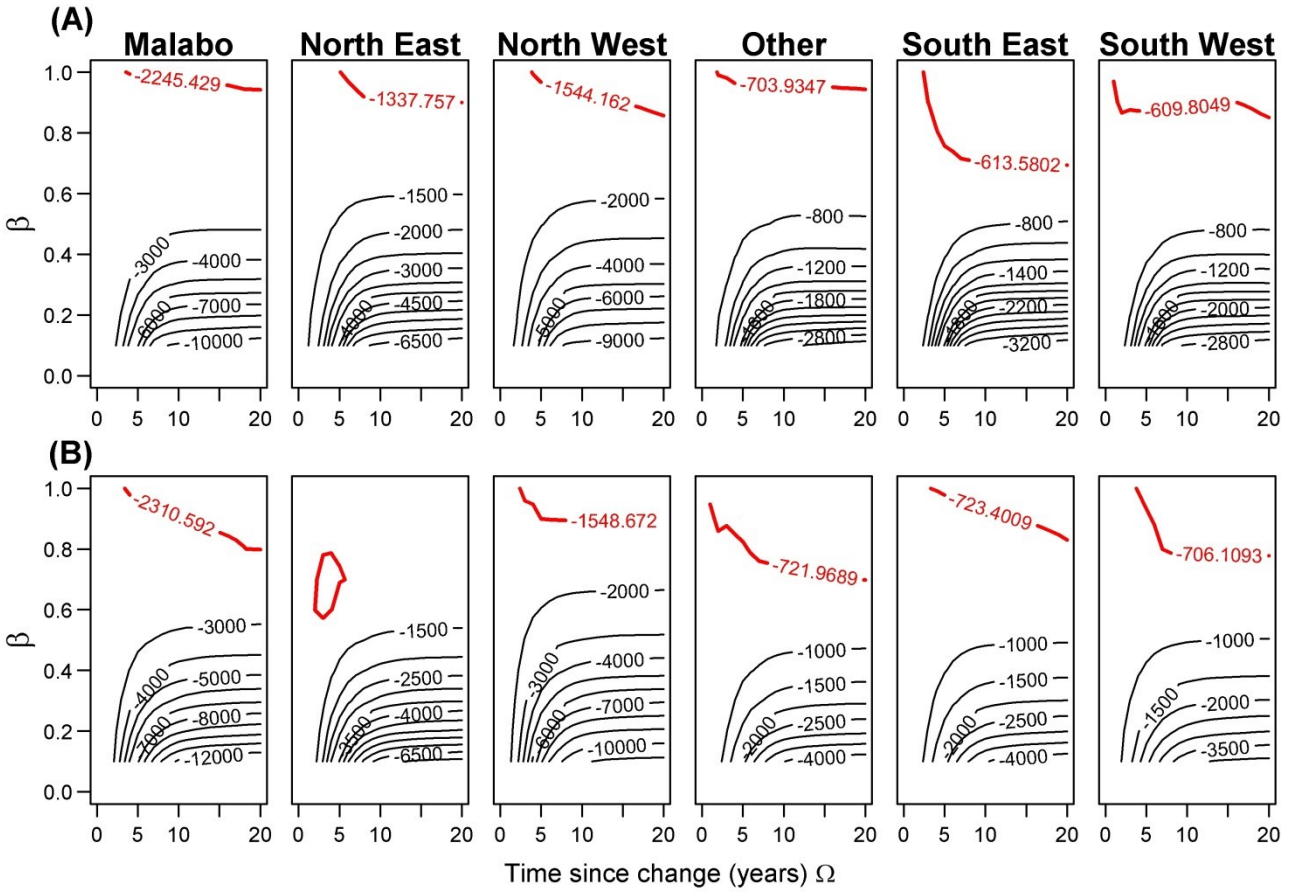


Figure 5.21: Log Likelihood computed for varying values of the exposure rate λ, scaling factor β (y-axis) and the time since the change of transmission intensity Ω (x-axis) for both MSP-1 (A) and AMA-1 (B) antigens for various regions in Bioko 95% confidence intervals are represented in red. Only the maximum likelihood for the different values of exposure rates is presented.

The results with MSP-1 antigen suggest that in most of the regions, it was not possible to detect any change in transmission by fitting to MSP-1 antibodies (scaling factor β is around 1). For the South East and South West regions, the results suggest that there might have been a change in transmission which happened at least 2 years before the survey when transmission dropped up to 30%. Results obtained with anti- AMA-1 antibodies show that a change in transmission happened between 2 and 7 years before the survey, resulting in a drop of transmission intensity between 20 and 40%. The results for AMA-1 also show that in Malabo, South East, South West, North West and Other regions, there might have been a change in transmission which happened at least 2 to 3 years before the survey when transmission dropped up to 30% in certain regions. In the study carried out by Cook and colleagues [133], a drop of transmission intensity was also recorded in the North East region. In addition, their reported magnitude for the drop in transmission intensity was around 83% when measured with SCR, 79% when measured with parasite rate and 65% when measured with under 5 mortality rate. Note that in their study, a change of transmission was detected in all regions except the North West, while my results suggest there might have been a change but the uncertainty around its date is rather high.

5.4.3.2 Assessing age-dependent exposure in Cambodia

In Cambodia, malaria exposure is thought to depend on age since adults working in the forest are at high risk through exposure to vector populations residing in the forest fringes. Therefore, to test whether age-dependent factors can be estimated from these data, I made the assumption that when children reach the age to work in the forest their exposure level changes. This model was fitted to a subset of data collected in Cambodia (previously presented in Chapter 2), with a domain known to be less exposed (Chris Drakeley – personal communication) disregarded.

I assumed that exposure was constant in the different areas and hence that only age determines a change in transmission. I made the assumption that children start to work in the forest between the age of 9 and 18 (used as uniform prior for ω) and children have lower exposure than adults, therefore I used a uniform prior on [0,1] for the scaling factor β. I performed 180,000 iterations for the MCMC algorithm with a burn-in period of 50,000 steps. The output was recorded every 100 iterations to generate a sample from the posterior distribution of size 1,300. The parameter estimates (See Table 5.11) were comparable to those in Chapter 3. However, an effect of age on transmission was recorded around 13 years old. Additionally, it was found that children are 10% less exposed than adults (See Figure 5.22). As suggested before, the change in exposure can be due to children starting to work in the forest or benefiting less from bednets.

Table 5.11: Parameter estimates for model using age effect in Cambodia

Parameter	Estimates Median (95% CrI)
Exposure for adults λ	5.4 (5.3-5.5)
Scaling factor for exposure for children β	0.9 (0.87-0.92)
Age of change ω	13 (9.7-15)
Maximum boost size a	0.6 (0.56-0.67)
Slope of dependence b	0.48 (0.46-0.5)
Standard deviation S	0.29 (0.26-0.33)
Maximum boost size for "seronegative" η	0.0029 (0.0014-0.0084)

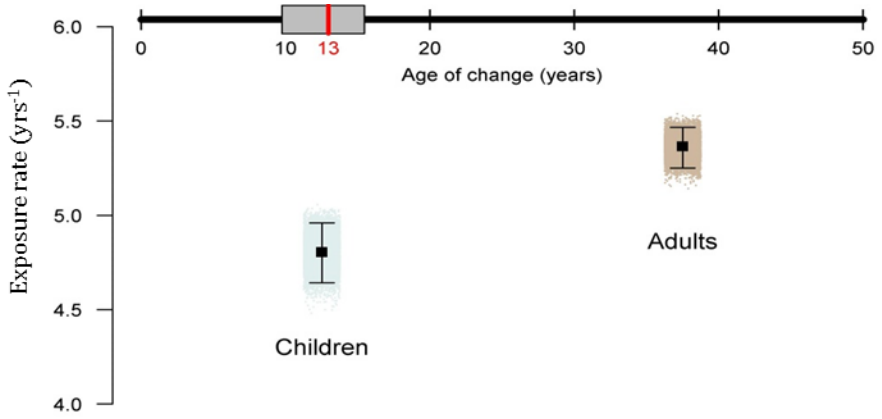


Figure 5.22: Exposure rate and age of change. Median estimates of exposure and their associated *posterior* 95% credible interval are presented for children and adults. The median age of change is presented in red and the grey box represents the *posterior* 95% credible interval.

5.5 Discussion

Several studies [95, 197, 243] have observed an association between blood stage antibodies and exposure to malaria. Different countries have different transmission intensities and this is reflected in antibody levels. The results presented here demonstrate that the density model developed in the previous chapters can be used to estimate malaria exposure across a range of different transmission settings.

My fitted model was able to reproduce anti MSP-1 and anti-AMA-1 antibody levels in all the studied countries. Additionally, no major differences in quality of model fits were observed between the different countries. In The Gambia, for instance, the fitted model underestimated the anti-MSP-1 antibody levels. However, the predictions for anti- AMA-1 were qualitatively good, suggesting that the poor fit is not due to the country or the endemicity level. Similarly, despite the fact that anti- AMA-1 antibody levels are systematically higher, no substantial differences in the quality of model fits were observed between AMA-1 and MSP-1 antigens. As noted, the model fit is better for anti-AMA-1 antibodies in The Gambia while in Bioko, the fit is worse with anti- AMA-1 antibodies, suggesting that the antibody type is not responsible for the marginal lack of fit. However, in Bioko and Uganda, the lack of fit for anti- AMA-1 antibody levels is very likely due to the decrease in antibody levels in adults, as the model cannot reproduce this phenomenon. Indeed, the model assumes that the exposure rate and the acquisition and decay of antibodies are constant over age (and time), which might not be valid. An age-dependent rate of exposure might be less likely as an explanation than age-dependency in the acquisition or decay as we would also expect to observe it with anti- MSP-1 antibodies. The available data does not allow us to identify the reason for this decrease of anti-AMA-1 antibodies over the age of the surveyed population in Bioko and Uganda.

My results demonstrate that the density models can be used to estimate transmission intensity from anti-MSP-1 antibody levels. A high degree of correlation was observed between the exposure rates estimated using the density model and seroconversion rates, typically used to estimate transmission intensity from serological data. The results showed a bimodal distribution for the results using AMA-1 antibodies. In other words, the results suggest that individuals have their current antibody level either because they have been often exposed but had a small boost each time or because they have less often been exposed but had a bigger boost each time. The available data on AMA-1 did not allow me to distinguish between two of those modes. However, either of the obtained modes was correlated with transmission intensity. This re-emphasizes that the results of the model should be regarded as relative results rather than absolute ones and in that case the model applied to AMA-1 data can also be used to compare transmission intensities.

Additionally, it is important to note that I found that the different antigens ranked Uganda and The Gambia in reverse order of transmission intensity, and the same result was obtained with SCR using catalytic

models. If antibody distribution is mainly determined by transmission intensity, we would not expect to observe a different result for different antibodies. Thus there must be some other reason that explains a low antibody level in The Gambia. Some unlikely explanations, such as antigenic variation or variation in laboratory assays could simply be at the origin of these results. However, a more likely explanation could be an immuno-regulation. A saturation of the immune system in high transmission setting, such as Uganda, could stop the production of antibodies. The reasons for such phenomenon remain unclear, it could be due to immune-suppression caused by malaria infection or the control of blood stage antigen might happen through other mechanisms and would control the infection without anti-merozoite responses. Individuals would therefore naturally lose their current antibodies and this would explain why anti- MSP-1 antibody level measurements are lower in Uganda than in The Gambia (where the transmission is lower). The same phenomenon was also observed in a more recent survey in Uganda (Chris Drakeley – personal communication). Another explanation could therefore be that there might be some cross-reactivity of the antibodies during co-infection with *Plasmodium malariae* and *Plasmodium ovale,* also reported in Uganda. Despite its low anti- MSP-1 level, the analysis using both antibody types correctly identifies Uganda as the country with the highest transmission level. As a result, using both antigens in a single model might reduce the bias introduced by any antibody-specific effects.

Furthermore, there is a growing interest in measuring simultaneously antibodies against multiple malarial antigens (rather than single antigens) with multiplex assays [244–246]. Such methods allow a faster assessment of the reactivity of antibodies to a panel of antigens and may provide novel information on the immune response that might not be identifiable when antigens are considered individually [247]. As a result, with its capability to analyse simultaneously multiple antigens, the density model might represent a useful tool to measure of transmission more accurately with the results of multiplex assays.

Here again, I have shown that the precision associated with exposure rates was better using density models in comparison with catalytic models. The full information contained in the continuous titre has the property to better inform the estimate of transmission in comparison with a binary measurement such as the seroprevalence used to derive SCR. The model here was not able to distinguish a single mode for exposure rate and therefore the precision associated with its estimate could not be assessed. However, In order to assess malaria transmission in low endemicity settings, AMA-1 antigen was shown to present some advantages (as it is highly immunogenic) while in high endemicity setting MSP-1 was shown to work better (as it is less immunogenic and provides better precision for estimates of exposure) [97]. Further work would therefore need to be conducted if the use of AMA-1 antibody increases. Indeed, a better understanding of the reason of the incapability of singularly identifying the exposure rate with AMA-1 would be essential.

The exact function of antibodies to merozoite surface antigens are not perfectly understood, but the role of anti- MSP-1 and anti- AMA-1 antigens have been characterised [248, 249] and their acquisition is therefore not expected to be similar. However, the results here demonstrated that, whether antibodies were analysed separately or simultaneously, the difference of acquisition of antibodies was mainly different due to the variability between individuals. Fitting the model to all countries simultaneously has allowed me to explore a wide range of antibody levels. However, by doing so, I have assumed no differences in acquisition of antibodies for all the individuals, ignoring ethnic and other differences that might influence the antibody response for individuals from each country. Somalia, The Gambia, Bioko and Uganda are countries spreading from West to East Africa and therefore there might be a high level of genetic variability between these countries that would affect antibody response [250]. Therefore conclusions on the biological parameters might not be completely valid.

Concerning the size of the average boost for "seronegative" individuals, estimates remain very low, indicating that "seronegative" individuals tend to stay below the limit of detection once they get exposed to infection. This result might not be a complete representation of reality as the model over-predicts the number of "seronegative" individuals in each country (with the exception of Somalia for the analyses with AMA-1). Such small values are very likely to be influenced by the number of "seronegative" individuals that never get exposed. In Chapter 3, I considered a model of the population of "seronegative" individuals as a mixture of individuals who never get exposed and those who are exposed and developed an antibody response. At the time, a more parsimonious model was chosen as model predictions were very similar between model results, in Cambodia, where endemicity is low. However, in areas where transmission intensity is much higher, such a model might present some advantages. Indeed, non-exposed populations would be accounted for in settings where the current model assumes everyone is highly exposed. Despite the high level of transmission in some countries, some individuals might never have been exposed, due to a number of different reasons including ecological and environmental factors such as variations in attractiveness to mosquitoes [251], damaged or unused bed nets [252]. Therefore taking into consideration some heterogeneity of exposure in high endemicity settings and accounting for non-exposed populations might be a better representation of what is really happening.

I also extended the original density model to take into consideration some specific characteristics of transmission. Heterogeneity in malaria transmission is relatively common. I therefore extended the model to account for heterogeneity in exposure which could indifferently be attributed to different settings such as spatial and temporal heterogeneity in mosquitoes distribution [253], heterogeneity due to age [78] or to vector control interventions [237]. The estimation method applied to 50 simulated cross sectional datasets showed that the specific parameters related to exposure, i.e. exposure rate, scaling factor for exposure rate

and proportion of the population exposed at a specific level can be correctly estimated. By setting the scaling factor to zero, this model would account for a proportion of the population that would never be exposed to malaria infection. This differs from the model presented in Chapter 3 as a specific proportion of the population, regardless of an individual's current antibody level, would not be exposed, and not only the "seronegative" individuals, as was assumed there.

Spatial heterogeneity can be modelled by applying the model to different regions, whereas heterogeneity due to an age effect can be represented by assuming a difference in exposure before and after a certain age. The model was therefore extended to account for an age effect on exposure. Simulation studies have shown that the parameters related to exposure, i.e. exposure rate, scaling factors and the age of change, can be correctly estimated. I fitted the extended model incorporating an age effect to a subset of the data from Cambodia. An age effect was detected around 13 years old. Children are expected to be 10% less exposed than adults. The results therefore imply that when children start working (in the forest), they increase their exposure level. Indeed, Cambodia is known to experience occupational malaria, i.e. individuals working in the forest represent a high risk group [254]. The results of the model have therefore corroborated with this hypothesis.

In the context of elimination of malaria and with the increasing number of malaria control interventions, it is of interest to assess any change in malaria transmission intensity. For this purpose, I have extended the model and tested it with simulated datasets. Results show that we can correctly estimate the parameters related to exposure, i.e. the exposure rate before the change in transmission, the scaling factor assessing the change and the time since the change. As an example, I applied the model to the data from Bioko. My results identify some change in transmission, in particular using AMA-1 antigen. These results corroborate with what was previously found using seroconversion rates for the same area [133]. However, in their study Cook and colleagues identified all the regions, with the exception of North West, as having experienced changes in transmission. The density model only clearly identified the North East as having a significant change in transmission. The other regions had a high uncertainty around the date of the change. This latter parameter could potentially be overestimated due to the babies born after the interventions who lose their antibodies relatively quickly. Indeed, the model had assumed a constant loss of antibodies that might not be a valid assumption. Additionally, the impact of the interventions measured by the ratio of seroprevalence in their study also ranked the North East region as the region that experienced the most substantial drop in transmission. The discrepancies in the magnitude of the drop of transmission intensity measured with SCR and exposure rate (using density model) can easily be interpreted as a difference in methodology. Indeed, SCR measures an incidence, the number of individuals that become seropositive i.e. that increase their antibody level above a cut-off value. In contrast, the exposure rate measures the

number of individuals that get their antibody level boosted and is therefore driven by the magnitude of the change in antibody distribution. As illustrated schematically in Figure 5.23, SCR would characterise the change presented in the figure as a substantial drop in transmission intensity while the extended density model might not even detect the small change in transmission.

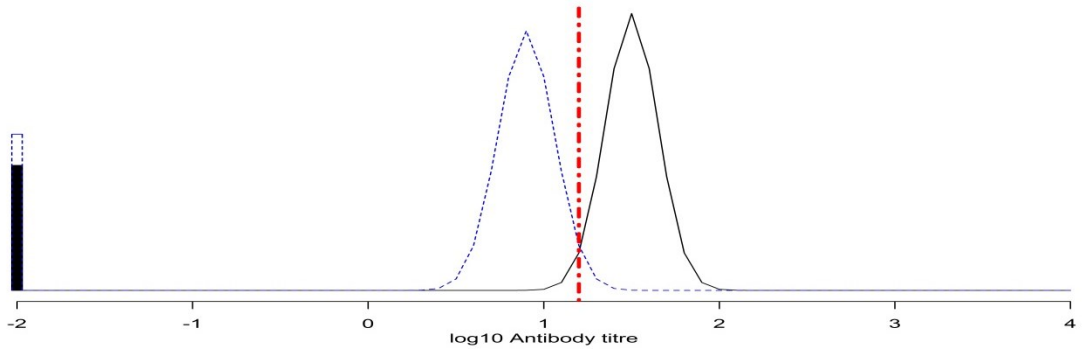


Figure 5.23: Schematic representation of antibody distribution during a decrease in transmission intensity. Antibody distribution before (black solid line) and after (blue dashed line) the change are discretised by the cutoff value (red vertical dashed line) that defines seroprevalence.

Historically, changes in transmission intensity have been assessed with multiple cross-sectional surveys or longitudinal surveys. Recently, there has been an increase in the use of a single survey that can detect changes in transmission using profile likelihood methods [98, 133, 135]. The results have shown that the density model can also be used to detect changes in transmission intensity from a single cross-sectional survey. Despite being computationally intensive, this method might more accurately capture the magnitude of the change in transmission. Additionally, these methods are more likely to detect a sudden change in transmission, with successful intense implemented interventions, and not if there has been a slow and continuous decrease in transmission over time.

5.6 Conclusion

In this chapter, I have applied the density model to four countries with very diverse endemicity settings and I was able to reproduce antibody levels for both anti-MSP-1 and anti- AMA-1 antibodies and to measure malaria transmission intensity. I then extended the original density model to explore three different scenarios: heterogeneity in exposure, an effect of age on exposure and a change in transmission intensity. The latter scenario was applied to data from Cambodia and Bioko, and identified respectively an age effect and some changes in transmission intensity. It would be beneficial to use longitudinal data to inform biological parameters so that the model can adequately assess changes in exposure. Some further validation might be required for the extended models but such methods have good potential for multiple applications in the context of malaria elimination.

Chapter 6: Discussion

In the current context of malaria control and elimination, measuring malaria transmission intensity is a key element for monitoring changes in transmission, planning and assessing the impact of the interventions. Recent advances in malaria control have highlighted the importance of serology as a method of quantifying malaria transmission, in particular for areas of low endemicity. In this thesis I have developed a new method for estimating malaria transmission intensity from antibody levels. This chapter summarises the key results, indicates the implications of the research, highlights the limitations of the method and finally points out some future directions for research in this area.

6.1 Summary of findings

In Chapter 2, I presented a method extended from the classic catalytic model as a "proof of concept" to show the association between levels of antibody and malaria exposure. I discretised the population into multiple arbitrary compartments according to the individual's level of antibodies. This discretised density model permitted the separation of the effect of exposure and antibody kinetics on the immune response. The model was fitted to data from Cambodia and showed a high correlation between seroconversion rates and exposure rates (from the density model) for both MSP-1 and AMA-1 antigens. Given the simplicity of the model only qualitative conclusions indicating a density- dependency of acquisition of antibodies could be established.

In Chapter 3, I further extended the model presented in Chapter 2 to include a larger number of compartments in order to approximate a continuous density of antibodies and therefore a continuous model. A variety of different models were tested to mimic individual's acquisition of antibodies and account for heterogeneity in the "seronegative" population. When these models were applied to Cambodian data, results indicated that age specific antibody titres, collected from cross-sectional surveys, could be reproduced using a density model that assumes a constant exposure rate, a constant loss of antibodies and density-dependent acquisition of antibodies. The model retained for the rest of the thesis assumes that individuals produce exponentially decreasing amounts of antibodies as their current antibody titre increase and the size of this boost is log-normally distributed. Additionally, "seronegative" individuals are assumed to have a single specific antibody response. Using this model, estimated malaria transmission intensity was correlated with distance to the forest in Cambodia, which can be thought of as a proxy for exposure in this setting.

In Cambodia the level of endemicity is very low, which might limit the applicability of the density model. Moreover, malariometric indices were not available to allow a direct comparison with transmission. To validate the density model as a tool for measuring malaria transmission intensity, in Chapter 4 I presented analyses of previously published data from Tanzania where an association between serological data and transmission intensity has already been identified across a wider range of endemicity settings [95]. The model was also validated by informing the biological parameters with informative priors. To provide an additional comparison, mixture methods were used to further developed catalytic models so that they do not require cut-off values and take into consideration misclassification of seropositive status of the individuals. The estimates obtained with the density model were consistent with estimates of seroconversion rates, obtained using European controls or mixture models, and provided better precision.

In Chapter 5, I further tested the density model by applying it to data from four countries with diverse endemicity settings. The model was fitted to data from Somalia and Uganda where transmission is respectively extremely low and high, from The Gambia, where malaria transmission is seasonal, and from Bioko Island which has a successful control programme. The model was able to reproduce antibody levels for both *P. falciparum* anti-MSP-1 and anti- AMA-1 antibodies and to measure malaria transmission intensity using anti-MSP-1 antibodies. The use of anti- AMA-1 antibodies to singularly determine malaria exposure would require further work. Estimates of exposure were indeed correlated with estimates of SCR. However, the data showed similar distributions of anti- MSP-1 antibodies for The Gambia and Uganda, resulting in similar estimates of exposure rates for both countries, whereas Uganda is known to experience a much higher intensity of transmission [230]. When both antigens were considered simultaneously in a single model, the estimates of exposure provided the correct ranking for estimates of transmission intensity. The results suggested that the combination of these two antigens offers higher precision for measuring malaria exposure and removed antibody specific effects. Finally, I further developed the density model to assess variations in exposure, to investigate an age dependent exposure (that could be associated with occupational malaria) and to assess the utility of the model in capturing changes in transmission following scaling up of malaria interventions. Results from simulation studies showed that the true parameters from these extended models could be retrieved. The application of an extension of the density model on data from Bioko captured a temporal change in transmission, confirming what was found in a previous study [133]. Additionally, the extension of the density model that accounts for an age effect on transmission was tested on data from Cambodia, presented in Chapter 2, and detected a significant difference in exposure between children and adults around the age of 13. Further work and validation of the extensions of the density model might however be required.

In summary, throughout the thesis, I have developed a density model capable of reproducing antibody levels reported in cross sectional surveys according to the intensity of parasite exposure. These model derived estimates of exposure appear to be a valid tool for measuring malaria transmission intensity in a wide range of transmission settings.

Overall interpretation of the findings

The model assumes that individuals lose their antibodies at a constant rate ρ, fixed during the thesis to 0.7 years^{-1} (half-life of 1 year), regardless of the antibody type. The model also makes the assumption that individuals produce exponentially decreasing amount of antibodies as their current antibody titre increase. Additionally, "seronegative" individuals are assumed to have a single specific antibody response η. As malaria transmission increases (measured with exposure rate λ), the number of "seronegative" individuals tend to decrease while the antibody distribution for "seropositive" is shifted towards higher antibody values (Figure 6.1).

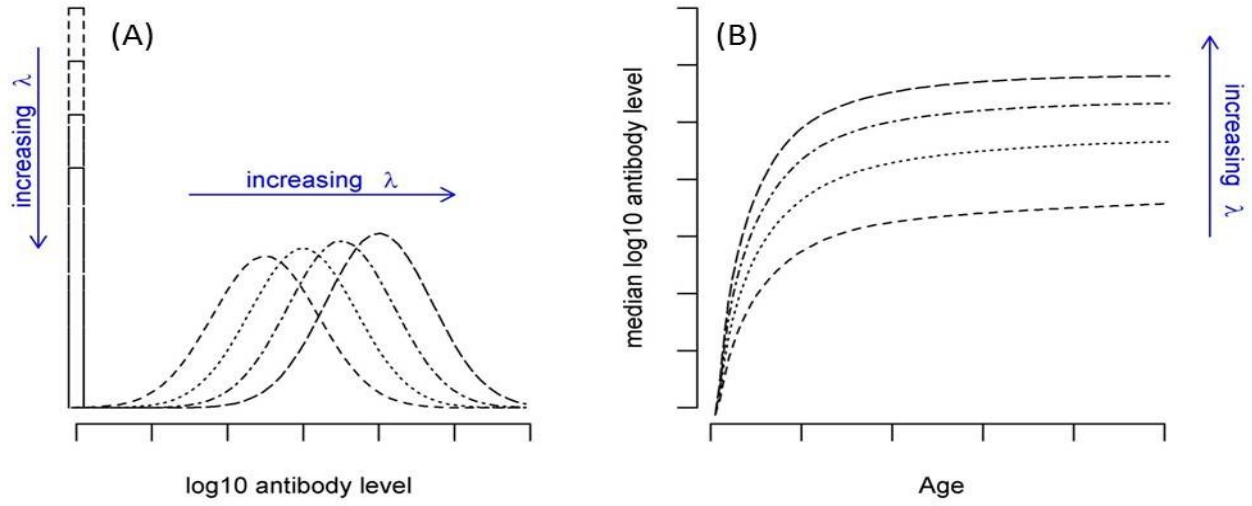


Figure 6.1: Effect of increasing malaria transmission intensity (λ) on overall antibody distribution (A) and age specific median antibody level (B).

In most settings tested the number of "seronegative" individuals appears to be an over-estimate. This result is due to low estimates of the parameter mainly driving the number of seronegative individuals, η, corresponding to the size of the boost for individuals who do not have circulating antibodies. These estimates might not provide an accurate description of the actual biological mechanism as this would infer a substantial difference of the size of the boost for "seronegative" individuals and individuals who have antibody levels just above the detection limit, corresponding to parameter a. A plausible explanation for this phenomenon might be the fact that the density model does not take into consideration the non-exposed individuals. Indeed the density model makes the assumption that all individuals are exposed while

individuals with no circulating antibodies ("seronegative") are very likely to be individuals who have not been exposed (rather than have acquired and lost antibodies). Kinyanjui and colleagues [200] had already suggested a similar hypothesis for clinical outcomes implying that individuals who do not get clinical episodes are very likely to be individuals who have not been exposed (rather than have mounted clinical immunity against malaria). However, it could be argued that in areas of moderate to high transmission intensity, there might be few individuals who are truly unexposed. An alternative explanation for individuals to be "seronegative" might be that there is another immune mechanism that reduces the number of blood stage parasites and therefore individuals do not mount an antibody response. In the same way that they suggest removing unexposed individuals, in areas of high endemicity (where submicroscopic infections are less frequent) estimates for the kinetics of antibodies for "seronegative" individuals in the density model could benefit from removing individuals who are not carrying parasites or did not experience a malaria episode (denoting exposure). The extended density model that considers heterogeneity in transmission could also be an alternative method to account for non-exposed individuals. Indeed, the density models I have further explored to assess variations in transmission (heterogeneity in transmission, effect of age on exposure and temporal change in transmission) are useful tools to measure malaria transmission intensity in more specific settings.

The density model developed during the thesis is a novel method for measuring malaria transmission intensity using serological data. To my knowledge, this is the first method that reproduces antibody levels from cross-sectional survey with mechanistic models based on exposure levels. Instead, antibody levels are often reproduced using longitudinal studies and statistical models [255], with no consideration of the biological mechanisms. The longevity of antibodies has been the focus of considerable research to better understand immune mechanisms. As a result, the decay of antibodies has been studied using longitudinal data [132, 192], whereas the acquisition of antibodies remains understudied mainly due to the logistic and ethical difficulties in pinpointing the time of an infection and measuring its associated boost of antibodies. Measuring transmission intensity from serological data, typically involves the use of seroprevalence data and rarely the full information contained in antibody titre. Moreover, I have shown that a better precision in estimates is obtained when continuous antibody titre are considered for exposure rate in comparison with estimates for SCR. Recently, Bretscher and colleagues developed a model to measure SCRs from individual-level longitudinal data on antibody titres [137]. Their method, as well as other mixture model approaches, use antibody titres but ignore the acquisition of antibody levels after seroconversion. Additionally, their method requires data from longitudinal studies which are logistically and ethically more demanding than cross-sectional studies and might not be useful for routine surveillance. I have also shown the possibility to extend the density model to account for more complex scenarios of exposure, which was

already possible with catalytic models [133, 134]. This makes the density model a useful tool for measuring malaria transmission intensity that can be used, for instance, to assess the impact of the interventions or to monitor elimination during routine surveillance.

6.2 Implications of the findings

In the current context of malaria elimination, there is a need to develop and use robust, accurate and precise tools for measuring malaria transmission intensity. Given the advantages that serological methods offer for measuring malaria transmission intensity and in particular in areas where transmission intensity has fallen to low levels, serological methods have attracted much attention. The renewed interest and importance of serological methods in the context of malaria elimination activities has already been discussed in detail in Chapter 1. As transmission intensity decreases, parasitological and entomological methods become less reliable and more logistically difficult. This makes measurements of anti-malarial antibodies, which integrate malaria exposure over time, a more preferred method for estimating exposure. Methods outlined in the PhD have the advantage over existing models in that they do not require an (arbitrary) cut-off value to differentiate between seropositive and seronegative individuals. As a result, these methods, as opposed to catalytic models that require European data for control [95, 97], do not need external data to be analysed and in high endemicity settings, the issue of differentiation between seropositive and seronegative individuals when mixture models are used (Chris Drakeley – personal communication), are not problematic with a density model.

In areas where malaria intensity is low / inexistent, collecting large number of blood samples becomes logistically and ethically challenging. Serology has the advantage of a less invasive alternative method of collecting saliva samples to determine antibody levels [256]. Serology could also be used to evaluate transmission blocking interventions which will shortly be entering phase II and III trials [257]. It is thought these will work best in areas of low transmission [258, 259], where standard metrics may not be sensitive enough and serology would be preferred. The number of applications and needs for serology is indeed increasing, in particular as malaria declines, therefore the improved precision of transmission intensity provided by the density model will complement the current techniques already in place. Serology can be used to define malaria endemicity and detect past exposure when parasite rates are zero [27] and as a consequence has an important role in surveillance, also considered to be an intervention to achieve malaria elimination [11]. The density models outlined in this thesis will complement existing serological methods which have already been used to confirm elimination of malaria, assess the impact of interventions and monitor changes in transmission intensity [8, 9, 133, 135].

The current methods applied to interrupt local transmission include killing the parasite with appropriate treatment and vector elimination activities [27]. If these interventions are successful, the absence of interaction between parasite/vectors and humans imply that individuals will lose their current antibody levels and young children will not produce anti-malarial antibodies [197]. The density models will be able to detect such phenomenon, indicating the effectiveness of the interventions, either during multiple cross sectional surveys (before and after the interventions have been put in place) or using an extended version that account for changes in transmission. An absence of significant decrease of antibody level where interventions have been put in place might indicate that the interventions have not worked.

In this thesis, I have developed a model that predicts antibody levels for a cross sectional survey based on exposure and kinetics of antibodies. Antibody responses to blood stage antigens induce protective immune response to natural infection and are associated with clinical protection. Antibodies are indeed crucial components of protective immune response. However, the strength of the association between antibodies and protection against malaria remains poorly understood [200]. Some studies have shown that high density of antibodies and notably against MSP-1 and AMA-1 antigens could play a major role in the immune responses [260, 261]. If further research validates the association between the antibody levels and the level of protection, then the density model might represent a useful tool for predicting antibody levels at a population-level and could therefore infer on the potential for protection in a population.

6.3 Limitations of the density model

The use of mathematical models typically involves a simplification of the actual phenomenon to only include characteristics of interest to answer specific scientific questions. In the context of the development of a model to reproduce antibody levels from cross-sectional surveys based on exposure levels and kinetics of antibodies, a number of biological and epidemiological assumptions have been made.

Simplification of complex biological processes

The process of loss of individual's antibodies assumes a constant rate of decay of antibodies fixed to 0.7 years^{-1} corresponding to a half-life around 360 days. This value, which originated from a longitudinal study [192], made the assumption that antibodies detected for serological studies produced by long lived plasma cells. However, short lived antibody responses to merozoite antigens are mostly observed [116, 117, 197]. Therefore by ignoring the short lived antibody response, the model might over-estimate antibody levels when estimates of the rate of decay are based on long lived antibodies. Additionally, the model assumes a constant rate of decay of antibodies; the reality is somehow different with the persistence of antibodies

differing with age [243, 262]. In the absence of reinfection, children have a more rapid decay in antibody levels in comparison with adults [225, 263]. As a result, the antibody levels for young children might be overestimated while they might be underestimated for adults. A similar assumption of age-independent seroreversion rate is, nevertheless, usually made with classic catalytic models. It is important to note that rate of seroreversion from the catalytic model and rate of decay of antibodies measure different quantities. Indeed, the seroreversion rate assumes the population is discretised between seropositive and seronegative and it therefore measures the rate at which individuals revert back to seronegative, i.e. their antibody level drops below the cut-off value. In the density model the rate of decay of antibodies corresponds to the decay of antibody levels experienced by individuals.

In the density model, the acquisition of antibodies is based on the simple assumptions that, as individuals increase their antibody levels, the amount of antibodies they produce upon infection decreases. In reality a number of immune mechanisms, which are not all modelled, are triggered by the presence of the parasite in humans, including pre-erythrocytic immune responses and non humoral immune responses for example. The immunological processes modelled here are assumed to be constant with time and age. However, there is evidence that the development of the antibody response is age-dependent due to the maturation of the immune system with age [264]. Additionally, the model does not account for current infections, which are generally associated with higher antibody levels [199]. Therefore, any subsequent infection is considered to be the same [265], ignoring the theory that superinfections might significantly boost the antibody response [266]. As the result, the levels of antibodies predicted by the density model might be underestimated in areas where superinfections are more likely (high endemicity).

A number of other biological assumptions have been made to simplify a complex reality. For instance, infants are known to be protected against malaria during at least the first few months of life due to the transfer of maternal antibodies [62, 267], which was ignored in the model. Also, there might be some biological differences in the acquisition and loss of antibodies between individuals or ethnic communities due to genetic makeup. Some studies have shown that unidentified genes contribute to variations in individual's inherent susceptibility to malaria [200, 268] and individuals also present differences in attractiveness to the mosquitoes [269]. Therefore the assumption of identical antibody kinetics between communities might not be the most accurate representation of reality. The model accounts for some variations between individuals with a lognormal distribution of the size of the boost though further work will be required to determine whether this is sufficient.

Epidemiological limitations

Variations in individual's antibody levels are mainly attributed to exposure. However, malaria exposure is rarely homogeneous. Heterogeneity in malaria transmission contributes to the challenges to achieve successful malaria elimination. It is believed that, similarly to other diseases [107], 20% of the endemic populations bear 80% of the transmission and more specifically more than 90% for malaria infections [270]. The models presented here were used to explore the heterogeneity of the seronegative population and an extended version of the density model also accounted for heterogeneity in transmission that could be attributed to identifiable subpopulations (according to, for example, spatial clusters or drug uptake) or temporal changes. However, heterogeneity in exposure can be attributed to unmeasurable determinants including non-random biting, travelling history, ethnicity or socio-economic differences which haven't been accounted for by the models presented in this thesis.

The age of an individual will influence their antibody level. In addition to its effect on maturation of the immune system, age also affects the force of infection [271]. However it is often difficult to distinguish between the actual role of age, the length of exposure and the difference in body size resulting in a larger exposed body surface [78]. The density model developed here assumes a constant rate of exposure which is likely to be an oversimplification. An age-dependent exposure rate was assessed with an extended version of the density model though only a non-continuous difference before and after a specific age was considered. A more realistic representation of reality would set the exposure rate as a function dependent on age (and time).

The increasing progress of malaria vaccines, notably the pre-erythrocytic vaccine RTS,S [272] might have some implications for the use of a model that considers levels of antibodies to determine malaria transmission intensity. In contrast to naturally acquired immunity that predominantly targets blood stage infections, the RTS,S vaccine induces pre-erythrocytic immune response, which in turn will influence blood stage immune responses. The observed reduction of levels of blood stage immune markers might be due to reduced number of merozoites invading red blood cells (due to pre-erythrocytic immunity induced by the vaccine). Indeed, this vaccine does not directly target blood-stage antigens and therefore does not directly trigger blood-stage antibody response[273]. The model would therefore correctly associate reduction in antibody levels to reduction of blood stage exposure. However, this would not be the case for blood stage vaccines, which would enhance the production of anti- merozoite antibodies, without necessarily reflecting the level of blood stage exposure. The variations of antibody levels due to the impact of other interventions such as vector control or treatment should correctly reflect variations in blood stage exposure.

Limitations for the use of serology to measure malaria transmission intensity

As for other methods that use serological data to infer the force of infection [95, 97, 139, 140, 274, 275], the density model assumes that the exposure rate is a measure for the force of infection, i.e. the rate at which an infectious mosquito bite successfully causes a blood stage infection. Seroconversion and exposure rates, determined respectively by catalytic and density models, have been used to indicate transmission intensity. However, it is important to note that serological markers strictly provide information on blood stage exposure but are not a direct measurement of mosquitoes exposure. These only approximate a mosquitos' exposure as a number of factors including pre-erythrocytic immune response could prevent injected sporozoites developing into blood stage infections [76]. Using the density model with pre-erythrocytic antibody data would provide a more accurate estimate of malaria exposure as it reflects exposure to sporozoites [120]. However, as pre-erythrocytic immune responses do not last as long as blood stage immune responses, this might be difficult to achieve.

As anti-malarial antibodies are highly determined by exposure then most people in areas of high endemicity will have high antibody levels. This makes it difficult to distinguish between seropositive and seronegative individuals and therefore limits the applicability of catalytic models. Despite the fact that antibody levels saturate at high values, the density model is capable of assessing exposure rate provided that enough information is collected for individuals who experience variations in antibody levels, i.e. young children. Therefore the design of the study in areas of high endemicity is of great importance in order to assess malaria transmission intensity accurately.

The density model was developed to predict the levels of antibodies produced upon exposure. However, the avidity of the antibodies is also of importance. Indeed, it is believed that antibody avidity increases with repeated exposure [276]. If this is the case, with increasing exposure fewer antibodies might be produced during each infection but would be of better quality. The antibody levels would therefore underestimate the level of exposure. However, the understanding of the avidity of the antibodies remains poorly understood with some studies finding no evidence of an association between antibody avidity and exposure [232].

Serological methods have been shown to represent useful tools for assessing malaria transmission in the context of experimental surveys [95]. However, some operational limitations remain for serology and therefore for the density model to be routinely used. The sample collection is easy but most malaria programs are currently not equipped to analyse serological samples [27]. An additional limitation comes from the computational complexity of the method. Despite having shown advantages of the density model over classic catalytic models, the implementation of the method using the full information in the antibody

levels is computationally intensive, which may limit its wider utility. However, this can to some extend be overcome given the wider availabilities of statistical packages that can be used to perform such analyses and through sharing of code. To this end the R code used for this analysis is available on request.

6.4 Further evaluations and potential applications

The significance of the developed density model will be determined by its usefulness in measuring malaria transmission intensity to help guide the prevention, control and surveillance of malaria. This thesis has shown that there is a wide scope for further developing models that could account for (1) heterogeneity of exposure; (2) changes in transmission intensity and (3) the use of multiple antigens as markers of exposure. Further studies including longitudinal data could further improve our understanding of antibody kinetics and help to better parameterise the model. Additionally the association between exposure rates could also be directly compared with measures such as the Entomological Inoculation Rate (EIR) to establish the relationship between mosquito and blood stage antigen exposures. Additionally, it would be important to calculate the sample size required for the model to provide valid estimates and to detect changes in transmission intensity of different magnitudes. The continuous model was mainly applied for *P. falciparum* antibodies. Despite some difference in the life cycle of the parasite, it would be interesting to further explore the applicability of the density model for *P. vivax.*

The density model was developed for anti-malarial antibodies though there is scope for applying this methodology to other diseases that present similar epidemiological and biological characteristics. Indeed, the model could be applied to infections with a relatively constant transmission rate and decay of immunity, antibody levels mainly depending on exposure level and antibody kinetics, and with a systematic acquisition of antibodies upon infection that last long enough to be assessed in cross-sectional survey.

References

1. Hume JCC, Lyons EJ, Day KP: **Human migration, mosquitoes and the evolution of Plasmodium falciparum.** *Trends Parasitol* 2003, **19**:144–149.

2. Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su X: **Early origin and recent expansion of Plasmodium falciparum.** *Science* 2003, **300**:318–21.

3. Coluzzi M, Sabatini A, Petrarca V, Di Deco MA: **Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex.** *Trans R Soc Trop Med Hyg* 1979, **73**:483–497.

4. Ross R: *Studies on Malaria*. London: Murray; 1928.

5. Ross R: *The Prevention of Malaria*. 2nd edition. London: Murray; 1911.

6. Nájera JA, González-Silva M, Alonso PL: **Some lessons for the future from the Global Malaria Eradication Programme (1955-1969).** *PLoS Med* 2011, **8**:e1000412.

7. Drakeley C, Cook J: *Chapter 5. Potential Contribution of Sero-Epidemiological Analysis for Monitoring Malaria Control and Elimination: Historical and Current Perspectives.* 1st edition. *Volume 69*. Elsevier Ltd.; 2009:299–352.

8. Bruce-Chwatt LJ, Draper CC, Konfortion P, Drapper CC: **Sero-epidemiological evidence of eradication of malaria from Mauritius**. *Lancet* 1973, **2**:547–51.

9. Bruce-Chwatt LJ, Draper CC, Avramidis D, Kazandzoglou O: **Sero-epidemiological surveillance of disappearing malaria in Greece.** *J Trop Med Hyg* 1975, **78**:194–200.

10. Roberts L, Enserink M: **Malaria. Did they really say ... eradication?** *Science* 2007, **318**:1544–5.

11. The malERA Consultative group on Monitoring E and S: **A research agenda for malaria eradication: monitoring, evaluation, and surveillance.** *PLoS Med* 2011, **8**:e1000400.

12. Guerra CA, Snow RW, Hay SI: **Mapping the global extent of malaria in 2005.** *Trends Parasitol* 2006, **22**:353–8.

13. Hay SI, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, Manh BH, Elyazar IRF, Brooker S, Smith DL, Moyeed RA, Snow RW: **A World Malaria Map : Plasmodium falciparum Endemicity in 2007**. *PLoS Med* 2009, **6**.

14. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, Tatem AJ, Hay SI: **A new world malaria map: Plasmodium falciparum endemicity in 2010.** *Malar J* 2011, **10**:378.

15. World Health Organization: **World Malaria Report**. 2012.

16. Roca-Feltrer A, Carneiro I, Armstrong Schellenberg JRM: **Estimates of the burden of malaria morbidity in Africa in children under the age of 5 years.** *Trop Med Int Health* 2008, **13**:771–83.

17. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, Hay SI, Snow RW: **The limits and intensity of Plasmodium falciparum transmission: implications for malaria control and elimination worldwide.** *PLoS Med* 2008, **5**:e38.

18. Mendis K, Sina BJ, Marchesini P, Carter R: **The neglected burden of Plasmodium vivax malaria**. *Am J Trop Med Hyg* 2001, **64**(1-2 Suppl):97–106.

19. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM: **Vivax Malaria: Neglected and Not Benign**. *Am J Trop Med Hyg* 2007, **77**:79–87.

20. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperley WH, Kabaria CW, Tatem AJ, Manh BH, Elyazar IRF, Baird JK, Snow RW, Hay SI: **The international limits and population at risk of Plasmodium vivax transmission in 2009.** *PLoS Negl Trop Dis* 2010, **4**:e774.

21. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, del Portillo HA: **Key gaps in the knowledge of Plasmodium vivax, a neglected human malaria parasite**. *Lancet Infect Dis* 2009, **9**:555–566.

22. Clinton Health Access Initiative: *The Health and Economic Benefits of Sustaining Control Measures*. 2011.

23. Greenwood B: **Can malaria be eliminated?** *Trans R Soc Trop Med Hyg* 2009, **103 Suppl** :S2–5.

24. Das P, Horton R: **Malaria elimination: worthy, challenging, and just possible.** *Lancet* 2010, **376**:1515–7.

25. Mendis K, Rietveld A, Warsame M, Bosman A, Greenwood B, Wernsdorfer WH: **From malaria control to eradication: The WHO perspective.** *Trop Med Int Health* 2009, **14**:802–9.

26. Greenwood BM: **Control to elimination: implications for malaria research.** *Trends Parasitol* 2008, **24**:449–54.

27. Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley CJ, Smith DL, Abeyasinghe RR, Rodriguez MH, Maharaj R, Tanner M, Targett G: **Operational strategies to achieve and maintain malaria elimination.** *Lancet* 2010, **376**:1592–603.

28. Feachem RG, Phillips A, Hwang J, Cotter C, Wielgosz B, Greenwood BM, Sabot O, Rodriguez MH, Abeyasinghe RR, Ghebreyesus TA, Snow RW: **Shrinking the malaria map: progress and prospects.** *Lancet* 2010, **376**:1566–78.

29. Atkinson J-A, Johnson M-L, Wijesinghe R, Bobogare A, Losi L, O'Sullivan M, Yamaguchi Y, Kenilorea G, Vallely A, Cheng Q, Ebringer A, Bain L, Gray K, Harris I, Whittaker M, Reid H, Clements A, Shanks D: **Operational research to inform a sub-national surveillance intervention for malaria elimination in Solomon Islands.** *Malar J* 2012, **11**:101.

30. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SLR, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya WMM, Lemnge MM, Cox J, Reyburn H, Riley EM: **Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure.** *Proc Natl Acad Sci U S A* 2005, **102**:5108–13.

31. World Health Organization: **global malaria control and elimination : report of a technical review global malaria control and elimination :** 2008:17–18.

32. Steketee RW, Campbell CC: **Impact of national malaria control scale-up programmes in Africa: magnitude and attribution of effects.** *Malar J* 2010, **9**:299.

33. Gosling RD, Okell L, Mosha J, Chandramohan D: **The role of antimalarial treatment in the elimination of malaria.** *Clin Microbiol Infect* 2011, **17**:1617–23.

34. Shanks GD, Oloo AJ, Aleman GM, Ohrt C, Klotz FW, Braitman D, Horton J, Brueckner R: **A new primaquine analogue, tafenoquine (WR 238605), for prophylaxis against Plasmodium falciparum malaria.** *Clin Infect Dis* 2001, **33**:1968–74.

35. Shekalaghe S, Drakeley C, Gosling R, Ndaro A, van Meegeren M, Enevold A, Alifrangis M, Mosha F, Sauerwein R, Bousema T: **Primaquine clears submicroscopic Plasmodium falciparum gametocytes that persist after treatment with sulphadoxine-pyrimethamine and artesunate.** *PLoS One* 2007, **2**:e1023.

36. Thwing JI, Perry RT, Townes DA, Diouf MB, Ndiaye S, Thior M: **Success of Senegal's first nationwide distribution of long-lasting insecticide-treated nets to children under five - contribution toward universal coverage.** *Malar J* 2011, **10**:86.

37. Aregawi MW, Ali AS, Al-Mafazy A-W, Molteni F, Katikiti S, Warsame M, Njau RJ, Komatsu R, Korenromp E, Hosseini M, Low-Beer D, Bjorkman A, D'Alessandro U, Marc Coosemans M, Otten M: **Reductions in malaria and anaemia case and death burden at hospitals following scale-up of malaria control in Zanzibar, 1999-2008.** *Malar J* 2011, **10**:46.

38. Mathanga D, Molyneux ME: **Bednets and malaria in Africa.** *Lancet* 2001, **357**:1219–20.

39. O'Meara WP, Mangeni JN, Steketee R, Greenwood B: **Changes in the burden of malaria in sub-Saharan Africa.** *Lancet Infect Dis* 2010, **10**:545–55.

40. Lengeler C: **Insecticide-treated bed nets and curtains for preventing malaria (Review)**. 2006.

41. Ceesay SJ, Casals-Pascual C, Erskine J, Anya SE, Duah NO, Fulford AJ, Sesay SS, Abubakar I, Dunyo S, Sey O, Palmer A, Fofana M, Corrah T, Bojang KA, Whittle HC, Greenwood BM, Conway DJ: **Changes in malaria indices between 1999 and 2007 in The Gambia: a retrospective analysis**. *Lancet* 2008, **372**:1545–1554.

42. Enserink M: **Redrawing Africa's malaria map**. *Science (80-)* 2010, **328**:842.

43. Nyarango PM, Gebremeskel T, Mebrahtu G, Mufunda J, Abdulmumini U, Ogbamariam A, Kosia A, Gebremichael A, Gunawardena D, Ghebrat Y, Okbaldet Y: **A steep decline of malaria morbidity and mortality trends in Eritrea between 2000 and 2004: the effect of combination of control methods.** *Malar J* 2006, **5**:33.

44. Kleinschmidt I, Schwabe C, Benavente L, Torrez M, Ridl FC, Segura JL, Ehmer P, Nchama GN: **Marked increase in child survival after four years of intensive malaria control.** *Am J Trop Med Hyg* 2009, **80**:882–8.

45. Bhattarai A, Ali AS, Kachur SP, Mårtensson A, Abbas AK, Khatib R, Al-Mafazy A-W, Ramsan M, Rotllant G, Gerstenmaier JF, Molteni F, Abdullah S, Montgomery SM, Kaneko A, Björkman A: **Impact of artemisinin-based combination therapy and insecticide-treated nets on malaria burden in Zanzibar.** *PLoS Med* 2007, **4**:e309.

46. Otten M, Aregawi M, Were W, Karema C, Medin A, Bekele W, Jima D, Gausi K, Komatsu R, Korenromp E, Low-Beer D, Grabowsky M: **Initial evidence of reduction of malaria cases and deaths in Rwanda and Ethiopia due to rapid scale-up of malaria prevention and treatment.** *Malar J* 2009, **8**:14.

47. Teklehaimanot HD, Teklehaimanot A, Kiszewski A, Rampao HS, Sachs JD: **Malaria in São Tomé and principe: on the brink of elimination after three years of effective antimalarial measures.** *Am J Trop Med Hyg* 2009, **80**:133–40.

48. O'Meara WP, Bejon P, Mwangi TW, Okiro E a, Peshu N, Snow RW, Newton CRJC, Marsh K: **Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya.** *Lancet* 2008, **372**:1555–62.

49. Girod R, Orlandi-Pradines E, Rogier C, Pages F: **Malaria transmission and insecticide resistance of Anopheles gambiae (Diptera: Culicidae) in the French military camp of Port-Bouët, Abidjan (Côte d'Ivoire): implications for vector control.** *J Med Entomol* 2006, **43**:1082–7.

50. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, Lwin KM, Ariey F, Hanpithakpong W, Lee SJ, Ringwald P, Silamut K, Imwong M, Chotivanich K, Lim P, Herdman T, An SS, Yeung S, Singhasivanon P, Day NPJ, Lindegardh N, Socheat D, White NJ: **Artemisinin resistance in Plasmodium falciparum malaria.** *N Engl J Med* 2009, **361**:455–67.

51. Dondorp AM, Yeung S, White L, Nguon C, Day NPJ, Socheat D, von Seidlein L: **Artemisinin resistance: current status and scenarios for containment.** *Nat Rev Microbiol* 2010, **8**:272–80.

52. Kochar DK, Saxena V, Singh N, Kochar SK, Kumar S V, Das A: **Plasmodium vivax malaria**. *Emerg Infect Dis* 2005, **11**:132–134.

53. Baird JK: **Neglect of Plasmodium vivax malaria**. *Trends Parasitol* 2007, **23**.

54. Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, Alpers MP, Muller I: **Plasmodium vivax and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea**. *PLoS Med* 2008, **5**:e127.

55. Trape JF, Rogier C, Konate L, Diagne N, Bouganali H, Canque B, Legros F, Badji A, Ndiaye G, Ndiaye P: **The Dielmo project: a longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community living in a holoendemic area of Senegal.** *Am J Trop Med Hyg* 1994, **51**:123–37.

56. Struik SS, Riley EM: **Does malaria suffer from lack of memory?** *Immunol Rev* 2004, **201**:268–90.

57. Snow RW, Gilles HM: **The epidemiology of malaria**. In *Essent Malariol*. 4th edition. Edited by Warrell DA, Gilles HM.; 2002:86–106.

58. Snow RW, Omumbo JA, Lowe B, Molyneux CS, Obiero JO, Palmer A, Weber MW, Pinder M, Nahlen B, Obonyo C, Newbold C, Gupta S, Marsh K: **Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa.** *Lancet* 1997, **349**:1650–4.

59. Rogier C, Ly a B, Tall a, Cissé B, Trape JF: **Plasmodium falciparum clinical malaria in Dielmo, a holoendemic area in Senegal: no influence of acquired immunity on initial symptomatology and severity of malaria attacks.** *Am J Trop Med Hyg* 1999, **60**:410–20.

60. Baird JKK: **Host age as a determinant of naturally acquired immunity to Plasmodium falciparum.** *Parasitol Today* 1995, **11**:105–11.

61. Aponte JJ, Menendez C, Schellenberg D, Kahigwa E, Mshinda H, Vountasou P, Tanner M, Alonso PL: **Age interactions in the development of naturally acquired immunity to Plasmodium falciparum and its clinical presentation.** *PLoS Med* 2007, **4**:e242.

62. Riley EMM, Wagner GEE, Akanmori BDD, Koram KAA: **Do maternally acquired antibodies protect infants from malaria infection?** *Parasite Immunol* 2001, **23**:51–59.

63. Rogier C, Trape JF: **Malaria attacks in children exposed to high transmission: who is protected?** *Trans R Soc Trop Med Hyg* 1993, **87**:245–6.

64. Baird JK: **Age-dependent characteristics of protection v. susceptibility to Plasmodium falciparum.** *Ann Trop Med Parasitol* 1998, **92**:367–90.

65. Langhorne J, Ndungu FM, Sponaas A-M, Marsh K: **Immunity to malaria: more questions than answers.** *Nat Immunol* 2008, **9**:725–32.

66. Bejon P, Warimwe G, Mackintosh CL, Mackinnon MJ, Kinyanjui SM, Musyoki JN, Bull PC, Marsh K: **Analysis of immunity to febrile malaria in children that distinguishes immunity from lack of exposure.** *Infect Immun* 2009, **77**:1917–23.

67. Smith T, Maire N, Dietz K, Killeen GF, Vounatsou P, Molineaux L, Tanner M: **Relationship between the entomologic inoculation rate and the force of infection for Plasmodium falciparum malaria.** *Trop Med* 2006, **75**(Suppl 2):11–18.

68. Trape JF, Rogier C: **Combating malaria morbidity and mortality by reducing transmission.** *Parasitol Today* 1996, **12**:236–40.

69. Smith TA, Killeen G, Lengeler C, Tanner M: **Relationships between the outcome of Plasmodium falciparum infection and the intensity of transmission in Africa.** *Am J Trop Med Hyg* 2004, **71**(2 Suppl):80–6.

70. Service MW, Townson H: **The Anopheles vector**. In *Essent Malariol*. 4th edition. Edited by Warrell DA, Gilles HM. London; 2002:59–84.

71. Clyde DF: **Recent trends in the epidemiology and control of malaria.** *Epidemiol Rev* 1987, **9**:219–43.

72. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, Otieno S, Carneiro I, Cox J, Msuya E, Kleinschmidt I, Maxwell C, Greenwood B, Riley E, Sauerwein R, Chandramohan D, Gosling R: **Identification of Hot Spots of Malaria Transmission for Targeted Malaria Control**. *Society* 2010, **201**:1764–1774.

73. Sinden RE, Gilles HM: **The malaria parasites**. In *Essent Malariol*. 4th edition. Edited by Warrell DA, Gilles HM. Arnold; 2002:9–34.

74. Dery DB, Brown C, Asante KP, Adams M, Dosoo D, Amenga-Etego S, Wilson M, Chandramohan D, Greenwood B, Owusu-Agyei S: **Patterns and seasonality of malaria transmission in the forest-savannah transitional zones of Ghana.** *Malar J* 2010, **9**:314.

75. Roper C, Elhassan IM, Hviid L, Giha H, Richardson W, Babiker H, Satti GM, Theander TG, Arnot DE: **Detection of very low level Plasmodium falciparum infections using the nested polymerase chain reaction and a reassessment of the epidemiology of unstable malaria in Sudan.** *Am J Trop Med Hyg* 1996, **54**:325–31.

76. Smith DL, Drakeley CJ, Chiyaka C, Hay SI: **A quantitative analysis of transmission efficiency versus intensity for malaria.** *Nat Commun* 2010, **1**:108.

77. Smith TA: **Estimation of heterogeneity in malaria transmission by stochastic modelling of apparent deviations from mass action kinetics.** *Malar J* 2008, **7**:12.

78. Carnevale P, Frezil JL, Bosseno MF, Le Pont F, Lancien J: **Study of agressivity of Anopheles gambiae A in relation to age and sex of human subjects**. *Bull WHO* 1978.

79. Smith DL, McKenzie FE, Snow RW, Hay SI: **Revisiting the basic reproductive number for malaria and its implications for malaria control.** *PLoS Biol* 2007, **5**:e42.

80. Anderson RM, May RM: *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press; 1992:768.

81. Dietz K: **The estimation of the basic reproduction number for infectious diseases.** *Stat Methods Med Res* 1993, **2**:23–41.

82. Hay SI, Smith DL, Snow RW: **Measuring malaria endemicity from intense to interrupted transmission.** *Lancet Infect Dis* 2008, **8**:369–78.

83. Beier JC, Killeen GF, Githure JI: **Short report: entomologic inoculation rates and Plasmodium falciparum malaria prevalence in Africa.** *Am J Trop Med Hyg* 1999, **61**:109–13.

84. Yekutiel P: **Problems of epidemiology in malaria eradication**. *Bull World Heal Organ* , **1960**.

85. Okell LC, Ghani AC, Lyons E, Drakeley CJ: **Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis.** *J Infect Dis* 2009, **200**:1509–17.

86. Wongsrichanalai C, Barcus MJ, Muth S, Sutamihardja A, Wernsdorfer WH: **A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT).** *Am J Trop Med Hyg* 2007, **77**(6 Suppl):119–27.

87. MacDonald G: **The analysis of malaria parasite rates in infants.** *Trop Dis Bull* 1950, **47**:915–38.

88. Riley E, Wagner G, Roper C: **Estimating the force of malaria infection.** *Parasitol Today* 1996, **12**:410–1; author reply 411.

89. Snow RW, Molyneux CS, Warn PA, Omumbo J, Nevill CG, Gupta S, Marsh K: **Infant parasite rates and immunoglobulin M seroprevalence as a measure of exposure to Plasmodium falciparum during a randomized controlled trial of insecticide-treated bed nets on the Kenyan coast.** *Am J Trop Med Hyg* 1996, **55**:144–9.

90. Dye C, Lines JD, Curtis CF: **A test of the malaria strain theory.** *Parasitol Today* 1996, **12**:88–9.

91. Smith DL, Dushoff J, Snow RW, Hay SI: **The entomological inoculation rate and Plasmodium falciparum infection in African children.** *Nature* 2005, **438**:492–5.

92. Killeen GF, McKenzie FE, Foy BD, Schieffelin C, Billingsley PF, Beier JC: **A simplified model for predicting malaria entomologic inoculation rates based on entomologic and parasitologic parameters relevant to control.** *Am J Trop Med Hyg* 2000, **62**:535–44.

93. Griffin JT, Hollingsworth TD, Okell LC, Churcher TS, White M, Hinsley W, Bousema T, Drakeley CJ, Ferguson NM, Basáñez M-G, Ghani AC: **Reducing Plasmodium falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies.** *PLoS Med* 2010, **7**.

94. Draper CC: **Malaria. Laboratory diagnosis.** *Br Med J* 1971, **2**:93–5.

95. Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SL, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya WM, Lemnge MM, Cox J, Reyburn H, Riley E: **Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure**. *Proc Natl Acad Sci U S A* 2005, **102**:5108–5113.

96. Wipasa J, Suphavilai C, Okell LC, Cook J, Corran PH, Liewsaree W, Riley EM, Hafalla JClR, Thaikla K: **Long-lived antibody and B Cell memory responses to the Human Malaria Parasites, Plasmodium falciparum and Plasmodium vivax**. *PLoS Pathog* 2010, **6**:e1000770.

97. Corran PH, Coleman PG, Riley E, Drakeley CJ: **Serology: a robust indicator of malaria transmission intensity?** *Trends Parasitol* 2007, **23**:575–582.

98. Stewart L, Gosling R, Griffin J, Gesase S, Campo J, Hashim R, Masika P, Mosha J, Bousema T, Shekalaghe S, Cook J, Corran P, Ghani A, Riley EM, Drakeley C: **Rapid Assessment of Malaria Transmission Using Age-Specific Sero-Conversion Rates**. *PLoS One* 2009, **4**:13.

99. Bousema T, Youssef RM, Cook J, Cox J, Alegana VA, Amran J, Noor AM, Snow RW, Drakeley C: **Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008**. *Emerg Infect Dis* 2010, **16**:392–399.

100. World Health Organization: *Report on the Malaria Conference in Equatorial Africa*. 1950.

101. Metselaar D, Thiel P Van: **Classification of malaria**. *Trop Geogr Med* 1959, **11**:178–183.

102. Craig MH, Snow RW, le Sueur D: **A climate-based distribution model of malaria transmission in sub-Saharan Africa.** *Parasitol Today* 1999, **15**:105–11.

103. Omumbo J a, Hay SI, Guerra CA, Snow RW: **The relationship between the Plasmodium falciparum parasite ratio in childhood and climate estimates of malaria transmission in Kenya.** *Malar J* 2004, **3**:17.

104. Mueller I, Schoepflin S, Smith TA, Benton KL, Bretscher MT, Lin E, Kiniboro B, Zimmerman PA, Speed TP, Siba P, Felger I: **Force of infection is key to understanding the epidemiology of Plasmodium falciparum malaria in Papua New Guinean children.** *Proc Natl Acad Sci U S A* 2012, **109**:10030–5.

105. Smith TA, Vounatsou P: **Estimation of infection and recovery rates for highly polymorphic parasites when detectability is imperfect, using hidden Markov models.** *Stat Med* 2003, **22**:1709–24.

106. Sama W, Owusu-Agyei S, Felger I, Dietz K, Smith TA: **Age and seasonal variation in the transition rates and detectability of Plasmodium falciparum malaria.** *Parasitology* 2006, **132**(Pt 1):13–21.

107. Woolhouse ME, Dye C, Etard JF, Smith TA, Charlwood JD, Garnett GP, Hagan P, Hii JL, Ndhlovu PD, Quinnell RJ, Watts CH, Chandiwana SK, Anderson RM: **Heterogeneities in the transmission of infectious agents: implications for the design of control programs.** *Proc Natl Acad Sci U S A* 1997, **94**:338–42.

108. Bejon P, Williams TN, Liljander A, Noor AM, Wambua J, Ogada E, Olotu A, Osier FHA, Hay SI, Färnert A, Marsh K: **Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya.** *PLoS Med* 2010, **7**:e1000304.

109. Sompayrac LM: *How the Immune System Works*. John Wiley & Sons; 2011:152.

110. Cohen S, Butcher G a, Mitchell GH: **Mechanisms of immunity to malaria.** *Bull World Health Organ* 1974, **50**:251–7.

111. Marsh K, Kinyanjui S: **Immune effector mechanisms in malaria.** *Parasite Immunol* 2006, **28**:51–60.

112. al-Yaman F, Genton B, Kramer KJ, Chang SP, Hui GS, Baisor M, Alpers MP: **Assessment of the role of naturally acquired antibody levels to Plasmodium falciparum merozoite surface protein-1 in protecting Papua New Guinean children from malaria morbidity.** *Am J Trop Med Hyg* 1996, **54**:443–8.

113. Wipasa J, Elliott S, Xu H, Good MF, Mai C: **Immunity to asexual blood stage malaria and vaccine approaches**. *Immunol Cell Biol* 2002:401–414.

114. Thomas AW, Trape JF, Rogier C, Goncalves A, Rosario VE, Narum DL: **High prevalence of natural antibodies against Plasmodium falciparum 83-kilodalton apical membrane antigen (PF83/AMA-1) as detected by capture-enzyme-linked immunosorbent assay using full-length baculovirus recombinant PF83/AMA-1.** *Am J Trop Med Hyg* 1994, **51**:730–40.

115. Achtman AH, Bull PC, Stephens R, Langhorne J: **Longevity of the immune response and memory to blood-stage malaria infection.** *Curr Top Microbiol Immunol* 2005, **297**:71–102.

116. Kinyanjui SM, Conway DJ, Lanar DE, Marsh K: **IgG antibody responses to Plasmodium falciparum merozoite antigens in Kenyan children have a short half-life.** *Malar J* 2007, **6**:82.

117. Akpogheneta OJ, Duah NO, Tetteh KK a, Dunyo S, Lanar DE, Pinder M, Conway DJ: **Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific.** *Infect Immun* 2008, **76**:1748–55.

118. Soares IS, da Cunha MG, Silva MN, Souza JM, Del Portillo H a, Rodrigues MM: **Longevity of naturally acquired antibody responses to the N- and C-terminal regions of Plasmodium vivax merozoite surface protein 1.** *Am J Trop Med Hyg* 1999, **60**:357–63.

119. Collins WE, Skinner JC, Jeffery GM: **Studies on the persistence of malarial antibody response.** *Am J Epidemiol* 1968, **87**:592–8.

120. Druilhe P, Pradier O, Marc JP, Miltgen F, Mazier D, Parent G: **Levels of antibodies to Plasmodium falciparum sporozoite surface antigens reflect malaria transmission rates and are persistent in the absence of reinfection.** *Infect Immun* 1986, **53**:393–7.

121. Bejon P, Lusingu J, Olotu A, Leach A, Lievens M, Vekemans J, Mshamu S, Lang T, Gould J, Dubois M-C, Demoitié M-A, Stallaert J-F, Vansadia P, Carter T, Njuguna P, Awuondo KO, Malabeja A, Abdul O, Gesase S, Mturi N, Drakeley CJ, Savarese B, Villafana T, Ballou WR, Cohen J, Riley EM, Lemnge MM, Marsh K, von Seidlein L: **Efficacy of RTS,S/AS01E vaccine against malaria in children 5 to 17 months of age.** *N Engl J Med* 2008, **359**:2521–32.

122. Riley EM, Wahl S, Perkins DJ, Schofield L: **Regulating immunity to malaria.** *Parasite Immunol* 2006, **28**:35–49.

123. Morais CG, Soares IS, Carvalho LH, Fontes CJF, Krettli antoniana U, Braga EM: **Antibodies to plasmodium vivax apical membrane antigen 1: persistence and correlation with malaria transmission intensity**. *Am J Trop Med Hyg* 2006, **75**:582–587.

124. Deloron P, Chougnet C: **Is immunity to malaria really short-lived?** *Parasitol Today* 1992, **8**:375–8.

125. Voller A, Draper CC: **Immunodiagnosis and sero-epidemiology of malaria.** *Br Med Bull* 1982, **38**:173–7.

126. Esposito F, Fabrizi P, Provvedi A, Tarli P, Habluetzel A, Lombardi S: **Evaluation of an ELISA kit for epidemiological detection of antibodies to Plasmodium falciparum sporozoites in human sera and bloodspot eluates.** *Acta Trop* 1990, **47**:1–10.

127. Molineaux L: **Essential parameters in seroepidemiological assessment: Epidemiological analysis of serological data**. 1981.

128. Akpogheneta OJ, Dunyo S, Pinder M, Conway DJ: **Boosting antibody responses to Plasmodium falciparum merozoite antigens in children with highly seasonal exposure to infection.** *Parasite Immunol* 2010, **32**:296–304.

129. Ambroise-Thomas P, Wernsdorfer WH, Grab B, Cullen J, Bertagna P: **Etude sero-seropidemiologique longitudinale sur le paludisme en Tunisie**. 1976, **54**.

130. Voller A, Storey J, Molineaux L: **A longitudinal study of Plasmodium falciparum malaria in the West African savanna using the ELISA technique ***. *World Health* 1980, **58**:429–438.

131. Voller A, Bruce-Chwatt LJ: **Serological malaria surveys in Nigeria.** *Bull World Health Organ* 1968, **39**:883–97.

132. Cornille-Brögger R, Mathews HM, Storey J, Ashkar TS, Brögger S, Molineaux L: **Changing patterns in the humoral immune response to malaria before, during, and after the application of control measures: a longitudinal study in the West African savanna.** *Bull World Health Organ* 1978, **56**:579–600.

133. Cook J, Kleinschmidt I, Schwabe C, Nseng G, Bousema T, Corran PH, Riley EM, Drakeley CJ: **Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, equatorial Guinea.** *PLoS One* 2011, **6**:e25137.

134. Cook J, Speybroeck N, Sochanta T, Somony H, Sokny M, Claes F, Lemmens K, Theisen M, Soares IS, D'Alessandro U, Coosemans M, Erhart A: **Sero-epidemiological evaluation of changes in Plasmodium falciparum and Plasmodium vivax transmission patterns over the rainy season in Cambodia.** *Malar J* 2012, **11**:86.

135. Cook J, Reid H, Iavro J, Kuwahata M, Taleo G, Clements A, Mccarthy J, Vallely A, Drakeley C: **Using serological measures to monitor changes in malaria transmission in Vanuatu**. *Malar J* 2010:1–15.

136. Proietti C, Pettinato DD, Kanoi BN, Ntege E, Crisanti A, Riley EM, Egwang TG, Drakeley C, Bousema T: **Continuing intense malaria transmission in northern Uganda.** *Am J Trop Med Hyg* 2011, **84**:830–7.

137. Bretscher MT, Supargiyono S, Wijayanti M a, Nugraheni D, Widyastuti AN, Lobo NF, Hawley W a, Cook J, Drakeley CJ: **Measurement of Plasmodium falciparum transmission intensity using serological cohort data from Indonesian schoolchildren.** *Malar J* 2013, **12**:21.

138. Hsiang MS, Hwang J, Kunene S, Drakeley C, Kandula D, Novotny J, Parizo J, Jensen T, Tong M, Kemere J, Dlamini S, Moonen B, Angov E, Dutta S, Ockenhouse C, Dorsey G, Greenhouse B: **Surveillance for malaria elimination in Swaziland: a national cross-sectional study using pooled PCR and serology.** *PLoS One* 2012, **7**:e29550.

139. Gatton M, Hogarth W, Saul A, Dayananda P: **A model for predicting the transmission rate of malaria from serological data**. *Math Biol* 1996:878–888.

140. Burattini MN, Massad E, Coutinho FA: **Malaria transmission rates estimated from serological data**. *Epidemiol Infect* 1993, **111**:503–523.

141. McKenzie FE: **Why model malaria?** *Parasitol Today* 2000, **16**:511–6.

142. Macdonald G: *The Epidemiology and Control of Malaria*. London, Oxford University Press; 1957.

143. **A research agenda for malaria eradication: modeling.** *PLoS Med* 2011, **8**:e1000403.

144. Dietz K, Molineaux L, Thomas A: **A malaria model tested in the African savannah.** *Bull World Health Organ* 1974, **50**:347–57.

145. Garrett-Jones C: **Prognosis for interuption of malaria transmission through the assessment of mosquito's vectorial capacity.** *Nature* 1964, **204**:1173–5.

146. Garrett-Jones C: **The human blood index of malaria in relation to epidemiological assessment**. *Bull World Health Organ* 1964, **30**:241–61.

147. Hasibeder G, Dye C: **Population dynamics of mosquito-borne disease: persistence in a completely heterogeneous environment.** *Theor Popul Biol* 1988, **33**:31–53.

148. Smith TA, Killeen GF, Maire N, Ross A, Molineaux L, Tediosi F, Hutton G, Utzinger J, Dietz K, Tanner M: **Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of Plasmodium falciparum malaria: Overview.** *Am J Trop Med Hyg* 2006, **75**(2 Suppl):1–10.

149. Aron JL: **Mathematical modelling of immunity to malaria**. *Math Biosci* 1988, **90**:385–396.

150. McKenzie FE, Wong RC, Bossert WH: **Discrete-Event Simulation Models of Plasmodium falciparum Malaria.** *Simulation* 1998, **71**:250–261.

151. Mandal S, Sarkar RR, Sinha S: **Mathematical models of malaria--a review.** *Malar J* 2011, **10**:202.

152. Gupta S, Swinton J, Anderson RM: **Theoretical studies of the effects of heterogeneity in the parasite population on the transmission dynamics of malaria.** *Proc Biol Sci* 1994, **256**:231–8.

153. White LJ, Maude RJ, Pongtavornpinyo W, Saralamba S, Aguas R, Van Effelterre T, Day NPJ, White NJ: **The role of simple mathematical models in malaria elimination strategy design.** *Malar J* 2009, **8**:212.

154. Johnston GL, Smith DL, Fidock DA: **Malaria's Missing Number: Calculating the Human Component of R0 by a Within-Host Mechanistic Model of Plasmodium falciparum Infection and Transmission**. *PLoS Comput Biol* 2013, **9**:e1003025.

155. Eckhoff PA: **A malaria transmission-directed model of mosquito life cycle and ecology.** *Malar J* 2011, **10**:303.

156. Antia R, Ganusov V V, Ahmed R: **The role of models in understanding CD8+ T-cell memory.** *Nat Rev Immunol* 2005, **5**:101–11.

157. Gupta S, Day KP: **A theoretical framework for the immunoepidemiology of Plasmodium falciparum malaria.** *Parasite Immunol* 1994, **16**:361–70.

158. Muench H: **Catalytic models in epidemiology**. 1959.

159. Draper CC, Voller A, Carpenter RG: **The epidemiologic interpretation of serologic data in malaria.** *Am J Trop Med Hyg* 1972, **21**:696–703.

160. Lobel HO, Nájera AJ, Ch'en WI, Munroe P, Mathews HM: **Seroepidemiologic investigations of malaria in Guyana.** *J Trop Med Hyg* 1976, **79**:275–84.

161. VanDruten JAM, Meuwissen JHET: *Models for Long-Term Retrospective Analysis in Seroepidemiology*. 1982.

162. Cook J, Griffin J, Corran P, Riley EM, Ghani AC, Drakeley CJ: **Use of mixture models to determine seroprevalence of antimalarial antibodies**. *Prep* .

163. Vyse AJ, Gay NJ, Hesketh LM, Pebody R, Morgan-Capner P, Miller E: **Interpreting serological surveys using mixture models: the seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of the 21st century.** *Epidemiol Infect* 2006, **134**:1303–12.

164. Ohuma EO, Okiro E a, Bett a, Abwao J, Were S, Samuel D, Vyse a, Gay N, Brown DWG, Nokes DJ: **Evaluation of a measles vaccine campaign by oral-fluid surveys in a rural Kenyan district: interpretation of antibody prevalence data using mixture models.** *Epidemiol Infect* 2009, **137**:227–33.

165. Drakeley CJ, Cook J, Griffin JT, Okell LC, Bousema T, Corran P, Ghani AC, Riley E: **Correlation between P. falciparum parasite prevalence and serological measures of malaria transmission**. *Prep* .

166. Molineaux L, Gramiccia G: *The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa, Chap 10: WHO*. 1980.

167. Burattini MN, Massad E, Coutinho F, Baruzzi RG: **Malaria prevalence amongst Brazilian Indians assessed by a new mathematical model**. *Epidemiol Infect* 1993:525–537.

168. Irion A, Beck HP, Smith TA: **Assessment of positivity in immuno-assays with variability in background measurements: a new approach applied to the antibody response to Plasmodium falciparum MSP2.** *J Immunol Methods* 2002, **259**:111–8.

169. Van der Kaay HJ: **Malaria in Surinam, a sero-epidemiological study.** *Acta Leiden* 1976, **43**:7–91.

170. Lobel HO, Mathews HM, Kagan IG: **Interpretation of IHA titres for the study of malaria epidemiology.** *Bull World Health Organ* 1973, **49**:485–92.

171. Kagan IG, Mathews H, Sulzer AJ: **The serology of malaria: recent applications.** *Bull N Y Acad Med* 1969, **45**:1027–42.

172. Herck K Van, Beutels P, Damme P Van, Beutels M, Dries J Van Den, Briantais P, Vidor E: **Mathematical Models for Assessment of Long-Term Persistence of Antibodies After Vaccination With Two Inactivated Hepatitis A Vaccines**. *J Med Virol* 2000, **7**(April 1999):1–7.

173. Fraser C, Tomassini JE, Xi L, Golm G, Watson M, Giuliano AR, Barr E, Ault KA: **Modeling the long-term antibody response of a human papillomavirus (HPV) virus-like particle (VLP) type 16 prophylactic vaccine.** *Vaccine* 2007, **25**:4324–33.

174. Wilson JN, Nokes DJ, Medley GF, Shouval D: **Mathematical model of the antibody response to hepatitis B vaccines: implications for reduced schedules.** *Vaccine* 2007, **25**:3705–12.

175. Teunis PFM, Van Der Heijden OG, De Melker HE, Schellekens JFP, Versteegh FGA, Kretzschmar MEE: **Kinetics of the IgG antibody response to pertussis toxin after infection with B. pertussis**. *Epidemiol Infect* 2003, **129**:479.

176. Versteegh FGA, Mertens PLJM, De Melker HE, Roord JJ, Schellekens JFP, Teunis PFM: **Age-specific long-term course of IgG antibodies to pertussis toxin after symptomatic infection with Bordetella pertussis**. *Epidemiol Infect* 2005, **133**:737–748.

177. Ferguson NM, Donnelly CA, Anderson RM: **Transmission dynamics and epidemiology of dengue: insights from age-stratified sero-prevalence surveys**. *Philos Trans R Soc L B Biol Sci* 1999, **354**:757–768.

178. Webster HK, Gingrich JB, Wongsrichanalai C, Tulyayon S, Suvarnamani A, Sookto P, Permpanich B: **Circumsporozoite antibody as a serologic marker of Plasmodium falciparum transmission.** *Am J Trop Med Hyg* 1992, **47**:489–97.

179. Incardona S, Vong S, Chiv L, Lim P, Nhem S, Sem R, Khim N, Doung S, Mercereau-Puijalon O, Fandeur T: **Large-scale malaria survey in Cambodia: novel insights on species distribution and risk factors.** *Malar J* 2007, **6**:37.

180. Dysoley L, Kaneko A, Eto H, Mita T, Socheat D, Börkman A, Kobayakawa T: **Changing patterns of forest malaria among the mobile adult male population in Chumkiri District, Cambodia.** *Acta Trop* 2008, **106**:207–12.

181. Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J, Cui L: **Plasmodium vivax transmission: chances for control?** *Trends Parasitol* 2004, **20**:192–8.

182. Konchom S, Singhasivanon P, Kaewkungwal J, Chupraphawan S, Thimasarn K, Kidson C, Rojanawatsirivet C, Yimsamran S, Looareesuwan S: **Trend of malaria incidence in highly endemic provinces along the Thai borders, 1991-2001.** *Southeast Asian J Trop Med Public Health* 2003, **34**:486–94.

183. Trung HD, Van Bortel W, Sochantha T, Keokenchanh K, Quang NT, Cong LD, Coosemans M: **Malaria transmission and major malaria vectors in different geographical areas of Southeast Asia.** *Trop Med Int Health* 2004, **9**:230–7.

184. **Development Trends in Cambodia Forest Cover 2006** [www.sithi.org]

185. Gilks WR, Richardson S, Spiegelhalter DJ: *Markov Chain Monte Carlo in Practice*. Chapman & Hall; 1996:486.

186. Wh P, Teukolsky S, Vetterling W, Flannery B: **Numerical recipes in C: the art of Scientific Computing**. 1992.

187. R Team Core: **R: A Language and Environment for Statistical Computing**. 2012.

188. Stoetaert K, Petzoldt T, Setzer RW: **Solving Differential Equations in R: Package deSolve**. *J Stat* 2010, **33**.

189. Cauchemez S, Carrat F, Viboud C, Valleron AJ, Boëlle PY: **A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data.** *Stat Med* 2004, **23**:3469–87.

190. Roberts GO, Rosenthal JS: **Optimal scaling for various Metropolis-Hastings algorithms**. *Stat Sci* 2001, **16**:351–367.

191. Bolker BM: *Ecological Models and Data in R*. Princeton University Press; 2008:396.

192. White M: **Models for measuring and predicting malaria vaccine efficacy Thesis submitted for PhD examination Department of Infectious Disease Epidemiology**. 2012.

193. Fowkes FJI, Richards JS, Simpson JA, Beeson JG: **The relationship between anti-merozoite antibodies and incidence of Plasmodium falciparum malaria: A systematic review and meta-analysis.** *PLoS Med* 2010, **7**:e1000218.

194. Bruce-Chwatt LJ, Dodge JS, Draper CC, Topley E, Voller A: **Sero-epidemiological studies on population groups previously exposed to malaria.** *Lancet* 1972, **1**:512–5.

195. Nebie I, Tiono AB, Diallo DA, Samandoulougou S, Diarra A, Konate AT, Cuzin-Ouattara N, Theisen M, Corradin G, Cousens S, Ouattara AS, Ilboudo-Sanogo E, Sirima BS: **Do antibody responses to malaria vaccine candidates influenced by the level of malaria transmission protect from malaria?** *Trop Med Int Health* 2008, **13**:229–37.

196. Askjaer N, Maxwell C, Chambo W, Staalsoe T, Nielsen M, Hviid L, Curtis C, Theander TG: **Insecticide-treated bed nets reduce plasma antibody levels and limit the repertoire of antibodies to Plasmodium falciparum variant surface antigens.** *Clin Diagn Lab Immunol* 2001, **8**:1289–91.

197. Cavanagh DR, Elhassan IM, Roper C, Robinson VJ, Giha H, Holder AA, Hviid L, Theander TG, Arnot DE, McBride JS: **A longitudinal study of type-specific antibody responses to Plasmodium falciparum merozoite surface protein-1 in an area of unstable malaria in Sudan.** *J Immunol* 1998, **161**:347–59.

198. Kinyanjui SM, Mwangi T, Bull PC, Newbold CI, Marsh K: **Protection against clinical malaria by heterologous immunoglobulin G antibodies against malaria-infected erythrocyte variant surface antigens requires interaction with asymptomatic infections.** *J Infect Dis* 2004, **190**:1527–33.

199. Bull PC, Lowe BS, Kaleli N, Njuga F, Kortok M, Ross A, Ndungu F, Snow RW, Marsh K: **Plasmodium falciparum infections are associated with agglutinating antibodies to parasite-infected erythrocyte surface antigens among healthy Kenyan children.** *J Infect Dis* 2002, **185**:1688–91.

200. Kinyanjui SM, Bejon P, Osier FH, Bull PC, Marsh K: **What you see is not what you get: implications of the brevity of antibody responses to malaria antigens and transmission heterogeneity in longitudinal studies of malaria immunity.** *Malar J* 2009, **8**:242.

201. Spiegelhalter DJ, Best NG, Carlin BP, Linde A van der: **Bayesian Measures of Model Complexity and Fit**. *J T Stat Soc B* 2002, **64**:583–639.

202. Durnez L, Mao S, Denis L, Roelants P, Sochantha T, Coosemans M: **Outdoor malaria transmission in forested villages of Cambodia.** *Malar J* 2013, **12**:329.

203. Collins WE, Jeffery GM, Skinner JC: **Fluorescent antibody studies in human malaria.I. Development of antibodies to plasmodium malariae.** *Am J Trop Med Hyg* 1964, **13**:1–5.

204. Stevenson MM, Riley EM: **Innate immunity to malaria.** *Nat Rev Immunol* 2004, **4**:169–80.

205. Ndungu FM, Cadman ET, Coulcher J, Nduati E, Couper E, Macdonald DW, Ng D, Langhorne J: **Functional memory B cells and long-lived plasma cells are generated after a single Plasmodium chabaudi infection in mice.** *PLoS Pathog* 2009, **5**:e1000690.

206. Dorfman JR, Bejon P, Ndungu FM, Langhorne J, Kortok MM, Lowe BS, Mwangi TW, Williams TN, Marsh K: **B cell memory to 3 Plasmodium falciparum blood-stage antigens in a malaria-endemic area.** *J Infect Dis* 2005, **191**:1623–30.

207. Luby JP, Collins WE, Kaiser RL: **Persistence of malarial antibody. Findings in patients infected during the outbreak of malaria in Lake Vera, California, 1952-1953.** *Am J Trop Med Hyg* 1967, **16**:255–7.

208. Filipe JAN, Riley EM, Drakeley CJ, Sutherland CJ, Ghani AC: **Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model**. *PLoS Comput Biol* 2007, **3**:e255.

209. **Malaria Indicators Surveys** [http://www.malariasurveys.org]

210. Modiano D, Petrarca V, Sirima BS, Nebié I, Diallo D, Esposito F, Coluzzi M: **Different response to Plasmodium falciparum malaria in west African sympatric ethnic groups.** *Proc Natl Acad Sci U S A* 1996, **93**:13206–11.

211. Kwiatkowski DP: **How malaria has affected the human genome and what human genetics can teach us about malaria.** *Am J Hum Genet* 2005, **77**:171–92.

212. Drakeley CJ, Carneiro I, Reyburn H, Malima R, Lusingu JPA, Cox J, Theander TG, Nkya WMMM, Lemnge MM, Riley EM: **Altitude-dependent and -independent variations in Plasmodium falciparum prevalence in northeastern Tanzania.** *J Infect Dis* 2005, **191**:1589–98.

213. Bødker R, Akida J, Shayo D, Kisinza W, Msangeni HA, Pedersen EM, Lindsay SW, Bødker AR: **Relationship between altitude and intensity of malaria transmission in the Usambara Mountains, Tanzania.** *J Med Entomol* 2003, **40**:706–17.

214. Bødker AR, Akida J, Shayo D, Kisinza W, Msangeni HA, Lindsay SW: **Relationship Between Altitude and Intensity of Malaria Transmission in the Usambara Mountains , Tanzania Relationship Between Altitude and Intensity of Malaria Transmission in the Usambara Mountains , Tanzania**. 2003, **40**:706–717.

215. Maxwell C a, Chambo W, Mwaimu M, Magogo F, Carneiro I a, Curtis CF: **Variation of malaria transmission and morbidity with altitude in Tanzania and with introduction of alphacypermethrin treated nets.** *Malar J* 2003, **2**:28.

216. Kulkarni AMA, Kweka E, Nyale E, Lyatuu E, Chandramohan D, Rau ME, Drakeley C: **Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District , Northeastern Tanzania Entomological Evaluation of Malaria Vectors at Different Altitudes in Hai District , Northeastern Tanzania**. 2006, **43**:580–588.

217. Snow RW, Amratia P, Kabaria CW, Noor AM, Marsh K: **The changing limits and incidence of malaria in Africa: 1939-2009.** *Adv Parasitol* 2012, **78**:169–262.

218. Bousema JT, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, Ghani A, Drakeley C, Gosling R: **Hitting hotspots: spatial targeting of malaria for control and elimination.** *PLoS Med* 2012, **9**:e1001165.

219. Ernst KC, Adoka SO, Kowuor DO, Wilson ML, John CC: **Malaria hotspot areas in a highland Kenya site are consistent in epidemic and non-epidemic years and are associated with ecological factors.** *Malar J* 2006, **5**:78.

220. Corran PH, Cook J, Lynch C, Leendertse H, Manjurano A, Griffin J, Cox J, Abeku T, Bousema JT, Ghani AC, Drakeley CJ, Riley E: **Dried blood spots as a source of anti-malarial antibodies for epidemiological studies.** *Malar J* 2008, **7**:195.

221. Noor AM, Clements ACA, Gething PW, Moloney G, Borle M, Shewchuk T, Hay SI, Snow RW: **Spatial prediction of Plasmodium falciparum prevalence in Somalia.** *Malar J* 2008, **7**:159.

222. Oesterholt MJAM, Bousema JT, Mwerinde OK, Harris C, Lushino P, Masokoto A, Mwerinde H, Mosha FW, Drakeley CJ: **Spatial and temporal variation in malaria transmission in a low endemicity area in northern Tanzania.** *Malar J* 2006, **5**:98.

223. Steenkeste N, Rogers WO, Okell L, Jeanne I, Incardona S, Duval L, Chy S, Hewitt S, Chou M, Socheat D, Babin F-X, Ariey F, Rogier C: **Sub-microscopic malaria cases and mixed malaria infection in a remote area of high malaria endemicity in Rattanakiri province, Cambodia: implication for malaria elimination.** *Malar J* 2010, **9**:108.

224. Harris I, Sharrock WW, Bain LM, Gray K-A, Bobogare A, Boaz L, Lilley K, Krause D, Vallely A, Johnson M-L, Gatton ML, Shanks GD, Cheng Q: **A large proportion of asymptomatic Plasmodium infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting.** *Malar J* 2010, **9**:254.

225. Taylor RR, Egan a, McGuinness D, Jepson a, Adair R, Drakely C, Riley E: **Selective recognition of malaria antigens by human serum antibodies is not genetically determined but demonstrates some features of clonal imprinting.** *Int Immunol* 1996, **8**:905–15.

226. Khosravi A, Hommel M, Sayemiri K: **Age-dependent antibody response to Plasmodium falciparum merozoite surface protein 2 (MSP-2).** *Parasite Immunol* 2011, **33**:145–57.

227. Kleinschmidt I, Sharp B, Benavente LE, Schwabe C, Torrez M, Kuklinski J, Morris N, Raman J, Carter J: **Reduction in infection with Plasmodium falciparum one year after the introduction of malaria control interventions on Bioko Island, Equatorial Guinea.** *Am J Trop Med Hyg* 2006, **74**:972–8.

228. Von Seidlein L, Walraven G, Milligan PJ, Alexander N, Manneh F, Deen JL, Coleman R, Jawara M, Lindsay SW, Drakeley C, De Martin S, Olliaro P, Bennett S, Schim van der Loeff M, Okunoye K, Targett G a, McAdam KP, Doherty JF, Greenwood BM, Pinder M: **The effect of mass administration of sulfadoxine-pyrimethamine combined with artesunate on malaria incidence: a double-blind, community-randomized, placebo-controlled trial in The Gambia.** *Trans R Soc Trop Med Hyg* 2003, **97**:217–25.

229. Njau JD, Stephenson R, Menon M, Kachur SP, McFarland DA: **Exploring the impact of targeted distribution of free bed nets on households bed net ownership, socio-economic disparities and childhood malaria infection rates: analysis of national malaria survey data from three sub-Saharan Africa countries.** *Malar J* 2013, **12**:245.

230. Okello PE, Van bortel W, Byaruhanga AM, Correwyn A, Roelants P, Talisuna A, D'Alessandro U, Coosemans M: **Variation in malaria transmission intensity in seven sites throughout Uganda**. *Am J Trop Med Hyg* 2006, **75**:219–225.

231. Uganda Bureau of Statistics (UBOS) and ICF Macro: *Uganda Malaria Indicator Survey 2009*. Calverton, Maryland, USA; 2010.

232. Ibison F, Olotu A, Muema DM, Mwacharo J, Ohuma E, Kimani D, Marsh K, Bejon P, Ndungu FM: **Lack of avidity maturation of merozoite antigen-specific antibodies with increasing exposure to Plasmodium falciparum amongst children and adults exposed to endemic malaria in Kenya.** *PLoS One* 2012, **7**:e52939.

233. Van Bortel W, Trung HD, Hoi LX, Van Ham N, Van Chut N, Luu ND, Roelants P, Denis L, Speybroeck N, D'Alessandro U, Coosemans M: **Malaria transmission and vector behaviour in a forested malaria focus in central Vietnam and the implications for vector control.** *Malar J* 2010, **9**:373.

234. Erhart A, Thang ND, Hung NQ, Toi L V, Hung LX, Tuy TQ, Cong LD, Speybroeck N, Coosemans M, D'Alessandro U: **Forest malaria in Vietnam: a challenge for control.** *Am J Trop Med Hyg* 2004, **70**:110–8.

235. Thang ND, Erhart A, Speybroeck N, Hung LX, Thuan LK, Hung CT, Ky P Van, Coosemans M, D'Alessandro U: **Malaria in central Vietnam: analysis of risk factors by multivariate analysis and classification tree models.** *Malar J* 2008, **7**:28.

236. Youssef RM, Alegana VA, Amran J, Noor AM, Snow RW: **Fever prevalence and management among three rural communities in the North West Zone, Somalia.** *East Mediterr Health J* 2010, **16**:595–601.

237. Kleinschmidt I, Torrez M, Schwabe C, Benavente L, Seocharan I, Jituboh D, Nseng G, Sharp B: **Factors influencing the effectiveness of malaria control in Bioko Island, equatorial Guinea.** *Am J Trop Med Hyg* 2007, **76**:1027–32.

238. Greenwood BM, Pickering H: **A review of the epidemiology and control of malaria in The Gambia, West Africa**. *Trans R Soc Trop Med Hyg* , **87**:3–11.

239. Lindsay SW, Shenton FC, Snow RW, Greenwood BM: **Responses of Anopheles gambiae complex mosquitoes to the use of untreated bednets in The Gambia.** *Med Vet Entomol* 1989, **3**:253–62.

240. Lulat A: **A study of humoral and cell-mediated immune responses to Plasmodium falciparum malaria gametocyte antigen in rural Gambia population**. University of London, UK; 1993.

241. Overgaard HJ, Reddy VP, Abaga S, Matias A, Reddy MR, Kulkarni V, Schwabe C, Segura L, Kleinschmidt I, Slotman MA: **Malaria transmission after five years of vector control on Bioko Island, Equatorial Guinea.** *Parasit Vectors* 2012, **5**:253.

242. Venzon DJ, Moolgaravkar SH: **A method for computing profile-likelihood-based confidence intervals**. *J R Stat Soc Ser C (Applied Stat* 1988, **37**:87–94.

243. Dobaño C, Quelhas D, Quintó L, Puyol L, Serra-Casas E, Mayor A, Nhampossa T, Macete E, Aide P, Mandomando I, Sanz S, Puniya SK, Singh B, Gupta P, Bhattacharya A, Chauhan VS, Aponte JJ, Chitnis CE, Alonso PL, Menéndez C: **Age-dependent IgG subclass responses to Plasmodium falciparum EBA-175 are differentially associated with incidence of malaria in Mozambican children.** *Clin Vaccine Immunol* 2012, **19**:157–66.

244. Fouda GG, Leke RFG, Long C, Druilhe P, Zhou A, Taylor DW, Johnson AH: **Multiplex assay for simultaneous measurement of antibodies to multiple Plasmodium falciparum antigens.** *Clin Vaccine Immunol* 2006, **13**:1307–13.

245. Ondigo BN, Park GS, Gose SO, Ho BM, Ochola LA, Ayodo GO, Ofulla A V, John CC: **Standardization and validation of a cytometric bead assay to assess antibodies to multiple Plasmodium falciparum recombinant antigens.** *Malar J* 2012, **11**:427.

246. Ambrosino E, Dumoulin C, Orlandi-Pradines E, Remoue F, Toure-Baldé A, Tall A, Sarr JB, Poinsignon A, Sokhna C, Puget K, Trape J-F, Pascual A, Druilhe P, Fusai T, Rogier C: **A multiplex assay for the simultaneous detection of antibodies against 15 Plasmodium falciparum and Anopheles gambiae saliva antigens.** *Malar J* 2010, **9**:317.

247. Gray JC, Corran PH, Mangia E, Gaunt MW, Li Q, Tetteh KKA, Polley SD, Conway DJ, Holder AA, Bacarese-Hamilton T, Riley EM, Crisanti A: **Profiling the antibody immune response against blood stage malaria vaccine candidates.** *Clin Chem* 2007, **53**:1244–53.

248. Mitchell GH, Thomas AW, Margos G, Dluzewski AR, Bannister LH: **Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells.** *Infect Immun* 2004, **72**:154–8.

249. Holder AA: **The carboxy-terminus of merozoite surface protein 1: structure, specific antibodies and immunity to malaria**. *Parasitology* 2009, **136**:1445–1456.

250. Lulli P, Mangano VD, Onori A, Batini C, Luoni G, Sirima BS, Nebie I, Chessa L, Petrarca V, Modiano D: **HLA-DRB1 and -DQB1 loci in three west African ethnic groups: genetic relationship with sub-Saharan African and European populations.** *Hum Immunol* 2009, **70**:903–9.

251. Mukabana WR, Kannady K, Kiama GM, Ijumba JN, Mathenge EM, Kiche I, Nkwengulila G, Mboera L, Mtasiwa D, Yamagata Y, van Schayk I, Knols BGJ, Lindsay SW, Caldas de Castro M, Mshinda H, Tanner M, Fillinger U, Killeen GF: **Ecologists can enable communities to implement malaria vector control in Africa.** *Malar J* 2006, **5**:9.

252. Mwangi TW, Ross A, Marsh K, Snow RW: **The effects of untreated bednets on malaria infection and morbidity on the Kenyan coast.** *Trans R Soc Trop Med Hyg* , **97**:369–72.

253. Mbogo CM, Mwangangi JM, Nzovu J, Gu W, Yan G, Gunter JT, Swalm C, Keating J, Regens JL, Shililu JI, Githure JI, Beier JC: **Spatial and temporal heterogeneity of anopheles mosquitoes and plasmodium falciparum transmission along the kenyan coast**. *Am J Trop Med Hyg* 2003, **68**:734–742.

254. Tongol-Rivera P, Kano S, Miguel E, Tongol P, Suzuki M: **Application of seroepidemiology in identification of local foci in a malarious community in Palawan, The Philippines.** *Am J Trop Med Hyg* 1993, **49**:608–12.

255. Bennett S, Riley EM: **The statistical analysis of data from immunoepidemiological studies.** *J Immunol Methods* 1992, **146**:229–239.

256. Estévez PT, Satoguina J, Nwakanma DC, West S, Conway DJ, Drakeley CJ: **Human saliva as a source of anti-malarial antibodies to examine population exposure to Plasmodium falciparum.** *Malar J* 2011, **10**:104.

257. Carter R: **Transmission blocking malaria vaccines**. *Vaccine* 2001, **19**:2309–2314.

258. Stowers A, Carter R: **Current developments in malaria transmission-blocking vaccines.** *Expert Opin Biol Ther* 2001, **1**:619–28.

259. Malkin EM, Durbin AP, Diemert DJ, Sattabongkot J, Wu Y, Miura K, Long CA, Lambert L, Miles AP, Wang J, Stowers A, Miller LH, Saul A: **Phase 1 vaccine trial of Pvs25H: a transmission blocking vaccine for Plasmodium vivax malaria.** *Vaccine* 2005, **23**:3131–8.

260. Dodoo D, Aikins A, Kusi KA, Lamptey H, Remarque E, Milligan P, Bosomprah S, Chilengi R, Osei YD, Akanmori BD, Theisen M: **Cohort study of the association of antibody levels to AMA1, MSP119, MSP3 and GLURP with protection from clinical malaria in Ghanaian children.** *Malar J* 2008, **7**:142.

261. Shi Y, Sayed U, Qari S, Roberts J, Udhayakumar V, Oloo A, Hawley W, Kaslow D, Nahlen B, Lal A: **Natural immune response to the C-terminal 19-kilodalton domain of Plasmodium falciparum merozoite surface protein 1**. *Infect Immun* 1996, **64**:2716–2723.

262. Branch OH, Oloo AJ, Nahlen BL, Kaslow D, Lal AA: **Anti-merozoite surface protein-1 19-kDa IgG in mother-infant pairs naturally exposed to Plasmodium falciparum: subclass analysis with age, exposure to asexual parasitemia, and protection against malaria. V. The Asembo Bay Cohort Project.** *J Infect Dis* 2000, **181**:1746–52.

263. Riley EM, Morris-Jones S, Blackman MJ, Greenwood BM, Holder AA: **A longitudinal study of naturally acquired cellular and humoral immune responses to a merozoite surface protein (MSP1) of Plasmodium falciparum in an area of seasonal malaria transmission.** *Parasite Immunol* 1993, **15**:513–24.

264. Tongren JE, Drakeley CJ, McDonald SLR, Reyburn HG, Manjurano A, Nkya WMM, Lemnge MM, Gowda CD, Todd JE, Corran PH, Riley EM: **Target antigen, age, and duration of antigen exposure independently regulate immunoglobulin G subclass switching in malaria.** *Infect Immun* 2006, **74**:257–64.

265. Morell A, Terry WD, Waldmann TA: **Metabolic properties of IgG subclasses in man.** *J Clin Invest* 1970, **49**:673–80.

266. Molineaux L, Wernsdorfer WH, I. M: **The epidemiology of human malaria as an explanation of its distribution, including some implications for its control.** 1988:913–998.

267. Branch OH, Udhayakumar V, Hightower AW, Oloo AJ, Hawley WA, Nahlen BL, Bloland PB, Kaslow DC, Lal AA: **A longitudinal investigation of IgG and IgM antibody responses to the merozoite surface protein-1 19-kiloDalton domain of Plasmodium falciparum in pregnant women and infants: associations with febrile illness, parasitemia, and anemia.** *Am J Trop Med Hyg* 1998, **58**:211–9.

268. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN: **Heritability of malaria in Africa.** *PLoS Med* 2005, **2**:e340.

269. Mukabana WR, Takken W, Coe R, Knols BGJ: **Host-specific cues cause differential attractiveness of Kenyan men to the African malaria vector Anopheles gambiae.** *Malar J* 2002, **1**:17.

270. Galvani AP, May RM: **Epidemiology: dimensions of superspreading.** *Nature* 2005, **438**:293–5.

271. Brooker S, Kolaczinski JH, Gitonga CW, Noor AM, Snow RW: **The use of schools for malaria surveillance and programme evaluation in Africa.** *Malar J* 2009, **8**:231.

272. Olotu A, Fegan G, Wambua J, Nyangweso G, Awuondo KO, Leach A, Lievens M, Leboulleux D, Njuguna P, Peshu N, Marsh K, Bejon P: **Four-Year Efficacy of RTS,S/AS01E and Its Interaction with Malaria Exposure**. *N Engl J Med* 2013, **368**:1111–1120.

273. Bejon P, Cook J, Bergmann-Leitner E, Olotu A, Lusingu J, Mwacharo J, Vekemans J, Njuguna P, Leach A, Lievens M, Dutta S, von Seidlein L, Savarese B, Villafana T, Lemnge MM, Cohen J, Marsh K, Corran PH, Angov E, Riley EM, Drakeley CJ: **Effect of the pre-erythrocytic candidate malaria vaccine RTS,S/AS01E on blood stage immunity in young children.** *J Infect Dis* 2011, **204**:9–18.

274. Riley E, Wagner G, Roper C: **Estimating the force of malaria infection.** *Parasitol Today* 1996, **12**:410–1; author reply 411.

275. Bretscher MT, Maire N, Chitnis N, Felger I, Owusu-Agyei S, Smith TA: **The distribution of Plasmodium falciparum infection durations.** *Epidemics* 2011, **3**:109–18.

276. Dörner T, Radbruch A: **Antibodies and B cell memory in viral immunity.** *Immunity* 2007, **27**:384–92.

Appendix I - Details on Metropolis Hastings algorithm

In a Metropolis Hasting algorithm, the probability of accepting a new value for a parameter θ is:

$$\text{Acceptance ratio: } \alpha = min\left(1, \frac{\Pi(\theta*)}{\Pi(\theta)} \frac{Q(\theta|\theta*)}{Q(\theta*|\theta)}\right)$$

with $\theta*$ the candidate value, Π and Q respectively the posterior and proposal distributions.

Here, I suggest deriving the ratio of the proposals: $\frac{\mathrm{Q}(\theta*/\theta)}{\mathrm{Q}(\theta/\theta*)}$

Let $Z*=\log(\theta*)$, hence $\mathrm{dZ}*=\mathrm{d}(\log(\theta*))=\frac{d\theta*}{\theta*}$

Therefore, if $Z* \sim Norm(Z, \sigma^2)$ then $Q(Z*/Z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Z*-Z)^2}{2\sigma^2}}$

The proposal is therefore:

$$\begin{aligned}
\mathrm{Q}(\theta*/\theta) &= \mathrm{Q}(\mathrm{Z}*/\theta)\frac{\mathrm{dZ}*}{\mathrm{d}\theta*} && \text{as } \mathrm{Q}(\theta*/\theta)\mathrm{d}\theta* = \mathrm{Q}(\mathrm{Z}*/\theta)\mathrm{dZ}* \\
&= \mathrm{Q}(\mathrm{Z}*/\theta)\times\frac{1}{\theta*} && \text{as } \mathrm{dZ}* = \frac{d\theta*}{\theta*} \\
&= \mathrm{Q}(\mathrm{Z}*/\mathrm{e}^{Z})\times\frac{1}{\theta*} && \text{as } \mathrm{Z}=\log(\theta) \\
&= \mathrm{Q}(\mathrm{Z}*/Z)\times\frac{1}{\theta*} \\
&= \frac{1}{\theta*}\times\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Z*-Z)^2}{2\sigma^2}} && \text{as } \mathrm{Q}(Z*/Z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Z*-Z)^2}{2\sigma^2}} \\
&= \frac{1}{\theta*}\times\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log(\theta*)-\log(\theta))^2}{2\sigma^2}} && \text{as } \mathrm{Z}=\log(\theta)
\end{aligned}$$

The proposal is log-normally distributed with parameters (θ, σ^2), mean and variance of the underlying normal distribution. The ratio of the proposal distributions is therefore

$$\frac{\mathrm{Q}(\theta*/\theta)}{\mathrm{Q}(\theta/\theta*)} = \frac{1}{\theta*}\times\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log(\theta*)-\log(\theta))^2}{2\sigma^2}} \times \frac{1}{\frac{1}{\theta}\times\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\log(\theta)-\log(\theta*))^2}{2\sigma^2}}} = \frac{\theta}{\theta*}$$

As a result, the acceptance ratio becomes:

$$\alpha = min\left(1, \frac{\Pi(\theta*)}{\Pi(\theta)} \frac{Q(\theta|\theta*)}{Q(\theta*|\theta)}\right) = \quad min\left(1, \frac{\Pi(\theta*)}{\Pi(\theta)} \frac{\theta*}{\theta}\right)$$

However, the posterior distribution is defined as: $\Pi(\theta) = P(D/\theta)P(\theta)$

with $P(D/\theta)$ the likelihood and $P(\theta)$ the prior. The acceptance ratio is then:

$$\alpha = \quad min\left(1, \frac{\Pi(\theta*)}{\Pi(\theta)} \frac{\theta*}{\theta}\right) = min\left(1, \frac{P(D/\theta*)P(\theta*)}{P(D/\theta)P(\theta)} \frac{\theta*}{\theta}\right)$$

We chose an uninformative prior for the parameters: $\theta \sim Unif[0, \max]$

Therefore, $P(\theta) = \begin{cases} \frac{1}{\max} & \text{if } \theta \in [0,\max] \\ 0 & \text{otherwise} \end{cases}$ and $\frac{P(\theta*)}{P(\theta)} = 1$

Finally, the acceptance ratio becomes:

$$\alpha = \quad min\left(1, \frac{P(D/\theta*)P(\theta*)}{P(D/\theta)P(\theta)} \frac{\theta*}{\theta}\right) = min\left(1, \frac{P(D/\theta*)}{P(D/\theta)} \frac{\theta*}{\theta}\right)$$

And, in the implementation, a new value is accepted if: $\frac{P(D/\theta*)}{P(D/\theta)} \frac{\theta*}{\theta} < \mathrm{U}$ where $U \sim Uniform[0,1]$

Or precisely if $\log\left\{\frac{P(D/\theta*)}{P(D/\theta)} \frac{\theta*}{\theta}\right\} < \log\{\mathrm{U}\}$

equivalent to $\log P(D/\theta*) - \log P(D/\theta) + \log(\theta*) - \log(\theta) < \log\{\mathrm{U}\}$

Appendix II - Deriving antibody levels from the outputs of the model

As I have discretised the continuous antibody level for the implementation of the density model, I want to retrieve an estimated density level from the proportion of individuals whose antibody level falls within the 51 defined intervals. This requires interpolation to determine the antibody level z_q associated with a percentile q of the population. In other words, I want to find the antibody level for which q% of the population has an antibody level below that level z_q (See Figure 7.1).

Let Q_i be the cumulative proportion of individuals that has a level below level $\bar{z}_i$, with level $\bar{z}_i$ being the median antibody level of the interval $(x_{i-1}, x_i]$ and x_{i-1}, x_i the lower and upper bounds of antibody class i, with $i \in \{2,...,N\}$ and $\bar{z}_1 = x_{\min} = -2$. Q_i is then defined as: $Q_i = \sum_{j=1}^{i} y_{j,t}$

Let Q_u and Q_{u+1} be the lower and upper cumulative proportion of individuals that contains the percentile of interest q therefore $q \in (Q_u; Q_{u+1}]$ for the intervals $(x_u, x_{u+1}]$ and $(x_{u+1}, x_{u+2}]$ and $\bar{z}_u, \bar{z}_{u+1}$ be the associated antibody level medians for those intervals.

The antibody level below which q% of the population is, is defined as:

$$\bar{z}_q = \bar{z}_u + (\bar{z}_{u+1} - \bar{z}_u) * \frac{q - Q_u}{Q_{u+1} - Q_u}$$

Note if we are interested in the median antibody level for an specific age range we use *q*=0.5

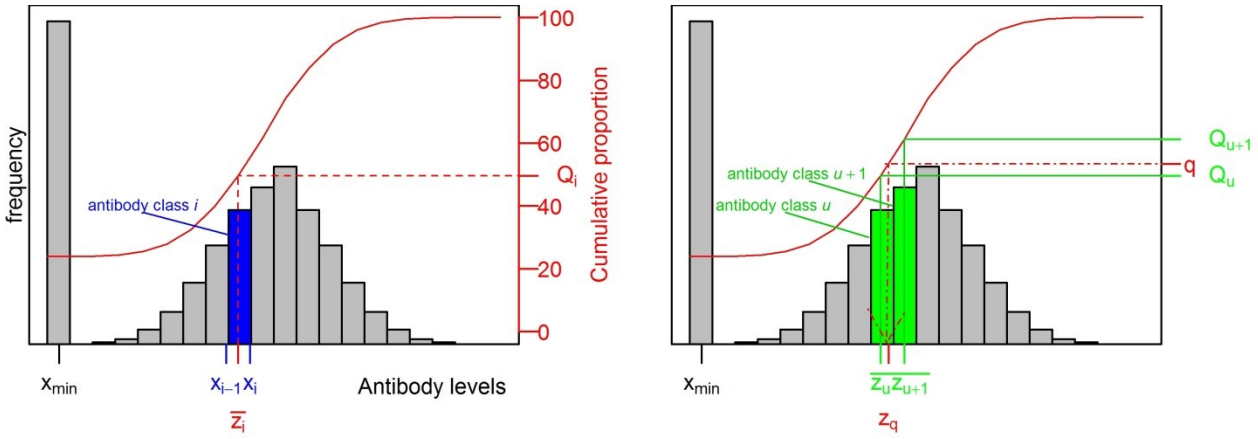


Figure 7.1 : (A) Schematic representation of the derivation of antibody levels from proportion of individuals in antibody class i. (B) Schematic representation of the interpolation to determine antibody level based associated with the percentile q.

Appendix III - Impact of the number of compartments for discretisation

The density model developed in Chapter 3 was a continuous model discretised by dividing the range of the antibody titre into *N=51* compartments. It is worth assessing whether changing the number of compartments and therefore the size of the intervals would affect the estimates of exposure rate. For this purpose, I carried out the parameter estimation using both *N=41* and *N=61*. The resulting estimates are shown in Figure 7.2. The change in the number of compartments appears to not significantly affect the any of the biological parameters (the maximum boost size a, the mean boost size for individuals with no circulating antibodies η and the variability amongst individuals, S). The exposure rate does not seem to be influenced by the number of compartments. As long as it is large enough, the number of compartments (and therefore the size of the intervals) does not significantly affect the exposure rate estimates.

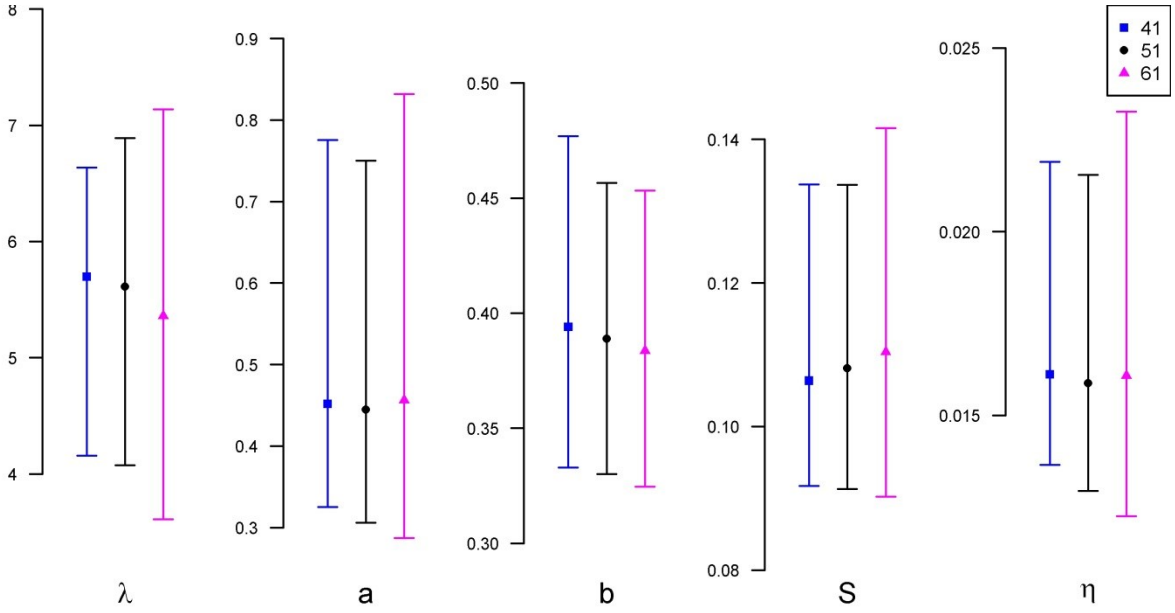


Figure 7.2 : *Posterior* 95% credible intervals for all the parameters of the model estimated with models using different numbers of compartments for the discretisation of the continuous antibody titre. The original discretisation being N=51(●), with results also shown for N=41(■) and N=61(▲).

Appendix IV -MCMC diagnostics for Tanzania dataset

In Chapter 4, I estimated simultaneously all exposure parameters assuming all other parameters were constant across villages. A total of 16 parameters were estimated, namely 12 for exposure rate for each village, the maximum boost size *a* ,the slope of the decrease *b,* the standard deviation *S* and the average antibody boost size for individuals with no current antibodies *η*. Figure 7.3 presents the MCMC outputs with the trace and the posterior distribution for each parameter. Visual inspection indicates convergence of the markov chains. The marginal *posterior* distributions obtained were informative with relatively narrow credible intervals and an approximately normal distribution.

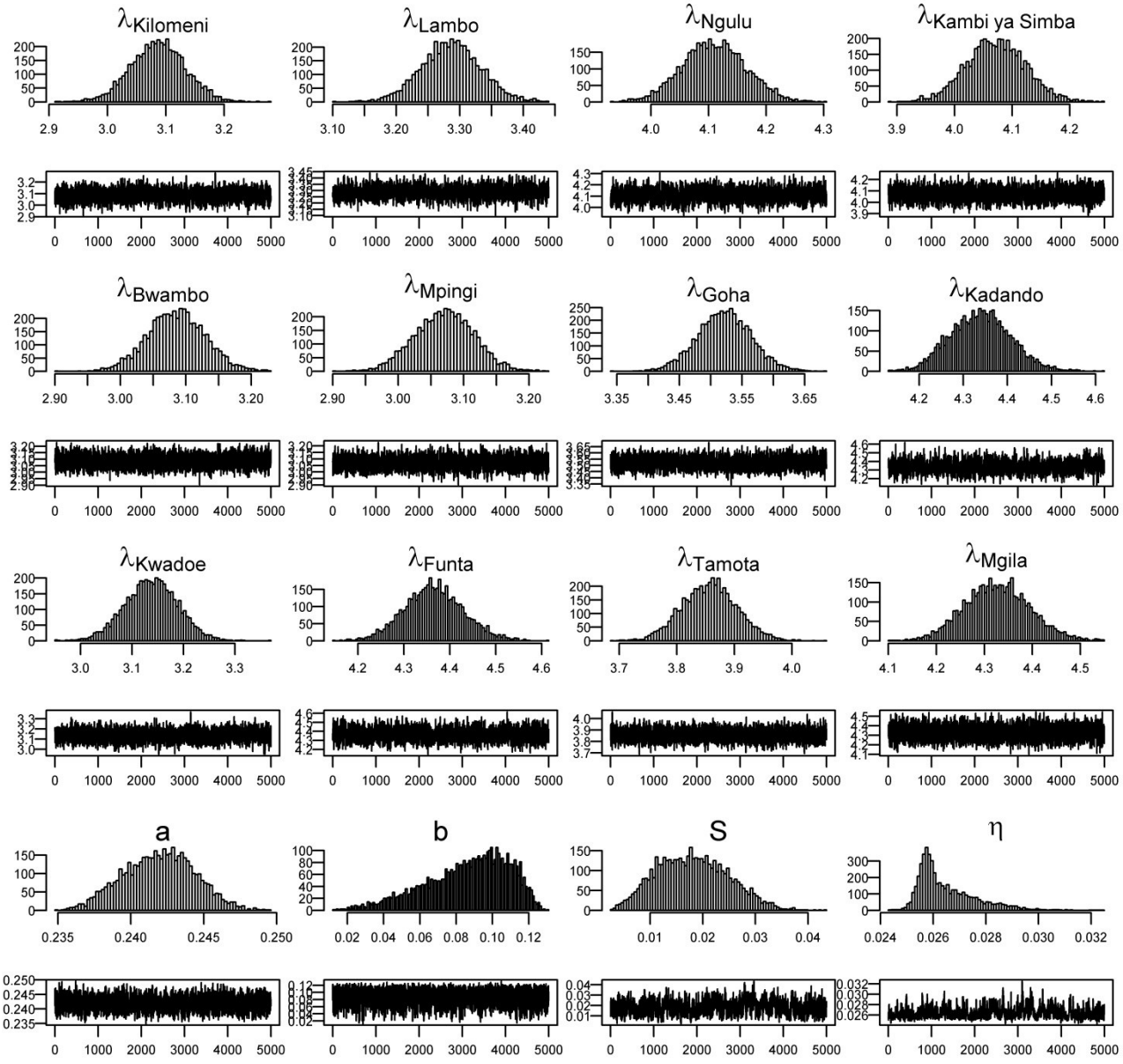


Figure 7.3: MCMC posterior distribution and trace for exposure rate (yrs-1) for each village and maximum boost size (a) , slope of the decrease (b), standard deviation (S) and the average antibody boost size for individuals with no current antibodies (η).

Appendix V - MCMC diagnostics for anti- AMA-1 antibodies

In Chapter 5, I performed the parameter estimation using anti- AMA-1 antibody data from Somalia, The Gambia, Bioko and Uganda. A total of 8 parameters were estimated, namely 4 for the exposure rate for each country, the maximum boost size *a* ,the slope of the decrease *b,* the standard deviation *S* and the average antibody boost size for individuals with no current antibodies *η*. Figure 7.4 presents the MCMC outputs with the trace and the posterior distribution for each parameter. Visual inspection indicates convergence of the Markov chains but the *posterior* distribution of some of the parameters appears to be bimodal. The marginal *posterior* distributions obtained around both modes are relatively well distinguishable. These results imply that, given the data, the model is not capable of distinguishing between the two modes. These modes appear to suggest that either the exposure is high and the maximum size of the boost is low or vice versa (correlation between *λ* and *a*).

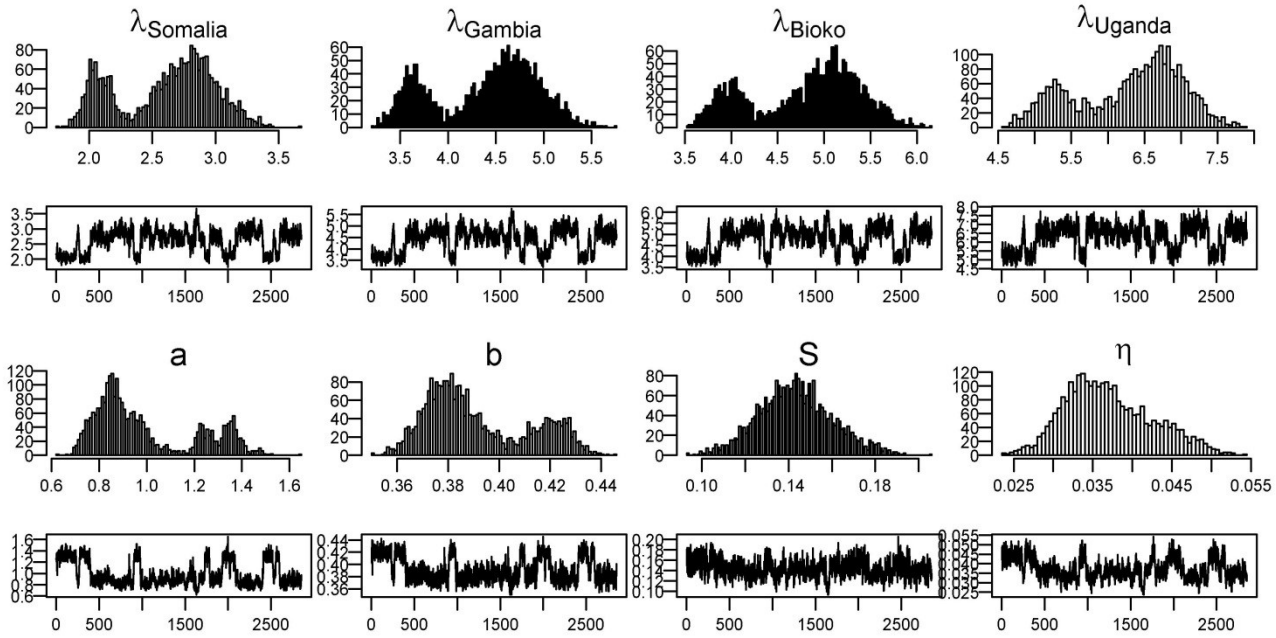


Figure 7.4: MCMC *posterior* distribution and trace for exposure rate (yrs^{-1}) for each country and maximum boost size (*a*) , slope of the decrease (*b*), standard deviation (*S*) and the average antibody boost size for individuals with no current antibodies (*η*).

Appendix VI -Additional results for change in transmission in Bioko

Figures 7.5 and 7.6 present the surface profile likelihood when a change in transmission is assessed for each of the six regions in Bioko. Each figure present two of the three parameters for which the likelihood was evaluated and the maximum value of the likelihood over the third parameter was reported.

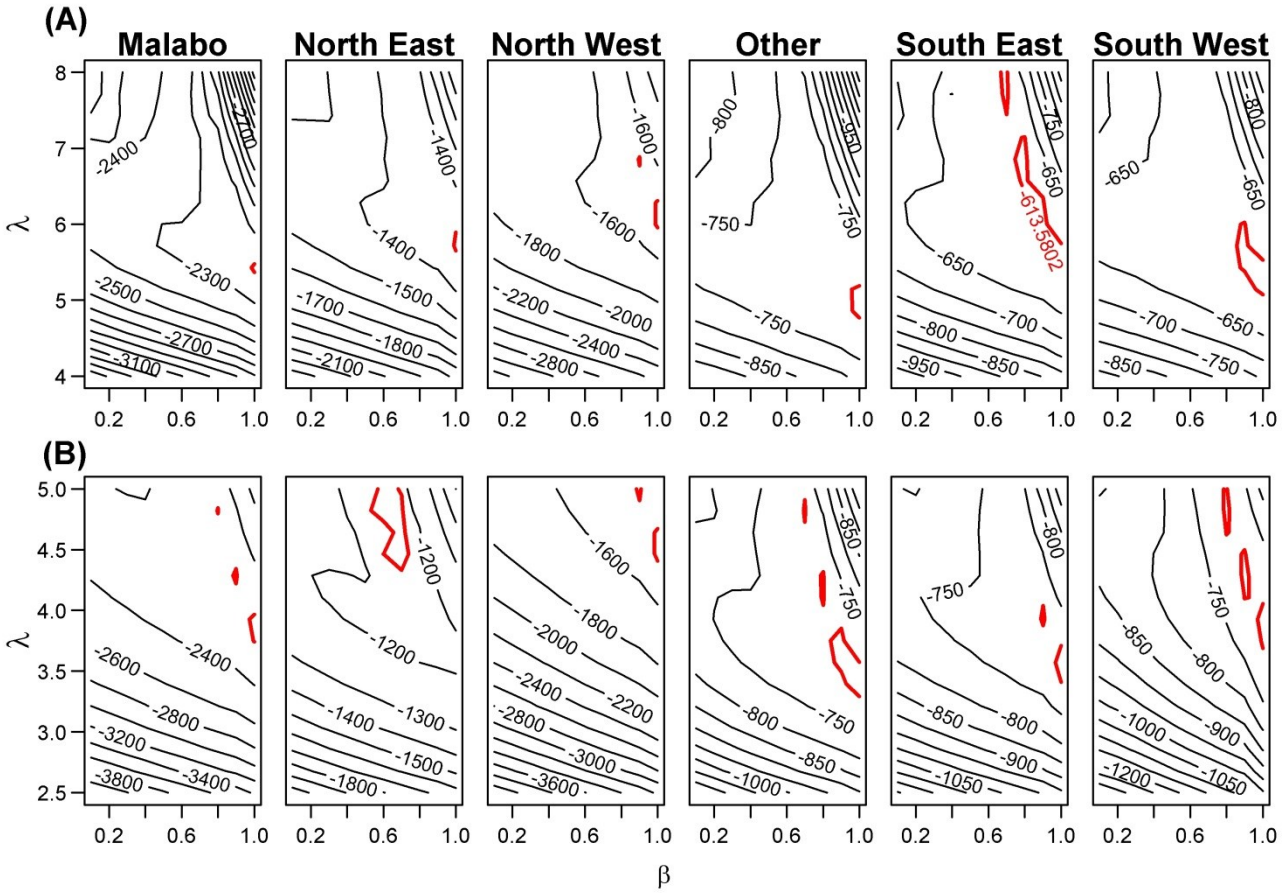


Figure 7.5: Log Likelihood computed for varying values of the exposure rate (x-axis), scaling factor (x-axis) and the time since the change of transmission intensity for both MSP-1 (A) and AMA-1 (B) antigens for various regions in Bioko. 95% confidence intervals are presented in red. Only the maximum likelihood for the different values of the time since exposure was presented.

The uncertainty around the parameters is relatively small for the exposure rate λ and the scaling factor β. However, the uncertainty around the date at which the change happens is relatively high for all regions except for the North East, when looking at AMA-1 results.

The multiple modes for the exposure rate could be simply explained by the fact that the grid requires more points to explore the likelihood surface. Or this result could also be explained by multiple or continuous drops in transmission that the model is trying to cast in a single change in transmission. For the most interesting result, in the North East, using the AMA-1 antigen, the distribution is however uni-modal.

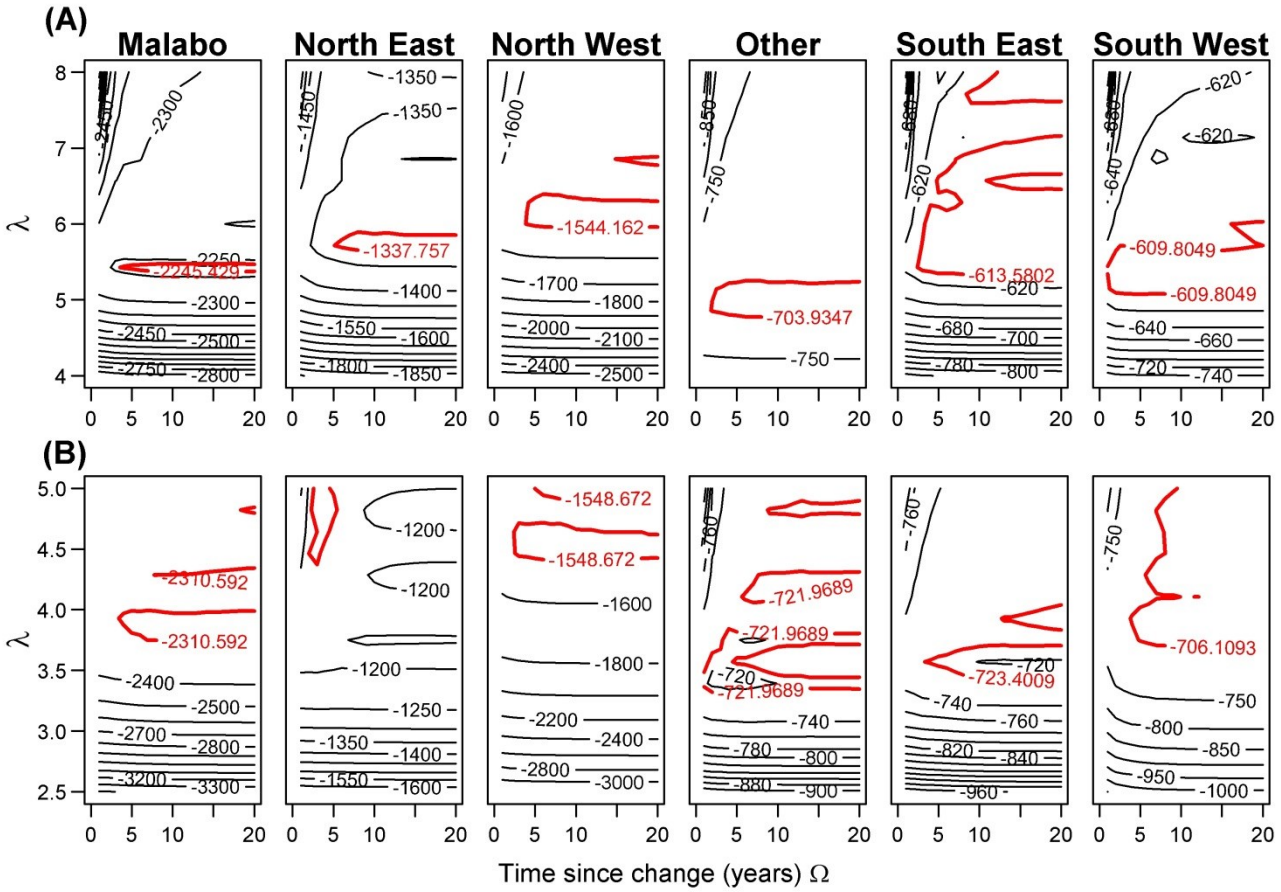


Figure 7.6: Log Likelihood computed for varying values of the exposure rate, scaling factor (y-axis) and the time since the change of transmission intensity (x-axis) for both MSP-1 (A) and AMA-1 (B) antigens for various regions in Bioko. 95% confidence intervals are presented in red. Only the maximum likelihood for the different values of the scaling factor was presented.

Appendix VII - Copyright

Page Number	Type of work: text, figure, map, etc.	Source work	Copyright holder & year	Work out of copyright	Permission to re-use
Page 20	figure	Malaria Journal (2011), vol 10, 378	© Malaria Journal 2011		✓
Page 23	figure	Nature reviews Microbiology (2010), vol 8,272-80	© Nature 2010		✓
Page 24	figure	Nature immunology (2008), vol 9, 725-732	© Nature 2008		✓
Page 25	figure	The American Journal of Tropical Medicine and Hygiene (2004), vol 71(2 Suppl), 80-6	© Am J Trop Med Hyg 2004		✓
Page 45	map	www.sithi.org	NA (CC-BY-NC-SA) http://creativecommons.org/licenses/by-nc-sa/3.0/.	✓	
Page 100	figure	Proc Natl Acad Sci USA (2005), vol 102, 5008-5113	© National Academy of Sciences, USA 2005		✓