Author/s:
Dharmaratne, Aluthwala Domingu Vithanage Tharkeshi Thanuja

Title:
Quantification of the dynamics of antibody response to malaria in pregnant women

Date:
2020

Persistent Link:
https://hdl.handle.net/11343/245492

Quantification of the dynamics of antibody response to malaria in pregnant women

A. D. V. Tharkeshi Thanuja Dharmaratne

ORCID identifier: 0000-0002-0788-7252

Master of Philosophy

April 2020

Centre for Epidemiology and Biostatistics

Melbourne School of Population and Global Health

The University of Melbourne

Submitted in total fulfilment of the requirements of the degree of Master of Philosophy

Produced on archival quality paper

Abstract

Malaria remains a major public health threat with pregnant women and young children the most vulnerable populations. Approximately 125 million pregnant women are at risk of malaria infection every year. Immunity to malaria can significantly reduce the severity of malaria symptoms. This has long motivated efforts to better understand the immune response to malaria in order to develop an efficacious antimalarial vaccine. However, the complex behaviour of the immune system, particularly in pregnant women, has hindered identifying the underpinning mechanisms of the immune response to malaria infection.

This project investigates antibody-mediated immunity in pregnant women who attended antenatal clinics located at the Thai-Myanmar border, where malaria transmission is low and *P. falciparum* and *P. vivax* (the most prevalent malaria species globally) coexist. Antibody responses during pregnancy to six parasite antigens were measured for 250 pregnant women in a median of 7 samples per woman (range 2 to 13) over the gestation period.

A multivariate mixture linear mixed model was fitted to longitudinal antibody data of 250 pregnant women to characterise the highly dynamic antibody responses using a Bayesian approach. The posterior distribution of the parameters estimated via Markov chain Monte Carlo (MCMC) simulations were used to classify the antibody responses. The results show that the infectious status of a woman during follow-up is a key factor influencing the classification of the joint behaviour of the antibody responses. Hence, the two malaria exposure categories (exposed to infection during pregnancy defined as a case and those non-exposed defined as controls) were represented by the two antibody profile clusters.

Using a manually developed code, entropy values were computed for each antibody, with which, the contribution made by each antibody on the classification was assessed. The antibodies *Pf*MSP3 and *Pv*AMA1 which maintained less dynamic antibody profiles significantly influenced the classification. However, these antibodies identified controls exceptionally well but did not perform well for the cases. For sero-surveillance, antibodies which best identify the cases are required, hence, the study was extended in performing classification based on all possible univariate and multivariate combinations of the six antibodies. The bivariate combination *Pf*AMA1 and *Pf*VAR2CSA resulted in identifying the majority of cases by contributing towards identification of potential

biomarker(s) for sero-surveillance of recent exposure to malaria during pregnancy. Therefore, the bivariate combination of antibodies, *Pf*AMA1 and *Pf*VAR2CSA should be used in the field, particularly for accurate malaria surveillance of pregnant women living in low malaria transmission settings. This could lead to the early detection and treatment of malaria infections in pregnant women, reducing transmission and thereby progressing towards elimination of malaria.

Declaration

This is to certify that:

(i) the thesis comprises only my original work towards the Master of Philosophy,

(ii) due acknowledgement has been made in the text to all other materials used,

(iii) the thesis is fewer than the maximum word limit in length, exclusive of tables, maps, bibliographies and appendices.

Signature: _____

Date: _____

Preface

This thesis makes use of a subset of data from the longitudinal study performed by Fowkes *et al.* (2012) [1] based on a group of pregnant women who attended the antenatal clinics of the Shoklo Malaria Research Unit (SMRU) located at the Thai-Myanmar border. This study is conducted in partnership between the SMRU and the Burnet Institute of Medical Research, Australia. I gained access to the data when I started my research in 2018.

Acknowledgements

I would like to express my sincere appreciation to my Principal Supervisor Prof. Julie Simpson and the Co-supervisors, Prof. Freya Fowkes, Dr. Sophie Zaloumis, Dr. Saber Dini and Dr. David Price for their continuous guidance and support over the process of this MPhil. I would also like to thank the members of my advisory committee, Prof. Shyamali Dharmage and Dr. Angela Devine for their valuable advice provided during the advisory committee meetings. Special thanks to Dr. Katherine O'Flaherty for reviewing the thesis and sharing valuable comments on the thesis. I would also like to extend my sincere thanks to the University of Melbourne for funding over the course by offering me a Melbourne Research Scholarship, a full fee remission scholarship.

I would also like to thank my friends and colleagues for helping me to get adjusted to the new environment, for being supportive and encouraging, especially during hard times. Last, but not least, I would like to acknowledge with gratitude, the support and love of my family. Thank you, my beloved husband, Nadeera Bandara, for your dedication, persistent support and love throughout the research. I owe special thanks to Kumara Dharmaratne (father), Renuka Dharmaratne (mother), Kushini Dharmaratne (sister) and Matilda Jayasekera (aunt) for encouraging me, keeping me harmonious and always making me feel that you always remain close to my heart even though we live physically miles apart. This accomplishment would not have been possible without them. Thank you.

Table of Contents

List of Tables

List of Figures

Abbreviations

ACTs	Artemisinin-based combination therapies
ADCI	Antibody-dependent cellular inhibition
AMA1	Apical membrane antigen 1
ANCs	Antenatal clinics
CDFs	Cumulative distribution functions
CQ	Chloroquine
CrI	Credible interval
CI	Confidence interval
CSA	Chondroitin sulphate A
DIC	Deviance information criteria
EBA	Erythrocyte binding antigens
EBL	Erythrocyte-binding-like
ELISA	Enzyme-linked immunosorbent assay
EM	Expectation-Maximisation
G6PD	Glucose-6-phosphate dehydrogenase
GA	Gestational age
GpA	Glycophorin A
GPI	Glycophosphatidylinositol
IEs	Infected erythrocytes
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IPTp	Intermittent preventive therapy in pregnancy
ITNs	Insecticide-treated bed nets
IUGR	Intrauterine growth retardation
IQR	Inter-quartile range
LBW	Low birth weight
LOESS	Locally estimated scatterplot smoothing
MCMC	Markov chain Monte Carlo
MLMM	Multivariate linear mixed model
MMLMM	Multivariate mixture linear mixed model
MSPs	Merozoite surface proteins
NAI	Naturally acquired immunity

NAT	Nucleic acid-based tests
OD	Optical density values
P.	*Plasmodium*
Pf	*P. falciparum*
Pv	*P. vivax*
PCR	Polymerase chain reaction
PED	Penalised expected deviance
RBCs	Red blood cells
RHS	Right-hand side
SMRU	Shoklo Malaria Research Unit
VSAs	Variant surface antigens

Chapter 1

Introduction

1.1 Background

Malaria is a major threat to public health. This vector-borne infectious disease caused 228 million clinical cases and 405,000 deaths in 2018 [2]. Malaria infection is caused by the *Plasmodium* spp. parasite and gets transmitted to humans through the bite of an infected female *Anopheles* mosquito. Of the six species of *Plasmodium* parasite that can infect humans, *P. falciparum* and *P. vivax* are the most prevalent [3]. *P. falciparum* can cause the most severe consequences compared to the other species and *P. vivax* is the most widespread malaria species globally. The highest burden of malaria has been reported in Africa, where *P. falciparum* malaria is holoendemic in some regions and high levels of malaria transmission have been observed [4]. Outside Africa, particularly in South Asia, Southeast Asia and South America, the level of malaria transmission is significantly lower and *P. falciparum* and *P. vivax* are co-endemic.

Every year, around 125 million pregnant women living in malaria endemic countries are at risk of malaria infection [5, 6]. The high susceptibility of pregnant women and children has led WHO to pay close attention to these two high risk groups, highlighted in the most recent world malaria report [2]. Pregnant women exposed to malaria may suffer from anaemia which strongly affects their health and accounts for around 10,000 maternal deaths every year, particularly in Sub-Saharan Africa [7-9]. Pregnant women may also suffer from severe complications of *P. falciparum* infection caused by sequestration of parasites in the placenta [10]. The burden of malaria on pregnant women living in areas where malaria transmission is high, tends to alter with the age of the mother [11-13] and parity [14], and the susceptibility to severe consequences of malaria is reported to be highest in younger pregnant women who are in their first pregnancy (primigravidae) compared to older pregnant women with multiple pregnancies (multigravida). This is mainly because with repeated exposure to the parasite over several pregnancies immunity that is protective against placental sequestration may develop [15, 16].

The adverse consequences of malaria during pregnancy are not limited to the mother but can also extend to the child. Maternal malaria impacts the development of the child during

the stages of foetus (unborn; more than eight weeks after conception throughout the gestation period), neonate (newborn) and infancy. The adverse effects related to malaria in pregnancy are miscarriage [17] and stillbirth of the foetus [18], neonatal and infant mortalities [9, 19]. Infants who survive have an increased risk for low-birth weight [8] which in turn impacts the development of infants and may cause several diseases (such as metabolic syndrome, non-communicable diseases and certain cancers) in adult life [20-22]. Of note, the possibility of sequestration of *P. falciparum* parasites in the placenta is identified as a leading cause of neonate and infant mortality [19]. Hence, developing effective treatments for malaria, particularly for placental malaria is of great importance to prevent the adverse consequences associated with malaria during pregnancy.

Various drug regimens and prevention strategies developed to treat and control malaria infection have successfully reduced the global incidence of malaria and have led many countries towards elimination of malaria [4, 23]. However, some of these efforts have begun to lose efficacy, particularly in low malaria transmission settings such as Southeast Asia, due to emerging parasite drug resistance [24] and insensitivity of standard surveillance methods in detecting submicroscopic and asymptomatic infections [25]. These impediments have therefore warranted the need for development of an efficacious population-level intervention, such as a vaccine.

An efficacious antimalarial vaccine for pregnant women is yet to be developed. The complex response of antibody-mediated immunity to malaria has hampered the complete understanding of how immunity to malaria can be acquired following repeated infections [26, 27] and why its provided protection persists only if the infections occur regularly, otherwise, the protection diminishes rapidly [28, 29]. Quantifying the antibody responses to malaria during pregnancy is more challenging due to the suppression of immunity during pregnancy and the great variations in antibody responses between and within pregnant women [1].

Characterising the behaviour of the immune system in response to malaria infection, particularly antibody-mediated immunity, will also provide a means to an accurate sero-surveillance tool. This surveillance technique, which measures serologic responses (anti-malarial antibodies) against the antigens expressed at different stages of the malaria parasite, is a powerful technique for detecting malaria infections, particularly where parasite density is significantly low [25]. The biomarkers provided by serological

surveillance (sero-surveillance) allows for monitoring recent exposure, transmission level and acquired immunity to malaria [30-32].

Antibodies targeting blood-stage antigens, predominantly the antigens expressed on the merozoites, have been the main focus of current vaccine development research. However, previous studies have commonly investigated a small selection of the many merozoite antigens [33] and only a few studies have examined the protective role of antibodies to multiple antigens simultaneously [34-38]. Of note, many of the candidate blood-stage antigens found by the previous studies are identified as biomarkers of protective immunity [30, 33, 39] while the biomarkers representing recent exposure are lacking.

This thesis is aimed at quantifying the dynamic antibody responses to multiple antigens, thereby, identifying the potential biomarkers of immunity and/or exposure. A multivariate model is developed in this work to study the behaviour of the antibody responses to multiple antigens jointly by accounting for the correlations between the antibodies. In addition, antibody responses to commonly studied merozoite antigens, considered important vaccine targets, are analysed in this work, along with the antibody responses to VAR2CSA, a pregnancy-specific erythrocytic antigen.

1.2 Aims of the thesis

The aims of this thesis, which centre around analysing multiple longitudinal measures of antibody levels simultaneously recorded from pregnant woman, are as follows:

i. Identify clusters of pregnant women with similar antibody responses to malaria.
ii. Identify the key maternal factor(s) that best discriminate the antibody groupings of the pregnant women.
iii. Determine the antibody response(s) which are potential biomarker(s) for sero-surveillance of recent exposure, transmission and acquired immunity to malaria.

1.3 Thesis structure

This thesis is structured as follows. Chapter 2 provides a literature review for malaria in general and the burden of malaria particularly in pregnant women. This chapter consists of three main sections. The first section introduces the global problem of malaria, interventions and prevention strategies developed to control and eliminate malaria globally, introduction of the burden of malaria in pregnant women and effects of malaria on the mother and the child. The second section introduces immunity; the natural defense

mechanism against malaria, the changes that occur in immunity to malaria during pregnancy and the factors which affect the immunity levels of pregnant women against malaria. In the third section, I explain the importance of antibody-mediated immunity in controlling malaria and briefly explain the function of the antigens, against which antibody measures were obtained in the dataset I analysed in this thesis.

Chapter 3 introduces the dataset analysed in this study. It briefly explains the population of the pregnant women involved in the primary study and the intensity of malaria in the study area, study design for the secondary study and the procedure of cleaning the secondary data to incorporate in the current study.

Chapter 4 introduces statistical methods available in the R software, which can cluster pregnant women with similar characteristics with respect to the joint behaviour of the trajectories of the six antibody responses. This chapter initially introduces the packages which perform clustering based on single antibody trajectories followed by the statistical methods which account for joint-trajectories to cluster the pregnant women. The multivariate classification methods include both distance-based methods as well as model-based clustering methods.

Chapter 5 summarises the characteristics of the participants described in Chapter 3 and presents the corresponding longitudinal antibody profiles across the six antibodies. The best statistical method introduced in Chapter 4 was then applied to classify the longitudinal antibody profiles. The importance of the antibodies in classifying the pregnant women into clusters that distinguish between the exposure categories (cases and controls) was also investigated.

Chapter 6 concludes the thesis with a summary of the findings and a discussion of the implications of this research. This chapter also contains a discussion of the strengths and limitations of this research and suggestions for the immunologists who work in this area as well as some extensions that can be made in the methodology.

Chapter 2

Background to malaria and the burden in pregnancy

2.1 Malaria

Malaria is a vector borne infectious disease with a large burden in 91 counties [23]. In 2018, around 228 million clinical cases and 405,000 deaths were attributable to malaria across the world [2]. This infection is caused by the *Plasmodium* parasite and transmitted to humans through a bite from an infected female *Anopheles* mosquito. Six species of *Plasmodium* spp. can infect humans (*P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale wallikeri*, *P. ovale curtisi* and *P. knowlesi* (a zoonotic infection)) [40, 41], among which *P. falciparum* and *P. vivax* are more prevalent [3].

The highest burden of malaria is in Sub-Saharan Africa, which observed 93% of global malaria cases and 94% of all malaria deaths recorded in 2018 [2]. *P. falciparum* predominately affects this region and solely accounted for 99.7% of estimated malaria cases that occurred in Africa in 2018 [2]. Outside Africa, *P. falciparum* and *P. vivax* often coexist while *P. vivax* is the most commonly found parasite in the Americas, causing 75% of the malaria cases in 2018. Furthermore, according to the World Malaria Report 2019, 53% of the global endemicity of malaria due to *P. vivax* infection is in the South-East Asia Region [2].

Malaria infection within a human commences when a mosquito bites a human for a blood meal and the *Plasmodium* spp. parasites, in the form of sporozoites, are inoculated into the bloodstream [42]. These sporozoites then migrate to the liver and invade hepatocytes [43], where they mature into merozoites and are released into the blood stream to invade red blood cells (RBCs). The merozoites that succeed to invade RBCs begin to develop through ring, trophozoite and schizont stages. Once the schizonts become mature they rupture and release on average 8-16 [44, 45] daughter merozoites into the blood, continuing the cycle of invasion and destruction of RBCs. Of note, unlike *P. falciparum*, a proportion of *P. vivax* sporozoites remain dormant in the liver as hypnozoites, which can reactivate after several weeks or months of the primary infection, causing infection-relapses [46].

Malaria symptoms, which commence with the release of merozoites into the blood stream, are shown to vary across different *Plasmodium* species. *P. falciparum* infections have been associated with more severe symptoms in contrast to *P. vivax* infections [47, 48]. The level of parasite density is significantly associated with severe complications of malaria [49, 50] and mortality [51]. High levels of parasitaemia is exhibited in falciparum malaria as this species is capable of infecting erythrocytes of any age [52]. According to the World Health Organization (WHO), hyperparasitaemia is a leading cause of severe falciparum malaria [53]. Severe symptoms are also caused by *P. falciparum* due to the adherence of the infected RBCs to the vascular endothelium, known as sequestration [54-57]. In contrast, *P. vivax* only invades the youngest RBCs, known as reticulocytes and the absence of the mechanism of sequestration leads *P. vivax* to cause less severe consequences compared to *P. falciparum* [58, 59].

2.1.1 Immunity: The natural defense mechanism against malaria

The immune system naturally offers protection to the human body against malaria infections and it mainly comprises three components: (i) innate immunity, (ii) acquired or adaptive immunity and (iii) passive immunity. Innate immunity is present from birth and reacts against all infections in a similar manner (non-specific immunity). It provides a first-line defense as soon as a pathogen is recognised, and if unable to provide protection, it mediates activation of subsequent acquired immunity responses. Acquired immunity is not present at birth but can develop later in life with repeated exposure to malaria. Antibodies are an important component of naturally acquired immunity (NAI) and are generated to attack specific antigens that mediate the infection. Lastly, passive immunity is not generated by one's immune system but instead antibodies are transferred from another source. Maternally derived antibodies are one such instance, where this naturally acquired protective immunity is transferred to the infant from mother through the placenta and is able to reduce the risk of clinical malaria [60]. However, maternally-derived protective immunity does not last long, and maternal antibodies offer protection only among young infants typically for the first five months of life [61-63].

Individuals living in malaria endemic areas can develop immunity against malaria parasites after repeated infections [26, 27], which can reduce the clinical symptoms [64]. Some studies have revealed that human immunity, especially the antibody-mediated immunity, can only provide long-term protection if individuals at risk are frequently

exposed to malaria as the antibody responses are relatively short-lived and its effect wanes rapidly, probably within a few months after parasite clearance [65, 66]. Opposingly, several other studies have argued that antibody responses are long-lived, mostly for years, in the absence of exposure to malaria [1, 67-69]. Of note, these contradictory findings are due to the biphasic decay of an antibody response. An antibody response usually tends to decay under two phases after approaching the peak; initially, with a rapid decay followed by a slower decay in the second phase (reviewed in [70]). Hence, the studies which investigated the first phase of the decay of an antibody have reported shorter antibody response half-life [66, 71] compared to that of second phase antibody decay studies [1, 67].

2.1.2 Control and elimination of malaria

Malaria control efforts over the past two decades have significantly reduced the malaria transmission intensity in many malaria-endemic countries [72-74]. Despite the significant progress that has been made to reduce the global malaria burden, a recent WHO report shows that the decline in burden and deaths have stalled [4]. Resistance to the artemisinin derivatives has now spread across South East Asia [24], and more recently there are reports of declining efficacy for the artemisinin-based combination therapies (ACTs), due to an emerging resistance of the parasites to the partner drug as well as the artemisinin derivative [75].

Various drug regimens and prevention strategies have been developed to treat and control malaria. Particularly, the use of insecticide-treated bed nets (ITN), ACTs and intermittent preventive therapy in pregnancy (IPTp) have successfully reduced the global incidence of malaria and have led many countries, towards elimination of malaria [4, 23].

Many of these interventions have primarily focused on controlling *P. falciparum* infections, resulting in a significant progress in reducing the burden of malaria in the areas where this species is prevalent, particularly in Africa. This, however, has led to an underestimation of the malignancy of *P. vivax* [76], which has caused significant morbidity and mortality in South Asia, Southeast Asia and South America [58, 77].

Currently, there is not a globally certified radical cure for *P. vivax* to completely clear the hypnozoites (liver-stage parasites) from the body [78] - ACTs kill only blood-stage parasites, thus, do not affect hypnozoites, which can induce progressive relapses of

infection. Although the antimalarial primaquine is capable of killing hypnozoites, its deployment has been disputed, as it can cause haemolysis in patients with glucose-6-phosphate dehydrogenase (G6PD) deficiency [79]. In addition, ITNs have shown to be less effective in reducing *P. vivax* transmission, further implying that the standard control strategies are not completely efficacious in controlling *P. vivax* [80].

In low transmission areas that are co-endemic for falciparum and vivax malaria, accurate surveillance is challenging, as submicroscopic, asymptomatic and sub-patent infections (particularly, infections with low parasite densities) are highly prevalent [81-84] but remain undetectable by the mostly commonly used diagnosis methods (such as microscopy and RDTs, which measure the parasite density in blood smears and detect malaria antigens, respectively) [85, 86]. This is a major impediment to malaria elimination efforts, as these asymptomatic infections silently contribute to human-to mosquito-transmission [25, 84, 87-89]. Hence, traditional diagnostic and surveillance techniques, which depend on clinical case reports from health services, entomological estimates and case estimates from cross-sectional surveys, become less accurate in capturing the true intensity of infection [89, 90].

Several researches have suggested that Polymerase chain reaction (PCR)-based diagnostic methods (such as conventional PCR and quantitative real-time PCR(qPCR)) [91, 92] and nucleic acid-based tests (NATs) [82, 93] perform well in detecting infections with low parasite densities compared to microscopy/RDT tests [89, 94]. qPCR is particularly accurate in diagnosing infections with as low as 0.5–10 parasites/μl of blood (sub-patent infections) [95] compared with microscopy that has a detection limit of 50–500 parasites/μl [25, 96]. However, despite the proven accuracy of PCR-based diagnostic methods [97-99], microscopy remains the standard diagnostic method for detecting exposure to malaria, particularly in remote areas due to scarce resources, the requirement of highly trained staff and the complexity and expense of PCR [100, 101].

Sero-surveillance, which measures anti-malarial antibodies, is an alternative method to assess whether recent infections have occurred and thus monitor transmission [25], particularly for *P. falciparum* and *P. vivax* [31]. Sero-surveillance is also a useful tool to monitor both clinical and asymptomatic infections and changes in malaria transmission intensity [102-104], particularly in low transmission settings [105, 106]. Antibody levels tend to play an important role under this circumstance as they are key measures of NAI

and tend to increase following repeated exposure to malaria infections. Hence, the antibodies could be sensitive marker(s) of current as well as previous exposure to malaria, particularly in low transmission areas where infection rates are low and are less likely to capture active infection [90, 104]. Studies have also introduced several antibodies as potential biomarkers for monitoring the risk of exposure to falciparum malaria such as *Pf*MSP2, *Pf*AMA1 and *Pf*EBAs [32, 107-109]. Moreover, the sero-conversion rates of the antibodies to *Pf*AMA1 and *Pf*MSP1-19 are being used to capture changes in malaria transmission intensity [110-113] and identify hotspots of malaria transmission [114-116].

For malaria control, serological studies have demonstrated that antibody levels can provide insights into both clinical immunity and recent exposure. Certain antibodies, particularly against relatively conserved merozoite antigens, are maintained at high levels and are associated with reduced clinical symptoms, hence, they are characterised as biomarkers of protective immunity [30, 33, 38, 117]. However, there are other antibodies which indicate individuals with recent exposure (i.e. who experienced clinical symptoms) instead of protection against malaria [32, 118]. Some of the antibodies, particularly, against merozoite antigens could play the dual role of biomarkers of protective immunity and recent exposure to malaria according to the transmission setting (i.e. perform as biomarkers of recent exposure to malaria in areas or populations with limited exposure such as in young children and in low transmission settings and perform as biomarkers of protective immunity in high transmission settings where exposure to malaria is frequent [32]). To date, serology has been widely examined as a potential marker of protective immunity, therefore, identifying the antibodies which are markers of recent exposure to malaria is important for informing elimination strategies of malaria in low-transmission areas.

2.1.3 Malaria in pregnancy

Every year, around 125 million pregnant women are at risk of malaria infection [5, 6], 24% of whom live in Africa, where *P. falciparum* is prevalent, while the rest reside out of Africa in areas where *P. falciparum* and *P. vivax* may coexist. (around 90 million live in the Asia-Pacific region [6, 119] and 3 million live in Latin America [6]). Pregnant women in malaria endemic regions are highly susceptible to malaria infection due to the immunological and hormonal changes that occur during pregnancy and the increased attraction of mosquitoes to them [120-122]. Consequently, they are more likely to get

infected with and suffer from malaria compared to non-pregnant women living in the same area [123].

The symptoms of malaria can be significantly exacerbated during pregnancy. During *P. falciparum* infection pregnancy is associated with increases in parasite density and sequestration of infected erythrocytes to the placental endothelium [10].

2.1.3.1 Effects of malaria on the mother

Pregnant women infected with *Plasmodium* spp. may suffer from acute lung injury, severe hypoglycemia, coma [124], and anaemia, which is the major cause of adverse outcomes in the mothers [7] - around 10,000 maternal deaths occurred every year in sub-Saharan Africa are due to severe maternal anaemia [9].

The adverse effects of malaria during pregnancy, particularly maternal anaemia, seem to vary with the transmission level in the region and gravidity. Pregnant women, primigravidae in particular, who live in areas with stable malaria transmission are more prone to experience severe anemic conditions [125-127]. In contrast, severe anaemia is less likely to be observed in areas with unstable transmission, irrespective of the prevalent *Plasmodium* spp. [128]. In the unstable transmission settings, maternal anaemia is more prevalent in primigravidae than in multigravida and in pregnant women infected with *P. falciparum* compared with *P. vivax* [129]. The latter is due to the difference in the mechanism of invading RBCs by the species; *P. falciparum* lessens the haemoglobin level by invading all RBCs, whereas *P. vivax* only invades reticulocytes, causing a lower density of parasitaemia [130]. However, it has been observed that anaemia is less likely to progress to its severe condition under this setting, irrespective of the type of malaria parasite.

The susceptibility to malaria during pregnancy in malaria endemic areas alters with the age of the mother [11-13] and parity [14]. There is growing evidence that vulnerability to malaria is highest in younger pregnant women, particularly, adolescent and young adult women compared with older women [131, 132]. Gravidity is also an important maternal factor associated with the burden of malaria. The risk of malaria infection is highest among primigravidae and declines with subsequent pregnancies [131-134]. However, gravidity has shown to be poorly associated with protective immunity in low malaria transmission settings as both primigravidae and multigravida are equally susceptible to

malaria [135-137]. The adverse effects of malaria during pregnancy are not limited to the mother but can also extend to the child [120, 138], detailed below.

2.1.3.2 Effects on foetus, neonate and in infancy

Children, particularly under 5 years of age are highly susceptible to malaria and accounted for 272, 000 (67% of the total malarial deaths) deaths in 2018, globally [2]. Malaria in pregnancy is a major cause of infant mortality and its adverse effects impact the foetus throughout the gestation period as well as the infant.

Maternal malaria increases the risk of miscarriage or spontaneous abortion [17] and accounts for the majority of pregnancy losses during the first two trimesters in malaria endemic regions [139, 140]. For example, McGready *et al.* (2012) showed that even one *P. vivax* infection during the first trimester increases the risk of miscarriage by four-fold [17]. Of note, pregnant women who live in unstable transmission settings are at higher risk of miscarriage [17].

Stillbirth is another outcome of maternal malaria which can occur to the foetus after 20 weeks of gestation. Between 126,109 and 207,971 were attributable to malaria during pregnancy worldwide in 2015 [18, 141-143]. The majority of the stillbirths associated with malaria during pregnancy were reported in sub-Saharan Africa, which is holoendemic with *P. falciparum* and accounts for 19.7% of the total stillbirths recorded in Africa [143]. A recent systematic review demonstrated that the risk of stillbirth is higher for pregnant women who are detected with falciparum malaria at delivery compared with the mothers who were detected and treated for falciparum malaria during pregnancy and the risk of stillbirth increases as the endemicity of malaria declines [18].

The harms of pregnancy-associated malaria extend to newborns. *P. falciparum* malaria in particular, is harmful to newborns due to the prevalence of placental malaria and excessive parasitaemia during the fetal stage, increasing the risk of neonatal death [19]. Infants who survive might still suffer from various illnesses as maternal malaria may cause intrauterine growth retardation (IUGR) and preterm delivery [144]. Low birth weight is the most common adverse effect on infants associated with malaria in pregnancy [8, 133, 137] and accounts for infant mortalities in malaria endemic areas each year (associated with 100,000 deaths in sub-Saharan Africa [8]) and by causing non-communicable diseases and metabolic diseases in older children [20, 22]. Low birth

weight is more probable in primigravidae and in mothers infected during the first two trimesters [133, 145, 146].

2.2 Immunity to malaria during pregnancy

Immunity is less efficacious for pregnant women as it appears to be substantially suppressed with pregnancy. Particularly, the decline of cell-mediated immunity [147, 148], which is required for clearing infections [149-151], elevates the susceptibility of pregnant women to malaria.

The level of acquired immunity varies significantly with the transmission level of the area where a pregnant woman resides. Women who live in high transmission areas, principally in Sub-Saharan Africa [152], gradually develop a strong NAI against various variant surface antigens (VSAs) expressed on the infected erythrocytes (IEs) [153]. Thus, this acquired immunity seems to be able to control the levels of parasitaemia and present fewer symptoms during the period of pregnancy, despite the pregnancy associated immunity suppression [26]. In contrast, pregnant women in low transmission areas develop little or no immunity, which makes them more susceptible to adverse outcomes of malaria [154, 155].

Gravidity is another factor that can significantly alter the humoral immunity, which is specific to pregnant women with malaria. During pregnancy, *P. falciparum* (*Pf*)- IEs can sequester in the placenta, evading pre-existing immunity. The pregnancy specific parasite antigen VAR2CSA, expressed on the IEs, facilitates placental sequestration by enabling IEs to bind to placental chondroitin sulphate A (CSA) receptors [156, 157]. Since the immune system will only encounter and recognise the VAR2CSA antigen when a woman gets pregnant, primigravidae living in malaria-endemic areas are highly susceptible to malaria as they have not acquired anti-VAR2CSA immunity to prevent the accumulation of the IEs in the placenta [15]. Conversely, multigravida women acquire antibody levels against the antigen VAR2CSA with successive pregnancies [158], which reduces their susceptibility to malaria by clearing the parasites accumulated in the placenta [159].

2.3 The antibody responses observed during pregnancy in the current study

To develop an effective vaccine, immunologists have targeted various antigens expressed by the *Plasmodium* spp. parasite [160], among which, antigens associated with the blood stage of the infection, merozoite antigens in particular, are of primary interest as they cause symptoms of malaria and are direct targets of acquired immunity [54, 161]. Understanding the behaviour of the antibody responses against merozoite antigens, is crucial to develop vaccines that disrupts the progression of merozoites to their subsequent stages. In the following sections, antibody response during pregnancy against five different merozoite antigens along with two erythrocytic antigens are described.

2.3.1 Merozoite surface proteins (MSPs)

Plasmodium spp. merozoites are covered with a thick coat of surface proteins which facilitate adhesion to and invasion of RBCs [162-165]. These surface proteins bind to merozoites in three different ways: glycophosphatidylinositol (GPI)-anchored proteins, integral membrane proteins and peripherally-associated proteins. The following subsections introduce two MSPs and their potential to be vaccine candidates.

2.3.1.1 *P. falciparum* merozoite surface protein 2 (*Pf*MSP2)

*Pf*MSP2 is a member of the GPI-anchored protein family [166], which mediate the initial attachment of merozoites to the surface of RBCs. *Pf*MSP2 is vital for invasion and remains on the surface of merozoites throughout the invasion process, but is cleaved afterwards [167]. However, studies have so far proven that only for certain variant specific antibodies to *Pf*MSP2 could provide protection if infected with the specific allele [168, 169]. Certain alleles of this antigen are considered as vaccine candidates, e.g. MSP2(3D7) allele is incorporated in a three-component blood-stage malaria vaccine [170], known as combination B, a vaccine which has successfully controlled *falciparum* malaria in malaria endemic areas during trials [171-173].

2.3.1.2 *P. falciparum* merozoite surface protein 3 (*Pf*MSP3)

*Pf*MSP3 is a peripherally-associated protein, and does not connect directly to the merozoite surface [174, 175]. Although the functionality of this family of antigens is not well understood, it has been shown to be involved in the process of RBC invasion [176]. *Pf*MSP3 is recognised as a polymorphic blood-stage vaccine candidate, and studies have

shown that acquired immunity to *Pf*MSP3 provides protection against malaria, through an antibody-dependent cellular inhibition fashion (ADCI) [177].

Apart from the proteins which are located on the surface of the merozoites, proteins which are contained within the secretory organelles of the merozoites known as rhoptries and micronemes are also involved in the invasion of RBCs. Two of the antigens which are initially located in the micronemes and are exposed on the merozoite surface during erythrocyte invasion are described in Subsections 2.3.2 and 2.3.3.

2.3.2 *P. falciparum* erythrocyte binding antigens (*Pf*EBA175)

EBA175 is an integral membrane protein, belonging to the Erythrocyte-binding-like (EBL) family of *Plasmodium falciparum* proteins, and binds to the receptor glycophorin A (GpA) on the RBCs, following the primary attachment to the RBC surface (reviewed in [178])[179-183]. *Pf*EBA175 plays an important role in the invasion process and is a leading vaccine candidate [184-187]. Particularly, region II (RII) within *Pf*EBA175 (*Pf*EBA175 RII) can bind to GpA and acquired immunity has been shown to be capable of inhibiting this mechanism [188].

2.3.3 Apical Membrane Antigen 1 (AMA1)

AMA1 is also an integral membrane protein which is common in all *Plasmodium* spp. [189, 190]. AMA1, unlike other antigens, involves invading two types of host cells: the liver cells (hepatocytes) and erythrocytes [191], by expressing on the surface of sporozoites [192] and merozoites respectively [193-195]. Specifically, AMA1 attaches to the rhoptry neck protein 2 and forms a tight junction between parasite and host, which completes the attachment to the host cells and commences invasion [196-198]. Studies conducted *in-vitro*, have shown that antibodies targeting this association are capable of inhibiting the invasion [199-201], making AMA1 a promising malaria vaccine candidate, against both *P. falciparum* and *P. vivax* infections [189, 202].

2.3.3.1 *P. falciparum* AMA1 (*Pf*AMA1)

*Pf*AMA1 is an important antigen for *P. falciparum* merozoites to invade RBCs [191, 203], hence it is considered as a candidate for vaccine against *P. falciparum* [161, 204]. *Pf*AMA1 also forms a tight junction prior to invasion of host cells [205] but the antibodies

to *Pf*AMA1 are capable of binding to this junction and inhibiting the invasion by blocking the interaction between the antigens [206].

2.3.3.2 *P. vivax* AMA1 (*Pv*AMA1)

*Pv*AMA1 is considered promising as a vaccine against *P. vivax* as its functions can be blocked by specific antibodies produced by the immune system [189]. Of note, formation of a tight junction between this antigen and RBCs during the invasion process is not observed, unlike *Pf*AMA1.

Moreover, a recent longitudinal study discovered that the performance of the antibodies of *Pv*AMA1 and *Pf*AMA1 during pregnancy are different, e.g.: anti-*Pv*AMA1 does not show a boosting profile, as opposed to anti-*Pf*AMA1, which boosts with every new infection [1].

2.3.4 Schizont extract

Schizont extract represents the proteins of infected erythrocytes at the schizont-stage of the infected RBC. Examining the behaviour of antibodies to schizont extract is very important as the proteome in this stage contains several membrane proteins, such as MSP1 and AMA1 which mediate invasion of RBCs [207, 208].

2.3.5 *Pf*VAR2CSA

Cytoadherence, which makes *P. falciparum* infection virulent, is mainly caused by the *P. falciparum* erythrocyte membrane protein 1 (*Pf*EMP1) family [209], expressed on the surface of the IEs during the trophozoite stage [210]. *Pf*VAR2CSA is the only variant of *Pf*EMP1 involved in pregnancy-associated malaria [211, 212]. This antigen facilitates IEs to sequester in placenta [213] by adhering to CSA receptors [214, 215].

*Pf*VAR2CSA can also be targeted by the antibody mediated immunity [159, 216-218], hence, *Pf*VAR2CSA antibodies that block their binding to CSA have been considered as a main vaccine candidate in pregnancy-associated malaria.

The following table (Table 1) provides an overview of the antigens, of which the antibody profiles were investigated and classified in this study.

Table 1 : Summary of the antigens investigated in the study.

Antigens analysed	*Plasmodium* spp. specific to the antigen	Stage(s) of the parasite the antigen presents	Effective/promising vaccine candidate
*Pf*MSP2	*P. falciparum*	Merozoite	Yes, certain alleles (eg: MSP2(3D7) allele)
*Pf*MP3	*P. falciparum*	Merozoite	Yes
*Pf*EBA175	*P. falciparum*	Merozoite	Yes
*Pf*AMA1	*P. falciparum*	Sporozoite, Merozoite	Yes
*Pv*AMA1	*P. vivax*	Sporozoite, Merozoite	Yes
*Pf*VAR2CSA	*P. falciparum*	Trophozoite (on IEs)	Yes

2.4 Summary

Malaria remains a major public health problem causing significant mortality and morbidity worldwide. This infectious disease is caused by the *Plasmodium* spp. parasite through a bite of the *Anopheles* mosquito. *P. falciparum* and *P. vivax* are the most prevalent species of this parasite in infecting humans. The highest burden of malaria is in Sub-Saharan Africa, holoendemic with *P. falciparum* infections, which have the highest risk of severe disease and death. Outside of Africa, malaria infection is co-endemic with *P. falciparum* and *P. vivax*.

Various global efforts have made tremendous progress in reducing the global incidence of malaria, progressing many countries towards elimination of malaria. Despite this progress, the decline in malaria incidence has been halted due to the following reasons. One of the main reasons is that the first-line treatments for uncomplicated malaria infections, such as the ACTs, are losing their efficacy due to emerging drug resistance, particularly in the Mekong region. Consequently, WHO has set the ambitious goal to eliminate malaria from at least 35 regions by 2030 before malaria control interventions are no longer effective. But eliminating malaria is challenging in areas where there are many asymptomatic malaria infections. These infections are commonly found in low malaria transmission areas and are not detected with the current standard surveillance methods, highlighting the need for new malaria surveillance methods. A promising surveillance method is sero-surveillance, which can detect recent exposure to malaria based on antibody levels.

Pregnant women in malaria endemic areas are highly susceptible to malaria and suffer from severe complications. The increased susceptibility to malaria in pregnant women is

mainly caused by: (i) cell-mediated immunity being substantially suppressed during pregnancy and (ii) the sequestration of the *P. falciparum* parasites in the placenta, evading pre-existing immunity.

The primary aim of this research is to characterise the dynamicity of the antibodies during the gestation period and investigate whether particular antigens can be used to distinguish between women exposed and not exposed to malaria during pregnancy. Such antibodies have the potential to be used as biomarkers for sero-surveillance of malaria transmission.

Antibody measurements obtained from a group of pregnant women from a low malaria transmission setting against six types of antigens (*Pf*AMA1, *Pv*AMA1, *Pf*EBA175, *Pf*MSP2, *Pf*MSP3 and *Pf*VAR2CSA) over the gestational period were analysed in this study. Of note among these antigens is *Pf*VAR2CSA, which is specific to pregnant women and it is expressed on the infected erythrocyte. The other antigens are expressed on the surface of the merozoite and facilitate the binding to and invasion of red blood cells by the parasite.

The selection of pregnant women for inclusion into the current study is described in detail in Chapter 3. Since longitudinal measures were available for each pregnant woman on multiple antibodies the clustering of pregnant women into antibody profile groupings was based on the joint behaviour of the six antibody responses. The statistical method for performing classification of multivariate longitudinal data into groupings is described in Chapter 4, and results are presented in Chapter 5.

Chapter 3

Study dataset: Longitudinal study of pregnant women from the Thai-Myanmar border

3.1 Population and study design

Data from pregnant women attending antenatal clinics (ANCs) at the Shoklo Malaria Research Unit (SMRU) [219, 220] were analysed. SMRU is located on the Thai-Myanmar border, a hilly forested area where malaria transmission is low and seasonal from May through September. In this area, infections with *P. falciparum* and *P. vivax* are the most prevalent and pregnant women are at higher risk of malaria infection due to single or co-infections of these two species [220].

Antenatal clinics

In 1986, SMRU ANCs were formed in the Maela refugee camp to reduce malaria associated maternal deaths on the Thai-Myanmar border. Since then, more than 90% of pregnant women who live in the camps have attended the ANC on a weekly basis [220]. All women are encouraged to present at an ANC as soon as they become aware of their pregnancy, after which they are frequently screened and treated for malaria until delivery. Finger pricked blood samples are taken from these women on a weekly basis to test for malaria by light microscopy. Additionally, the women are tested for anaemia every fortnight using their haematocrit level. These women are also encouraged to deliver at the SMRU delivery unit, though their tradition is to deliver at home.

3.1.1 Study setting of the primary study

The antibody measurements for the current study were obtained from a subset of women [1] included in a randomised double-blind, placebo-controlled trial conducted from November 1998 to January 2000 [221]. The main objective of the trial was to determine the safety of chloroquine (CQ) as a prophylaxis against *P. vivax* infection during

pregnancy. A total of 1000 Karen[1] pregnant women (lived either in Maela refugee camp or in the neighbourhood of Maw Ker Tai village [221]) who presented at the ANCs of the SMRU and had a negative blood smear to malaria at the time of enrolment were randomised to receive CQ or placebo. A complete physical examination was performed on each pregnant woman at enrolment and the medical history was self-reported by the pregnant women on the questionnaires provided. Of note, the number of times exposed to malaria prior to enrolment was calculated based on the records available at the SMRU before enrollment into the study.

3.1.1.1 Cases and controls selected for the longitudinal antibody measurement study and antibody determination

During the trial, all pregnant woman diagnosed with a *Plasmodium* spp. infection, were treated according to SMRU guidelines [221]. Some pregnant women who were exposed to malaria presented with fever. Of the 1000 Karen women, all 136 pregnant women who were detected with *Plasmodium* spp. infection during the trial follow-up period were assigned to the cases group in the antibody study of Fowkes *et al.* (2012) [1]. Of the 864 women who had no malaria episode detected during the study, 331 women were randomly selected to represent the control group of this study.

Antibody levels for the longitudinal analysis were then measured from all available samples of 136 cases (n=733) and a subset of the control group. The selection of controls was made based on the total IgG responses to Schizont extract from all control samples collected at enrolment (n=320, see supplementary material of Fowkes *et al.* (2012) [1]). In brief, 78 women in the control group who were identified with relatively high antibody responses at enrolment were classified as "high schizont lysate responder controls". From the rest of the controls (242), 37 pregnant women who had relatively low antibody responses at enrolment were randomly selected and denoted as "low schizont lysate responder controls" [1]. Subsequently, the antibody levels were determined at multiple time points during pregnancy for the 136 cases and the subset of 115 controls. The number

[1] Karen: A term used to define natives of eastern Burma (Myanmar) and western Thailand

of blood samples obtained from a pregnant woman ranged between 2 and 13 with a median of 7 samples.

The total IgG titer was used to detect and quantify AMA1 (*Pf*AMA1, *Pv*AMA1), *Pf*EBA175 RIII-V, *Pf*MSP2, *Pf*MSP3, schizont extract and *Pf*VAR2CSA (DBL5ε domain) using high-throughput enzyme-linked immunosorbent assay (ELISA) (supplementary material of Fowkes *et al.* (2012) [1]).

Figure 1 provides a flow chart for selecting the sample of pregnant women included in the current analysis of longitudinal antibody measures.

3.1.1.2 Determination of serostatus for each antibody

The serostatus measure, a dichotomous variable which is defined based on the magnitude of the optical density (OD) value measured against each antigen is widely used across the laboratories as a measure of presence (seropositive) or absence (seronegative) of detectable levels of a specific antibody within the serum. In Fowkes *et al.* (2012) [1] study, the cut-off for serostatus for each antigen [90] was the mean plus three standard deviations of OD values for the corresponding antigen from the serum samples of 17 negative controls (non-exposed Melbourne donors). A sample was seropositive if the antibody level was above the cut-off value and seronegative if it was below (supplementary material of Fowkes *et al.* (2012) [1]).

3.1.1.3 Estimated gestational age (EGA)

The gestational age (GA) at delivery was estimated using the Dubowitz method [222]. However, if the delivery occurred at home, gestational age at delivery was determined from fundal height based on a formula derived from an analysis of a cohort of Karen pregnant women whose gestational age was measured using both fundal height and the Dubowitz method [219].

3.1.2 Study setting of the current study

All the measured antibody responses except for Schizont extract (the response based upon which the subset of controls was selected), were included in this analysis. Records with missing values for all six antibody responses (*Pf*AMA1, *Pf*EBA175, *Pf*MSP2, *Pf*MSP3, *Pv*AMA1, *Pf*VAR2CSA) were excluded from the analyses (105 (5.8%) samples). Duplicated measurements at certain gestational weeks per woman were averaged and the

case which had only one record (enrolment sample) was removed. Finally, 1692 samples were used in the current analyses, of which 727 samples were from 135 cases and 965 samples from 115 controls (322 samples were from low schizont lysate responder controls and 643 samples were from high schizont lysate responder controls).

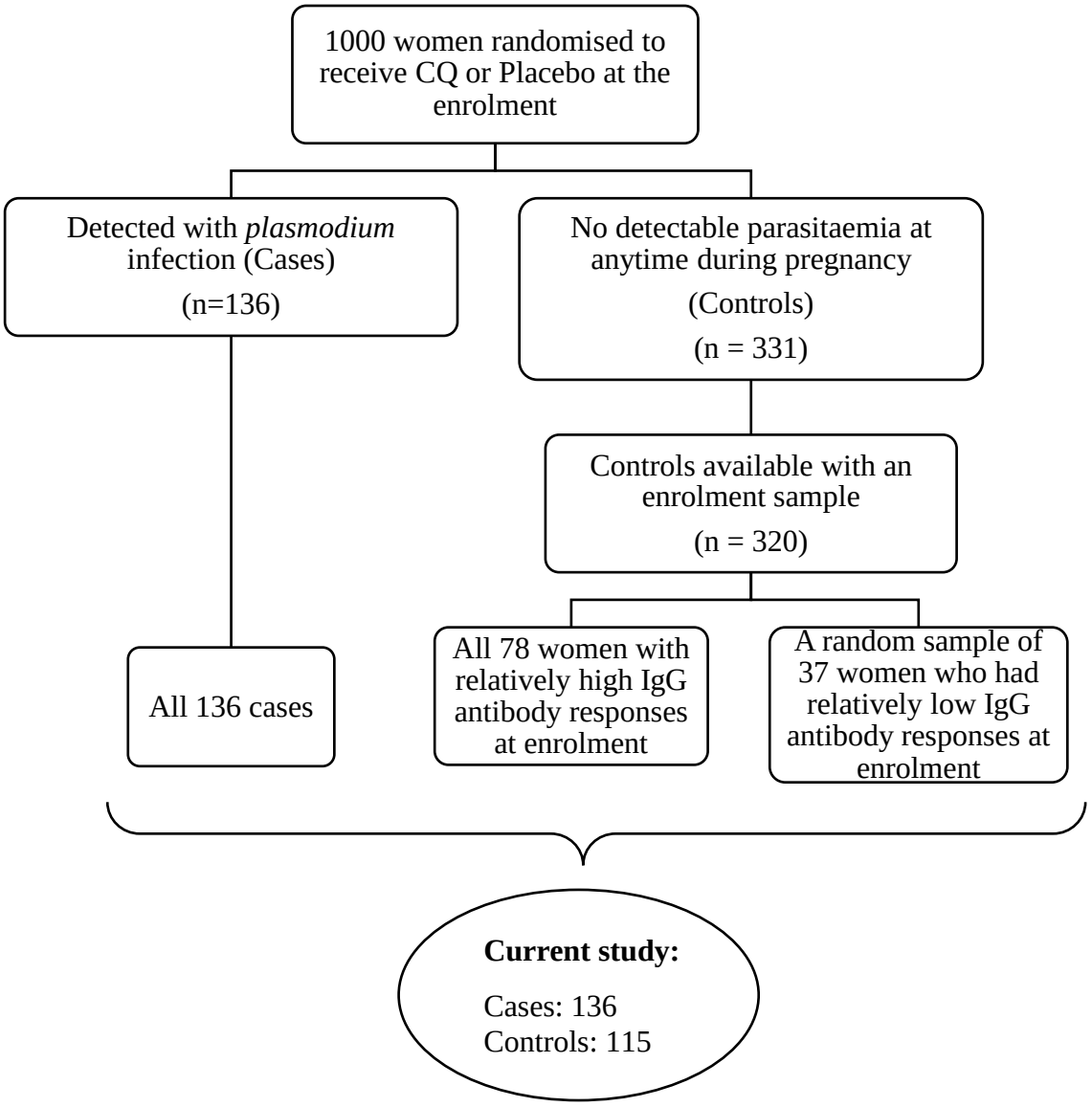


Figure 1: Flow diagram of cohort selection for the current study.

3.2 Summary

A better understanding of the functionality of the antibody responses against malaria parasite antigens is essential to develop effective malaria vaccines. Investigating the antibody measures longitudinally provides more comprehensive insights into the behaviour of the antibodies compared to cross-sectional studies.

The behaviour of antibody responses to malaria during pregnancy is not well known. It has been shown that antibody responses tend to largely vary between as well as within pregnant women [1]. The aim of this study is to quantify the dynamic antibody responses during pregnancy and identify the antibodies that detect exposure to infection in pregnant women.

The antibody measurements of two groups of pregnant women (exposed and unexposed to malaria) living in the Thai-Myanmar border – a low malaria transmission setting –are analysed in this work. The studied data includes longitudinal measurements which are irregularly sampled from the followed up pregnant women; the number of samples for a pregnant woman varies between 2 and 13. The measurements were obtained on 6 antibodies that are identified as promising vaccine candidates or markers for sero-surveillance: *Pf*AMA1, *Pv*AMA1, *Pf*EBA175, *Pf*MSP2, *Pf*MSP3 and *Pf*VAR2CSA. The total IgG titer on the Schizont extract was used to detect and quantify each antibody using ELISA. Subsequently, based on the cut-off value determined for each antibody, each antibody measurement was characterised as seropositive or seronegative. The statistical method, allowing the joint behaviour of the six antibody responses to be studied, is introduced in the next chapter.

Chapter 4

Statistical methods for clustering pregnant women based on multivariate continuous longitudinal antibody measurements

The aim of this thesis was to quantify the dynamic response of the antibody levels over the gestational period by clustering the pregnant women into groups based on the collective behaviour of the antibody responses *Pf*AMA1, *Pf*EBA175, *Pf*MSP2, *Pf*MSP3, *Pv*AMA1 and *Pf*VAR2CSA. This chapter discusses in detail the packages which perform classifications based on longitudinal data both in univariate and multivariate contexts, including a justification for the method selected to analyse the antibody data.

Notation

A total of 250 pregnant women were included in this analysis. The follow-up times, $\boldsymbol{t}_i = \left(t_{i,1}, \ldots, t_{i,n_i}\right)^T$, correspond to the i^{th} woman with n_i ANC visits from enrolment to delivery; the number of ANC visits and the visit times can vary between the women.

Let $\boldsymbol{Y}_{i,r} = \left(Y_{i,r,1}, \ldots, Y_{i,r,n_i}\right)^T$, r = 1, …, 6, represent a random vector of longitudinal measurements of the i^{th} woman on the r^{th} antibody. Then $\boldsymbol{Y_i} = \left(\boldsymbol{Y}_{i,1}{}^T, \ldots, \boldsymbol{Y}_{i,6}{}^T\right)^T$ denotes the entire set of multivariate longitudinal data for the i^{th} woman. $\boldsymbol{Y_i}$ will then be classified into K groups, where K may be unknown.

4.1 Available statistical methods

The goal of this analysis is to cluster the women into groups such that women within a group exhibit similar characteristics with respect to the trajectories of their antibody levels. Furthermore, the clustering should be based on the joint behaviour of the trajectories of the six antibody responses. Since all statistical analyses of this study were performed using **R** software (version 3.5.2), only packages which are implemented in this software were explored.

Several methods have been developed to cluster individuals based on longitudinal data, but the majority of them perform clustering with respect to a single trajectory in the individuals [223-227]. Examples of such packages are **kml** [228, 229] and **traj** [230].

Clustering joint-trajectories (multiple single-trajectories) can be performed by initially clustering the single-trajectories independently and then aggregating their results by considering the similarities across their suggested clusters. However, this method is not reliable as it does not take account of the correlation between the trajectories.

`kml3d` [231, 232] is a clustering algorithm capable of classifying multivariate longitudinal data with several features, such as providing a number of standard imputation methods to deal with missing values and several quality criteria to select the best number of clusters. The package also allows users to select a particular clustering outcome using a graphical interface that shows the possible results when different numbers of clusters are chosen [228, 229].

The clustering algorithm implemented in kml3d utilises the *k*-means algorithm [233] in the context of multivariate longitudinal data. K-means is a non-parametric hill-climbing algorithm belonging to the Expectation-Maximisation (EM) class [234]. Initially, each individual is randomly (alternative approaches are also available) classified into a cluster and the optimal clustering is yielded by performing the "Expectation" and "Maximisation" phases alternately. The centre of each cluster is measured in the Expectation phase while each individual is assigned to its "nearest cluster" during the Maximisation phase. This process is repeated until the membership of individuals within the clusters does not change anymore [228]. In the longitudinal data perspective, "cluster centre" represents the mean of the joint-trajectories of the individuals classified into the specific cluster (mean trajectory). Then, for the i^{th} individual, the distance between the joint-trajectory and the mean trajectory of each cluster is determined and the cluster *k* which minimises this distance is determined as the nearest cluster. Of note, the kml3d clustering algorithm requires each individual to have measurements for each outcome at the same time points and if there is incomplete data it needs to be imputed (for more details about computing distance between joint-trajectories, refer Genolini *et al.* [235]).

kml3d is mainly an exploratory clustering method, hence, it is unable to provide an optimal number of clusters that is statistically justifiable. In addition, because the clustering is not model-based, performing a goodness-of-fit test is not possible. The package is also limited to classifying only continuous data [235].

Model-based clustering methods are superior to exploratory methods by producing statistically justifiable results. In these methods, data are assumed to be generated from a

probability distribution which is created from a mixture of two or more components. These components that correspond to the clusters are represented by different probabilistic models [236-238] (i.e. each cluster is represented by a density function and a probability or "weight" associated with the density function).

Initially, the number of clusters (*K*) is assumed to be unknown and the unobservable component allocations $U_1, \ldots, U_N \in \{1, \ldots, K\}$, which denote the assignment of the i^{th} woman to a cluster are introduced with cluster allocation probabilities

$$P(U_i = k; \boldsymbol{w}) = w_k, \qquad i = 1, \ldots, N, \;\; k = 1, \ldots K,$$

where $\boldsymbol{w} = (w_1, \ldots, w_K)^T$ represents the vector of cluster allocation probabilities, satisfying $w_i > 0$ and $\sum_{i=1}^{K} w_i = 1$.

The marginal density of $\boldsymbol{Y}_i$ is then represented as a mixture of cluster specific densities:

$$f_i(\boldsymbol{Y}_i; \boldsymbol{\xi}) = \sum_{k=1}^{K} w_k \, f_{i,k}(\boldsymbol{Y}_i; \boldsymbol{\psi}, \boldsymbol{\theta}_k), \qquad (1)$$

where $\boldsymbol{\xi} = (\boldsymbol{w}^T, \boldsymbol{\psi}^T, \boldsymbol{\theta}_1^T, \ldots \boldsymbol{\theta}_K^T)^T$ denotes the vector of model parameters to be estimated. The model density $f_{i,k}(\boldsymbol{Y}_i; \boldsymbol{\psi}, \boldsymbol{\theta}_k)$ represents the conditional density of $\boldsymbol{Y}_i$ given $U_i = k$ and the subscript *i* in $f_{i,k}$ indicates that the density may depend on subject-specific factors, i.e. covariates such as the visit times $\boldsymbol{t}_i$. $\boldsymbol{\theta}_k$ is the vector of parameters specific to the k^{th} cluster and $\boldsymbol{\psi}$ is the vector of model parameters common across all clusters (the two types of model parameters are explained in detail later in this chapter).

The individual component probabilities ($\widehat{p_{i,k}}$) are estimated using

$$\widehat{p_{i,k}} = \widehat{p_{i,k}}(\boldsymbol{\xi}) = \mathrm{P}(U_i = k | \boldsymbol{Y}_i = \boldsymbol{y}_i; \boldsymbol{\xi}) = \frac{w_k f_{i,k}(\boldsymbol{Y}_i; \boldsymbol{\psi}, \boldsymbol{\theta}_k)}{f_i(\boldsymbol{Y}_i, \boldsymbol{\xi})}, \quad \mathrm{i} = 1, \ldots, N, k = 1, \ldots, K. \; (2)$$

A woman is allocated to the cluster with the largest probability.

Various R packages have been introduced with different model-based clustering methods. Some packages are capable of performing clustering only for longitudinal data which are sampled regularly ($n_1 = \cdots = n_N = n, \boldsymbol{t}_1 = \cdots = \boldsymbol{t}_N = \boldsymbol{t}$). Most packages also assume that the vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$ are not only independent but also identically distributed (i.i.d.). For instance, packages such as **`Mclust`** [238], **`teigen`** [239, 240] and the earlier versions of the **`mixAK`** package assume the mixture components follow either a multivariate normal or *t*-distribution. In **`Rmixmod`** package**,** the mixture components

fitted to a dataset could also follow multinomial probability distribution functions without limiting to multivariate Gaussian distributions. The package **HDclassif** [241] facilitates clustering of high-dimensional data.

However, all of the aforementioned packages do not allow the mixture parameters to depend on covariates (for instance the visit time vectors $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N$, necessary for longitudinal data with varying measurement times across individuals), which limits their usability. The packages **long-clust** [242] or **MFDA** [243] are capable of incorporating the dependence of the mixture parameters on variables like the visit times $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N$, but this advantage is constrained by using only regularly sampled continuous longitudinal data.

However, in many applications including the current study, the longitudinal data are irregularly sampled with different number of visit times $n_1, \ldots, n_N$ and different visit time vectors $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_N$.

Model-based clustering methods for irregularly sampled data build upon a mixture of appropriate regression models, which allow each outcome variable $\boldsymbol{Y}_i$ to vary with covariates. Fitting a mixture of linear mixed models is the basis of the **mixtools** package [244], which can perform clustering only on continuous longitudinal data. The **lcmm** [245], however, can handle either continuous or discrete response variables; note, this method cannot be applied to data that has both types of response variables concurrently. **flexmix** [246] is an R package which does not have this limitation, i.e. it can cluster multivariate longitudinal data with responses of different data types (both continuous and discrete). However, **flexmix** assumes independence between the responses, which is undesirable, as it is likely that the malaria antibody responses in this study are highly correlated.

The **mixAK** [247, 248] package is the most appropriate method for this study as it allows classification of irregularly sampled multivariate longitudinal data, both continuous and discrete responses can be used, and it is not constrained by the independence assumption between multiple responses for an individual. This package assumes that the densities $f_{i,k}$ correspond to multivariate generalised linear mixed models, hence, the density f_i (157 1) can be represented by a mixture of multivariate generalised linear mixed models.

Comprehensive details of the package can be found in Komarek and Komarkova (2013) [249].

The **`mixAK`** package was selected to cluster the pregnant women based on the behaviour of the longitudinal measurements across all six antibodies.

4.2 Multivariate mixture linear mixed modelling and clustering procedure

4.2.1 Multivariate linear mixed model (MLMM) for a single outcome

To construct the multivariate linear mixed effects model [250] first a standard linear mixed effects model was specified for each of the six antibody responses with the following covariates as fixed effects (X): age (a continuous variable, years), primigravidae (a dichotomous variable assigned 1 if primigravidae), treatment arm (a dichotomous variable assigned 1 if given chloroquine (CQ) as prophylaxis at enrolment) and having a history of malaria prior to enrolment (a dichotomous variable assigned 1 if exposed to malaria at least once prior to enrolment). Antibody measures at conception (intercept) and the rate of change of the antibody response over gestational time (in weeks) were included as random effects to allow these parameters to vary between women. Therefore, the linear mixed effects model specified for each antibody including four fixed effects and two random effects is given by:

$$h_r^{-1}\{E(\boldsymbol{Y}_{i,r}|\boldsymbol{b}_{i,r},\boldsymbol{\alpha}_r)\} = \boldsymbol{X}_{i,r}^T\boldsymbol{\alpha}_r + \boldsymbol{Z}_{i,r}^T\boldsymbol{b}_{i,r} \quad i = 1,\dots,250,\ r = 1,\dots,6, \qquad (3)$$

where,

h_r^{-1}: A chosen link function

$\boldsymbol{X}_{i,r}^T$: $n_{i,r}$ × 4 covariate matrix for fixed effects

$\boldsymbol{Z}_{i,r}^T$: $n_{i,r}$ × 2 covariate matrix for random effects

$\boldsymbol{\alpha}_r = (\alpha_{r,age}, \alpha_{r,primigravidae}, \alpha_{r,CQ}, \alpha_{r,history})^T \in \mathbb{R}^{p_r}$: vector of fixed effects for antibody *r* with p_r =4

$\boldsymbol{b}_{i,r} = (b_{i,r,intercept}, b_{i,r,slope})^T \in \mathbb{R}^{q_r}$: vector of random effects for antibody *r* specific for the i^{th} woman with q_r =2

4.2.2 Multivariate linear mixed model for multiple antibodies

A multivariate linear mixed model is fitted to the data of all six antibody measures simultaneously to perform clustering and is expressed as follows:

$$h_r^{-1}\{E(\boldsymbol{Y}_i|\boldsymbol{b}_i,\boldsymbol{\alpha})\} = \boldsymbol{X}_i^T\boldsymbol{\alpha} + \boldsymbol{Z}_i^T\,\boldsymbol{b_i}, \quad (4)$$

where,

$\boldsymbol{X}_i^T$: n_i × 18 block-diagonal matrix with $\boldsymbol{X}_{i,1}, \dots, \boldsymbol{X}_{i,6}$ matrices on the diagonal

$\boldsymbol{Z}_i^T$: n_i × 12 block-diagonal matrix with $\boldsymbol{Z}_{i,1}, \dots, \boldsymbol{Z}_{i,6}$ matrices on the diagonal

$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_6^T)^T$: vector of fixed effects for all antibody responses

$\boldsymbol{b}_i$: $\left(\boldsymbol{b_{i,1}}^T, \dots, \boldsymbol{b_{i,6}}^T\right)^T \epsilon \mathbb{R}^q$:The vector of random effects of the i^{th} woman on all antibody responses with $q = \sum_{r=1}^{6} q_r = 12$

4.2.3 Multivariate mixture linear mixed model (MMLMM) implemented in the `mixAK` package

The mixture model (Equation 1) fitted by the **`mixAK`** package to cluster the $\boldsymbol{Y}_i = \left(Y_{i,1,1}, \dots, Y_{i,6,n_i}\right)^T$, $i = 1, \dots, 250$, makes the following assumptions.

1. Each $Y_{i,r,j}$ $(i = 1, \dots, 250,\ r = 1, \dots, 6$ and each $j = 1, \dots, n_i)$, follows a distribution D_r from the exponential family with the dispersion parameter ϕ_r (in this study refers to the error variance) and the mean given by;

$$h_r^{-1}\left\{E\left(Y_{i,r,j}|\boldsymbol{b}_{i,r},\boldsymbol{\alpha}_r\right)\right\} = X_{i,r,j}^T\,\boldsymbol{\alpha}_r + Z_{i,r,j}^T\boldsymbol{b}_{i,r}, \quad (5)$$

   where,

   $X_{i,r,j}^T$: The j^{th} fixed effects covariate of the vector $\boldsymbol{X}_{i,r}$

   $Z_{i,r,j}^T$: The j^{th} random effects covariate of the vector $\boldsymbol{Z}_{i,r}$

2. **The distribution of random effects**

   For the purpose of clustering antibody response profiles, the random effects are assumed to follow a mixture of K heteroscedastic normal distributions:

$$\boldsymbol{b}_i \sim \sum_{k=1}^{K} w_k\, N_q(\boldsymbol{\mu}_k, \mathbb{D}_k), \quad (6)$$

where the conditional distribution of the joint random effects, $\boldsymbol{b_i}$ given $U_i = k$, is a multivariate (q-variate = 12) normal distribution with unknown mean $\boldsymbol{\mu_k}$, unknown $q \times q$ positive definite covariance matrix $\mathbb{D}_{\boldsymbol{k}}$ with an unknown proportion of women being classified into each cluster, w_K, $k = 1, \dots, K$.

3. For each $i = 1, \dots, 250$, the random variables $Y_{i,1,1}, \dots, Y_{i,6,n_i}$ are conditionally independent given the random effects vector $\boldsymbol{b}_i$.
4. $Y_1, \dots, Y_{250}$ are independent.
5. $\boldsymbol{b}_1, \dots, \boldsymbol{b}_{250}$ are independent.

4.2.3.1 Model specification

Initially, cluster-specific densities $f_{i,k}$, $i = 1, \dots, 250$, $k = 1, \dots, K$ were defined:

$$f_{i,k}(\boldsymbol{Y}_i; \boldsymbol{\psi}, \boldsymbol{\theta}_k) = \int_{\mathbb{R}^q} \left\{ \prod_{r=1}^{6} \prod_{j=1}^{n_i} f_{D_r}\left(Y_{i,r,j}; \boldsymbol{\alpha}_r, \phi_r, \boldsymbol{b}_{i,r}\right) \right\} \varphi(\boldsymbol{b}_i; \boldsymbol{\mu}_k, \mathbb{D}_k) db_i, \quad (7)$$

where, $\varphi(.; \boldsymbol{\mu}_k, \mathbb{D}_k)$ is a density of the multivariate normal distribution with a mean $\boldsymbol{\mu}_k$ and a covariance matrix $\mathbb{D}_k$, $k = 1, \dots, K$ according to the second assumption.

The i^{th} subject's likelihood contribution is then obtained by substituting $f_{i,k}$ into Equation 1:

$$f_i(\boldsymbol{Y}_i; \boldsymbol{\xi}) = \sum_{k=1}^{K} w_k \int_{\mathbb{R}^q} \left\{ \prod_{r=1}^{6} \prod_{j=1}^{n_i} f_{D_r}\left(Y_{i,r,j}; \boldsymbol{\alpha}_r, \phi_r, \boldsymbol{b}_{i,r}\right) \right\} \varphi(\boldsymbol{b}_i; \boldsymbol{\mu}_k, \boldsymbol{D}_k) db_i, \quad (8)$$

$$= \int_{\mathbb{R}^q} \left\{ \prod_{r=1}^{6} \prod_{j=1}^{n_i} f_{D_r}\left(Y_{i,r,j}; \boldsymbol{\alpha}_r, \phi_r, \boldsymbol{b}_{i,r}\right) \right\} \{ \sum_{k=1}^{K} w_k \varphi(\boldsymbol{b}_i; \boldsymbol{\mu}_k, \boldsymbol{D}_k) \} db_i. \quad (9)$$

It follows from Equations (8 and 9) that the observable random vectors $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_{250}$ can be interpreted either as a mixture of multivariate generalised linear mixed models with random effects following a normal distribution (Equation 8), or as a multivariate generalised linear mixed model with a normal mixture in the random effects distribution (Equation 9), with the overall mean and the overall covariance matrix of the random effects $\boldsymbol{b}_i$, $i = 1, \dots, 250$,

$$\boldsymbol{\beta} = \mathrm{E}(\boldsymbol{b}_i; \boldsymbol{\xi}) = \sum_{k=1}^{K} w_k \boldsymbol{\mu}_k, \quad (10)$$

$$\mathbb{D} = \mathrm{var}(\boldsymbol{b}_i; \boldsymbol{\xi}) = \sum_{k=1}^{K} w_k \left\{ \mathbb{D}_k + \left(\boldsymbol{\mu}_k - \sum_{j=1}^{K} w_j \boldsymbol{\mu}_j\right) \left(\boldsymbol{\mu}_k - \sum_{j=1}^{K} w_j \boldsymbol{\mu}_j\right)^T \right\}. \quad (11)$$

In this study, pregnant women will be classified based on their optical density measures for the six antibodies and then on their binary seropositivity measures for each antibody. Each of these analyses are described in more detail below in Scenarios 1 and 2.

Scenario 1 The optical density (OD) measure of each antibody is considered as a continuous variable and follows a Gaussian distribution. According to the first assumption of the model, the link function h_r^{-1} is an identity. Furthermore, the dispersion parameter ϕ_r is the unknown residual ($\varepsilon_{i,r,j}$) variance and was assumed to be mutually independent and normally distributed, with a specific variance for each antibody (ϕ_r), defined by:

$$\varepsilon_{i,r,j} \sim N(0, \phi_r).$$

A linear mixed model is therefore fitted to each antibody and the MMGLMM reduces to a multivariate mixture linear mixed model (MMLMM).

Scenario 2: Each antibody measure is converted to either seropositive or seronegative and each antibody is converted to a binary variable. The pregnant women are then classified based on the seropositivity values of the antibodies, collectively. In this situation each measure follows a Bernoulli distribution with the logit link and the ϕ_r =1.

4.2.3.2 Types of model parameters

In this study, the distribution of the joint random effects was represented by a mixture of two 12-dimensional multivariate normal distributions (the criteria of selecting 2 mixture components is explained in the Appendix A). Hence, Equation 6:

$$\boldsymbol{b}_i \sim \sum_{k=1}^{2} w_k \, N_{12}(\boldsymbol{\mu}_k, \mathbb{D}_k), \qquad (12)$$

$$\boldsymbol{b}_i | \, U_i = k \sim N_{12}(\boldsymbol{\mu_k}, \mathbb{D}_k), \qquad k = 1, 2,.$$

Therefore, the MMLMM comprises two types of model parameters (i.e. model parameters common to all clusters and cluster-specific model parameters), which need to be estimated to perform the cluster analysis.

- **Model parameters common to all clusters**

Model parameters common to all clusters include the fixed effects parameters (i.e. the parameters for age, primigravidity (being pregnant for the first time), CQ prophylaxis and history of malaria) and the dispersion parameter (residual error variance) introduced for each antibody response:

$$\boldsymbol{\psi} = (\boldsymbol{\alpha_1}^{\boldsymbol{T}}, \ldots, \boldsymbol{\alpha_6}^{\boldsymbol{T}}, \boldsymbol{\phi_1}, \ldots, \boldsymbol{\phi_6})^{\boldsymbol{T}}.$$

Since four fixed effects were introduced for each antibody response, $\boldsymbol{\psi}$ is comprised of 24 fixed effects and 6 residual error variances (limited to Scenario 1).

- **Cluster-specific model parameters**

  The parameters specific to each cluster is comprised of the w_k, $\boldsymbol{\mu_k}$ and $\mathbb{D}_{\boldsymbol{k}}$ of the distribution of the random effects. Since two clusters were identified as the optimal number of mixtures for the data (see the Appendix A), the parameters of the distributions of the random effects within each cluster are:

$$\boldsymbol{\theta} = \left(\boldsymbol{w}^T, \boldsymbol{\mu_1}^T, \boldsymbol{\mu_2}^T, \mathbf{vec}(\mathbb{D}_1), \mathbf{vec}(\mathbb{D}_2)\right)^T,$$

where, $\mathbf{vec}(\mathbb{D}_k)$ represents the vector of elements in the lower triangle of the matrix $\mathbb{D}_k$.

4.2.3.3 Parameter estimation

A Bayesian approach, using Markov chain Monte Carlo (MCMC) simulations, was employed to derive posterior distributions of the unknown model parameters $\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\theta})$.

Initially, to determine the prior distributions, it was assumed independence between MMLMM parameters ($\boldsymbol{\psi}$) and mixture related parameters ($\boldsymbol{\theta}$), i.e. p ($\boldsymbol{\psi}$, $\boldsymbol{\theta}$) = p($\boldsymbol{\psi}$)×p($\boldsymbol{\theta}$). Then for $p(\boldsymbol{\theta})$, the multivariate version of the priors proposed by Richardson and Green (1997) [251], which is defined for univariate normal mixtures (e.g., $\mathbb{D}_k^{-1}$ follows a gamma distribution), was selected, and for $p(\boldsymbol{\psi})$, we adopted priors used typically for the fixed-effects and residual error in univariate linear mixed-effect modelling (e.g., $\boldsymbol{\alpha}$ is assumed to be normally distributed, see Fong, Rue and Wakefield (2010) [252]).

Hyperparameters were then introduced in such a way that the prior distribution of p ($\boldsymbol{\psi}$, $\boldsymbol{\theta}$) will be weakly informative (see Appendix A of the supplementary material of Komarek and Komarkova (2013) [253] for a detailed explanation of hyperparameter specifications).

The likelihood of the MMLMM is defined by,

$$L(\boldsymbol{\psi}, \boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\psi}, \boldsymbol{\theta}) = \prod_{i=1}^{250}\left(\sum_{k=1}^{2} w_k L_{i,k}(\boldsymbol{\psi}, \boldsymbol{\theta}_k)\right), \quad (13)$$

where,

$$L_{i,k}(\boldsymbol{Y}_i;\boldsymbol{\psi},\boldsymbol{\theta}_k)=\int\left\{\prod_{r=1}^{6}\prod_{j=1}^{n_i}f_{D_r}\left(y_{i,r,j};\boldsymbol{\alpha}_r,\phi_r,\boldsymbol{b}_{i,r}\right)\right\}\varphi(\boldsymbol{b}_i;\boldsymbol{\mu}_k,\mathbb{D}_k)db_i, i=1,\ldots,250\,, k=1{,}2$$

was the contribution of the i^{th} woman to the likelihood assuming that the woman was classified into the k^{th} mixture component and was obtained by substituting terms specific to the study in $f_{i,k}$ (Equation 7).

The **MixAK** package assumes that the antibody data are missing at random (i.e. missing depend only on observed covariates). Therefore, the missingness of any antibody response is accounted for by modifying the likelihood contribution of the i^{th} woman for a given r =1, … ,6, by taking the density product over only the available measurements.

Subsequently, the MCMC method was used to generate samples of model parameters,

$$S_M=\left\{\left(\boldsymbol{\psi}^{(m)},\boldsymbol{\theta}^{(m)}\right):m=1,\ldots,M\right\},$$

from the joint posterior distribution, $p(\boldsymbol{\psi},\boldsymbol{\theta}|\mathbf{y})\ \propto\ L(\boldsymbol{\psi},\boldsymbol{\theta})\ p(\boldsymbol{\psi},\boldsymbol{\theta})$.

A common problem with mixture models is that the model likelihood remains invariant against the switching of the labels of the clusters. This problem was addressed by applying the relabeling algorithm of Stephens (2000) [254], which re-calculates the posterior samples which are invariant towards label switching. Solving this identifiability problem is vital as the mixture model is the basis for the clustering. The details for the Bayesian inference were extracted from Komarek and Komarkova (2014) [248].

The parameters of the MMLMM with 2 mixture components fitted to the study dataset were estimated using 50,000 posterior samples of the joint posterior distribution obtained by running two chains of MCMC algorithm after a burn-in period of 500 iterations and 1:50 thinning. The burn-in period (a set of initial samples that are not expected to be converged to the target distribution hence excluded), thinning (to reduce the autocorrelation in the sequence of iterations, every l^{th} iteration after the burn-in is involved in the study sample where in this study l=50) and the number of iterations were determined by performing multiple simulations allocating varying combinations for the three values and using convergence diagnostics as detailed in Subsection 4.2.3.4. Furthermore, the MCMC algorithms were automatically initiated with different sets of initial values by the program (eg for chain 1 on which the results of this study was based on, the initial values for the fixed effects are the maximum-likelihood estimates computed

using **lmer** function while initial weights are all equal to K^{-1} i.e. 0.5 in this study) [247]. Further details about the initial values used for the two chains are provided in Komarek (2009) [255] and Komarek (2014) [247].

4.2.3.4 Convergence of the chains

Convergence of the MCMC was evaluated using the **coda** package available in the **R** software [256]. Among the available methods, the convergence of this study was determined based on the visual assessment of the autocorrelations, autocorrelation plots, traceplots of the parameters and two convergence diagnostic methods (Geweke's convergence diagnostic and Gelman and Rubin's convergence diagnostic) applied on the deviance, as explained below.

- **Autocorrelations and the autocorrelation plot**

CODA package computes autocorrelations within each chain for a specified variable. High autocorrelations within a chain indicates that both the mixing and convergence are slow.

- **Traceplot**

A traceplot shows a plot of the sampled values for the selected parameter computed at each iteration, where no clustering in different areas indicates good mixing and convergence around a particular value.

- **Geweke's convergence diagnostic**

The Geweke (1992) [257] diagnostic tests for the convergence of a Markov chain (in this study, chain 1) involves testing for the equality of the two means computed using the estimates of a particular parameter at the first and last part of the chain (by default the first 10% and the last 50%). In this study, as 50,000 samples were incorporated, the means computed using the first 5000(50,000*10%) estimates and last 25,000 (50,000*50%) estimates were compared. If the entire chain is stationary, then the means of the estimates should be similar. The Geweke's statistic is a standard $\mathcal{Z}$ score (that is the standardized mean difference) and the sampling distribution of $\mathcal{Z}$ converges to $N(0,1)$ as $n \to \infty$ if the chain has converged. Therefore, the values of $\mathcal{Z}$ which are closer to 0 indicate that the chain has converged and suggests that the burn-in period and thinning are sufficient to converge the chain while the values of $\mathcal{Z}$ which fall in the extreme tails of $N(0,1)$ indicate

that the chain has not converged. (For further information refer Smith (2007) [258] and Plummer *et al* (2009) [256]).

- **Gelman and Rubin's convergence diagnostic**

The Gelman and Rubin (1992) [259] method computes the convergence of the estimates (n where in this study 50,000 iterations) of a particular parameter computed under multiple parallel chains (m) which start the MCMC sampling from different starting points (in this study m =2 chains which start off with different initial values). In this method, the variance of all the chains are decomposed ($\hat{\sigma}^2$) into a variance within each chain (W) and between-chain variance (B/n) as follows:

$$\hat{\sigma}^2 = \frac{(n-1)W}{n} + \frac{B}{n}.$$

Chains become converged when the between variance of the chains is not larger than the variance within each individual chain and as $n \to \infty$, both $\hat{\sigma}^2$ and W converges to σ^2. This method assesses the convergence by computing potential scale reduction or shrink factor (R) by considering n samples in each of m parallel chains:

$$R = \sqrt{\frac{(n-1)}{n} + \frac{(m+1)}{mn}\frac{B}{W}}.$$

As the simulation converges, the shrink factor becomes close to 1, meaning that the parallel Markov chains are overlapping while values above 1 ($R > 1.05$) indicate non-convergence. In **CODA**, the 50% and 97.5% quantiles of the sampling distribution are computed as the estimate and the upper limit of the shrink factor, respectively. The convergence of multiple parameters simultaneously was proposed by Brooks and Gelman (1998) [260] (For further information refer Smith (2007) [258] and Plummer *et al* (2009) [256]).

4.2.3.5 Selection of the number of mixture components

In order to determine the optimal number of mixture distributions that best explains the overall distribution of the random effects of the study dataset, mixture models with different numbers of clusters (*K*) were fitted using the **mixAK** package. The optimal number of mixture components was then selected using two approaches: penalised

expected deviance and the plot of posterior cumulative distribution functions (CDFs) of the deviances, explained below.

- **Penalised expected deviance**

Deviance information criteria (DIC) [261] is one of the most common Bayesian-model choice methods and works both as a measure of fit and as a measure of model complexity. However, Celeux *et al.* (2006) [262] has shown that its use in mixture models is controversial. Plummer (2008) [263], therefore, introduced an alternative model selection criterion by incorporating penalised loss functions and cross-validating arguments. In the method, the deviance is defined in a different form, known as penalised expected deviance (PED), given by

$$PED = E\{ D(\theta)|Y = y\} + P_{opt},$$

where the first term on the right-hand side (RHS) is the expected deviance, given Y=y, which is computed by averaging the data deviance $D(\psi, \theta)$ observed at each MCMC simulation. The second term on the RHS, P_{opt}, is a penalty term called optimism. This penalty term is introduced to test the adequacy of the model, and captures the amount by which the model adequacy is overstated by each woman ($\boldsymbol{Y_i}$) (for further details see [263]).

In the **`mixAK`** package, P_{opt} was estimated using importance sampling and the two parallel chains (refer to Komarek [255] for detailed information). The **`mixAK`** package also uses this measure to compare models and identify the best fitting model which has the lowest PED value.

- **The plot of posterior cumulative distribution functions (CDFs) of the deviances**

This method [264, 265] compares models on the basis of the full posterior distribution of the deviances. The uncertainty in the comparison, which tends to increase with the number of model parameters, is accounted for in this approach, which is not possible using single measure approaches, such as DIC and PED.

Suppose the two models to be compared comprises of clusters 1 (K=K_1) and 2 (K=K_2) with deviances $D_{K_1}(\theta)$ and $D_{K_2}(\theta)$ respectively. Therefore, the posterior probability is:

$$P_{K_2,K_1} = P\{D_{K_2}(\theta) - D_{K_1}(\theta) < 0 \mid Y = y\}.$$

According to Aitkin (2010) [265], the optimal cluster size is determined by comparing the distribution of the cumulative posterior distributions (CDFs) of the deviances, defined by

$$F_{D_j}(d|Y = y) = P\{D_j(\theta) \leq d|Y = y\},\ j = 1,2.$$

The selection of the optimal number of clusters based on PED values and the plot of posterior CDFs of the deviances is provided in the Appendix A.

Both approaches suggested that a mixture of two multivariate normal distributions is optimal to explain the overall distribution of the random effects corresponding to the study dataset, hence, the pregnant women were classified between two clusters. Each woman was classified into the cluster with a specific multivariate normal distribution for the random effects under which the joint random effect ($\boldsymbol{b_i}$) corresponding to the i^{th} woman is highly likely to occur.

Of note, due to computational time, the selection of the number of mixture components was based on a sample of 10,000 iterations obtained after a burning period of 500 and thinning of 1:10.

4.2.3.6 Clustering procedure

Classification of the participants into 2 clusters was based on their posterior probability of belonging to each of the 2 components as defined in Equation 2:

$$p_{i,k} = p_{i,k}(\boldsymbol{\theta}) = \mathrm{P}(U_i = k|\boldsymbol{Y_i} = \boldsymbol{y_i};\ \boldsymbol{\theta}) = \frac{w_k f_{i,k}(\boldsymbol{Y_i};\boldsymbol{\theta_k},\boldsymbol{\psi})}{f_i(\boldsymbol{Y_i},\boldsymbol{\theta})},\ i = 1,\ldots,250,\ \mathrm{k} = 1,2. \quad (14)$$

Estimates of the $p_{i,k}$ are generated from the 1000 posterior samples of the model parameters. Specifically, during the MCMC sampling, at each iteration, each individual gets assigned a probability for falling into cluster 1 and cluster 2 based on Equation 7. Next, the posterior median, posterior mean and the credible interval (CrI) of the chain of 1000 probabilities for each cluster were calculated for each woman. The classification was then performed based on these summary measures under two steps, as follows.

Step 1:

Each pregnant woman was assigned to the cluster (k=1,2) with the higher mean/median of the estimated probabilities, $p_{i,k}$, across the 1000 posterior samples. In this analysis, the classification was performed based on both estimates of mean and median. If the classification result differs between the two estimates, median will be used; in our results, the classification outcomes were the same for mean and median.

Step 2:

To determine the uncertainty of the classification, 95% CrIs of the probabilities derived from the 1000 MCMC samples were employed.

Specifically, for each woman, if the lower limit of the 95% CrI (2.5th percentile) of the probability of belonging to the nominated cluster exceeds 0.5, the woman remains in the assigned cluster, otherwise she is considered unclassified.

4.2.4 Contribution of each antibody on the classification of pregnant women

A secondary aim was to identify the relative contribution made by each antibody to clustering the pregnant women. To achieve this, the importance level of each antibody in clustering the pregnant women was determined by computing the univariate entropy value (E_r) for each antibody [266] as follows:

$$E_r = 1 + \frac{1}{N \log(K)} \left(\sum_{i=1}^{N} \sum_{k=1}^{K} P(U_i = k|\boldsymbol{C}_{ir}) \log\left(P(U_i = k|\boldsymbol{C}_{ir})\right) \right) \; r = 1, \dots, 6, \quad (15)$$

where N is the sample size (here, N = 250 pregnant women), K is the number of classes and $\boldsymbol{C}_i$ represents the longitudinal trajectories of all six antibody responses corresponding to the i^{th} woman, upon which the classification was made. Furthermore, the probability of classifying the i^{th} woman to the k^{th} cluster considering only the r^{th} antibody $\left(P(U_i = k|\boldsymbol{C}_{ir})\right)$, where $\boldsymbol{C}_{ir}$ is the subset of $\boldsymbol{C}_i$ corresponding to the r^{th} antibody, was computed by [253]:

$$P\left(U_i = k|\boldsymbol{C}_{ir}\right) = \frac{w_k \varphi(\boldsymbol{b}^*_{ir}|\boldsymbol{\mu}^*_{kr}, \mathbb{D}^*_{kr})}{\sum_{l=1}^{2} w_l \varphi(\boldsymbol{b}^*_{ir}|\boldsymbol{\mu}^*_{lr}, \mathbb{D}^*_{lr})} \qquad k = 1,2, \quad (16)$$

where $\boldsymbol{b}^*_i$ is a standardised version of $\boldsymbol{b}_i$ used for improving the mixing and numerical stability of the MCMC algorithm [253].

Of note, the entropy is a value between 0 and 1 and an antibody with an entropy value closer to 1 indicates that the classification is highly influenced by the specific antibody.

Since the **mixAK** package does not provide entropy values, code was developed to calculate this quantity for each antibody (see Appendix B).

4.3 Summary

This chapter described in detail the statistical methods available in R for clustering the pregnant women into groups that have similar antibody responses. Many of the methods reviewed were not appropriate for the current study due to several limitations, e.g. require clustering longitudinal data which are sampled regularly – note, the antibody measures in this study were obtained at different time points for each pregnant woman and the number of measurements for a given antibody was not the same for all pregnant women (unbalanced and irregularly sampled dataset).

Even methods that allow clustering irregularly sampled data either cannot handle both continuous and discrete variables simultaneously or assumes independence between the responses (although not applicable to the current study as only continuous variables were incorporated).

The **mixAK** package was identified as the most suitable and flexible method for clustering multivariate longitudinal data, as it allows analysing irregularly sampled data consisting of both continuous and discrete responses and is capable of taking account of the dependence between the responses. The **mixAK** package was therefore employed in the current study as it addresses many other packages' limitations.

This package performs clustering by fitting a multivariate mixture linear mixed model (MMLMM) to the data. The age, status of gravidity (primigravidae or not), the treatment arm and having a history of malaria were assumed to have fixed effects on the measurements of each antibody while the antibody measures at conception (intercept) and the rate of change of the antibody response over gestational time (in weeks) within each antibody were assumed to have random effects that allow these parameters to differ between pregnant women. The random effects (or random intercept and slope) for each woman were assumed to follow a mixture of K heteroscedastic multivariate normal distributions. This assumption allows both the correlation between the antibody measurements from a single woman to be accounted for and clustering of the pregnant

women into groups where the intercepts and slopes of their six antibody trajectories are similar after adjusting for the covariates included as fixed effects.

The **mixAK** package does not have an inbuilt method to determine the individual contribution of each antibody to the classification of pregnant women to clusters. This problem was addressed by computing the univariate entropy value for each antibody; the antibodies with high univariate entropy values were identified to have the highest influence on the classification.

Chapter 5

Application of the `mixAK` package to classify the longitudinal antibody responses of pregnant women from the Thai-Myanmar border

5.1 Descriptive statistics at enrolment

The descriptive statistics at enrolment data for the 135 cases (women infected with *P. falciparum* and/or *P. vivax* during pregnancy) and 115 controls (women without detected *Plasmodium* infection) are provided in Table 2.

The proportion of cases who were in their first pregnancy (22.2%) at enrolment was around twice of that in the control group (13.9%). Almost a quarter of cases were anemic (23.5%) at enrolment compared to only 10% of controls. The proportion of women who had received chloroquine prophylaxis was slightly higher for the control group, likely due to prevention of vivax episodes [221]. The majority of the pregnant women in both groups were enrolled in the trial during their first trimester (76% and 85% of cases and controls, respectively). Nearly all controls were residing at refugee camps (99.1%) while less than half of the cases (44.9%) were living at refugee camps during pregnancy.

Furthermore, more than half of the cases (55.6%) had been exposed to malaria prior to enrolment, compared with 40% of the controls. Considering the history of malaria specific to the *Plasmodium* spp., almost a similar proportion in the case (37.9%) and control (27.8%) groups had a history of falciparum infection(s) prior to enrolment. However, the proportion of cases (24.2%) who were infected with vivax malaria prior to enrolment was more than twice that of the control group (11.3%).

Table 2: Distribution of patient characteristics at enrolment for cases and controls

Characteristic	Cases (n=135)	Controls (n=115)
Age (years), median (IQR)	24 (20, 30)	26 (22, 32)
Gravidity, median (IQR)	3 (2.0, 4.5)	3 (2.0, 5.0)
Primigravida, n (%)	30 (22.2)	16 (13.9)
Multigravida, n (%)	105 (77.8)	99 (86.1)
Parity, median (IQR)	1 (0, 3)	2 (1, 4)
Haematocrit (%), median (IQR)	32.5 (30, 35)	34.0 (32, 36)
Anaemia[a], n (%)	32 (23.7)	11 (9.6)
Residence in refugee camp	61 (45.2)	114 (99.1)
Receiving chloroquine prophylaxis, n (%)	56 (41.5)	55 (47.8)
Estimated Gestational Age[b], median (IQR)	9.7 (7, 14)	9.4 (7.6, 11.6)
Trimester		
1 (<14wks), n (%)	102 (75.6)	98 (85.2)
2 (14 to <28wks), n (%)	31 (23)	17 (14.8)
3 (28 wks or more), n (%)	2 (1.5)	0 (0.0)
Plasmodium spp. before enrolment[c], n (%)	75 (55.6)	42 (36.5)
P. falciparum		
Proportion women infected, n (%)	50 (37.9)	32 (27.8)
P. vivax		
Proportion women infected, n (%)	32 (24.2)	13 (11.3)
Follow up (weeks), median (range)	28.9 (22.6, 32.1)	30.9 (28.3, 32.4)

[a] Haematocrit <30%.

[b] Determined at enrolment.

[c] Any microscopically confirmed *Plasmodium* infection documented at SMRU before enrolment into the study.

5.2 Graphical representation of six types of antibody responses

5.2.1 Optical density values (antibody levels) versus gestation

Overall, in all spaghetti plots, the antibody profiles of the pregnant women who were infected with malaria (cases) were maintained at comparatively higher levels and/or fluctuated significantly compared to the profiles of the women who were free from malaria (controls). However, substantial fluctuations were observed in the profiles of antibodies *Pf*AMA1, *Pf*EBA175, *Pf*MSP2 and *Pf*VAR2CSA, irrespective of the exposure group (Figure 2 A, B, C and F). The levels of *Pf*AMA1, *Pf*EBA175 and *Pf*VAR2CSA for a considerable number of women without malaria remained at a low level over gestation, unlike a significant number of *Pf*MSP2 responses of controls which remained at a high level. Only a few responses of the cases remained high throughout the study period for *Pf*AMA1 and *Pf*EBA175, whereas for *Pf*MSP2, the majority of the cases remained high. None of the samples remained at high levels for *Pf*VAR2CSA, and the responses which remained at low levels were maintained by majority of controls and a considerable number of cases. Of note, the antibody profiles of *Pf*MSP3 and *Pv*AMA1 were overall quite low and less noisy regardless of gestation weeks (Figure 2 D and E) and the majority of the responses of the controls remained at low levels while the profiles of the cases fluctuated.

The LOESS average response of each antibody indicates overall a positive linear association between the antibody level and gestation week during the first half of the gestation, even though there is great variability within a pregnant woman. This linear association plateaus (or slightly drops in some cases) at around 20 weeks of gestation. The patterns of the LOESS curve constructed for each exposure group within each antibody were almost similar to the overall LOESS curve (except for the control groups of *Pf*EBA175 and *Pf*MSP2 which showed a negative linear association between the antibody level and gestation week during the first half of the gestation). For *Pf*AMA1 and *Pv*AMA1, the average antibody response of the cases remained higher compared to the controls over the gestational period while for other antibodies the average response of a case started off with a value lower from that of the controls up to around 5 gestation weeks (for *Pf*EBA175 the break-even was almost 7.5 weeks) and then remained higher. It is also worth noting that the average antibody response of *Pf*MSP2 was the highest at conception

and remained the highest during any time point compared to the other five antibodies, irrespective of the exposure group.

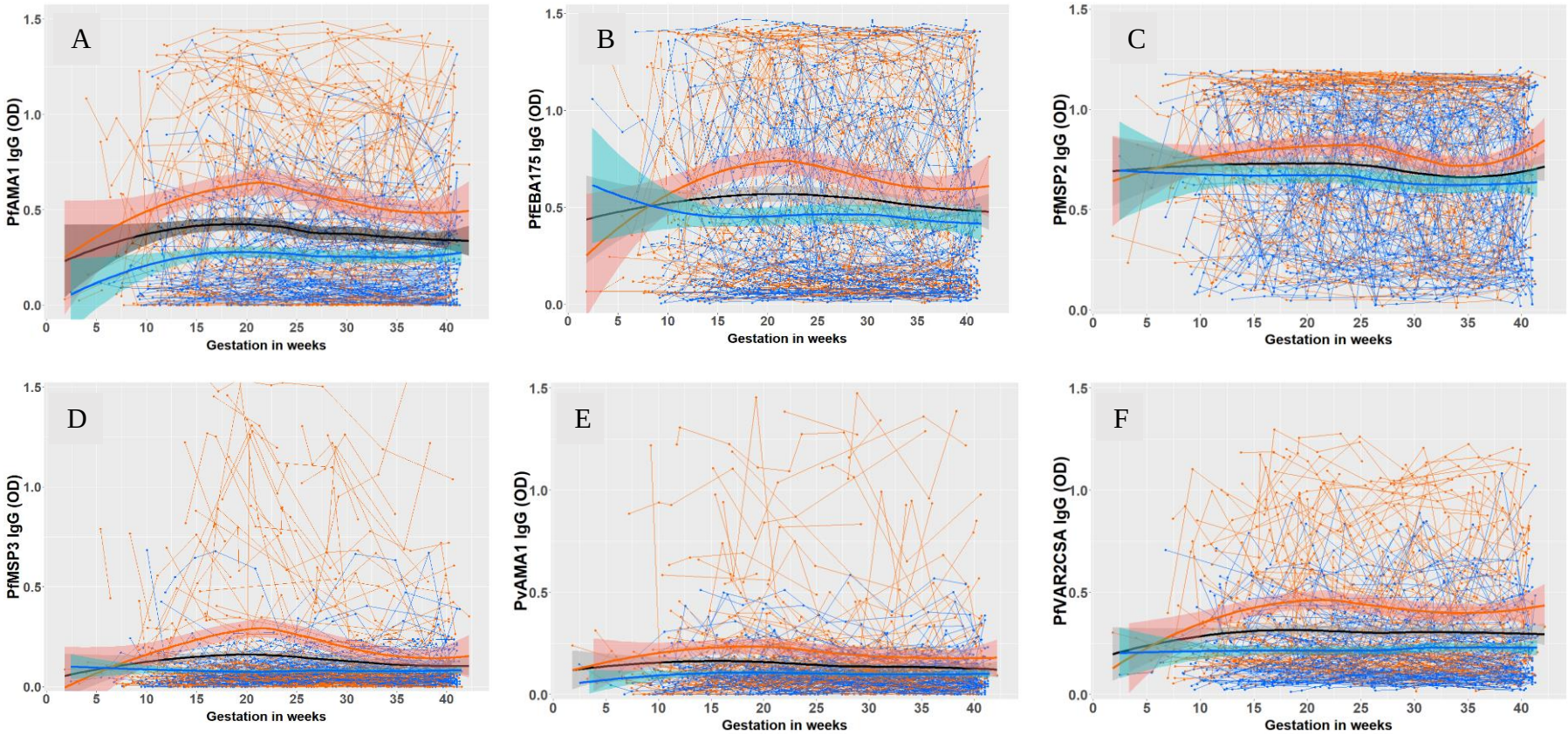


Figure 2: The spaghetti plots from A to F represent the longitudinal antibody levels of *Pf*AMA1, *Pf*EBA175, *Pf*MSP2, *Pf*MSP3, *Pv*AMA1 and *Pf*VAR2CSA, respectively, for all women. The antibody levels of the pregnant women exposed to malaria (cases) and free from malaria (controls) are represented by orange and blue, respectively. LOESS curves for all pregnant women (in black) and for each exposure group are superimposed on each spaghetti plot. The shaded area around each LOESS curve represents the 95% confidence interval (CI). Of note, the Y axes of the plots of A, B and D are truncated at 0; the CIs did extend to negative values due to limited information in the early period of gestation.

5.2.2 Serostatus of each antibody versus gestation

As explained in the Subsection 3.1.1.2, the OD values of each pregnant woman were categorised as seropositive, if the value was above a cut-off for each antigen derived from unexposed individuals from Melbourne. The stability of the serostatus throughout the study period was examined for each antibody in two ways: (i) considering the status of the samples measured initially and at the last sample before or at delivery and (ii) incorporating all serostatus measures available for each pregnant woman longitudinally.

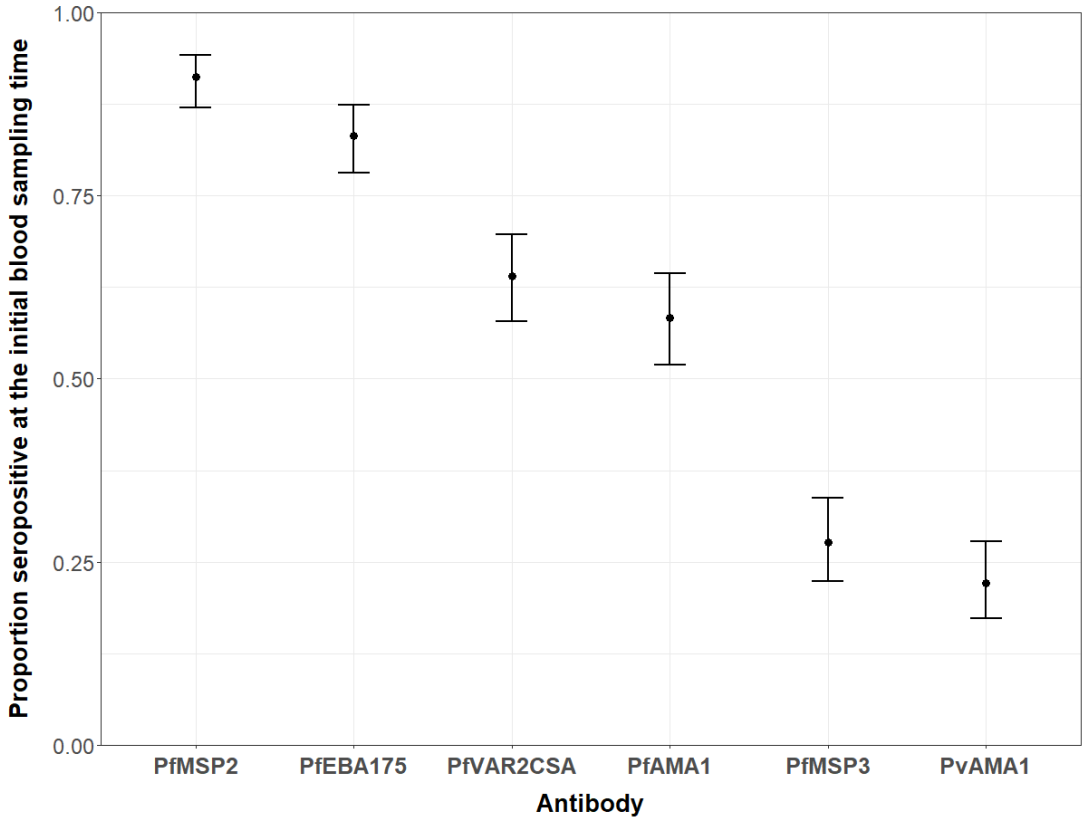


Figure 3: Proportion of pregnant women seropositive at enrolment for each antibody. The error bars represent the 95% CI for the proportion estimated for each antibody.

Figure 3 shows that the highest proportion of pregnant women were seropositive for *Pf*MSP2 at enrolment (0.91, 95% CI: 0.87,0.94), followed by *Pf*EBA175 (0.83, 95% CI:0.78,0.87). In contrast, *Pv*AMA1 (0.22, 95% CI: 0.17,0.28) and *Pf*MSP3 (0.28, 95% CI: 0.22,0.34) had the lowest proportion of seropositive measures at baseline.

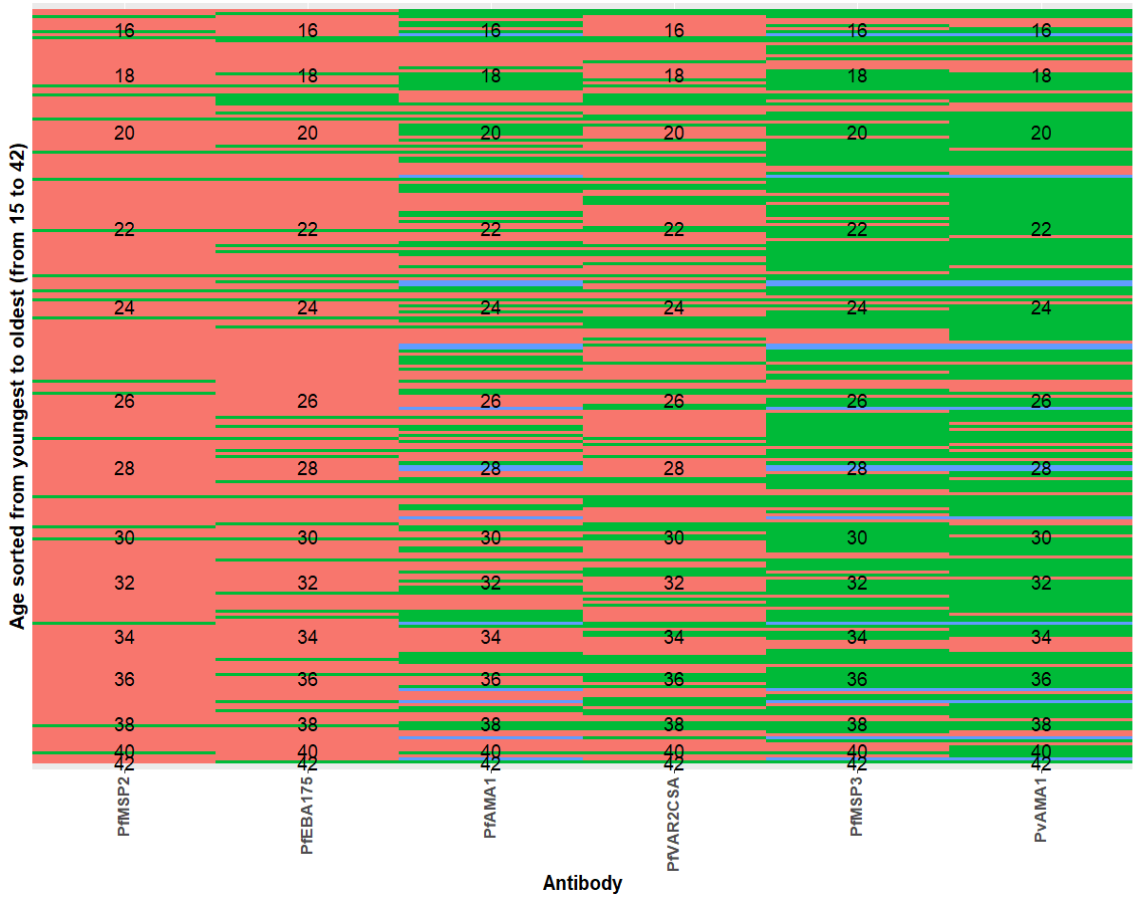


Figure 4: Heatmap of serostatus considering the first sample of each pregnant woman, ordered by age (ranging from 15 to 42) for each antibody (seropositive: red, seronegative: green and record unavailable: blue), age of women (in 2-year intervals) is displayed.

For each pregnant woman the antibody-specific seropositivity status of the first sample measured is presented in

Figure 4, ordered from the youngest to the oldest. Overall, serostatus was not associated with age for any of the six included responses. However, the majority of pregnant women were seropositive for the antibodies *Pf*MSP2 and *Pf*EBA175, while seronegative for *Pf*MSP3 and *Pv*AMA1 at first sample, irrespective of age. Furthermore, pregnant women of the same age also maintained different serostatus patterns across the antibodies.

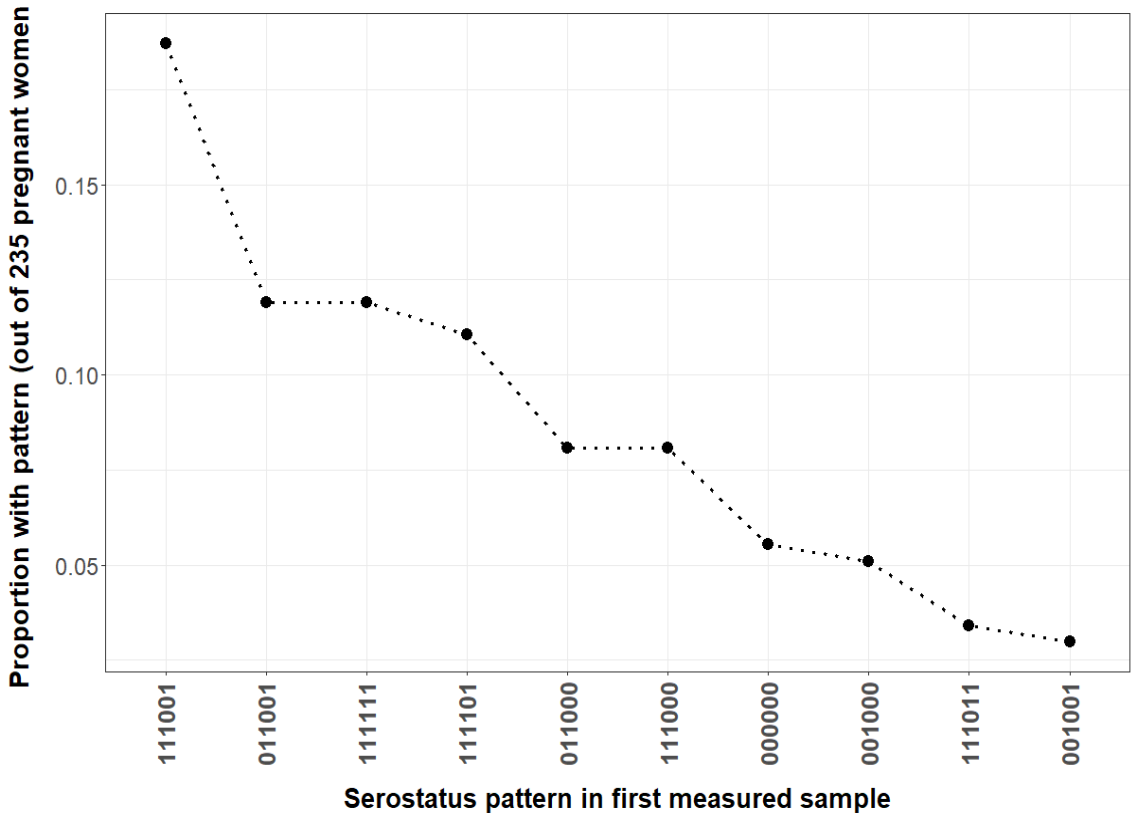


Figure 5: Proportion of joint-serostatus patterns of the pregnant women (235) who had antibody levels measured for all six antibodies at enrolment (note, 15 pregnant women were omitted in this analysis as their records were missing for at least one antibody). The joint serostatus patterns are ordered as: *Pf*AMA1, *Pf*EBA175, *Pf*MSP2, PfMSP3, *Pv*AMA1 and *Pf*VAR2CSA, respectively, where a value of 1 represents seropositive.

The seropositivity patterns represent different seropositivity statuses exhibited by each pregnant woman on all antibodies at enrolment. Of the 64 (=2^6) unique patterns, the joint seropositivity statuses of the pregnant women in this study (235) represented 28 different patterns, where the top ten patterns observed for 204 (87%) of the pregnant women are displayed in Figure 5. The most prevalent joint seropositivity pattern (19% (n=44) of pregnant women) is 111001, where the pregnant women are seropositive for all the antibodies except for *Pf*MSP3 and *Pv*AMA1. The second most equally prevalent patterns (12% n=28) correspond to the pregnant women being seropositive for all the antibodies except for *Pf*AMA1, *Pf*MSP3 and *Pv*AMA1 and being seropositive for all the antibodies.

The remaining 13% (n=31) of pregnant women (not displayed in Figure 5) maintained one of the other 18 patterns where at most 4 women were observed to have that serostatus pattern.

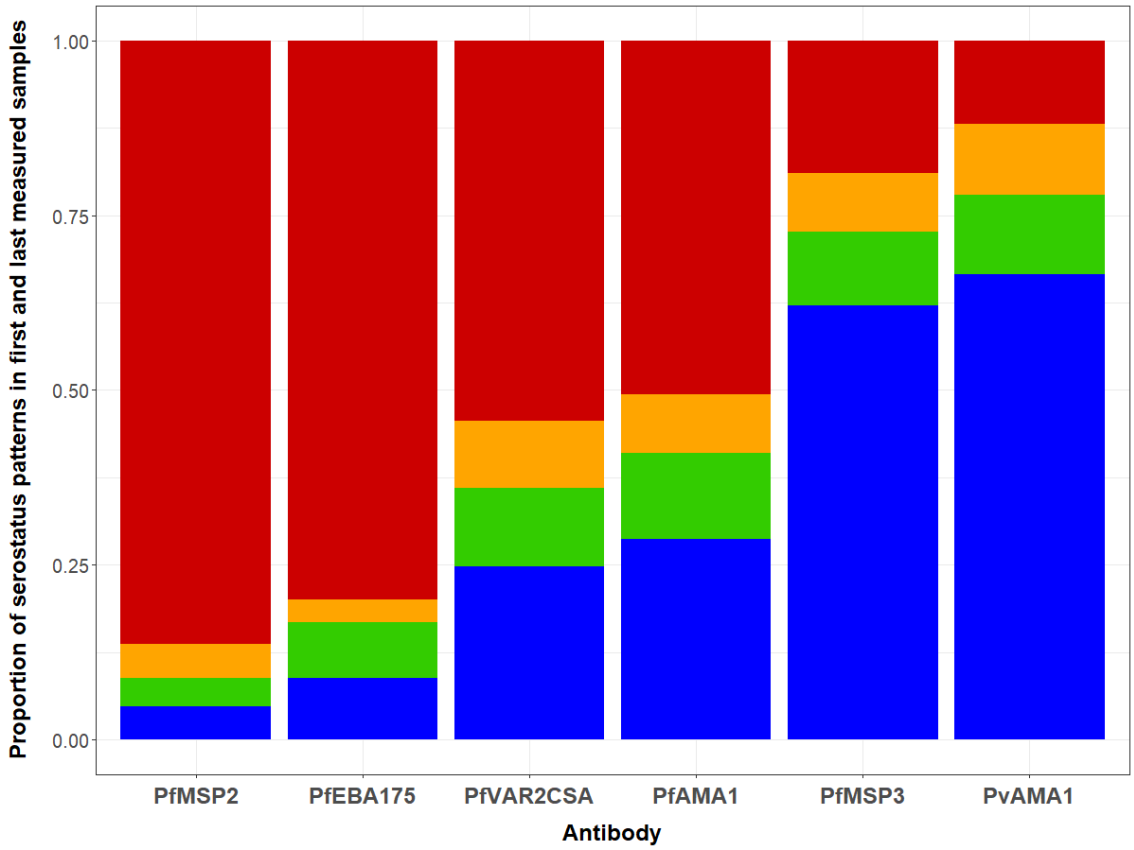


Figure 6: Proportion of serostatus patterns considering first and last measured samples for each antibody. Both first and last samples were available for all 250 pregnant women for the antibodies *Pf*EBA175, *Pf*MSP2 and *Pf*VAR2CSA, while 227 for the antibodies *Pf*AMA1, *Pf*MSP3 and *Pv*AMA1. Within each antibody, based on the serostatus recorded at first and last sample, each woman was categorised into one of the four groups: both measures seronegative (blue), seronegative at first but seropositive in the last sample (green), initially seropositive but seronegative at last (orange), and both were seropositive (red).

The total number of pregnant women who had records for both first and last samples were not the same for all the antibodies, hence, Figure 6 presents the proportion of pregnant women for each antibody. Overall, for more than 75% of the pregnant women, both the first and last samples remained either seropositive or seronegative. The highest number of pregnant women whose both samples were seropositive was observed in *Pf*MSP2 (86% (216/250)) followed by *Pf*EBA175 (80% (200/250)). Contrarily, the highest number of pregnant women whose both samples were seronegative (67% (151/227)) was observed for *Pv*AMA1, followed by *Pf*MSP3 (62% (141/227)).

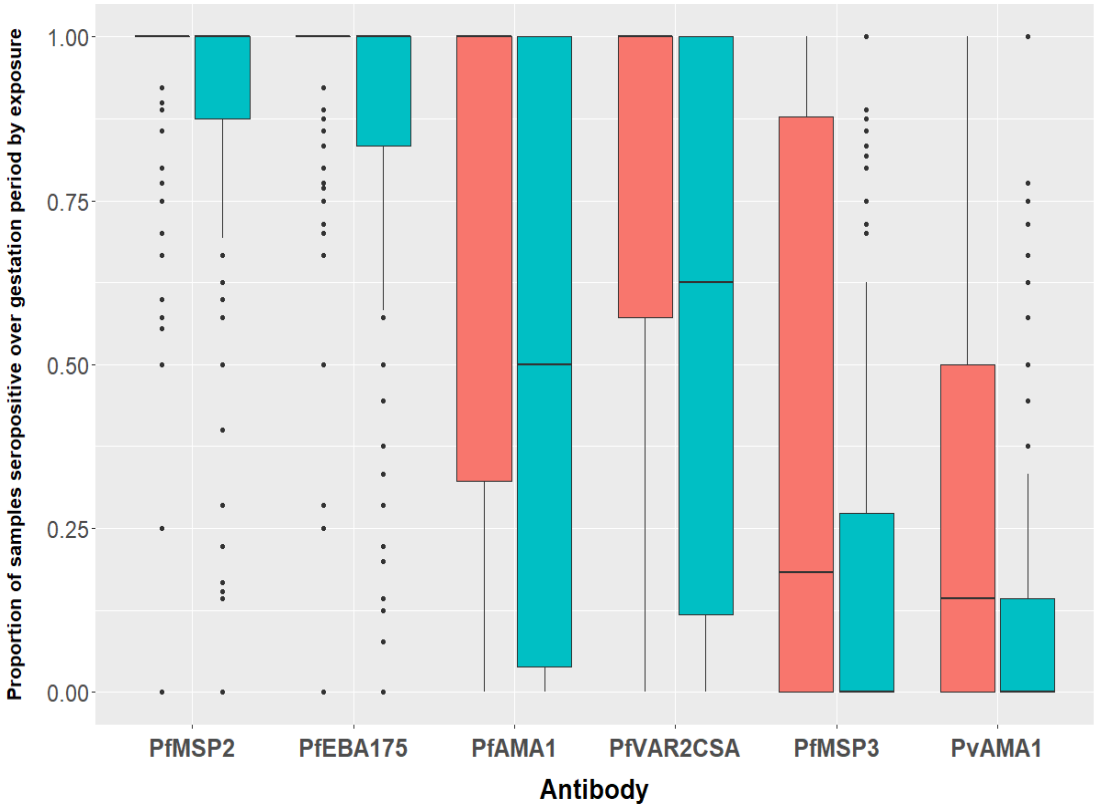


Figure 7: Boxplots for the proportion of samples seropositive for each antibody available for each pregnant woman longitudinally over the gestation period by the exposure category (cases in pink and controls in blue). Overall, all 250 pregnant women in the study had longitudinal measurements for the antibodies *Pf*EBA175, *Pf*MSP2 and *Pf*VAR2CSA, while longitudinal measurements for antibodies *Pf*AMA1, *Pf*MSP3 and *Pv*AMA1 were available for only 247 (missing for 3 cases) of the pregnant women. Of note, the number of longitudinal measurements available for each woman varied between 2 and 13.

The proportion of samples seropositive in the longitudinal data of each woman measured for each antibody is summarised by the exposure category in Figure 7. Overall, for *Pf*MSP2 and *Pf*EBA175, all available longitudinal measurements of half of the pregnant women in the control group remained seropositive while all available measurements were seropositive for almost all the cases. In contrast, for *Pv*AMA1 and *Pf*MP3, all available measurements of half of the controls remained seronegative while half of the cases had at most 14% and 18% of their measured samples seropositive on the respective antibody. Of note, significant variations in maintaining the serostatus over the trial was observed in *Pf*AMA1 as the overall serostatus of the pregnant women ranged from all seronegative to all seropositive, irrespective of the exposure category (however, for the majority of the

pregnant women in the case group, most of the available measures were seropositive, median of 100%).

Of note, the maintenance of the serostatus measures of the majority of pregnant women as either seropositive or seronegative throughout the gestation resulted in the parameter estimates being a singular fit (i.e. estimates were the same at each iteration), once the model was fitted using the serostatus data. Therefore, the model-based clustering was performed only considering the optical density measures of the antibodies, which captures more subtle differences that occur in antibody responses and the outcomes are as follows.

5.3 Cluster pregnant women based on individual component probabilities

A model with 2 clusters provided the best fit to the data (refer to the Appendix A for the diagnostics used to determine the optimal number of clusters). Women were classified into the group that had the highest individual component probability (see Section 0 of methods chapter). The antibody profiles of the pregnant women were classified into two clusters: 186 into cluster 1 and 55 pregnant women into cluster 2, with 9 pregnant women unclassified. The classification outcome remained the same irrespective of the summary statistic (the posterior mean or median used to estimate the component probabilities) used for the initial classification procedure.

The clustering procedure has clearly classified the trajectories of all 6 antibodies into a group of women who maintain comparatively low levels of antibodies across all 6 antibodies (i.e. Cluster 1) and a group of women who maintain relatively higher antibody levels over time (Cluster 2) (Figure 8). Based on the different antibody maintenance levels of the two clusters, clusters 1 and 2 are named “low immune group” and “high immune group”, respectively.

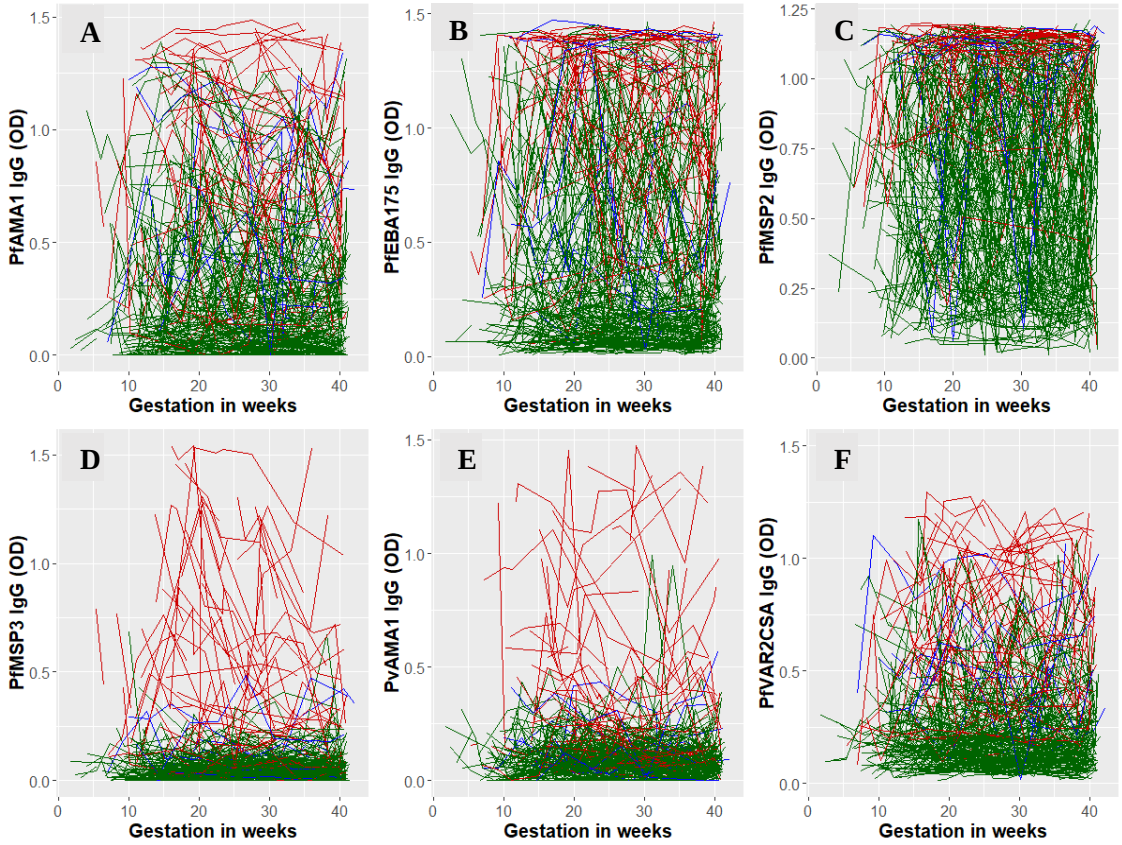


Figure 8: Trajectory plot of the antibody responses versus gestation (weeks) for each cluster (cluster 1-green, cluster 2 -red). The trajectories in blue are the pregnant women who were not classified with certainty into either cluster.

Association between the maternal factors and the antibody responses

The multivariate mixture linear mixed model showed that age and receiving chloroquine as a prophylaxis were positively associated with the responses for the majority of antibodies; the exceptions were *Pv*AMA1 for age (estimated mean change of -0.001 per year of change (95% CrI -0.002,0.0004)) and *Pf*MSP3 for chloroquine (-0.003 (95% CrI -0.019,0.013)). Negative associations were observed between gravidity and the antibodies of *Pf*AMA1, *Pf*MSP2 and *Pf*MSP3, contrary to the association between gravidity and the other antigens (Table 3). Of note, having a history of malaria was only associated with the antibody *Pv*AMA1 (estimated mean level was 0.012 (95% CrI -0.005, 0.027) higher in a woman with a history of malaria compared to a woman without prior infection(s)). The absence of influence of having prior infection(s) on maintaining higher immunity levels with respect to other antibodies could be due to the recording of history of malaria which was limited to infections prior to pregnancy documented at SMRU clinics. Hence cannot guarantee the accuracy of the recordings nor the exact time of exposure. As expected, the average antibody level at the start of pregnancy was higher in the high immune group compared to the low immune group in all six antibodies; mean antibody levels at conception ranged from 0.448 (*Pv*AMA1) to 1.076 (*Pf*EBA175) higher in the high immune group than in the low immune group. The antibody responses to *Pf*AMA1 and *Pv*AMA1 antigens had a decreasing trend over gestation time in both the high and low immune groups. Average antibody responses of *Pf*EBA175, *Pf*MSP2, *Pf*VAR2CSA and *Pf*MSP3 over gestation week showed opposite trends by immune group. The antibodies of *Pf*EBA175, *Pf*MSP2 and *Pf*VAR2CSA had declining trends over gestation time in the low immune group but had increasing trends in the high immune group; the associations were reversed for *Pf*MSP3 antibodies.

Table 3: Multivariate linear mixed-effects modelling of the combined IgG responses to *Plasmodium falciparum* and *Plasmodium vivax* recombinant antigens

Variable	*Pf*AMA1	*Pf*EBA175	*Pf*MSP2	*Pf*MSP3	*Pv*AMA1	*Pf*VAR2CSA
Fixed effects						
Age (years)	0.001 (-0.006,0.008)	0.007 (-0.001,0.015)	0.005 (-0.001,0.01)	0.001 (0.0001,0.003)	-0.001 (-0.002,0.0004)	0.003 (-0.0001,0.007)
Gravidity[a]						
Primigravidae	0.064 (-0.05,0.179)	-0.017 (-0.153,0.118)	0.013 (-0.075,0.102)	0.004 (-0.018,0.027)	-0.005 (-0.029, 0.02)	-0.012 (-0.069,0.046)
Intervention group[b]						
Chloroquine	0.054 (-0.021,0.13)	0.048 (-0.041,0.136)	0.009 (-0.05,0.067)	-0.003 (-0.019,0.013)	0.019 (0.002,0.036)	0.031 (-0.007,0.069)
History of malaria						
Yes	-0.04 (-0.116,0.036)	-0.024 (-0.113,0.065)	-0.02 (-0.08,0.039)	-0.002 (-0.018,0.013)	0.012 (-0.005,0.028)	-0.014 (-0.053, 0.024)
Average random effects for a woman						
Low immune						
Constant[c]	0.31 (0.231,0.388)	0.469 (0.373,0.564)	0.704 (0.631,0.777)	0.069 (0.048,0.09)	0.095 (0.073,0.117)	0.235 (0.189,0.281)
Gestation (weeks)[d]	-0.001 (-0.002,0.0001)	-0.002 (-0.003, -0.0004)	-0.003 (-0.004, -0.001)	0.00004 (-0.001,0.001)	-0.0003 (-0.001,0.0003)	-0.001 (-0.002,0.0004)
High immune						
Constant[c]	0.864 (0.67,1.062)	1.076 (0.897,1.264)	1.000 (0.893,1.111)	0.603 (0.404,0.808)	0.448 (0.283,0.608)	0.634 (0.498,0.766)
Gestation (weeks)[d]	-0.0004 (-0.006,0.005)	0.0001 (-0.005,0.005)	0.001 (-0.002,0.005)	-0.006 (-0.012, -0.001)	-0.003 (-0.008,0.002)	0.002 (-0.003,0.006)

Values are coefficients (95% CrI) of optical density values per unit change in covariate.

[a] Vs multigravida.

[b] Vs Placebo.

[c] The mean antibody level for a woman at conception who is at mean age (26.16 years), multigravida, did not receive chloroquine prophylaxis and absence of exposure to malaria prior to enrolment.

[d] The rate of change of the specific antibody per gestation week.

5.4 Group-specific characteristics of the pregnant women

Table 4 includes summaries of key factors for both clusters. The high immunity group was dominated by the cases (92.7%), particularly the cases who were only exposed to falciparum malaria (28 of 51 cases) and the low immunity group by the controls (57.5%). This is not surprising given that it is known that malaria infections temporarily boost antibody levels. As such, the proportion of pregnant women in the high immunity group who were anemic at enrolment (34.5%) was three-fold greater compared with the low immunity group (12.4%). Furthermore, the proportion of pregnant women who resided in refugee camps was two-fold higher in the low immunity group (78.5%) compared to the high immunity group (41.8%). Almost a similar proportion in the low (32.1%) and high immunity groups (38.9%) had a history of *P. falciparum* infection/s prior to enrolment. However, the proportion of pregnant women who were infected with *P. vivax* malaria prior to enrolment in the low immunity group (21.2%) was around twice that in the high immunity group (11.1%). Of note, a similar pattern was observed in the classification of pregnant women who were only exposed to vivax malaria during the study period, 48.1% and 9.8% in the low and high immunity groups respectively.

Table 4: Characteristics of the pregnant women by low and high immune antibody response profile clusters.[a]

Characteristic	Cluster 1 Low immune group (n=186)	Cluster 2 High immune group (n=55)
Exposed to malaria during pregnancy		
Case, n (%)	79 (42.5)	51 (92.7)
P. falciparum only n (%)	23 (29.1)	28 (54.9)
P. vivax only n (%)	38 (48.1)	5 (9.8)
Both *P. falciparum* and *P. vivax* n (%)	16 (20.3)	13 (25.5)
P. falciparum, P. vivax and mixed infection n (%)	2 (2.5)	2 (3.9)
Other three infections[b] n (%)	0 (0)	3 (5.9)
Control, n (%)	107 (57.5)	4 (7.3)
Age (years), median (IQR)	25 (21, 31)	24 (20, 30.5)
Gravidity, median (IQR)	3 (2, 5)	3(2, 5.5)
Primigravida, n (%)	33 (17.7)	13 (23.6)
Multigravida, n (%)	153 (82.3)	42 (76.4)
Parity, median (IQR)	2 (1, 3)	1 (0, 3.5)
Haematocrit (%), median (IQR)	33.8 (31.5, 36)	31.5 (27.6, 33.8)
Anaemia, n (%)	23 (12.4)	19 (34.5)
Residence in refugee camp, n (%)	146 (78.5)	23(41.8)
Receiving chloroquine prophylaxis, n (%)	85 (45.5)	24 (42.9)
Estimated Gestational Age (weeks), median (IQR)	9.3 (7.3, 12.3)	10.9 (7.8, 16.8)
Trimester		
1 (<14wks), n (%)	157 (84.4)	36 (65.5)
2 (14 to <28wks), n (%)	28 (15.1)	18 (32.7)
3 (28 wks or more), n (%)	1 (0.5)	1 (1.8)
***Plasmodium* spp. before enrolment, n (%)**	89 (47.8)	26 (47.3)
Infected with *P. falciparum*, n (%)	59 (32.1)	21 (38.9)
Infected with *P. vivax*, n (%)	39 (21.2)	6 (11.1)
Follow up (weeks), median (range)	30.8 (26.6, 32.4)	26.3 (19,31.1)

[a]9 women who were not classified into any cluster were excluded.

[b]3 other infections include *P. falciparum* and *P. malariae* infection, *P. vivax* and *P. malariae* infection and *P. falciparum, P. vivax* and *P. malariae* infection with one pregnant woman in each type of infection.

Influence of specific antibodies on the classification into high and low immunity profiles

Table 5 shows the entropy values for each antibody arranged in descending order. The high entropy values for *Pf*MSP3 (0.949) and *Pv*AMA1 (0.935) indicate that the profiles from these antigens drove the classification of pregnant women among the clusters, compared to the other antigens, whereas *Pf*MSP2 was least able to discriminate between the two groups (entropy = 0.763).

Table 5: Influence of specific antibodies on the classification into high and low immunity profiles.

Antibody	Entropy value[a]
*Pf*MSP3	0.949
*Pv*AMA1	0.935
*Pf*VAR2CSA	0.873
*Pf*AMA1	0.844
*Pf*EBA175	0.817
*Pf*MSP2	0.763

[a]The entropy is a value between 0 and 1 and values closer to 1 indicate that an antibody highly contributes towards the classification of women to cluster 1 or 2. The influence on classification declines as the entropy values declines.

5.5 Sensitivity analysis

5.5.1 Importance of including *Pf*MSP3 and *Pv*AMA1

According to the entropy values in Table 5, *Pf*MSP3 and *Pv*AMA1 have the highest contribution to classifying the pregnant women into the two immunity profile groups. The biological mechanism of the antigens associated with these two antibodies are quite different from the other antibodies (i.e. *Pv*AMA1 is a *P. vivax* antigen while *Pf*MSP3 does not connect directly to the merozoite surface unlike the four other *P. falciparum* related antibodies). The importance of incorporating these two antibodies in separating the low and high immune antibody response profiles was assessed by repeating the analyses with and without the inclusion of these two antibodies.

Table 6: Assessment the importance of combining *Pv*AMA1 with other *P. falciparum* antibodies

		Cluster groupings obtained considering only *P. falciparum* antibodies (i.e. dropping *Pv*AMA1)			
		Cluster 1	Cluster 2	Unclassified	Total
Cluster groupings obtained considering all antibodies	Cluster 1	178	00	08	186
	Cluster 2	03	48	04	55
	Unclassified	02	04	03	09
	Total	183	52	15	250

According to Table 6, 178 of 186 (96%) pregnant women who were originally classified into Cluster 1 (classified including *Pv*AMA1 with the *P. falciparum* antibodies) was identified similarly even after dropping *Pv*AMA1. Furthermore, 87% (48/55) of the pregnant women who were originally classified into Cluster 2 could be identified without incorporating *Pv*AMA1.

Table 7: Assess the importance of combining *Pf*MSP3 with other antibodies

		Cluster groupings obtained considering all antibodies without *Pf*MSP3			
		Cluster 1	Cluster 2	Unclassified	Total
Cluster groupings obtained considering all antibodies	Cluster 1	173	01	12	186
	Cluster 2	03	49	03	55
	Unclassified	03	03	03	09
	Total	179	53	18	250

Similarly to Table 6, the majority of the pregnant women who were originally classified (considering all six antibodies) among the two clusters were identified in a similar manner after dropping *Pf*MSP3 (173 of 186 (93%) and 49 of 55 (89%), who were originally classified into Cluster 1 and Cluster 2 respectively).

Table 8: Assess the importance of including at least *Pf*MSP3 or *Pv*AMA1 with other *Plasmodium falciparum* antibodies (*Pf*AMA1, *Pf*EBA175, *Pf*MSP2 and *Pf*VAR2CSA).

		Cluster groupings excluding *Pv*AMA1 and *Pf*MSP3			
		Cluster 1	Cluster 2	Unclassified	Total
Cluster groupings obtained considering all antibodies	Cluster 1	87	79	20	186
	Cluster 2	02	53	00	55
	Unclassified	01	08	00	09
	Total	90	140	20	250

Table 8 examines the possibility of clustering pregnant women omitting both *Pv*AMA1 and *Pf*MSP3. Even though the *P. falciparum* antibodies identified almost all the pregnant women who were classified into Cluster 2 without *Pf*MSP3 (and *Pv*AMA1) (53 of 55 (96%)), only 87 of the original 186 (47%) Cluster 1 women were classified into Cluster 1.

Hence, dropping one of *Pv*AMA1 (Table 6) or *Pf*MSP3 (Table 7) does not have a major impact on the classification outcomes obtained originally (considering all six antibodies), but removing both these antibodies caused significant changes in the classification outcomes, especially in Cluster 1 (Table 8). Hence, it is vital to consider at least one of these two antibodies (*Pf*MSP3 is preferred as it has the highest entropy value and is a *P. falciparum* antibody) in classification in order to more precisely capture the pregnant women who categorise into Cluster 1 with comparatively low antibody profiles. As a further analysis, *Pf*MSP3 and *Pv*AMA1 were explored in a univariate model to identify how well each antibody could identify and classify the pregnant women with similar exposure (case or control) and concur with the findings above (see Appendix A).

5.5.2 Antibodies which best identify the malaria cases

Overall, the joint levels of the six antibodies identified the controls exceptionally well by classifying them into Cluster 1. However, the cases have been poorly identified, as they are classified almost equally between the two clusters. To further investigate this, the capability of different combinations of antibodies in distinguishing between the pregnant women was evaluated. For this purpose, a series of univariate and multivariate analyses were undertaken. Table 9 shows the capability of each antibody response in

discriminating pregnant women into the expected cluster according to the malaria exposure status.

Table 9: Performance of classifying controls to Cluster 1 (low immunity group) and cases to Cluster 2 (high immunity group) based on univariate analyses. Antibodies are ordered from the highest to the lowest percentage of classifying pregnant women into the expected cluster.

	Number of controls[a] classified into Cluster 1 (%)	**Number of cases[b] classified into Cluster 2 (%)**	**Total women classified into the expected Cluster[c] (%)**
***Pf*AMA1**	76 (66.1)	74 (54.8)	150 (60)
***Pf*VAR2CSA**	86 (74.8)	57 (42)	143 (57.2)
***Pf*EBA175**	61 (53)	81 (60)	142 (56.8)
***Pv*AMA1**	109 (94.8)	28 (20.7)	137 (54.8)
***Pf*MSP3**	103 (89.6)	33 (24.4)	136 (54.4)
***Pf*MSP2**	63 (54.8)	67 (49.6)	130 (52)

[a] Out of 115 controls
[b] Out of 135 cases
[c] Computed based on the decision that Cluster 1 would represent controls and Cluster 2 would represent cases. The sum of controls classified into Cluster 1 and cases classified into Cluster 2 were then divided by the total pregnant women, i.e. 250 to obtain the percentage.

The results show that *Pf*AMA1 provided the best overall discrimination by classifying 60% of the pregnant women into the expected clusters (controls in Cluster 1 and cases in Cluster 2). *Pf*VAR2CSA had the second-best performance by classifying 57.2% of the pregnant women into the expected clusters.

The antibody responses to *Pv*AMA1 best identified the controls (109 of 115) by classifying them into Cluster 1, followed by *Pf*MSP3 as the second-best performer. Opposingly, *Pf*EBA175 best identified a majority of the cases (81 of 135 cases) by classifying them into Cluster 2 and *Pf*AMA1 was the second-best antibody in classifying the cases (74 of 135). Of note, even though *Pv*AMA1 and *Pf*EBA175 performed best at categorising one exposure group (controls and cases, respectively), they performed poorly with respect to identifying the pregnant women in the alternative exposure group.

Subsequent clustering analyses using combinations of antibody responses showed that the performance of classifying the cases into Cluster 2 was highest once the classification was performed using the joint antibody responses to *Pf*AMA1 and *Pf*VAR2CSA. This bivariate combination classified 64% of the total controls into Cluster 1 and 63% of the cases into Cluster 2; see Appendix A for the details of classification of the pregnant women using the combination of *Pf*AMA1 and *Pf*VAR2CSA.

Table 10: Classification of cases by the type of infection, based on antibody responses to *Pf*AMA1 and *Pf*VAR2CSA

Type of infection(s) recorded during pregnancy	**Cluster 1**	**Cluster 2**	**Unclassified**	**Total**
P. falciparum only	05	43	05	53
Both *P. falciparum* and *P. vivax*	06	24	01	31
P. falciparum, P. vivax and mixed infection	01	03	00	04
P. falciparum and *P. malariae* infection	00	01	00	01
P. falciparum, P. vivax and *P. malariae*	00	01	00	01
P. vivax and *P. malariae*	00	01	00	01
P. vivax only	29	12	03	44
Total	41	85	09	135

The majority of the cases who had at least one episode of falciparum infection (first 5 infection types) were classified into Cluster 2 (women categorised into Cluster 2 of total women with one of the first 5 infection types; 80%) (Table 10). However, the pregnant women who were infected with only *P. vivax* infection(s) were poorly categorised, i.e. a majority of them were classified into Cluster 1 (women categorised into Cluster 1 of total women with only vivax infection; 66%). Surprisingly, classification of pregnant women with only *P. vivax* infection using *Pv*AMA1 resulted in the majority of the cases into Cluster 1 which is more representative of controls (refer to the Appendix A).

Chapter 6

Discussion

Antibody response to the malaria parasite is highly dynamic and can vary greatly within and between pregnant women. This dynamicity has challenged the identification of key antibodies as potential vaccine candidates and/or sero-surveillance biomarkers of malaria infections in pregnant women.

A majority of previous studies quantifying antibody responses to malaria only used cross-sectional antibody data, or analysed longitudinal antibody data in a univariate framework, disregarding the correlations between the antibodies measured. This thesis used a statistical approach to model the longitudinal antibody data of six different antibodies from a group of 250 pregnant women in a multivariate context which accommodates for the correlation between the measured antibodies. Ultimately, the model was used to classify the pregnant women into groups based on the joint behaviour of their antibody responses. The key findings of this thesis are outlined in Section 6.1. Section 6.2 describes the strengths and limitations of this thesis, followed by some suggestions for future work in Section 6.3. In Section 6.4, some concluding remarks are presented.

6.1 Key findings

The results showed that classifying the pregnant women into two groups provided the best segregation of the antibody profiles; overall, the antibody profiles in one group remained at low levels (low immune group) compared to the other group (high immune group) across all six antibodies. A key finding of this classification analysis was that the maintenance level of the joint longitudinal antibody measures is central to separate the controls from the cases (exposed to malaria during pregnancy) so that the majority of women who were free from malaria during pregnancy were grouped into the low immune group.

The classification outcome was shown to be significantly driven by the antibodies to *Pf*MSP3 and *Pv*AMA1. The profiles of these two antibodies mostly remained at low levels with less fluctuations, irrespective of the exposure to malaria during pregnancy, contrary to the responses of the other four antibodies. The difference in the maintenance levels of these two antibodies compared to other antibodies could be a main cause for

grouping the cases, particularly those who maintained considerably lower immunity levels against these two antigens with the majority of controls. The model-based clustering using the six antibody responses performed exceptionally well for the uninfected women as 93% (107 of 115 controls) were classified into the low immunity group but poorly performed in clustering the women infected with malaria into the high immune group (51 of 135 cases (38%)). Therefore, even though all the antibodies are incorporated in the analysis, classification outcomes rely significantly on those which consist of profiles that are relatively stable over time. These results suggest that the antibodies which do not boost with each successive infection are not suitable as sero-surveillance markers of exposure to malaria during pregnancy as they hamper separation of cases from controls.

Another important finding was that the combination of anti-*Pf*AMA1 and anti-*Pf*VAR2CSA antibodies, among all possible combinations, provides the best classification result for cases, i.e. it classifies the majority of the cases (63%) into the high-immune group. This suggests that these two antibodies may be good candidates for sero-surveilling exposure to malaria during pregnancy. These results indicate the importance of evaluating the performance of a classifier when different combinations of antibodies are used.

However, the reduced clustering model of the antibodies *Pf*AMA1 and *Pf*VAR2CSA was unable to identify all the cases, as 41 of them (i.e. 30% of the cases) were classified in the low immune group. Further investigation showed that the majority of these pregnant women (29 of 41 cases (71%)) had only been infected with vivax malaria. This implies that this combination of antibodies enables identification of pregnant women who had at least one falciparum episode in their infection history. On the other hand, the antibody response to *Pv*AMA1 could not be used to distinguish these cases, as it remained at low levels throughout the pregnancy for the majority of women. Therefore, it is crucial to look for a new antibody to vivax malaria which boosts with each successive vivax infection, for identifying recent exposure to vivax malaria.

Recently, serology has been considered as a promising epidemiological tool, incorporated in studies as a proxy measure for the detection and surveillance of malaria [32, 38, 267, 268]. Interestingly, the immune surveillance studies have recommended anti-*Pf*AMA1 antibody as a biomarker of protective immunity to symptomatic malaria [168] and as an

important vaccine candidate [269, 270]. Antibody responses to *Pf*AMA1 and *Pv*AMA1 have also been considered for assessing malaria transmission intensity and for identifying hotspots of malaria transmission, specifically in low transmission settings [114]. There has been limited investigation of pregnant women in these studies, many simply stating that pregnant women could be used as a sentinel population to estimate burden of malaria [271, 272]. The available serosurveillance studies on pregnant women also only investigated the *Pf*VAR2CSA antibody, perhaps as it is the antibody specific to pregnant women. Almost all these studies have strongly recommended *Pf*VAR2CSA as a candidate for vaccine development [273, 274], where some VAR2CSA-based vaccines are being currently tested in clinical trials (https://clinicaltrials.gov/ct2/show/NCT02647489, https://clinicaltrials.gov/ct2/show/NCT02658253, https://www.clinicaltrials.gov). However, two recent studies have suggested *Pf*VAR2CSA to be used to monitor the changes in malaria transmission [275] and as a biomarker of recent exposure to malaria during pregnancy [275, 276]. Few studies have incorporated non-pregnancy specific antibodies, particularly anti-*Pf*AMA1 antibody in pregnancy related serology studies on malaria, particularly in low transmission settings [1, 277] and discovered that the anti-*Pf*VAR2CSA antibody was associated with reduced risk for high placental parasitemia but was not associated with the anti-*Pf*AMA1 antibody [277]. None of the above studies in pregnant women have considered the combined effects of the antibodies. Hence, the key finding of this study that recommends the combination of *Pf*AMA1 and *Pf*VAR2CSA antibodies as an indicator of recent exposure to malaria during pregnancy is novel and creates a new direction for serosurveillance studies.

This study also assessed the association of maternal age and gravidity with the immunity of pregnant women. Consistent with previous studies, the results show that overall, the antibody responses corresponding to *P. falciparum* parasite are positively associated with increasing age [32, 278] of a pregnant woman. However, no strong correlation with gravidity was observed, particularly in the pregnancy-specific immunity (*Pf*VAR2CSA), even though this association was shown in many previous studies [15, 216, 279, 280]. Surprisingly administration of chloroquine prophylaxis was associated with a higher *P. vivax*-specific antibody response (*Pv*AMA1), even though those receiving chloroquine prophylaxis had less *P. vivax* malaria infections, thereby limiting activation of an immune response [221]. This finding suggests that exploration of better *P. vivax* antibodies for measures of protection and sero-surveillance are required.

6.2 Strengths and Limitations

Modelling six antibody responses jointly is a considerable advantage of this study as the antibodies are inter-related in regard to their action, e.g. anti-merozoite. Examining acquired immunity to malaria longitudinally offers further advantages, as antibodies differ in terms of half-life and stability, and therefore longitudinal immunity profiles are needed to identify potential targets of protective immunity [33, 38] and indicators of recent exposure to malaria. The capability of identifying some vivax cases (particularly, coinfected with falciparum malaria) further strengthens modelling multiple falciparum antigens as it may include cross species activity which could be further explored.

Furthermore, a multivalent vaccine which consists of multiple serotypes is believed to be more efficacious as it could target multiple species and multiple strains of a parasite ([281, 282], reviewed in [283]), which was lacking in many of the past malaria vaccine candidates. This again indicates the importance of joint modelling of antibodies and taking the correlation between them into account.

Almost all of the immune surveillance studies focus only on one type of infection, falciparum [32, 38, 107] or vivax infection [284, 285] and the antigens specific to that parasite. However, in low transmission settings (e.g. many Asian countries), it would be more practical to assess the performance of the antibodies against both parasites, as pregnant woman in these areas may be infected with vivax and falciparum malaria during pregnancy (36 of 135 (27%) as in this study). Therefore, including pregnant women exposed to any infection is another strength of this study, where a combination of antibodies was discovered to perform well in identifying women who were exposed to both infections during pregnancy (28 of 36 (78%)).

This study also contains several limitations as follows. First, only one antibody marker, i.e. *Pv*AMA1, was available for the vivax malaria, limiting the understanding of immunity response in pregnant women exposed to vivax malaria. A considerably low level of parasite density – which is believed to have a proportional relationship with antibody maintenance and boosting – in vivax infections compared to falciparum could possibly explain the maintenance of *Pv*AMA1 antibody responses at low levels. However, this could not be ascertained in this study due to unavailability of the factors such as parasite density [286] and frequency of past exposures [65], which were identified by previous research to impact the level of acquired immunity. Administration of chloroquine,

preventing vivax episodes, could be another significant confounding factor in disrupting antibody responses to *Pv*AMA1 [221]. The antibody responses to the only vivax related antigen (*Pv*AMA1) remained relatively low (based on LOESS curves) compared to other antibodies (except for *Pf*MSP3), particularly in the case category and in such instance analysing all *P. falciparum* and *P. vivax* cases together and not separately could have led to the poor identification of vivax cases.

Another limitation of this study was that the detection of exposure to malaria was based on microscopy, thereby, failing to detect infections with low levels of parasitaemia. Indeed, as detailed in Chapter 3, there were some controls (78 of 115) who recorded high responses to schizont lysate ("high schizont lysate responder controls") suggesting potential submicroscopic infections that were not detected by light microscopy at fortnightly routine testing.

Although exposure to infection(s) prior to their first antibody measurement was incorporated in the analysis, the previous infections for the cohort were not limited to a certain timeframe and the exact time(s) of exposure(s) was unknown. Considering exposure to malaria prior to enrolment was just based on data from SMRU clinics (as the women could have possibly attended other clinics) and submicroscopic infections were not detected, may partly explain the negative association between immunity levels (except anti-*Pv*AMA1 antibodies) and prior infection. Another limitation was that antibody measures were only assessed during the gestation period, with the median length of time between the first and last antibody measurement being 22.5 weeks. However, Fowkes *et al*. (2012) [1] computed that at the population level, the half-life of the merozoite antibody responses could range from 0.8 to 7.6 years while the half-life for *Pf*VAR2CSA could range from 36 to 157 years. Hence, the follow-up period of this study, i.e. the gestational period, was not long enough. Another limitation of this study was the low number of measurements for some pregnant women (2 or 3 measurements), which challenged capturing seropositivity patterns of these cases. Lastly, the dataset only included pregnant women from one region (Thai-Myanmar border), thus, to validate the results and attribute the findings to all low-transmission areas, this analysis must be conducted on data collected from other regions as well.

6.3 Recommendations for future work

This research can be extended and improved in terms of immunology, data collection and statistical methodology; some of my suggestions are as follows. Firstly, the data used in this work were collected almost two decades ago, hence, attributing the findings of this work to the current immunity status of pregnant women residing on the Thai-Myanmar border might not be completely accurate. Furthermore, the study is currently limited to the women residing on the Thai-Myanmar border, hence this framework needs to be repeated on multiple datasets across different transmission settings in order to generalise the findings to all pregnant women. It is also recommended to incorporate PCR-based diagnostic methods to detect sub microscopic infections in future serology studies.

Immuno-epidemiological studies in malaria are currently designed based on convenience and cost, that is, the sampling designs are not developed based on the purpose of a study; e.g., it is questionable whether the frequency of collecting the blood samples in this study is adequate to capture changes in the antibody profiles over time. Hence, future studies should develop an optimal sampling design for quantifying antibody responses of pregnant women.

Using longitudinal data, our study demonstrated that the antibodies to *Pf*AMA1 and *Pf*VAR2CSA best separates the pregnant women with a recent exposure to malaria from those who were free from malaria. However, using longitudinal data is not practical out in the field as the status of exposure to malaria is normally identified using a test performed at a particular time. Therefore, further serostatus studies need to be undertaken, so that based on the seropositivity patterns of the anti-*Pf*AMA1 and anti-*Pf*VAR2CSA antibodies at a particular time, the status of exposure to malaria could be revealed – note, selecting the right antibodies for analysing serostatus is extremely important, as our results showed that a majority of measurements remained either seropositive or seronegative over time, specifically for *Pf*MSP2, *Pf*EBA175 and *Pv*AMA1, irrespective of exposure category. Furthermore, since the prevalence of submicroscopic infections are high, particularly in low transmission settings, it would be worth investigating the capability of these two antibodies in detecting submicroscopic infections by conducting a longitudinal study where sensitivity PCR has been used to detect malaria infection in pregnant women.

Future work is also required to improve identification of the pregnant women who were exposed only to vivax malaria. For this purpose, an antibody specific to *P. vivax* infection must be found which boosts with exposure to vivax malaria and remains at low levels otherwise. Finding a suitable *P. vivax* antigen is challenging as there is a limited understanding of the *P. vivax* antigens, compared with *P. falciparum*. For example, a 19 kDa fragment of merozoite surface protein ($Pv\text{MSP}{-}1_{19}$) [287] is a possible alternative to *Pv*AMA1 antigen, as the natural acquired immunity against $Pv\text{MSP}{-}1_{19}$ antigen tends to be higher compared to anti-*Pv*AMA1 antibody responses [288]. Another possible approach to capture *P. vivax* infections is to incorporate antigens of other stages of the parasite life cycle (apart from the proteins in the blood stage), such as sporozoite and liver stage proteins (for instance, circumsporozoite protein (CSP), liver-stage antigen-1 (LSA-1) and thrombospondin-related adhesive protein (TRAP)), where CSP of both falciparum and vivax parasites has been identified as a potential biomarker for the surveillance of exposure to malaria in low transmission settings of Brazilian Amazon [287].

Discovering novel variants of antigens may be worthwhile, as the corresponding antibody responses may better segregate the pregnant women. For instance, although most immunity studies on the EBA antigen use the antibodies against the variant EBA175, it was discovered that EBA140 and EBA181 are also important [289] mediators of protection against malaria [290]. Furthermore, the present study was limited to antigen combinations of merozoite surface (GPI-anchored and non-GPI-anchored), micronemes and an erythrocytic antigen. New classification studies can be undertaken in future using microneme and rhoptry antigens, as these are considerably informative about the prospective risk of malaria infection; e.g., the antibody combinations of EBA and PfRh antigens are strongly associated with protective immunity to falciparum malaria [38].

Incorporating Immunoglobulin M (IgM) antibody responses in the analysis rather than limiting to IgG antibodies might also contribute to identifying novel serosurveillance markers of malaria infections during pregnancy. IgM immunity, which is the first antibody to be induced at the early stage of infection but rapidly decays and switches to IgG after several weeks, has been investigated in recent serology studies. Some studies have found IgM to be a potential biomarker of protective immunity [291, 292] whereas another study has suggested IgM as a marker of exposure to malaria [293].

A few improvements can be made in the **mixAK** package as follows. Initially, a direct method should be developed to illustrate the importance of each variable in performing classification. Secondly, as explained in Kom*á*rek & Kom*á*rkov*á* in 2014 [248], the classification outcomes would be more informative if the fixed effects vectors (e.g. the variables, age and primigravity in this study) be cluster specific, as these could impact the immunity of individuals differently across the clusters. Similarly, antibody-specific fixed effects can be used instead of the same fixed effects for all antibodies (e.g.: once linear mixed effect models were fitted to each antibody separately incorporating the four fixed effects, using the **lmerTest** package in the R software, age had a significant impact on all the antibodies and for *Pf*MSP3 it is the only significant fixed effect at 10% significance level. Also, history of malaria was associated only with *Pv*AMA1 at 10% significance level).

6.4 Conclusion

This study demonstrates that the combination of anti-*Pf*AMA1 and anti-*Pf*VAR2CSA antibodies could be a potential biomarker of recent exposure to malaria during pregnancy. In addition, these analyses showed that the antibodies, *Pf*MSP3 and *Pv*AMA1, did not boost with an infection, and therefore, these antibodies are not suitable markers for sero-surveillance of recent exposure to malaria during pregnancy.

A significant number of malaria-endemic regions are transitioning to low transmission levels where a larger proportion of infections are of low parasite densities, hence undetected by many standard surveillance tools. The combined antibody diagnostic tool (using anti-*Pf*AMA1 and anti-*Pf*VAR2CSA) proposed in this study may facilitate the detection of microscopic infections as an alternative to standard diagnostic methods, particularly in low transmission settings. This proposed malaria serosurveillance tool may enhance malaria control and progress efforts to eliminate malaria.

References

1. Fowkes FJ, McGready R, Cross NJ, Hommel M, Simpson JA, Elliott SR, Richards JS, Lackovic K, Viladpai-Nguen J, Narum D *et al*: **New insights into acquisition, boosting, and longevity of immunity to malaria in pregnant women**. *J Infect Dis* 2012, **206**(10):1612-1621. doi:10.1093/infdis/jis566.
2. World Health Organization: **World malaria report 2019**. Geneva; 2019.
3. World Health Organization: **World malaria report 2014**. Geneva; 2014.
4. World Health Organization: **World malaria report 2018**. Geneva; 2018.
5. Rogerson SJ, Desai M, Mayor A, Sicuri E, Taylor SM, van Eijk AM: **Burden, pathology, and costs of malaria in pregnancy: new developments for an old problem**. *The Lancet Infectious Diseases* 2018, **18**(4):e107-e118.
6. Dellicour S, Tatem AJ, Guerra CA, Snow RW, ter Kuile FO: **Quantifying the number of pregnancies at risk of malaria in 2007: a demographic study**. *PLoS Med* 2010, **7**(1):e1000221. doi:10.1371/journal.pmed.1000221.
7. Nussenblatt V, Semba RD: **Micronutrient malnutrition and the pathogenesis of malarial anemia**. *Acta Trop* 2002, **82**(3):321-337.
8. Guyatt HL, Snow RW: **Impact of malaria during pregnancy on low birth weight in sub-Saharan Africa**. *Clin Microbiol Rev* 2004, **17**(4):760-769.
9. Steketee RW, Nahlen BL, Parise ME, Menendez C: **The burden of malaria in pregnancy in malaria-endemic areas**. *The American Journal of Tropical Medicine and Hygiene* 2001, **64**(1_suppl):28-35.
10. Hviid L: **The immuno-epidemiology of pregnancy-associated Plasmodium falciparum malaria: a variant surface antigen-specific perspective**. *Parasite Immunol* 2004, **26**(11-12):477-486.
11. Saute F, Menendez C, Mayor A, Aponte J, Gomez-Olive X, Dgedge M, Alonso P: **Malaria in pregnancy in rural Mozambique: the role of parity, submicroscopic and multiple Plasmodium falciparum infections**. *Trop Med Int Health* 2002, **7**(1):19-28.
12. Rogerson SJ, Van den Broek N, Chaluluka E, Qongwane C, Mhango C, Molyneux ME: **Malaria and anemia in antenatal women in Blantyre, Malawi: a twelve-month survey**. *The American Journal of Tropical Medicine and Hygiene* 2000, **62**(3):335-340.
13. Dicko A, Mantel C, Thera MA, Doumbia S, Diallo M, Diakité M, Sagara I, Doumbo OK: **Risk factors for malaria infection and anemia for pregnant women in the Sahel area of Bandiagara, Mali**. *Acta Trop* 2003, **89**(1):17-23.
14. Brabin BJ: **An analysis of malaria in pregnancy in Africa**. *Bull World Health Organ* 1983, **61**(6):1005-1016.
15. Staalsoe T, Shulman CE, Bulmer JN, Kawuondo K, Marsh K, Hviid L: **Variant surface antigen-specific IgG and protection against clinical consequences of pregnancy-associated Plasmodium falciparum malaria**. *The Lancet* 2004, **363**(9405):283-289.
16. Beeson JG, Mann EJ, Elliott SR, Lema VM, Tadesse E, Molyneux ME, Brown GV, Rogerson SJ: **Antibodies to variant surface antigens of Plasmodium falciparum–infected erythrocytes and adhesion inhibitory antibodies are associated with placental malaria and have overlapping and distinct targets**. *The Journal of Infectious Diseases* 2004, **189**(3):540-551.
17. McGready R, Lee S, Wiladphaingern J, Ashley E, Rijken M, Boel M, Simpson J, Paw M, Pimanpanarak M, Mu O: **Adverse effects of falciparum and vivax**

malaria and the safety of antimalarial treatment in early pregnancy: a population-based study. *The Lancet Infectious Diseases* 2012, **12**(5):388-396.
18. Moore KA, Simpson JA, Scoullar MJL, McGready R, Fowkes FJI: **Quantification of the association between malaria in pregnancy and stillbirth: a systematic review and meta-analysis**. *The Lancet Global Health* 2017, **5**(11):e1101-e1112. doi:10.1016/s2214-109x(17)30340-6.
19. Bardají A, Sigauque B, Sanz S, Maixenchs M, Ordi J, Aponte JJ, Mabunda S, Alonso PL, Menéndez C: **Impact of malaria at the end of pregnancy on infant mortality and morbidity**. *J Infect Dis* 2011, **203**(5):691-699.
20. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, Lalli M, Bhutta Z, Barros AJ, Christian P: **Every Newborn: progress, priorities, and potential beyond survival**. *The Lancet* 2014, **384**(9938):189-205.
21. Umbers AJ, Aitken EH, Rogerson SJ: **Malaria in pregnancy: small babies, big problem**. *Trends in Parasitology* 2011, **27**(4):168-175.
22. Christensen DL, Kapur A, Bygbjerg IC: **Physiological adaption to maternal malaria and other adverse exposure: low birth weight, functional capacity, and possible metabolic disease in adult life**. *International Journal of Gynecology & Obstetrics* 2011, **115**:S16-S19.
23. World Health Organization: **World malaria report 2017**. In. Geneiva; 2017.
24. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Mao S, Sam B: **Spread of artemisinin resistance in Plasmodium falciparum malaria**. *N Engl J Med* 2014, **371**(5):411-423.
25. Diagnostics TmCGoDa: **A research agenda for malaria eradication: diagnoses and diagnostics**. *PLoS Med* 2011, **8**(1):e1000396. doi:10.1371/journal.pmed.1000396.
26. Langhorne J, Ndungu FM, Sponaas A-M, Marsh K: **Immunity to malaria: more questions than answers**. *Nat Immunol* 2008, **9**(7):725-732.
27. Marsh Ka, Kinyanjui S: **Immune effector mechanisms in malaria**. *Parasite Immunol* 2006, **28**(1-2):51-60.
28. Ryg-Cornejo V, Ly A, Hansen DS: **Immunological processes underlying the slow acquisition of humoral immunity to malaria**. *Parasitology* 2016, **143**(2):199-207.
29. Achtman A, Bull P, Stephens R, Langhorne J: **Longevity of the immune response and memory to blood-stage malaria infection**. In: *Immunology and immunopathogenesis of malaria.* edn.: Springer; 2005: 71-102.
30. Cutts JC, Powell R, Agius PA, Beeson JG, Simpson JA, Fowkes FJ: **Immunological markers of Plasmodium vivax exposure and immunity: a systematic review and meta-analysis**. *BMC Med* 2014, **12**(1):150. doi:10.1186/s12916-014-0150-1.
31. Sturrock HJ, Hsiang MS, Cohen JM, Smith DL, Greenhouse B, Bousema T, Gosling RD: **Targeting asymptomatic malaria infections: active surveillance in control and elimination**. *PLoS Med* 2013, **10**(6):e1001467. doi:10.1371/journal.pmed.1001467.
32. Stanisic DI, Fowkes FJ, Koinari M, Javati S, Lin E, Kiniboro B, Richards JS, Robinson LJ, Schofield L, Kazura JW *et al*: **Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response**. *Infect Immun* 2015, **83**(2):646-660. doi:10.1128/IAI.02398-14.
33. Fowkes FJ, 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**(1):e1000218. doi:10.1371/journal.pmed.1000218.

34. Crompton PD, Kayala MA, Traore B, Kayentao K, Ongoiba A, Weiss GE, Molina DM, Burk CR, Waisberg M, Jasinskas A: **A prospective analysis of the Ab response to Plasmodium falciparum before and after a malaria season by protein microarray**. *Proceedings of the National Academy of Sciences* 2010, **107**(15):6958-6963.
35. Osier FH, Fegan G, Polley SD, Murungi L, Verra F, Tetteh KK, Lowe B, Mwangi T, Bull PC, Thomas AW: **Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria**. *Infect Immun* 2008, **76**(5):2240-2248.
36. Greenhouse B, Ho B, Hubbard A, Njama-Meya D, Narum DL, Lanar DE, Dutta S, Rosenthal PJ, Dorsey G, John CC: **Antibodies to Plasmodium falciparum antigens predict a higher risk of malaria but protection from symptoms once parasitemic**. *J Infect Dis* 2011, **204**(1):19-26.
37. Agak GW, Bejon P, Fegan G, Gicheru N, Villard V, Kajava AV, Marsh K, Corradin G: **Longitudinal analyses of immune responses to Plasmodium falciparum derived peptides corresponding to novel blood stage antigens in coastal Kenya**. *Vaccine* 2008, **26**(16):1963-1971.
38. Richards JS, Arumugam TU, Reiling L, Healer J, Hodder AN, Fowkes FJ, Cross N, Langer C, Takeo S, Uboldi AD: **Identification and prioritization of merozoite antigens as targets of protective human immunity to Plasmodium falciparum malaria for vaccine and biomarker development**. *The Journal of Immunology* 2013, **191**(2):795-809.
39. Chan J-A, Fowkes FJ, Beeson JG: **Surface antigens of Plasmodium falciparum-infected erythrocytes as immune targets and malaria vaccine candidates**. *Cell Mol Life Sci* 2014, **71**(19):3633-3657.
40. Antinori S, Galimberti L, Milazzo L, Corbellino M: **Plasmodium knowlesi: the emerging zoonotic malaria parasite**. *Acta Trop* 2013, **125**(2):191-201.
41. Ansari HR, Templeton TJ, Subudhi AK, Ramaprasad A, Tang J, Lu F, Naeem R, Hashish Y, Oguike MC, Benavente ED: **Genome-scale comparison of expanded gene families in Plasmodium ovale wallikeri and Plasmodium ovale curtisi with Plasmodium malariae and with other Plasmodium species**. *Int J Parasitol* 2016, **46**(11):685-696.
42. Celli A: **A History of malaria in the Italian campagna from ancient times**. *London* 1933.
43. Stephens JWW: **Blackwater Fever. A Historical Survey and Summary of Observations made over a Century**; 1937.
44. Simpson J, Aarons L, Collins W, Jeffery G, White N: **Population dynamics of untreated Plasmodium falciparum malaria within the adult human host during the expansion phase of the infection**. *Parasitology* 2002, **124**(3):247-263.
45. Wockner LF, Hoffmann I, O'Rourke P, McCarthy JS, Marquart L: **Comparison of statistical models to estimate parasite growth rate in the induced blood stage malaria model**. *Malar J* 2017, **16**(1):352. doi:10.1186/s12936-017-1999-1.
46. Galinski MR, Meyer EV, Barnwell JW: **Plasmodium vivax: modern strategies to study a persistent parasite's life cycle**. In: *Adv Parasitol. Volume 81*, edn.: Elsevier; 2013: 1-26.
47. World Health Organization: **World Malaria Report 2013**. Geneva; 2013.

48. Svenson JE, MacLean JD, Gyorkos TW, Keystone J: **Imported malaria: clinical presentation and examination of symptomatic travelers**. *Arch Intern Med* 1995, **155**(8):861-868.
49. Tangpukdee N, Krudsood S, Kano S, Wilairatana P: **Falciparum malaria parasitemia index for predicting severe malaria**. *Int J Lab Hematol* 2012, **34**(3):320-327.
50. Phillips A, Bassett P, Szeki S, Newman S, Pasvol G: **Risk factors for severe disease in adults with falciparum malaria**. *Clin Infect Dis* 2009, **48**(7):871-878.
51. Murthy G, Sahay R, Srinivasan V, Upadhaya A, Shantaram V, Gayatri K: **Clinical profile of falciparum malaria in a tertiary care hospital**. *J Indian Med Assoc* 2000, **98**(4):160-162, 169.
52. Mollá S, de La Rubia J, Arriaga F, Fernández MJ, Carpio N, Marty ML: **Role of exchange transfusion in patients with severe Falciparum malaria: report of six cases**. *Haematologica* 2001, **86**(2):208-209.
53. Organization WH: **Severe and complicated malaria**. *Trans R Soc Trop Med Hyg* 1990, **84**:1-65.
54. Miller LH, Baruch DI, Marsh K, Doumbo OK: **The pathogenic basis of malaria**. *Nature* 2002, **415**(6872):673-679. doi:10.1038/415673a.
55. Pongponratn E, Riganti M, Punpoowong B, Aikawa M: **Microvascular sequestration of parasitized erythrocytes in human falciparum malaria: a pathological study**. *The American Journal of Tropical Medicine and Hygiene* 1991, **44**(2):168-175.
56. Turner GD, Morrison H, Jones M, Davis TM, Looareesuwan S, Buley ID, Gatter KC, Newbold CI, Pukritayakamee S, Nagachinta B *et al*: **An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration**. *Am J Pathol* 1994, **145**(5):1057-1069.
57. Milner Jr DA, Whitten RO, Kamiza S, Carr R, Liomba G, Dzamalala C, Seydel KB, Molyneux ME, Taylor TE: **The systemic pathology of cerebral malaria in African children**. *Frontiers in Cellular and Infection Microbiology* 2014, **4**:104. doi:10.3389/fcimb.2014.00104.
58. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM: **Vivax malaria: neglected and not benign**. *The American Journal of Tropical Medicine and Hygiene* 2007, **77**(6_Suppl):79-87.
59. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P *et al*: **Comparative genomics of the neglected human malaria parasite Plasmodium vivax**. *Nature* 2008, **455**(7214):757-763. doi:10.1038/nature07327.
60. Hart C: **Transplacental antibody transfer and neonatal immunity**. *Br J Hosp Med* 1997, **58**(7):317-319.
61. Dobbs KR, Dent AE: **Plasmodium malaria and antimalarial antibodies in the first year of life**. *Parasitology* 2016, **143**(2):129-138.
62. Harrington WE, Kakuru A, Jagannathan P: **Malaria in pregnancy shapes the development of foetal and infant immunity**. *Parasite Immunol* 2019, **41**(3):e12573. doi:10.1111/pim.12573.
63. Riley E, Wagner G, Akanmori B, Koram K: **Do maternally acquired antibodies protect infants from malaria infection?** *Parasite Immunol* 2001, **23**(2):51-59.

64. Ademolue TW, Awandare GA: **Evaluating antidisease immunity to malaria and implications for vaccine design**. *Immunology* 2018, **153**(4):423-434. doi:10.1111/imm.12877.
65. Weiss GE, Traore B, Kayentao K, Ongoiba A, Doumbo S, Doumtabe D, Kone Y, Dia S, Guindo A, Traore A *et al*: **The Plasmodium falciparum-specific human memory B cell compartment expands gradually with repeated malaria infections**. *PLoS Pathog* 2010, **6**(5):e1000912. doi:10.1371/journal.ppat.1000912.
66. Bousema T, Roeffen W, Meijerink H, Mwerinde H, Mwakalinga S, van Gemert G-J, van de Vegte-Bolmer M, Mosha F, Targett G, Riley EM: **The dynamics of naturally acquired immune responses to Plasmodium falciparum sexual stage antigens Pfs230 & Pfs48/45 in a low endemic area in Tanzania**. *PLoS One* 2010, **5**(11).
67. Wipasa J, Suphavilai C, Okell LC, Cook J, Corran PH, Thaikla K, Liewsaree W, Riley EM, Hafalla JCR: **Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax**. *PLoS Pathog* 2010, **6**(2).
68. Clark EH, Silva CJ, Weiss GE, Li S, Padilla C, Crompton PD, Hernandez JN, Branch OH: **Plasmodium falciparum malaria in the Peruvian Amazon, a region of low transmission, is associated with immunologic memory**. *Infect Immun* 2012, **80**(4):1583-1592.
69. 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**(6):905-915.
70. Amanna IJ, Slifka MK: **Mechanisms that determine plasma cell lifespan and the duration of humoral immunity**. *Immunol Rev* 2010, **236**(1):125-138.
71. 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**(1):82. doi:10.1186/1475-2875-6-82.
72. O'Meara WP, Mangeni JN, Steketee R, Greenwood B: **Changes in the burden of malaria in sub-Saharan Africa**. *The Lancet Infectious Diseases* 2010, **10**(8):545-555.
73. Kalayjian BC, Malhotra I, Mungai P, Holding P, King CL: **Marked decline in malaria prevalence among pregnant women and their offspring from 1996 to 2010 on the south Kenyan Coast**. *The American Journal of Tropical Medicine and Hygiene* 2013, **89**(6):1129-1134.
74. Snow RW, Kibuchi E, Karuri SW, Sang G, Gitonga CW, Mwandawiro C, Bejon P, Noor AM: **Changing malaria prevalence on the Kenyan coast since 1974: climate, drugs and vector control**. *PLoS One* 2015, **10**(6):e0128792. doi:10.1371/journal.pone.0128792.
75. Dondorp AM, Fairhurst RM, Slutsker L, MacArthur JR, Guerin PJ, Wellems TE, Ringwald P, Newman RD, Plowe CV: **The threat of artemisinin-resistant malaria**. *N Engl J Med* 2011, **365**(12):1073-1075.
76. Picot S: **Is Plasmodium vivax still a paradigm for uncomplicated malaria?** *Med Mal Infect* 2006, **36**(8):406-413.
77. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW: **The global distribution and population at risk of malaria: past, present, and future**. *The Lancet Infectious Diseases* 2004, **4**(6):327-336.

78. Baird JK: **Neglect of Plasmodium vivax malaria**. *Trends in Parasitology* 2007, **23**(11):533-539.
79. Beutler E, Duparc S: **Glucose-6-phosphate dehydrogenase deficiency and antimalarial drug development**. *The American Journal of Tropical Medicine and Hygiene* 2007, **77**(4):779-789.
80. Bockarie MJ, Dagoro H: **Are insecticide-treated bednets more protective against Plasmodium falciparum than Plasmodium vivax-infected mosquitoes?** *Malar J* 2006, **5**(1):15. doi:10.1186/1475-2875-5-15.
81. Zaw MT, Thant M, Hlaing TM, Aung NZ, Thu M, Phumchuea K, Phusri K, Saeseu T, Yorsaeng R, Nguitragool W: **Asymptomatic and sub-microscopic malaria infection in Kayah State, eastern Myanmar**. *Malar J* 2017, **16**(1):138. doi:10.1186/s12936-017-1789-9.
82. Harris I, Sharrock WW, Bain LM, Gray KA, Bobogare A, Boaz L, Lilley K, Krause D, Vallely A, Johnson ML *et al*: **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**(1):254. doi:10.1186/1475-2875-9-254.
83. Tsoka-Gwegweni JM, Okafor U: **Asymptomatic malaria in refugees living in a non-endemic South African city**. *PLoS One* 2014, **9**(9):e107693. doi:10.1371/journal.pone.0107693.
84. Baum E, Sattabongkot J, Sirichaisinthop J, Kiattibutr K, Jain A, Taghavian O, Lee M-C, Davies DH, Cui L, Felgner PL: **Common asymptomatic and submicroscopic malaria infections in Western Thailand revealed in longitudinal molecular and serological studies: a challenge to malaria elimination**. *Malar J* 2016, **15**(1):1-15.
85. Bell DR, Wilson DW, Martin LB: **False-positive results of a Plasmodium falciparum histidine-rich protein 2–detecting malaria rapid diagnostic test due to high sensitivity in a community with fluctuating low parasite density**. *The American Journal of Tropical Medicine and Hygiene* 2005, **73**(1):199-203.
86. Kilian A, Metzger W, Mutschelknauss E, Kabagambe G, Langi P, Korte R, Von Sonnenburg F: **Reliability of malaria microscopy in epidemiological studies: results of quality control**. *Trop Med Int Health* 2000, **5**(1):3-8.
87. Tadesse FG, Slater HC, Chali W, Teelen K, Lanke K, Belachew M, Menberu T, Shumie G, Shitaye G, Okell LC: **The relative contribution of symptomatic and asymptomatic Plasmodium vivax and Plasmodium falciparum infections to the infectious reservoir in a low-endemic setting in Ethiopia**. *Clin Infect Dis* 2018, **66**(12):1883-1891.
88. Slater HC, Ross A, Felger I, Hofmann NE, Robinson L, Cook J, Gonçalves BP, Björkman A, Ouedraogo AL, Morris U: **The temporal dynamics and infectiousness of subpatent Plasmodium falciparum infections in relation to parasite density**. *Nature communications* 2019, **10**(1):1-16.
89. Okell LC, Bousema T, Griffin JT, Ouédraogo AL, Ghani AC, Drakeley CJ: **Factors determining the occurrence of submicroscopic malaria infections and their relevance for control**. *Nature Communications* 2012, **3**:1237. doi:10.1038/ncomms2241.
90. Drakeley C, Cook J: **Potential contribution of sero-epidemiological analysis for monitoring malaria control and elimination: historical and current perspectives**. *Adv Parasitol* 2009, **69**:299-352.

91. Mfuh KO, Achonduh-Atijegbe OA, Bekindaka ON, Esemu LF, Mbakop CD, Gandhi K, Leke RGF, Taylor DW, Nerurkar VR: **A comparison of thick-film microscopy, rapid diagnostic test, and polymerase chain reaction for accurate diagnosis of Plasmodium falciparum malaria**. *Malar J* 2019, **18**(1):73. doi:10.1186/s12936-019-2711-4.
92. Baum E, Sattabongkot J, Sirichaisinthop J, Kiattibutr K, Davies DH, Jain A, Lo E, Lee M-C, Randall AZ, Molina DM: **Submicroscopic and asymptomatic Plasmodium falciparum and Plasmodium vivax infections are common in western Thailand-molecular and serological evidence**. *Malar J* 2015, **14**(1):95. doi:10.1186/s12936-015-0611-9.
93. Stresman GH, Kamanga A, Moono P, Hamapumbu H, Mharakurwa S, Kobayashi T, Moss WJ, Shiff C: **A method of active case detection to target reservoirs of asymptomatic malaria and gametocyte carriers in a rural area in Southern Province, Zambia**. *Malar J* 2010, **9**(1):265. doi:10.1186/1475-2875-9-265.
94. de AW Gunasekera WKT, Abeyasinghe RR, Premawansa S, Fernando SD: **Usefulness of polymerase chain reaction to supplement field microscopy in a pre-selected population with a high probability of malaria infections**. *The American journal of tropical medicine and hygiene* 2011, **85**(1):6-11.
95. Hofmann N, Mwingira F, Shekalaghe S, Robinson LJ, Mueller I, Felger I: **Ultra-sensitive detection of Plasmodium falciparum by amplification of multi-copy subtelomeric targets**. *PLoS Med* 2015, **12**(3):e1001788. doi:10.1371/journal.pmed.1001788.
96. Britton S, Cheng Q, McCarthy JS: **Novel molecular diagnostic tools for malaria elimination: a review of options from the point of view of high-throughput and applicability in resource limited settings**. *Malar J* 2016, **15**(1):88. doi:10.1186/s12936-016-1158-0.
97. Bashir IM, Otsyula N, Awinda G, Spring M, Schneider P, Waitumbi JN: **Comparison of PfHRP-2/pLDH ELISA, qPCR and microscopy for the detection of plasmodium events and prediction of sick visits during a malaria vaccine study**. *PLoS One* 2013, **8**(3):e56828. doi:10.1371/journal.pone.0056828.
98. Pakalapati D, Garg S, Middha S, Kochar A, Subudhi AK, Arunachalam BP, Kochar SK, Saxena V, Pareek R, Acharya J: **Comparative evaluation of microscopy, OptiMAL® and 18S rRNA gene based multiplex PCR for detection of Plasmodium falciparum & Plasmodium vivax from field isolates of Bikaner, India**. *Asian Pac J Trop Med* 2013, **6**(5):346-351.
99. Fançony C, Sebastião YV, Pires JE, Gamboa D, Nery SV: **Performance of microscopy and RDTs in the context of a malaria prevalence survey in Angola: a comparison using PCR as the gold standard**. *Malar J* 2013, **12**(1):284. doi:10.1186/1475-2875-12-284.
100. Shillcutt S, Morel C, Goodman C, Coleman P, Bell D, Whitty CJ, Mills A: **Cost-effectiveness of malaria diagnostic methods in sub-Saharan Africa in an era of combination therapy**. *Bull World Health Organ* 2008, **86**:101-110.
101. Uneke C: **Diagnosis of Plasmoduim falciparum malaria in pregnancy in sub-Saharan Africa: the challenges and public health implications**. *Parasitol Res* 2008, **102**(3):333-342.
102. 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**(9):e25137. doi:10.1371/journal.pone.0025137.

103. Drakeley C, Corran P, Coleman P, Tongren J, McDonald S, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya W: **Estimating medium-and long-term trends in malaria transmission by using serological markers of malaria exposure**. *Proceedings of the National Academy of Sciences* 2005, **102**(14):5108-5113.
104. Corran P, Coleman P, Riley E, Drakeley C: **Serology: a robust indicator of malaria transmission intensity?** *Trends in Parasitology* 2007, **23**(12):575-582.
105. Wu L, van den Hoogen LL, Slater H, Walker PG, Ghani AC, Drakeley CJ, Okell LC: **Comparison of diagnostics for the detection of asymptomatic Plasmodium falciparum infections to inform control and elimination strategies**. *Nature* 2015, **528**(7580):S86-93. doi:10.1038/nature16039.
106. Okell LC, Ghani AC, Lyons E, Drakeley CJ: **Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis**. *The Journal of Infectious Diseases* 2009, **200**(10):1509-1517.
107. Helb DA, Tetteh KK, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, Mayanja-Kizza H, Ssewanyana I, Kamya MR, Beeson JG: **Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities**. *Proceedings of the National Academy of Sciences* 2015, **112**(32):E4438-E4447.
108. Baum E, Badu K, Molina DM, Liang X, Felgner PL, Yan G: **Protein microarray analysis of antibody responses to Plasmodium falciparum in western Kenyan highland sites with differing transmission levels**. *PLoS One* 2013, **8**(12).
109. Ondigo BN, Hodges JS, Ireland KF, Magak NwG, Lanar DE, Dutta S, Narum DL, Park GS, Ofulla AV, John CC: **Estimation of recent and long-term malaria transmission in a population by antibody testing to multiple Plasmodium falciparum antigens**. *The Journal of Infectious Diseases* 2014, **210**(7):1123-1132.
110. Yeka A, Nankabirwa J, Mpimbaza A, Kigozi R, Arinaitwe E, Drakeley C, Greenhouse B, Kamya MR, Dorsey G, Staedke SG: **Factors associated with malaria parasitemia, anemia and serological responses in a spectrum of epidemiological settings in Uganda**. *PLoS One* 2015, **10**(3):e0118901. doi:10.1371/journal.pone.0118901.
111. Bretscher MT, Supargiyono S, Wijayanti MA, Nugraheni D, Widyastuti AN, Lobo NF, Hawley WA, Cook J, Drakeley CJ: **Measurement of Plasmodium falciparum transmission intensity using serological cohort data from Indonesian schoolchildren**. *Malar J* 2013, **12**(1):21. doi: 10.1186/1475-2875-12-21.
112. 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**(1):1-8.
113. Stewart L, Gosling R, Griffin J, Gesase S, Campo J, Hashim R, Masika P, Mosha J, Bousema T, Shekalaghe S: **Rapid assessment of malaria transmission using age-specific sero-conversion rates**. *PLoS One* 2009, **4**(6):e6083. doi:10.1371/journal.pone.0006083.
114. Surendra H, Wijayanti MA, Murhandarwati EH, Irnawati, Yuniarti T, Mardiati, Herdiana, Sumiwi ME, Hawley WA, Lobo NF *et al*: **Analysis of serological data to investigate heterogeneity of malaria transmission: a community-based cross-sectional study in an area conducting elimination in Indonesia**. *Malar J* 2019, **18**(1):227. doi:10.1186/s12936-019-2866-z.

115. 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**(3):392-399. doi:10.3201/eid1603.090732.
116. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F, Otieno S, Carneiro I, Cox J, Msuya E: **Identification of hot spots of malaria transmission for targeted malaria control**. *The Journal of Infectious Diseases* 2010, **201**(11):1764-1774.
117. Osier FH, Mackinnon MJ, Crosnier C, Fegan G, Kamuyu G, Wanaguru M, Ogada E, McDade B, Rayner JC, Wright GJ: **New antigens for a multicomponent blood-stage malaria vaccine**. *Sci Transl Med* 2014, **6**(247):247ra102-247ra102.
118. Kobayashi T, Chishimba S, Shields T, Hamapumbu H, Mharakurwa S, Thuma PE, Glass G, Moss WJ: **Temporal and spatial patterns of serologic responses to Plasmodium falciparum antigens in a region of declining malaria transmission in southern Zambia**. *Malar J* 2012, **11**(1):438. doi:10.1186/1475-2875-11-438.
119. Rijken MJ, McGready R, Boel ME, Poespoprodjo R, Singh N, Syafruddin D, Rogerson S, Nosten F: **Malaria in pregnancy in the Asia-Pacific region**. *The Lancet Infectious Diseases* 2012, **12**(1):75-88.
120. Rogerson SJ, Hviid L, Duffy PE, Leke RFG, Taylor DW: **Malaria in pregnancy: pathogenesis and immunity**. *The Lancet Infectious Diseases* 2007, **7**(2):105-117. doi:10.1016/s1473-3099(07)70022-1.
121. Ansell J, Hamilton K, Pinder M, Walraven G, Lindsay S: **Short-range attractiveness of pregnant women to Anopheles gambiae mosquitoes**. *Trans R Soc Trop Med Hyg* 2002, **96**(2):113-116.
122. Raghupathy R: **Th 1-type immunity is incompatible with successful pregnancy**. *Immunol Today* 1997, **18**(10):478-482.
123. Kourtis AP, Read JS, Jamieson DJ: **Pregnancy and infection**. *N Engl J Med* 2014, **370**(23):2211-2218.
124. Menendez C, Romagosa C, Ismail MR, Carrilho C, Saute F, Osman N, Machungo F, Bardaji A, Quinto L, Mayor A *et al*: **An autopsy study of maternal mortality in Mozambique: the contribution of infectious diseases**. *PLoS Med* 2008, **5**(2):e44. doi:10.1371/journal.pmed.0050044.
125. Van Eijk AM, Ayisi JG, Slutsker L, Ter Kuile FO, Rosen DH, Otieno JA, Shi YP, Kager PA, Steketee RW, Nahlen BL: **Effect of haematinic supplementation and malaria prevention on maternal anaemia and malaria in western Kenya**. *Trop Med Int Health* 2007, **12**(3):342-352.
126. Guyatt HL, Snow RW: **The epidemiology and burden of Plasmodium falciparum-related anemia among pregnant women in sub-Saharan Africa**. *The American journal of tropical medicine and hygiene* 2001, **64**(1_suppl):36-44.
127. Tarimo DS: **Appraisal on the prevalence of malaria and anaemia in pregnancy and factors influencing uptake of intermittent preventive therapy with sulfadoxine-pyrimethamine in Kibaha district, Tanzania**. *East African Journal of Public Health* 2007, **4**(2):80-83.
128. O'Donnell A, Raiko A, Clegg JB, Weatherall DJ, Allen SJ: **Southeast Asian ovalocytosis and pregnancy in a malaria-endemic region of Papua New Guinea**. *The American Journal of Tropical Medicine and Hygiene* 2007, **76**(4):631-633.
129. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Warikar N, Seal A, McGready R, Sugiarto P, Tjitra E, Anstey NM: **Adverse pregnancy outcomes in**

an area where multidrug-resistant Plasmodium vivax and Plasmodium falciparum infections are endemic. *Clin Infect Dis* 2008, **46**(9):1374-1381.
130. White NJ: **Manson's Tropical Diseases**. In. Edited by Cook GC, Zumla AI, 22 edn: Edinburgh: Saunders Ltd.; 2009.
131. Ayoola OO, Whatmore A, Balogun WO, Jarrett OO, Cruickshank JK, Clayton PE: **Maternal malaria status and metabolic profiles in pregnancy and in cord blood: relationships with birth size in Nigerian infants**. *Malar J* 2012, **11**(1):75. doi:10.1186/1475-2875-11-75.
132. Gontie GB, Wolde HF, Baraki AG: **Prevalence and associated factors of malaria among pregnant women in Sherkole district, Benishangul Gumuz regional state, West Ethiopia**. *BMC Infect Dis* 2020, **20**(1):573. doi:10.1186/s12879-020-05289-9.
133. Valea I, Tinto H, Drabo MK, Huybregts L, Sorgho H, Ouedraogo JB, Guiguemde RT, van Geertruyden JP, Kolsteren P, D'Alessandro U *et al*: **An analysis of timing and frequency of malaria infection during pregnancy in relation to the risk of low birth weight, anaemia and perinatal mortality in Burkina Faso**. *Malar J* 2012, **11**(1):71. doi:10.1186/1475-2875-11-71.
134. Cisse M, Sangare I, Lougue G, Bamba S, Bayane D, Guiguemde RT: **Prevalence and risk factors for Plasmodium falciparum malaria in pregnant women attending antenatal clinic in Bobo-Dioulasso (Burkina Faso)**. *BMC Infect Dis* 2014, **14**(1):631. doi:10.1186/s12879-014-0631-z.
135. Chandrasiri UP, Randall LM, Saad AA, Bashir AM, Rogerson SJ, Adam I: **Low antibody levels to pregnancy-specific malaria antigens and heightened cytokine responses associated with severe malaria in pregnancy**. *The Journal of infectious diseases* 2013, **209**(9):1408-1417.
136. Adam I, Adam GK, Mohmmed AA, Salih MM, Ibrahuim SA, Ryan CA: **Placental malaria and lack of prenatal care in an area of unstable malaria transmission in eastern Sudan**. *J Parasitol* 2009, **95**(3):751-752.
137. Desai M, ter Kuile FO, Nosten F, McGready R, Asamoa K, Brabin B, Newman RD: **Epidemiology and burden of malaria in pregnancy**. *The Lancet Infectious Diseases* 2007, **7**(2):93-104. doi:10.1016/s1473-3099(07)70021-x.
138. McGready R, Boel M, Rijken MJ, Ashley EA, Cho T, Moo O, Paw MK, Pimanpanarak M, Hkirijareon L, Carrara VI: **Effect of early detection and treatment on malaria related maternal mortality on the north-western border of Thailand 1986–2010**. *PLoS One* 2012, **7**(7):e40244. doi:10.1371/journal.pone.0040244.
139. Moore KA, Fowkes FJI, Wiladphaingern J, Wai NS, Paw MK, Pimanpanarak M, Carrara VI, Raksuansak J, Simpson JA, White NJ *et al*: **Mediation of the effect of malaria in pregnancy on stillbirth and neonatal death in an area of low transmission: observational data analysis**. *BMC Med* 2017, **15**(1):98. doi:10.1186/s12916-017-0863-z.
140. Moore KA, Simpson JA, Paw MK, Pimanpanarak M, Wiladphaingern J, Rijken MJ, Jittamala P, White NJ, Fowkes FJI, Nosten F *et al*: **Safety of artemisinins in first trimester of prospectively followed pregnancies: an observational study**. *The Lancet Infectious Diseases* 2016, **16**(5):576-583. doi:10.1016/s1473-3099(15)00547-2.
141. Fowkes FJ, Davidson E, Moore KA, McGready R, Simpson JA: **The invisible burden of malaria-attributable stillbirths**. *The Lancet* 2020, **395**(10220):268. doi:10.1016/S0140-6736(19)33011-9.

142. Weiss DJ, Lucas TC, Nguyen M, Nandi AK, Bisanzio D, Battle KE, Cameron E, Twohig KA, Pfeffer DA, Rozier JA: **Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study**. *The Lancet* 2019, **394**(10195):322-331.
143. Lawn JE, Blencowe H, Waiswa P, Amouzou A, Mathers C, Hogan D, Flenady V, Frøen JF, Qureshi ZU, Calderwood C: **Stillbirths: rates, risk factors, and acceleration towards 2030**. *The Lancet* 2016, **387**(10018):587-603.
144. Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B, Group ICDS: **Developmental potential in the first 5 years for children in developing countries**. *The Lancet* 2007, **369**(9555):60-70.
145. Kalilani L, Mofolo I, Chaponda M, Rogerson SJ, Meshnick SR: **The effect of timing and frequency of Plasmodium falciparum infection during pregnancy on the risk of low birth weight and maternal anemia**. *Trans R Soc Trop Med Hyg* 2010, **104**(6):416-422.
146. Huynh B-T, Fievet N, Gbaguidi G, Dechavanne S, Borgella S, Guézo-Mévo B, Massougbodji A, Ndam NT, Deloron P, Cot M: **Influence of the timing of malaria infection during pregnancy on birth weight and on maternal anemia in Benin**. *The American Journal of Tropical Medicine and Hygiene* 2011, **85**(2):214-220.
147. Jamieson DJ, Theiler RN, Rasmussen SA: **Emerging infections and pregnancy**. *Emerg Infect Dis* 2006, **12**(11):1638.
148. Robinson DP, Klein SL: **Pregnancy and pregnancy-associated hormones alter immune responses and disease pathogenesis**. *Horm Behav* 2012, **62**(3):263-271.
149. Kurup SP, Butler NS, Harty JT: **T cell-mediated immunity to malaria**. *Nature Reviews Immunology* 2019, **19**(7):457-471.
150. Zajac A, Harrington L: **Immune response to viruses: cell-mediated immunity**. 2014.
151. Olive AJ, Gondek DC, Starnbach MN: **CXCR3 and CCR5 are both required for T cell-mediated protection against C. trachomatis infection in the murine genital mucosa**. *Mucosal Immunol* 2011, **4**(2):208-216.
152. Osier F, Polley S, Mwangi T, Lowe B, Conway D, Marsh K: **Naturally acquired antibodies to polymorphic and conserved epitopes of Plasmodium falciparum merozoite surface protein 3**. *Parasite Immunol* 2007, **29**(8):387-394.
153. Nielsen MA, Staalsoe T, Kurtzhals JA, Goka BQ, Dodoo D, Alifrangis M, Theander TG, Akanmori BD, Hviid L: **Plasmodium falciparum variant surface antigen expression varies between isolates causing severe and nonsevere malaria and is modified by acquired immunity**. *The Journal of Immunology* 2002, **168**(7):3444-3450.
154. White NJ: **Declining malaria transmission and pregnancy outcomes in Southern Mozambique**. In.: Mass Medical Soc; 2015.
155. Mayor A, Bardají A, Macete E, Nhampossa T, Fonseca AM, González R, Maculuve S, Cisteró P, Rupérez M, Campo J: **Changing trends in P. falciparum burden, immunity, and disease in pregnancy**. *N Engl J Med* 2015, **373**(17):1607-1617.
156. Khunrae P, Dahlbäck M, Nielsen MA, Andersen G, Ditlev SB, Resende M, Pinto VV, Theander TG, Higgins MK, Salanti A: **Full-length recombinant Plasmodium falciparum VAR2CSA binds specifically to CSPG and induces**

potent parasite adhesion-blocking antibodies. *J Mol Biol* 2010, **397**(3):826-834.
157. Doritchamou J, Sossou-tchatcha S, Cottrell G, Moussiliou A, Hounton Houngbeme C, Massougbodji A, Deloron P, Ndam NT: **Dynamics in the cytoadherence phenotypes of Plasmodium falciparum infected erythrocytes isolated during pregnancy**. *PLoS One* 2014, **9**(6):e98577. doi:10.1371/journal.pone.0098577.
158. O'Neil-Dunne I, Achur RN, Agbor-Enoh ST, Valiyaveettil M, Naik RS, Ockenhouse CF, Zhou A, Megnekou R, Leke R, Taylor DW: **Gravidity-dependent production of antibodies that inhibit binding of Plasmodium falciparum-infected erythrocytes to placental chondroitin sulfate proteoglycan during pregnancy**. *Infect Immun* 2001, **69**(12):7487-7492.
159. Fried M, Nosten F, Brockman A, Brabin BJ, Duffy PE: **Maternal antibodies block malaria**. *Nature* 1998, **395**(6705):851-852. doi:10.1038/27570.
160. Targett GA, Greenwood BM: **Malaria vaccines and their potential role in the elimination of malaria**. *Malar J* 2008, **7**(1):S10. doi:10.1186/1475-2875-7-S1-S10.
161. Richards JS, Beeson JG: **The future for blood-stage vaccines against malaria**. *Immunol Cell Biol* 2009, **87**(5):377-390. doi:10.1038/icb.2009.27.
162. Ladda R, Aikawa M, Sprinz H: **Penetration of erythrocytes by merozoites of mammalian and avian malarial parasites**. *The Journal of Parasitology* 1969:633-644.
163. Bannister L, Butcher G, Dennis E, Mitchell G: **Structure and invasive behaviour of Plasmodium knowlesi merozoites in vitro**. *Parasitology* 1975, **71**(3):483-491.
164. Aikawa M, Miller LH, Johnson J, Rabbege J: **Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite**. *The Journal of Cell Biology* 1978, **77**(1):72-82.
165. Langreth SG, Nguyen-Dinh P, Trager W: **Plasmodium falciparum: merozoite invasion in vitro in the presence of chloroquine**. *Exp Parasitol* 1978, **46**(2):235-238.
166. Sanders PR, Gilson PR, Cantin GT, Greenbaum DC, Nebl T, Carucci DJ, McConville MJ, Schofield L, Hodder AN, Yates JR: **Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of Plasmodium falciparum**. *J Biol Chem* 2005.
167. Boyle MJ, Langer C, Chan J-A, Hodder AN, Coppel RL, Anders RF, Beeson JG: **Sequential processing of merozoite surface proteins during and after erythrocyte invasion by Plasmodium falciparum**. *Infect Immun* 2014, **82**(3):924-936.
168. Stanisic DI, Richards JS, McCallum FJ, Michon P, King CL, Schoepflin S, Gilson PR, Murphy VJ, Anders RF, Mueller I: **Immunoglobulin G subclass-specific responses against Plasmodium falciparum merozoite antigens are associated with control of parasitemia and protection from symptomatic illness**. *Infect Immun* 2009, **77**(3):1165-1174.
169. Scopel KK, Da Silva-Nunes M, Malafronte RS, Braga EM, Ferreira MU: **Variant-specific antibodies to merozoite surface protein 2 and clinical expression of Plasmodium falciparum malaria in rural Amazonians**. *The American Journal of Tropical Medicine and Hygiene* 2007, **76**(6):1084-1091.
170. Lawrence G, Cheng Q, Reed C, Taylor D, Stowers A, Cloonan N, Rzepczyk C, Smillie A, Anderson K, Pombo D: **Effect of vaccination with 3 recombinant**

asexual-stage malaria antigens on initial growth rates of Plasmodium falciparum in non-immune volunteers. *Vaccine* 2000, **18**(18):1925-1931.

171. Genton B, Betuela I, Felger I, Al-Yaman F, Anders RF, Saul A, Rare L, Baisor M, Lorry K, Brown GV: **A recombinant blood-stage malaria vaccine reduces Plasmodium falciparum density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea**. *The Journal of Infectious Diseases* 2002, **185**(6):820-827.
172. Genton B, Al-Yaman F, Betuela I, Anders RF, Saul A, Baea K, Mellombo M, Taraika J, Brown GV, Pye D *et al*: **Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against Plasmodium falciparum in Papua New Guinean children**. *Vaccine* 2003, **22**(1):30-41. doi:10.1016/s0264-410x(03)00536-x.
173. Sallenave-Sales S, Faria CP, Zalis MG, Daniel-Ribeiro CT, Ferreira-da-Cruz MdF: **Merozoite surface protein 2 allelic variation influences the specific antibody response during acute malaria in individuals from a Brazilian endemic area**. *Mem Inst Oswaldo Cruz* 2007, **102**(3):421-425.
174. Singh S, Soe S, Weisman S, Barnwell JW, Perignon JL, Druilhe P: **A conserved multi-gene family induces cross-reactive antibodies effective in defense against Plasmodium falciparum**. *PLoS One* 2009, **4**(4):e5410. doi:10.1371/journal.pone.0005410.
175. Lin CS, Uboldi AD, Epp C, Bujard H, Tsuboi T, Czabotar PE, Cowman AF: **Multiple Plasmodium falciparum merozoite surface protein 1 complexes mediate merozoite binding to human erythrocytes**. *J Biol Chem* 2016, **291**(14):7703-7715.
176. Sirima SB, Nébié I, Ouédraogo A, Tiono AB, Konaté AT, Gansané A, Dermé AI, Diarra A, Ouédraogo A, Soulama I: **Safety and immunogenicity of the Plasmodium falciparum merozoite surface protein-3 long synthethic peptide (MSP3-LSP) malaria vaccine in healthy, semi-immune adult males in Burkina Faso, West Africa**. *Vaccine* 2007, **25**(14):2723-2732.
177. Oeuvray C, Bouharoun-Tayoun H, Gras-Masse H, Bottius E, Kaidoh T, Aikawa M, Filgueira M-C, Tartar A, Druilhe P: **Merozoite surface protein-3: a malaria protein inducing antibodies that promote Plasmodium falciparum killing by cooperation with blood monocytes**. *Blood* 1994, **84**(5):1594-1602.
178. Tham W-H, Healer J, Cowman AF: **Erythrocyte and reticulocyte binding-like proteins of Plasmodium falciparum**. *Trends in Parasitology* 2012, **28**(1):23-30.
179. Orlandi PA, Klotz FW, Haynes JD: **A malaria invasion receptor, the 175-kilodalton erythrocyte binding antigen of Plasmodium falciparum recognizes the terminal Neu5Ac (alpha 2-3) Gal-sequences of glycophorin A**. *The Journal of Cell Biology* 1992, **116**(4):901-909.
180. Adams JH, Sim B, Dolan SA, Fang X, Kaslow DC, Miller LH: **A family of erythrocyte binding proteins of malaria parasites**. *Proceedings of the National Academy of Sciences* 1992, **89**(15):7085-7089.
181. Camus D, Hadley TJ: **A Plasmodium falciparum antigen that binds to host erythrocytes and merozoites**. *Science* 1985, **230**(4725):553-556.
182. Sim B, Chitnis C, Wasniowska K, Hadley T, Miller L: **Receptor and ligand domains for invasion of erythrocytes by Plasmodium falciparum**. *Science* 1994, **264**(5167):1941-1944.
183. Sim B, Orlandi PA, Haynes JD, Klotz FW, Carter JM, Camus D, Zegans ME, Chulay JD: **Primary structure of the 175K Plasmodium falciparum erythrocyte binding antigen and identification of a peptide which elicits**

antibodies that inhibit malaria merozoite invasion. *The Journal of Cell Biology* 1990, **111**(5):1877-1884.

184. Jiang L, Gaur D, Mu J, Zhou H, Long CA, Miller LH: **Evidence for erythrocyte-binding antigen 175 as a component of a ligand-blocking blood-stage malaria vaccine**. *Proceedings of the National Academy of Sciences* 2011, **108**(18):7553-7558. doi:10.1073/pnas.1104050108.
185. Ord RL, Rodriguez M, Yamasaki T, Takeo S, Tsuboi T, Lobo CA: **Targeting sialic acid dependent and independent pathways of invasion in Plasmodium falciparum**. *PLoS One* 2012, **7**(1):e30251. doi:10.1371/journal.pone.0030251.
186. El Sahly H, Patel S, Atmar R, Lanford T, Dube T, Thompson D, Sim B, Long C, Keitel W: **Safety and immunogenicity of a recombinant nonglycosylated erythrocyte binding antigen 175 Region II malaria vaccine in healthy adults living in an area where malaria is not endemic**. *Clin Vaccine Immunol* 2010, **17**(10):1552-1559.
187. Sim BK, Narum DL, Chattopadhyay R, Ahumada A, Haynes JD, Fuhrmann SR, Wingard JN, Liang H, Moch JK, Hoffman SL: **Delineation of stage specific expression of Plasmodium falciparum EBA-175 by biologically functional region II monoclonal antibodies**. *PLoS One* 2011, **6**(4):e18393. doi:10.1371/journal.pone.0018393.
188. Okenu DM, Riley EM, Bickle QD, Agomo PU, Barbosa A, Daugherty JR, Lanar DE, Conway DJ: **Analysis of human antibodies to erythrocyte binding antigen 175 of Plasmodium falciparum**. *Infect Immun* 2000, **68**(10):5559-5566.
189. Remarque EJ, Faber BW, Kocken CH, Thomas AW: **Apical membrane antigen 1: a malaria vaccine candidate in review**. *Trends in Parasitology* 2008, **24**(2):74-84.
190. Bai T, Becker M, Gupta A, Strike P, Murphy VJ, Anders RF, Batchelor AH: **Structure of AMA1 from Plasmodium falciparum reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket**. *Proceedings of the National Academy of Sciences* 2005, **102**(36):12736-12741.
191. Yap A, Azevedo MF, Gilson PR, Weiss GE, O'neill MT, Wilson DW, Crabb BS, Cowman AF: **Conditional expression of apical membrane antigen 1 in P lasmodium falciparum shows it is required for erythrocyte invasion by merozoites**. *Cell Microbiol* 2014, **16**(5):642-656.
192. Silvie O, Franetich J-F, Charrin S, Mueller MS, Siau A, Bodescot M, Rubinstein E, Hannoun L, Charoenvit Y, Kocken CH: **A role for apical membrane antigen 1 during invasion of hepatocytes by Plasmodium falciparum sporozoites**. *J Biol Chem* 2004, **279**(10):9490-9496.
193. Healer J, Crawford S, Ralph S, McFadden G, Cowman AF: **Independent translocation of two micronemal proteins in developing Plasmodium falciparum merozoites**. *Infect Immun* 2002, **70**(10):5751-5758.
194. Treeck M, Zacherl S, Herrmann S, Cabrera A, Kono M, Struck NS, Engelberg K, Haase S, Frischknecht F, Miura K *et al*: **Functional analysis of the leading malaria vaccine candidate AMA-1 reveals an essential role for the cytoplasmic domain in the invasion process**. *PLoS Pathog* 2009, **5**(3):e1000322. doi:10.1371/journal.ppat.1000322.
195. Mitchell G, Thomas A, Margos G, Dluzewski A, Bannister L: **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**(1):154-158.

196. Lamarque M, Besteiro S, Papoin J, Roques M, Vulliez-Le Normand B, Morlon-Guyot J, Dubremetz J-F, Fauquenoy S, Tomavo S, Faber BW: **The RON2-AMA1 interaction is a critical step in moving junction-dependent invasion by apicomplexan parasites**. *PLoS Pathog* 2011, **7**(2):e1001276. doi:10.1371/journal.ppat.1001276.
197. Richard D, MacRaild CA, Riglar DT, Chan J-A, Foley M, Baum J, Ralph SA, Norton RS, Cowman AF: **Interaction between Plasmodium falciparum apical membrane antigen 1 and the rhoptry neck protein complex defines a key step in the erythrocyte invasion process of malaria parasites**. *J Biol Chem* 2010, **285**(19):14815-14822. doi:10.1074/jbc.M109.080770.
198. Alexander DL, Mital J, Ward GE, Bradley P, Boothroyd JC: **Identification of the moving junction complex of Toxoplasma gondii: a collaboration between distinct secretory organelles**. *PLoS Pathog* 2005, **1**(2):e17. doi:10.1371/journal.ppat.0010017.
199. Deans J, Knight A, Jean W, Waters A, Cohen S, Mitchell G: **Vaccination trials in rhesus monkeys with a minor, invariant, Plasmodium knowlesi 66 kD merozoite antigen**. *Parasite Immunol* 1988, **10**(5):535-552.
200. Hamid MMA, Remarque EJ, van Duivenvoorde LM, van der Werff N, Walraven V, Faber BW, Kocken CH, Thomas AW: **Vaccination with Plasmodium knowlesi AMA1 formulated in the novel adjuvant co-vaccine HT™ protects against blood-stage challenge in rhesus macaques**. *PLoS One* 2011, **6**(5):e20547. doi:10.1371/journal.pone.0020547.
201. Gentil F, Bargieri DY, Leite JA, Françoso KS, Patricio MB, Espíndola NM, Vaz AJ, Palatnik-de-Sousa CB, Rodrigues MM, Costa FT: **A recombinant vaccine based on domain II of Plasmodium vivax Apical Membrane Antigen 1 induces high antibody titres in mice**. *Vaccine* 2010, **28**(38):6183-6190.
202. Galinski MR, Barnwell JW: **Plasmodium vivax: who cares?** *Malar J* 2008, **7 Suppl 1**(1):S9. doi:10.1186/1475-2875-7-S1-S9.
203. Triglia T, Healer J, Caruana SR, Hodder AN, Anders RF, Crabb BS, Cowman AF: **Apical membrane antigen 1 plays a central role in erythrocyte invasion by Plasmodium species**. *Mol Microbiol* 2000, **38**(4):706-718.
204. Stowers AW, Kennedy MC, Keegan BP, Saul A, Long CA, Miller LH: **Vaccination of monkeys with recombinant Plasmodium falciparum apical membrane antigen 1 confers protection against blood-stage malaria**. *Infect Immun* 2002, **70**(12):6961-6967.
205. Vulliez-Le Normand B, Tonkin ML, Lamarque MH, Langer S, Hoos S, Roques M, Saul FA, Faber BW, Bentley GA, Boulanger MJ: **Structural and functional insights into the malaria parasite moving junction complex**. *PLoS Pathog* 2012, **8**(6):e1002755. doi:10.1371/journal.ppat.1002755.
206. Collins CR, Withers-Martinez C, Bentley GA, Batchelor AH, Thomas AW, Blackman MJ: **Fine mapping of an epitope recognized by an invasion-inhibitory monoclonal antibody on the malaria vaccine candidate apical membrane antigen 1**. *J Biol Chem* 2007, **282**(10):7431-7441.
207. Roobsoong W, Roytrakul S, Sattabongkot J, Li J, Udomsangpetch R, Cui L: **Determination of the Plasmodium vivax schizont stage proteome**. *J Proteomics* 2011, **74**(9):1701-1710. doi:10.1016/j.jprot.2011.03.035.
208. Pachebat JA, Ling IT, Grainger M, Trucco C, Howell S, Fernandez-Reyes D, Gunaratne R, Holder AA: **The 22 kDa component of the protein complex on the surface of Plasmodium falciparum merozoites is derived from a larger**

precursor, merozoite surface protein 7. *Mol Biochem Parasitol* 2001, **117**(1):83-89.
209. Kraemer SM, Smith JD: **A family affair: var genes, PfEMP1 binding, and malaria disease**. *Curr Opin Microbiol* 2006, **9**(4):374-380.
210. Smith JD, Subramanian G, Gamain B, Baruch DI, Miller LH: **Classification of adhesive domains in the Plasmodium falciparum erythrocyte membrane protein 1 family**. *Mol Biochem Parasitol* 2000, **110**(2):293-310.
211. Salanti A, Staalsoe T, Lavstsen T, Jensen AT, Sowa MK, Arnot DE, Hviid L, Theander TG: **Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria**. *Mol Microbiol* 2003, **49**(1):179-191.
212. Tuikue Ndam NG, Salanti A, Bertin G, Dahlbäck M, Fievet N, Turner L, Gaye A, Theander T, Deloron P: **High level of var2csa transcription by Plasmodium falciparum isolated from the placenta**. *The Journal of Infectious Diseases* 2005, **192**(2):331-335.
213. Salanti A, Dahlbäck M, Turner L, Nielsen MA, Barfod L, Magistrado P, Jensen AT, Lavstsen T, Ofori MF, Marsh K: **Evidence for the involvement of VAR2CSA in pregnancy-associated malaria**. *J Exp Med* 2004, **200**(9):1197-1203.
214. Fried M, Duffy PE: **Adherence of Plasmodium falciparum to chondroitin sulfate A in the human placenta**. *Science* 1996, **272**(5267):1502-1504.
215. Beeson JG, Brown GV, Molyneux ME, Mhango C, Dzinjalamala F, Rogerson SJ: **Plasmodium falciparum isolates from infected pregnant women and children are associated with distinct adhesive and antigenic properties**. *The Journal of Infectious Diseases* 1999, **180**(2):464-472.
216. Brolin KJ, Persson KE, Wahlgren M, Rogerson SJ, Chen Q: **Differential recognition of P. falciparum VAR2CSA domains by naturally acquired antibodies in pregnant women from a malaria endemic area**. *PLoS One* 2010, **5**(2):e9230. doi:10.1371/journal.pone.0009230.
217. Travassos MA, Coulibaly D, Bailey JA, Niangaly A, Adams M, Nyunt MM, Ouattara A, Lyke KE, Laurens MB, Pablo J: **Differential recognition of terminal extracellular Plasmodium falciparum VAR2CSA domains by sera from multigravid, malaria-exposed Malian women**. *The American Journal of Tropical Medicine and Hygiene* 2015, **92**(6):1190-1194.
218. Tutterrow YL, Salanti A, Avril M, Smith JD, Pagano IS, Ako S, Fogako J, Leke RG, Taylor DW: **High avidity antibodies to full-length VAR2CSA correlate with absence of placental malaria**. *PLoS One* 2012, **7**(6):e40049. doi:10.1371/journal.pone.0040049.
219. Nosten F, McGready R, Simpson JA, Thwai KL, Balkan S, Cho T, Hkirijaroen L, Looareesuwan S, White NJ: **Effects of Plasmodium vivax malaria in pregnancy**. *Lancet* 1999, **354**(9178):546-549. doi:10.1016/s0140-6736(98)09247-2.
220. Nosten F, Ter Kuile F, Maelankirri L, Decludt B, White N: **Malaria during pregnancy in an area of unstable endemicity**. *Trans R Soc Trop Med Hyg* 1991, **85**(4):424-429.
221. Villegas L, McGready R, Htway M, Paw MK, Pimanpanarak M, Arunjerdja R, Viladpai-Nguen SJ, Greenwood B, White NJ, Nosten F: **Chloroquine prophylaxis against vivax malaria in pregnancy: a randomized, double-blind, placebo-controlled trial**. *Trop Med Int Health* 2007, **12**(2):209-218. doi:10.1111/j.1365-3156.2006.01778.x.

222. Dubowitz LM, Dubowitz V: **Gestational age of the newborn: a clinical manual**: Addison Wesley Publishing Company; 1977.
223. Tarpey T, Kinateder KK: **Clustering functional data**. *Journal of Classification* 2003, **20**(1):093-114.
224. Rossi F, Conan-Guez B, El Golli A: **Clustering functional data with the SOM algorithm**. In: *ESANN: 2004*; 2004: 305-312.
225. Abraham C, Cornillon P-A, Matzner-Løber E, Molinari N: **Unsupervised curve clustering using B-splines**. *Scandinavian Journal of Statistics* 2003, **30**(3):581-595.
226. James GM, Sugar CA: **Clustering for sparsely sampled functional data**. *Journal of the American Statistical Association* 2003, **98**(462):397-408.
227. Nagin DS, NAGIN D: **Group-based modeling of development**: Harvard University Press; 2005.
228. Genolini C, Falissard B: **KmL: a package to cluster longitudinal data**. *Comput Methods Programs Biomed* 2011, **104**(3):e112-121. doi:10.1016/j.cmpb.2011.05.008.
229. Genolini C, Falissard B: **KmL: k-means for longitudinal data**. *Computational Statistics* 2010, **25**(2):317-328.
230. Sylvestre M-P, Vatnik D, Vatnik MD: **Package 'traj'**. 2014.
231. Genolini C, Falissard B, Pingault J-B, Genolini MC: **Package 'kml3d'**; 2017.
232. Genolini C, Pingault J-B, Driss T, Côté S, Tremblay RE, Vitaro F, Arnaud C, Falissard B: **KmL3D: a non-parametric algorithm for clustering joint trajectories**. *Comput Methods Programs Biomed* 2013, **109**(1):104-111.
233. Hartigan JA, Wong MA: **Algorithm AS 136: A k-means clustering algorithm**. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 1979, **28**(1):100-108.
234. Celeux G, Govaert G: **A classification EM algorithm for clustering and two stochastic versions**. *Computational Statistics & Data Analysis* 1992, **14**(3):315-332. https://doi.org/10.1016/0167-9473(92)90042-E.
235. Genolini C, Alacoque X, Sentenac M, Arnaud C: **kml and kml3d: R packages to cluster longitudinal data**. *Journal of Statistical Software* 2015, **65**(4):1-34.
236. Bock HH: **Probabilistic models in cluster analysis**. *Computational Statistics & Data Analysis* 1996, **23**(1):5-28.
237. Fraley C, Raftery AE: **Model-based clustering, discriminant analysis, and density estimation**. *Journal of the American Statistical Association* 2002, **97**(458):611-631.
238. Fraley C, Raftery AE, Murphy TB, Scrucca L: **mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation**. In.: Technical report; 2012.
239. Andrews JL, Mcnicholas PD: **Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t-distributions**. *Statistics and Computing* 2012, **22**(5):1021-1029.
240. Andrews JL, Wickins JR, Boers NM, McNicholas PD: **teigen: an R package for model-based clustering and classification via the multivariate T distribution**. *J Stat Softw* 2018.
241. Bergé L, Bouveyron C, Girard S: **HDclassif: An R package for model-based clustering and discriminant analysis of high-dimensional data**. 2012.
242. McNicholas P, Jampani K, Subedi S: **longclust: Model-Based Clustering and Classification for Longitudinal Data**. *R package version* 2012, **1**.
243. Zhong W, Ma P: **Model Based Functional Data Analysis**. 2007.

244. Benaglia T, Chauveau D, Hunter D, Young D: **mixtools: An R package for analyzing finite mixture models**. 2009.
245. Proust-Lima C, Philipps V, Liquet B: **Estimation of extended mixed models using latent classes and latent processes: the R package lcmm**. *arXiv preprint arXiv:150300890* 2015.
246. Grun B, Leisch F: **FlexMix version 2: finite mixtures with concomitant variables and varying and constant parameters**. 2008.
247. Komárek A: **mixAK: Multivariate Normal Mixture Models and Mixtures of Generalized Linear Mixed Models Including Model Based Clustering**. *R package version* 2014, **3**.
248. Komárek A, Komárková L: **Capabilities of R package mixAK for clustering based on multivariate continuous and discrete longitudinal data**. *Journal of Statistical Software* 2014, **59**(12):1-38.
249. Komárek A, Komárková L: **Clustering for multivariate continuous and discrete longitudinal data**. *The Annals of Applied Statistics* 2013, **7**(1):177-200.
250. Morrell CH, Brant LJ, Sheng S, Metter EJ: **Using multivariate mixed-effects models to predict prostate cancer**. *Proceedings of the American Statistical Association, Biometrics Section Alexandria, American Statistical Association* 2005:332-337.
251. Richardson S, Green PJ: **On Bayesian analysis of mixtures with an unknown number of components (with discussion)**. *Journal of the Royal Statistical Society: series B (statistical methodology)* 1997, **59**(4):731-792.
252. Fong Y, Rue H, Wakefield J: **Bayesian inference for generalized linear mixed models**. *Biostatistics* 2010, **11**(3):397-412.
253. Komárek A, Komárková L: **Supplement to "Clustering for multivariate continuous and discrete longitudinal data."**. 2013. doi:10.1214/12-AOAS580SUPP.
254. Stephens M: **Dealing with label switching in mixture models**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2000, **62**(4):795-809.
255. Komárek A: **A new R package for Bayesian estimation of multivariate normal mixtures allowing for selection of the number of components and interval-censored data**. *Computational Statistics & Data Analysis* 2009, **53**(12):3932-3947. doi:10.1016/j.csda.2009.05.006.
256. Plummer M, Best N, Cowles K, Vines K: **Coda: Output Analysis and Diagnostics for MCMC**. 2009, R package version 0.13-4.
257. Geweke J: **Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments**. *Bayesian statistics* 1992, **4**:641-649.
258. Smith BJ: **boa: an R package for MCMC output convergence assessment and posterior inference**. *Journal of statistical software* 2007, **21**(11):1-37.
259. Gelman A, Rubin DB: **Inference from iterative simulation using multiple sequences**. *Statistical science* 1992, **7**(4):457-472.
260. Brooks SP, Gelman A: **General methods for monitoring convergence of iterative simulations**. *Journal of computational and graphical statistics* 1998, **7**(4):434-455.
261. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A: **Bayesian measures of model complexity and fit**. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002, **64**(4):583-639.
262. Celeux G, Forbes F, Robert CP, Titterington DM: **Deviance information criteria for missing data models**. *Bayesian Analysis* 2006, **1**(4):651-673.

263. Plummer M: **Penalized loss functions for Bayesian model comparison**. *Biostatistics* 2008, **9**(3):523-539.
264. Aitkin M, Liu CC, Chadwick T: **Bayesian model comparison and model averaging for small-area estimation**. *The Annals of Applied Statistics* 2009, **3**(1):199-221.
265. Aitkin M: **Statistical inference: an integrated Bayesian/likelihood approach**: Chapman and Hall/CRC; 2010.
266. Asparouhov T, Muthén B: **Variable-specific entropy contribution**. *Recuperado de http://www statmodel com/download/UnivariateEntropy pdf* 2014.
267. Elliott SR, Fowkes FJ, Richards JS, Reiling L, Drew DR, Beeson JG: **Research priorities for the development and implementation of serological tools for malaria surveillance**. *F1000prime reports* 2014, **6**:100-113.
268. Tran TM, Ongoiba A, Coursen J, Crosnier C, Diouf A, Huang C-Y, Li S, Doumbo S, Doumtabe D, Kone Y: **Naturally acquired antibodies specific for Plasmodium falciparum reticulocyte-binding protein homologue 5 inhibit parasite growth and predict protection from malaria**. *The Journal of infectious diseases* 2014, **209**(5):789-798.
269. Remarque EJ, Roestenberg M, Younis S, Walraven V, van der Werff N, Faber BW, Leroy O, Sauerwein R, Kocken CH, Thomas AW: **Humoral immune responses to a single allele PfAMA1 vaccine in healthy malaria-naive adults**. *PLoS One* 2012, **7**(6):e38898. doi:10.1371/journal.pone.0038898.
270. Thera MA, Doumbo OK, Coulibaly D, Laurens MB, Ouattara A, Kone AK, Guindo AB, Traore K, Traore I, Kouriba B: **A field trial to assess a blood-stage malaria vaccine**. *N Engl J Med* 2011, **365**(11):1004-1013.
271. Brunner NC, Chacky F, Mandike R, Mohamed A, Runge M, Thawer SG, Ross A, Vounatsou P, Lengeler C, Molteni F *et al*: **The potential of pregnant women as a sentinel population for malaria surveillance**. *Malar J* 2019, **18**(1):370. doi:10.1186/s12936-019-2999-0.
272. Kitojo C, Gutman JR, Chacky F, Kigadye E, Mkude S, Mandike R, Mohamed A, Reaves EJ, Walker P, Ishengoma DS: **Estimating malaria burden among pregnant women using data from antenatal care centres in Tanzania: a population-based study**. *The Lancet Global Health* 2019, **7**(12):e1695-e1705. doi:10.1016/s2214-109x(19)30405-x.
273. Gbédandé K, Cottrell G, Vianou B, Ibitokou S, Fernando A, Troye-Blomberg M, Salanti A, Moutairou K, Massougbodji A, Ndam NT: **Infections with Plasmodium falciparum during pregnancy affect VAR2CSA DBL-5 domain-specific T cell cytokine responses**. *Malar J* 2016, **15**(1):1-11.
274. Fried M, Duffy PE: **Designing a VAR2CSA-based vaccine to prevent placental malaria**. *Vaccine* 2015, **33**(52):7483-7488.
275. Fonseca AM, Gonzalez R, Bardaji A, Jairoce C, Ruperez M, Jimenez A, Quinto L, Cistero P, Vala A, Sacoor C *et al*: **VAR2CSA Serology to Detect Plasmodium falciparum Transmission Patterns in Pregnancy**. *Emerg Infect Dis* 2019, **25**(10):1851-1860. doi:10.3201/eid2510.181177.
276. Cutts JC, Agius PA, Zaw L, Powell R, Moore K, Draper B, Simpson JA, Fowkes FJI: **Pregnancy-specific malarial immunity and risk of malaria in pregnancy and adverse birth outcomes: a systematic review**. *BMC Med* 2020, **18**(1):14. doi:10.1186/s12916-019-1467-6.
277. Lloyd YM, Fang R, Bobbili N, Vanda K, Ngati E, Sanchez-Quintero MJ, Salanti A, Chen JJ, Leke RGF, Taylor DW: **Association of Antibodies to VAR2CSA**

and Merozoite Antigens with Pregnancy Outcomes in Women Living in Yaounde, Cameroon. *Infect Immun* 2018, **86**(9). doi:10.1128/IAI.00166-18.

278. 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**(12).
279. Fried M, Kurtis JD, Swihart B, Morrison R, Pond-Tor S, Barry A, Sidibe Y, Keita S, Mahamar A, Andemel N: **Antibody levels to recombinant VAR2CSA domains vary with Plasmodium falciparum parasitaemia, gestational age, and gravidity, but do not predict pregnancy outcomes**. *Malar J* 2018, **17**(1):106. doi:10.1186/s12936-018-2258-9.
280. Ricke CH, Staalsoe T, Koram K, Akanmori BD, Riley EM, Theander TG, Hviid L: **Plasma antibodies from malaria-exposed pregnant women recognize variant surface antigens on Plasmodium falciparum-infected erythrocytes in a parity-dependent manner and block parasite adhesion to chondroitin sulfate A**. *The Journal of Immunology* 2000, **165**(6):3309-3316.
281. Lopaticki S, Maier AG, Thompson J, Wilson DW, Tham W-H, Triglia T, Gout A, Speed TP, Beeson JG, Healer J: **Reticulocyte and erythrocyte binding-like proteins function cooperatively in invasion of human erythrocytes by malaria parasites**. *Infect Immun* 2011, **79**(3):1107-1117.
282. Dobaño C, Ubillos I, Jairoce C, Gyan B, Vidal M, Jiménez A, Santano R, Dosoo D, Nhabomba AJ, Ayestaran A: **RTS, S/AS01E immunization increases antibody responses to vaccine-unrelated Plasmodium falciparum antigens associated with protection against clinical malaria in African children: a case-control study**. *BMC Med* 2019, **17**(1):157. doi:10.1186/s12916-019-1378-6.
283. Barry AE, Arnott A: **Strategies for designing and monitoring malaria vaccines targeting diverse antigens**. *Front Immunol* 2014, **5**:359. doi:10.3389/fimmu.2014.00359.
284. Longley RJ, White MT, Takashima E, Brewster J, Morita M, Harbers M, Robinson LJ, Matsuura F, Liu S-JZ, Li-Wai-Suen CS: **Development and validation of serological markers for detecting recent exposure to Plasmodium vivax infection**. *bioRxiv* 2018:481168. doi:10.1101/481168.
285. Rosas-Aguirre A, Patra KP, Calderón M, Torres K, Gamboa D, Arocutipa E, Málaga E, Garro K, Fernández C, Trompeter G: **Anti–MSP-10 IgG indicates recent exposure to Plasmodium vivax infection in the Peruvian Amazon**. *JCI insight* 2020, **5**(1). doi:10.1172/jci.insight.130769.
286. 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**(9):1527-1533.
287. Folegatti PM, Siqueira AM, Monteiro WM, Lacerda MV, Drakeley CJ, Braga EM: **A systematic review on malaria sero-epidemiology studies in the Brazilian Amazon: insights into immunological markers for exposure and protection**. *Malar J* 2017, **16**(1):107. doi:10.1186/s12936-017-1762-7.
288. Kale S, Yadav CP, Rao PN, Shalini S, Eapen A, Srivasatava HC, Sharma SK, Pande V, Carlton JM, Singh OP *et al*: **Antibody responses within two leading Plasmodium vivax vaccine candidate antigens in three geographically diverse**

malaria-endemic regions of India. *Malar J* 2019, **18**(1):425. doi:10.1186/s12936-019-3066-6.

289. Persson KE, Fowkes FJ, McCallum FJ, Gicheru N, Reiling L, Richards JS, Wilson DW, Lopaticki S, Cowman AF, Marsh K: **Erythrocyte-binding antigens of Plasmodium falciparum are targets of human inhibitory antibodies and function to evade naturally acquired immunity**. *The Journal of Immunology* 2013, **191**(2):785-794.
290. Richards JS, Stanisic DI, Fowkes FJ, Tavul L, Dabod E, Thompson JK, Kumar S, Chitnis CE, Narum DL, Michon P: **Association between naturally acquired antibodies to erythrocyte-binding antigens of Plasmodium falciparum and protection from malaria and high-density parasitemia**. *Clin Infect Dis* 2010, **51**(8):e50-e60.
291. Boyle M, Chan J, Handayuni I, Reiling L, Feng G, Hilton A, Kurtovic L, Oyong D, Piera K, Barber B: **IgM in human immunity to Plasmodium falciparum malaria**. *Science advances* 2019, **5**(9):eaax4489. doi:10.1126/sciadv.aax4489.
292. Dodoo D, Aikins A, Kusi KA, Lamptey H, Remarque E, Milligan P, Bosomprah S, Chilengi R, Osei YD, Akanmori BD: **Cohort study of the association of antibody levels to AMA1, MSP1 19, MSP3 and GLURP with protection from clinical malaria in Ghanaian children**. *Malar J* 2008, **7**(1):1-11.
293. Mayor A, Dobaño C, Nhabomba A, Guinovart C, Jiménez A, Manaca MN, Aguilar R, Barbosa A, Rodríguez MH, Cisteró P: **IgM and IgG against Plasmodium falciparum lysate as surrogates of malaria exposure and protection during pregnancy**. *Malar J* 2018, **17**(1):182. doi:10.1186/s12936-018-2331-4.

Appendix

Appendix A

Selection of the optimal number of clusters

The optimal number of clusters was determined based on the PED value and the cumulative distribution of the observed deviance values over one to five clusters.

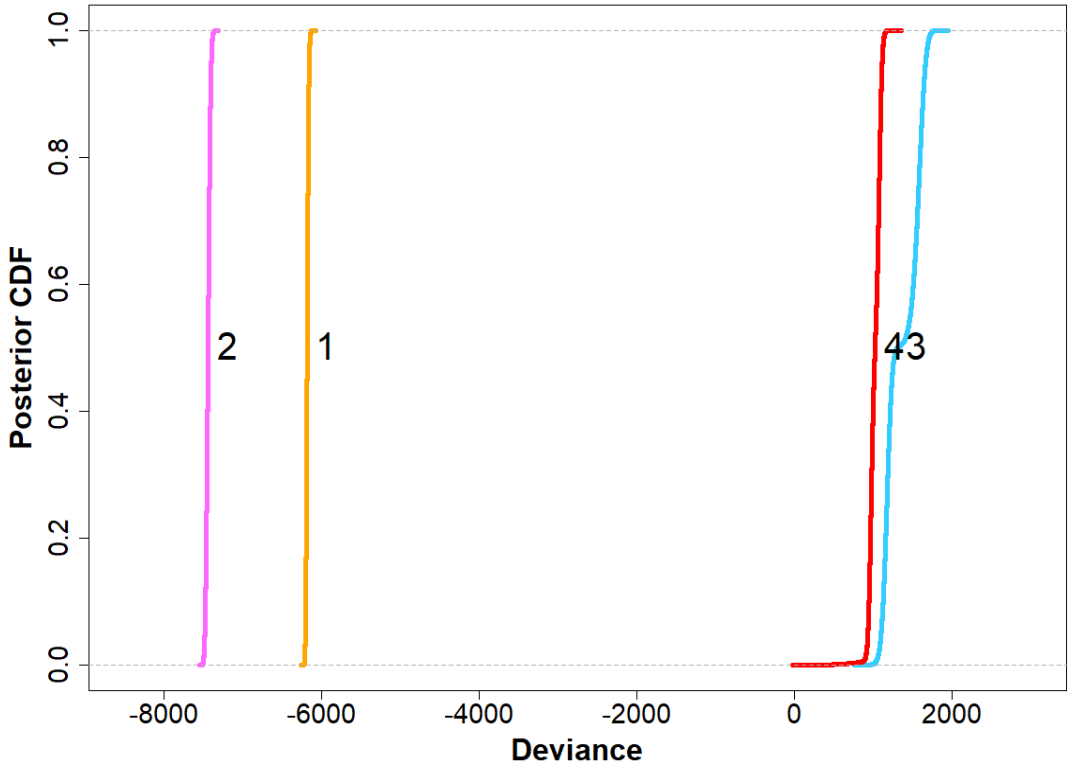
Penalised Expected Deviance (PED) estimated for the models with the number of mixture distributions (i.e. clusters) ranging from one to four

K	Estimate of average deviance	P_{opt}	Penalised Expected Deviance (PED)
1	-6183.12	220.8	-5962.31
2	-7442.12	473.06	-6969.06
3	1374.62	2005.26	3379.87
4	1026.26	2570.99	3597.25

PED – sum of average deviance and P_{opt}

K =2 provided the minimum PED value, i.e., a multivariate linear mixed model with a mixture of two random effects distributions best explains variation in the antibody profiles of the pregnant women. Hence, it was determined to classify the pregnant women into two groups.

Posterior cumulative distribution functions of the observed data deviances for models with K = 1, 2, 3 and 4

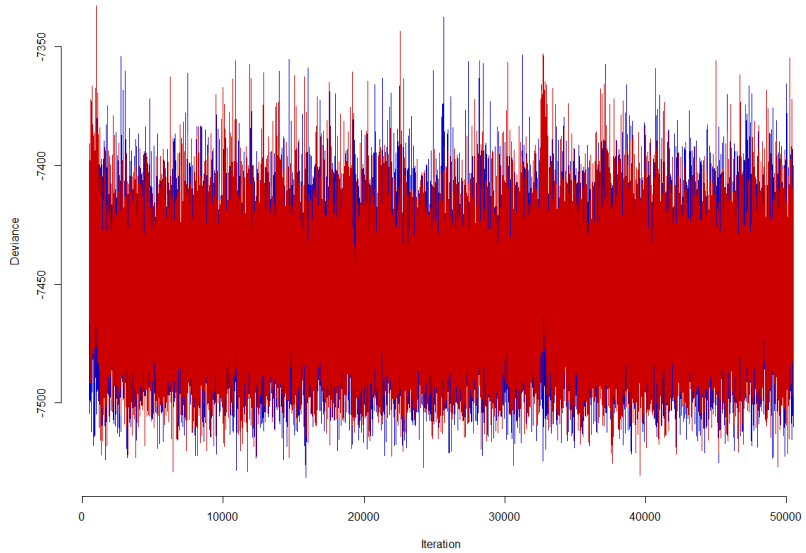


Legend: The posterior cumulative distribution of the deviance for the models with K = 1, 2, 3 and 4 components are represented by orange, pink, blue and red line segments respectively.

The posterior cumulative distribution of the deviance for the model with K = 2 components remained low throughout the 10,000 MCMC samples (remaining at the left corner of the plot). Hence in agreement with the PED value, it is optimal to classify the pregnant women into two clusters.

Convergence diagnostics of multivariate linear mixed effects modelling of antibody levels

- **Traceplot of the posterior deviances under two chains**



Legend: The traceplot of the posterior deviances obtained under 50,000 iterations where the deviances of chain 1 is represented in red while chain 2 in blue.

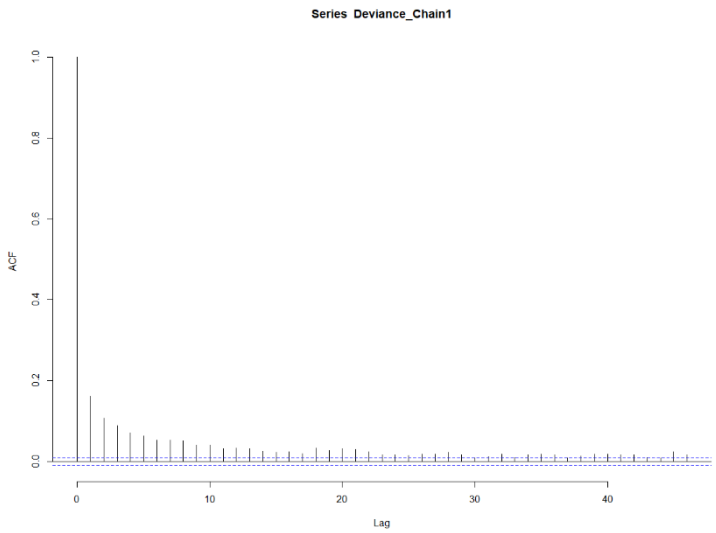
The traceplot of the posterior deviances displays the iterations vs. sampled values for the deviance in the two chains. It is clear that the sequence of sample deviances of both chains is overlapping and mixing well. Hence, it indicates that the estimates generated under both chains are similar, hence the classification outcomes could be made based on the estimates of the estimates of one chain.

- **Autocorrelations and autocorrelation function (ACF) plot of the deviances**

Table: Autocorrelations of the sample deviances in chain 1

Lag	Autocorrelation
0	1.000
1	0.161
5	0.063
10	0.041
50	0.017

It is clear that the autocorrelation of the deviances in chain 1 rapidly declines with the increment of the lag indicating good mixing and high convergence.



Legend: Autocorrelation function (ACF) plot of the series of posterior deviances of chain 1 for ear lag term.

ACF plot further showcases the convergence of the autocorrelations of the sample deviance in chain 1 to the stationarity.

- **Geweke's convergence diagnostic**

Table of Geweke's convergence diagnostics for the given parameters computed under chain 1

Parameter	**$\mathcal{Z}$ score**	
Deviance	0.361	
	Cluster 1	**Cluster 2**
Mean estimates of the random effects		
*Pf*AMA1 intecept	-1.836	-1.546
*Pf*AMA1 slope	-0.04	-0.127
*Pf*EBA175 intecept	-1.441	-1.158
*Pf*EBA175 slope	-0.067	-0.415
*Pf*MSP2 intecept	-1.265	-0.975
*Pf*MSP2 slope	0.425	-0.801
*Pf*MSP3 intecept	-1.077	-0.21
*Pf*MSP3slope	0.151	-1.447
*Pv*AMA1 intecept	-0.493	-0.037
*Pv*AMA1 slope	-0.864	0.097
*Pf*VAR2CSA intecept	-1.113	-0.68
*Pf*VAR2CSA slope	-0.189	-0.671

For precision, the Geweke estimates for the parameters deviance and the means of random effects were computed. According to the above table it is clear that all $\mathcal{Z}$ scores are lying close to 0 indicating the chain has converged and suggests that the burn-in period of 500 and thinning of 50 are sufficient to converge the chain.

- **Gelman and Rubin's convergence diagnostic**

Table of potential scale reduction factors considering both chains

Parameter	**Point estimate (50%)**		**Upper limit (97.5%)**	
Deviance	1		1	
	Cluster 1		**Cluster 2**	
	Point estimate (50%)	**Upper limit (97.5%)**	**Point estimate (50%)**	**Upper limit (97.5%)**
Mean estimates of the random effects				
*Pf*AMA1 intecept	1.01	1.03	1	1.02
*Pf*AMA1 slope	1	1	1	1
*Pf*EBA175 intecept	1.01	1.03	1	1.01
*Pf*EBA175 slope	1	1	1	1
*Pf*MSP2 intecept	1	1.01	1	1.01
*Pf*MSP2 slope	1	1	1	1
*Pf*MSP3 intecept	1	1.01	1	1
*Pf*MSP3slope	1	1	1	1
*Pv*AMA1 intecept	1	1	1	1
*Pv*AMA1 slope	1	1	1	1
*Pf*VAR2CSA intecept	1	1.01	1	1
*Pf*VAR2CSA slope	1	1	1	1
Multivariate potential scale reduction factor: 1.01				

According to the table, all potential scale reduction factors of deviance and means of random effects under both clusters individually as well multivariate manner are closer to 1 and below 1.05. Hence indicates that parallel Markov chains are overlapping and are converged.

Univariate classification (considering *Pf*MSP3 and *Pv*AMA1) of pregnant women

	Considering *Pf*MSP3 only			**Considering *Pv*AMA1 only**			Total
	Cluster 1	Cluster 2	Unclassified	Cluster 1	Cluster 2	Unclassified	
Case	97	33	5	100	28	7	135
Control	103	4	8	109	5	1	115
Total	200	37	13	209	33	8	250

Both antibodies are capable of identifying the pregnant women with no infections (control) and classify them into Cluster 1 (*Pf*MSP3 and *Pv*AMA1 alone recognises 103 (90%) and 109 (95%) of the controls respectively). However, those with infection (case) could be in either cluster where *Pf*MSP3 clearly captures 33 (24%) while *Pv*AMA1 alone identifies 28 (21%) and classify into Cluster 2. Hence, need to incorporate some of the other antibodies to drive the cases to be classified into Cluster 2.

Classification of pregnant women based on the joint antibody responses of *Pf*AMA1 and *Pf*VAR2CSA

Exposure	**Cluster 1**	**Cluster 2**	**Unclassified**	**Total**
Case	41	85	09	135
Control	73	26	16	115
Total	114	111	25	250

It is evident that the classification of cases into Cluster 2 (85 of 135; 63%) is considerably high when only considered *Pf*AMA1 and *Pf*VAR2CSA compared to that of considering all six antibody responses, collectively (51 of 135; 37.8%). However, the classification of controls into Cluster 1 (73 of 115; 63%) was quite low compared to the joint analysis of six antibodies (107 of 115; 93%). In order to improve classifying controls into Cluster 1, the responses of *Pf*MSP3 antibody were combined with *Pf*AMA1 and *Pf*VAR2CSA, but as in six antibody study, controls were then clearly classified into Cluster 1 but hindered classifying cases into Cluster 2.

Classification of cases based on anti-*Pv*AMA1 antibody responses by the type of infection

Type of infection(s) recorded during pregnancy	**Cluster 1**	**Cluster 2**	**Unclassified**	**Total**
P. falciparum only	34	17	02	53
Both *P. falciparum* and *P. vivax*	25	04	02	31
P. falciparum, P. vivax and mixed infection	02	01	01	04
P. falciparum and *P. malariae* infection	00	01	00	01
P. falciparum, P. vivax and *P. malariae*	01	00	00	01
P. vivax and *P. malariae*	00	01	00	01
P. vivax only	38	04	02	44
Total	100	28	07	135

Classification of pregnant women based on anti-*Pv*AMA1 antibody responses highlights that almost all the cases were classified into Cluster 1 with majority of the controls, irrespective of the infection type.

Appendix B

R code for the model-based clustering using the `mixAK` package

Below is the R code written to classify the group of pregnant women based on the joint behaviour of the six antibody responses using the MMLMM implemented in the **`mixAK`** package in Chapter 5.

#Following the instructions from Komarek 2014 paper.

```
library("mixAK")

mod <- GLMM_MCMC (y = df [, c (“pfama1”, “pfeba175”, “pfmsp2”, “pfmsp3”,
                                “pvama1”, “pfvar2csa”)],
                  dist = c (“gaussian”, “gaussian”, “gaussian”, “gaussian”,
                            “gaussian”, “gaussian”),
                  id = df [, “cpcode”],
                  x = list (pfama1=df[c("age","primigravidae","CQ","malhistory")],
                            pfeba175=df [, c("age","primigravidae","CQ","malhistory")],
                            pfmsp2=df [, c("age","primigravidae","CQ","malhistory")],
                            pfmsp3=df [, c("age","primigravidae","CQ","malhistory")],
                            pvama1=df [, c("age","primigravidae","CQ","malhistory")],
                            pfvar2csa=df [, c("age","primigravidae","CQ","malhistory")]),
                  z = list (pfama1=df [, "gestation"], pfeba175=df [, "gestation"],
                            pfmsp2=df [, "gestation"], pfmsp3=df [, "gestation"],
                            pvama1=df [, "gestation"], pfvar2csa=df [, "gestation"]),
                  random. intercept = c (pfama1=TRUE, pfeba175=TRUE, pfmsp2=TRUE,
                                         pfmsp3=TRUE, pvama1=TRUE, pfvar2csa=TRUE),
                  prior. b = list (Kmax = 2),
                  nMCMC = c (burn = 500, keep = 50000, thin = 50, info = 500), store =
                            c (b = TRUE), parallel = TRUE)


mod <- NMixRelabel (mod, type = "stephens", keep. comp. prob = TRUE)
```

Convergence diagnostics considering chain 1

```
library(coda)

DevChains <- mcmc.list(mcmc(mod[[1]]$Deviance), mcmc(mod[[2]]$Deviance))

autocorr(DevChains)

acf(mod[[1]]$Deviance)

estimates<- as.matrix(cbind(mod[[1]]$Deviance, mod[[1]]$mu_b))

estimates <-as.mcmc(estimates)
```

```
geweke.diag(estimates, frac1=0.1, frac2=0.5)

mod4<-mcmc.list(as.mcmc(mod[[1]]$Deviance),as.mcmc(mod[[2]]$Deviance))

mod5<-mcmc.list(as.mcmc(mod[[1]]$mu_b),as.mcmc(mod[[2]]$mu_b))

gelman.diag(mod4)

gelman.diag(mod5)
```

#Estimates generated under chain 1 was considered

#Step 1 of classification
##Classification based on posterior mean

```
groupMean <- apply (mod [[1]] $poster. comp. prob, 1, which. max)
table(groupMean)
```

##Classification based on posterior median

```
groupMed <- apply (mod [[1]] $quant. comp. prob [["50%"]], 1, which. max)
table(groupMed)
library("coda")
```

#Step 2 of classification

```
pHPD <- HPDinterval (mcmc (mod [[1]] $comp. prob))
pHPDlower <- matrix (pHPD [, "lower"], ncol = 2, byrow = TRUE)
pHPDupper <- matrix (pHPD [, "upper"], ncol = 2, byrow = TRUE)
rownames (pHPDlower) <- rownames (pHPDupper) <- unique (df $cpcode)
groupHPD <- groupMed
groupHPD [groupHPD == 1 & pHPDlower [, 1] <= 0.5] <- NA
groupHPD [groupHPD == 2 & pHPDlower [, 2] <= 0.5] <- NA
table (groupHPD, useNA = "ifany")
```

#Store the classification outcomes

```
TAB <- table (df$cpcode)
df$groupMed <- factor (rep (groupMed, TAB))
df$groupHPD <- factor (rep (groupHPD, TAB))
```

Computation of individual component probabilities under each antibody under each cluster manually

---

 # Store cluster-specific model parameters

##Store mixture weights

```
weights< -as.data.Frame (mod [[1]]$w_b)
```

##Store mixture means of the random effects

```
mu_b<- as.data.Frame (mod [[1]]$mu_b)
```

##Store lower triangles of mixture covariance matrices
sigma_b<- as.data.frame(mod[[1]]$Sigma_b)

#Scaling of random effects $(\boldsymbol{b_i^*})$

##Store random errors $(\boldsymbol{b_i})$
rand_b<- as.data.Frame (mod [[1]]$b)

##Store components required to compute shifted-scaled random effects
$(\boldsymbol{b_i^* = S^{-1}(b_i - s)})$

#Extract $\boldsymbol{12 \times 1}$ **Shift vector (s)**
shift<-mod [[1]]$scale.b$shift

#Extract $\boldsymbol{12 \times 12}$ **diagonal scale matrix (S)**
scale<-diag(mod [[1]] $scale.b$scale)
# $\boldsymbol{S^{-1}}$
in_scale<-solve(scale)

 # Compute shifted-scaled random effects $(\boldsymbol{b_i^*})$ **for each pregnant woman at each iteration**

```
scaled_b<-data.frame()
for (i in 1:50000){
        for (j in 1:250){
                k<-12*j-11
                b<-unname(as.numeric(rand_b[i,k:(k+11)]))
                x<-(b-shift)
                d<-as.vector(in_scale %*%x)
                scaled_b[i,k:(k+11)]<-d
        }
}
```

#remove scaled_b values of the 9 pw who were not classified based on HPD interval

```
library(dplyr)
```

Identify the ultimate classification of each pregnant woman (Based on HPD interval)

```
pw<- df%>%
      group_by(cpcode)%>%
      summarise(group=head(group,1), groupHPD=head(groupHPD,1))
```

##Identify the position of the pregnant women who were not classified into a group with certainty

```
unclass_id<-as.vector(which(is.na(pw$groupHPD)))
```

##b_i of each pregnant woman is a 12 × 1 vector. Hence, identify the starting and ending columns corresponding to the 7 women excluded.

```
for (i in 1:9) {
        j<-unclass_id[i]
        k1<-12*j-11
        k2<-k1+11
       cat (paste (k1, ":", k2), \ 'n')
}
```

Exclude the random effects measures correspond to the 7 women excluded from the classification.

```
new_scaled_b<-scaled_b[,c(553:564,793:804,877:888,913:924,1093:1104,1897:1908,
                          1993:2004,2545:2556,2869:2880)]
```

 # Compute cluster wise individual densities under each antibody

##Create a function to generate the mixture matrix using the lower triangles

###Only 241 pw were classified precisely. In density data frame, each woman has density measures for six antibodies. Since, 1 column represents a density measure for one antibody, 6 columns represent the densities of 1 woman at each iteration (in each row). Hence, each cluster specific density table consists of 1446 (241*6) columns and 50,000 rows.

```
library(mvtnorm)

new_cluster1<-data.frame()
new_cluster2<-data.frame()

for (i in 1:50000){

      mean1<-unname(as.numeric(mu_b[i,1:12]))
        d1<-unname(as.numeric(sigma_b[i,1:78]))
       S1 <- diag(12)
       S1[lower.tri(S1, diag=TRUE)] <- d1
       S1[upper.tri(S1)] <- t(S1)[upper.tri(S1)]

      mean2<-unname(as.numeric(mu_b[i,13:24]))
        d2<-unname(as.numeric(sigma_b[i,79:156]))
      S2 <- diag(12)
        S2[lower.tri(S2, diag=TRUE)] <- d2
        S2[upper.tri(S2)] <- t(S2)[upper.tri(S2)]


      for (j in 1:241){
            j<-j
            l<-j*6-5
```

```
                k1<-12*j-11
                k2<-k1+2
                k3<-k2+2
                k4<-k3+2
                k5<-k4+2
                k6<-k5+2

                x1<-unname(as.numeric(new_scaled_b[i,k1:(k1+1)]))
                x2<-unname(as.numeric(new_scaled_b[i,k2:(k2+1)]))
                x3<-unname(as.numeric(new_scaled_b[i,k3:(k3+1)]))
                x4<-unname(as.numeric(new_scaled_b[i,k4:(k4+1)]))
                x5<-unname(as.numeric(new_scaled_b[i,k5:(k5+1)]))
                x6<-unname(as.numeric(new_scaled_b[i,k6:(k6+1)]))

                density1.1<-dmvn(x1, mean1[1:2], S1[1:2,1:2])
                density2.1<-dmvn(x2, mean1[3:4], S1[3:4,3:4])
                density3.1<-dmvn(x3, mean1[5:6], S1[5:6,5:6])
                density4.1<-dmvn(x4, mean1[7:8], S1[7:8,7:8])
                density5.1<-dmvn(x5, mean1[9:10], S1[9:10,9:10])
                density6.1<-dmvn(x6, mean1[11:12], S1[11:12,11:12])

                density1.2<-dmvn(x1, mean2[1:2], S2[1:2,1:2])
                density2.2<-dmvn(x2, mean2[3:4], S2[3:4,3:4])
                density3.2<-dmvn(x3, mean2[5:6], S2[5:6,5:6])
                density4.2<-dmvn(x4, mean2[7:8], S2[7:8,7:8])
                density5.2<-dmvn(x5, mean2[9:10], S2[9:10,9:10])
                density6.2<-dmvn(x6, mean2[11:12], S2[11:12,11:12])

        density1<-c (density1.1, density2.1, density3.1, density4.1, density5.1,
                     density6.1)
        density2<-c (density1.2, density2.2, density3.2, density4.2, density5.2,
                     density6.2)

            new_cluster1[i,l:(l+5)]<-density1
            new_cluster2[i,l:(l+5)]<-density2
        }
}

new_cluster1$weight1<-weights[,1]
new_cluster2$weight2<-weights[,2]


colnames(new_cluster1[1447])<-paste("weight1")
colnames(new_cluster2[1447])<-paste("weight2")
```

#Weighted densities for each woman under each iteration each cluster

```
new_weighted_density_C1<-data.frame()
new_weighted_density_C2<-data.frame()
```

```
for (i in 1:50000){
  d1<-unname(as.numeric(new_cluster1[i,1:1446]*new_cluster1$weight1[i]))
  d2<-unname(as.numeric(new_cluster2[i,1:1446]*new_cluster2$weight2[i]))

  new_weighted_density_C1[i, 1:1446] <- d1
  new_weighted_density_C2[i, 1:1446] <- d2
}
```

#Total likelihood of each individual under each iteration for each cluster

```
new_tot_likelihood<-data.frame()
for(i in 1:50000){
d<-   unname(as.numeric(new_weighted_density_C1[i,1:1446]+
                  new_weighted_density_C2[i,1:1446]))

new_tot_likelihood[i, 1:1446] <- d
}
```

#Individual component probabilities for each cluster under each antibody

```
new_ind_prob_it_C1<-data.frame()
new_ind_prob_it_C2<-data.frame()

for(i in 1:50000){
  like<-new_tot_likelihood[i,1:1446]
  d1<-unname(as.numeric(new_weighted_density_C1[i,1:1446]/like))
  d2<-unname(as.numeric(new_weighted_density_C2[i,1: 1446]/like))

  new_ind_prob_it_C1[i,1: 1446]<-d1
  new_ind_prob_it_C2[i,1: 1446]<-d2
}
```

#Average of the individual component probabilities under each antibody for each cluster

```
new_ind_comp_prob_C1<-data.frame()
new_ind_comp_prob_C2<-data.frame()

colnames(new_ind_comp_prob_C1)<-colnames(new_ind_comp_prob_C2)<-
c("PfAMA1","PfEBA175","PfMSP2","PfMSP3","PvAMA1","PfVAR2CSA")

for (i in 1:241){
  k<-6*i-5
  d1<-unname(as.numeric(colSums(new_ind_prob_it_C1[,k:(k+5)],
        na.rm=TRUE)/50000))
  d2<-unname(as.numeric(colSums(new_ind_prob_it_C2[,k:(k+5)],
        na.rm=TRUE)/50000))
```

```
  new_ind_comp_prob_C1[i,1:6]<-d1
  new_ind_comp_prob_C2[i,1:6]<-d2
}
```

Computation of entropy values (Equation 15)

```
PfAMA1<-0
PfEBA175<-0
PfMSP2<-0
PfMSP3<-0
PvAMA1<-0
PfVAR2CSA<-0

Entropy<-data.frame(matrix(nrow = 1, ncol = 6))

colnames(Entropy)<-
c("PfAMA1","PfEBA175","PfMSP2","PfMSP3","PvAMA1","PfVAR2CSA")

for (i in 1:241){

  x1<-(new_ind_comp_prob_C1[i,1]*log(new_ind_comp_prob_C1[i,1]))+
        (new_ind_comp_prob_C2[i,1]*log(new_ind_comp_prob_C2[i,1]))

  x2<-(new_ind_comp_prob_C1[i,2]*log(new_ind_comp_prob_C1[i,2]))+
       (new_ind_comp_prob_C2[i,2]*log(new_ind_comp_prob_C2[i,2]))

  x3<-(new_ind_comp_prob_C1[i,3]*log(new_ind_comp_prob_C1[i,3]))+
       (new_ind_comp_prob_C2[i,3]*log(new_ind_comp_prob_C2[i,3]))

  x4<-sum(na.omit(new_ind_comp_prob_C1[i,4]*log(new_ind_comp_prob_C1[i,4])),
       (new_ind_comp_prob_C2[i,4]*log(new_ind_comp_prob_C2[i,4])))

  x5<-sum(na.omit(new_ind_comp_prob_C1[i,5]*log(new_ind_comp_prob_C1[i,5])),
       (new_ind_comp_prob_C2[i,5]*log(new_ind_comp_prob_C2[i,5])))

  x6<-sum(na.omit(new_ind_comp_prob_C1[i,6]*log(new_ind_comp_prob_C1[i,6])),
       (new_ind_comp_prob_C2[i,6]*log(new_ind_comp_prob_C2[i,6])))

  PfAMA1<-PfAMA1+x1
  PfEBA175<-PfEBA175+x2
  PfMSP2<-PfMSP2+x3
  PfMSP3<-PfMSP3+x4
  PvAMA1<-PvAMA1+x5
  PfVAR2CSA<-PfVAR2CSA+x6

  if(i==241){

    Entropy[,1]<-1+(PfAMA1/241*log(2))
```

```
    Entropy[,2]<-1+(PfEBA175/241*log(2))
    Entropy[,3]<-1+(PfMSP2/241*log(2))
    Entropy[,4]<-1+(PfMSP3/241*log(2))
    Entropy[,5]<-1+(PvAMA1/241*log(2))
    Entropy[,6]<-1+(PfVAR2CSA/241*log(2))

  }
}
```