Spatial epidemiology of *Plasmodium knowlesi* in Sabah, Malaysia.

KIMBERLY MICHELE FORNACE

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the University of London

AUGUST 2018

Department of Immunology and Infection

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Medical Research Council, Natural Environment Research Council, Economic and Social Research Council, and Biotechnology and Biosciences Research Council through the Environment and Social Ecology of Human Infectious Diseases Initiative

I, Kimberly Michele Fornace, confirm that the work presented within this thesis is my own. Where information has been derived from other sources, I confirm this has been indicated in the thesis.

Table of Contents

List of Figures

List of Tables

1 Abstract

Since identification of a large number of human infections in 2004, the zoonotic malaria species *Plasmodium knowlesi* is currently the main cause of human malaria in Malaysian Borneo. Carried by long and pig-tailed macaques, deforestation and associated environmental and population changes have been hypothesised to be the main drivers of this emergence. This thesis aims to describe the epidemiology of *P. knowlesi* in Northern Sabah, Malaysia at a range of spatial scales and evaluate how environmental change and behaviour affect human infection risks. Satellite-based and aerial remote sensing technologies were utilised with GPS tracking and field surveys to characterise dynamic interactions between the environment, human, macaque and mosquito populations (Chapter 4).

A retrospective analysis of reported *P. knowlesi* cases found marked spatial heterogeneity in village-level *P. knowlesi* incidence, with village-level numbers of cases positively associated with both forest cover and forest loss in surrounding areas (Chapter 5). To explore how these people used these environments, mobility patterns and resource utilisation were mapped using GPS tracking devices; these data were integrated with predictions of mosquito biting rates to estimate individual and location-specific exposure risks (Chapter 6). Additional surveys were conducted in households and villages of symptomatic cases and within case study communities to evaluate levels of infection and exposure within the wider population. These identified a substantial proportion of asymptomatic infections not detected by hospital-based surveillance systems (Chapter 7) and found positive associations between knowlesi sero-positivity and environmental factors at a fine spatial scale (Chapter 8). Results from these studies were used to design an ecologically-stratified cross-sectional survey across four districts in Northern Sabah; this study was used to identify individual and environmental variables associated with exposure and infection risk across a wider geographical area (Chapter 9).

Together, these studies indicate a wide distribution of *P. knowlesi* infection and exposure in demographic groups underrepresented in clinical reports and highlight the role of local environmental change in *P. knowlesi* risk. Further research is needed to refine diagnostic methods and understand the longer-term impacts of ecological changes on disease dynamics.

Statement of Contributions and Additional Publications

The work described within this thesis was conducted as part of the MONKEYBAR research programme under the UK Research Council Environmental and Social Ecology of Human Infectious Diseases Initiative (ESEI), Grant Number: G1100796. All field work activities were conducted in Sabah, Malaysia through the London School of Hygiene and Tropical Medicine in collaboration with partners at the Universiti Sabah Malaysia, Danau Girang Field Centre, the Infectious Diseases Society Kota Kinabalu, the Malaysian Ministry of Health, University of Malaya, Royal Veterinary College, University of Glasgow and Menzies School of Health Research. Additional work in the Philippines was conducted in partnership with the Research Institute of Tropical Medicine and the University of the Philippines Los Banos. This project was initially developed by Chris Drakeley and Jon Cox in collaboration with partners in the UK, Malaysia, the Philippines and Australia.

My role and the contributions of my collaborators to the studies included in this thesis are summarised below.

Chapter 2:
This chapter describes background on the epidemiology of *Plasmodium knowlesi* and outlines key knowledge gaps. I wrote all the content for this chapter; however, sections of this chapter were later incorporated into a World Health Organisation report on land use change and malaria.

Chapter 3:
This chapter outlines the study design and objectives as well as the study sites where this work was undertaken.

Chapter 4:
This chapter summarises methods of environmental classification utilised throughout this thesis and includes a published paper on the uses of unmanned aerial vehicles (UAVs) in epidemiological studies. The initial approach to land use mapping using UAVs was conceived of by Jon Cox and permissions for this work were obtained by Timothy William, Fe Espino, JD Marcos and Albert Lim. I supervised the field work and conducted UAV mapping activities in collaboration with Tommy Rowel Abidin, Dellroy Donny and Redley Yambun. I wrote the manuscript, analysed the data and conducted all subsequent analysis of satellite-based remote sensing data and developmental of regional land cover maps. Albert Lim, Tommy Rowel Abidin and Lina Marlina assisted with additional ground-truthing of land cover classification and identification of spatial data sources.

FORNACE, K. M., DRAKELEY, C. J., WILLIAM, T., ESPINO, F. & COX, J. 2014. Mapping infectious disease landscapes: unmanned aerial vehicles and epidemiology. *Trends Parasitol,* 30, 514-519.

Chapter 5:

This chapter includes a published paper on the spatial patterns of *P. knowlesi* in Sabah and associations with land use change. I designed this study, analysed the data and wrote the manuscript. Tommy Rowel Abidin, Amanda Murphy and Matthew Grigg contributed to collecting and geo-locating records and Timothy William and Jayaram Menon secured permissions. Neal Alexander, Jon Cox, Paddy Brock and Chris Drakeley advised on analysis methods and reviewed the manuscript.

FORNACE, K. M., ABIDIN, T. R., ALEXANDER, N., BROCK, P., GRIGG, M. J., MURPHY, A., WILLIAM, T., MENON, J., DRAKELEY, C. J. & COX, J. 2016. Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. *Emerg Infect Dis*, 22**,** 201-9.

Chapter 6:

Chapter 6 includes an unpublished chapter on the effect of local human movement patterns on *P. knowlesi* exposure. The initial use of GPS tracking devices was conceived of by Chris Drakeley and Jon Cox. I designed the protocol for this study with help from Jon Cox, supervised the study, conducted all analysis of the data and wrote the chapter. Data was collected by the field teams in Sabah coordinated by Albert Lim and Tommy Rowel Abidin. Entomological data included in this chapter was collected by collaborators at University of Malaya and Universiti Malaysia Sabah (references included in additional publications).

Chapter 7:

This chapter includes a published paper on asymptomatic carriage of *P. knowlesi*. Samples analysed within this study were collected under a case control study led by Matthew Grigg (see additional publications). All laboratory work was conducted by Afizah Nuin, Martha Betson and Lau Tiek Ying. I analysed the data and wrote the manuscript.

FORNACE, K. M., NUIN, N. A., BETSON, M., GRIGG, M. J., WILLIAM, T., ANSTEY, N. M., YEO, T. W., COX, J., YING, L. T. & DRAKELEY, C. J. 2016. Asymptomatic and Submicroscopic Carriage of Plasmodium knowlesi Malaria in Household and Community Members of Clinical Cases in Sabah, Malaysia. *J Infect Dis*, 213**,** 784-7.

Chapter 8:

Chapter 8 describes comprehensive surveys of case study communities in Sabah and the Philippines. This work relies heavily on the development of species-specific antigens for *P. knowlesi* and serological analysis, led by Lou Herman and Kevin Tetteh (additional publications, Annex 1). I wrote the protocol for this study in collaboration with Jon Cox, Chris Drakeley, Timothy William and Fe Espino. I supervised field activities in Sabah and helped to coordinate activities in the Philippines with Fe Espino and Joy Lorenzo. Surveys were conducted by the field teams in Sabah and the Philippines and microscopy and PCR analysis was completed by colleagues at the Universiti Sabah Malaysia, Infectious Diseases Society Kota Kinabalu and Research Institute for Tropical Medicine. I analysed all data for these surveys and wrote the manuscript.

FORNACE, K. M., HERMAN, L., ABIDIN, T. R., CHUA, T. H., DAIM, S., LORENZO, P. J., GRIGNARD, L. NUIN, N. A., YING, L. T., GRIGG, M. J., WILLIAM, T., ESPINO, F., COX, J., TETTEH, K. & DRAKELEY, C. J. 2018. Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, the Philippines. *PLoS Negl Trop Dis.* 12: e0006432

Chapter 9:

Chapter 9 describes a population-based cross-sectional survey conducted across Northern Sabah. This study was designed with Jon Cox, Chris Drakeley and other project partners. I supervised the field activities and data collection and management, analysed the data and wrote the chapter. Molecular analysis was completed by Lynn Grignard and serological analysis was completed by Kevin Tetteh, Lou Herman and colleagues at LSHTM.

Chapter 10:

This chapter summarises findings from this work and highlights additional areas where further research is needed.

Additional publications:

I additionally contributed to other publications which were not part of my PhD work but completed during the MONKEYBAR project.

BROCK, P. M., FORNACE, K. M., PARMITER, M., COX, J., DRAKELEY, C. J., FERGUSON, H. M. & KAO, R. R. 2016. Plasmodium knowlesi transmission: integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology,* 143**,** 389-400.

BROCK, P. M., FORNACE, K. M., GRIGG, M. J., ANSTEY, N. M., WILLIAM, T., COX, J., DRAKELEY, C. J., FERGUSON, H. M. & KAO, R. R. (submitted). Data mining across spatial scales links zoonotic malaria to deforestation.

GRIGG, M. J., COX, J., WILLIAM, T., JELIP, J., FORNACE, K. M., BROCK, P. M., VON SEIDLEIN, L., BARBER, B. E., ANSTEY, N. M., YEO, T. W. & DRAKELEY, C. J. 2017. Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case control study. *Lancet Planetary Health,* 1**,** e97-104.

HERMAN, L. S., FORNACE, K. M., PHELAN, J., GRIGG, M. J., ANSTEY, N. M., WILLIAM, T., BLACKMAN, M., DRAKELEY, C. J. & TETTEH, K. K. A. 2018. Identification and validation of a novel panel of *Plasmodium knowlesi* biomarkers of serological exposure. *PLoS Negl Trop Dis*, 12:e0006457 (included in **Annex 1**)

MANIN, B. O., FERGUSON, H. M., VYTHILINGAM, I., FORNACE, K., WILLIAM, T., TORR, S. J., DRAKELEY, C. & CHUA, T. H. 2016. Investigating the Contribution of Peri-domestic Transmission to Risk of Zoonotic Malaria Infection in Humans. *PLoS Negl Trop Dis,* 10**,** e0005064.

STARK, D. J.*, FORNACE, K. M.*, BROCK, P. M., ABIDIN, T. R., GILHOOLY, L., JALIUS, C., GOOSSENS, B., DRAKELEY, C. J. & SALGADO-LYNN, M. (submitted). Response of a group of long-tailed macaques (*Macaca fascicularis*) to habitat clearing in Sabah, Malaysian Borneo: implications for *Plasmodium knowlesi* risk.
*contributed equally

WONG, M. L., CHUA, T. H., LEONG, C. S., KHAW, L. T., FORNACE, K., WAN-SULAIMAN, W. Y., WILLIAM, T., DRAKELEY, C., FERGUSON, H. M. & VYTHILINGAM, I. 2015. Seasonal and Spatial Dynamics of the Primary Vector of Plasmodium knowlesi within a Major Transmission Focus in Sabah, Malaysia. *PLoS Negl Trop Dis,* 9**,** e0004135.

2 Background

Anthropogenic environmental change has been widely proposed as a main driver of infectious disease emergence and transmission [1-3]. Changes such as deforestation, urbanisation and agricultural expansion affect the distribution of people, animal reservoirs and disease vectors, impacting infectious disease risks [4, 5]. These have been linked to altered dynamics and geographical distribution of malaria and other vector-borne diseases globally [6-8]. Understanding how ecological changes driven by human activity can modify the epidemiology of these diseases is vital to predicting future disease risks and designing effective public health measures.

The zoonotic malaria, *Plasmodium knowlesi*, is an emerging disease hypothesised to be affected by these changing patterns of land use. Maintained by long and pig-tailed macaques (*Macaca fascicularis* and *M. nemestrina*) and transmitted by the *Anopheles leucosphyrus* group of mosquitoes, the geographical range of *P. knowlesi* is limited by the distribution of the mosquito vectors and simian hosts [9, 10]. Since the identification of a large number of human infections with *P. knowlesi* in Malaysian Borneo in 2004, *P. knowlesi* is now the most common cause of human malaria in Sabah, Malaysia [11-13]. Land use changes, resulting in increased spatial overlap between people, macaques and mosquitoes, have been proposed as the main driver of this apparent emergence [10, 14, 15]. However, despite these potential linkages, detailed spatial and environmental risk factors for human infection with *P. knowlesi* are unknown. This thesis will characterise associations between environmental change, landscape factors and the risk of human *P. knowlesi* infection in Sabah, Malaysia.

2.1 Epidemiology of *Plasmodium knowlesi*

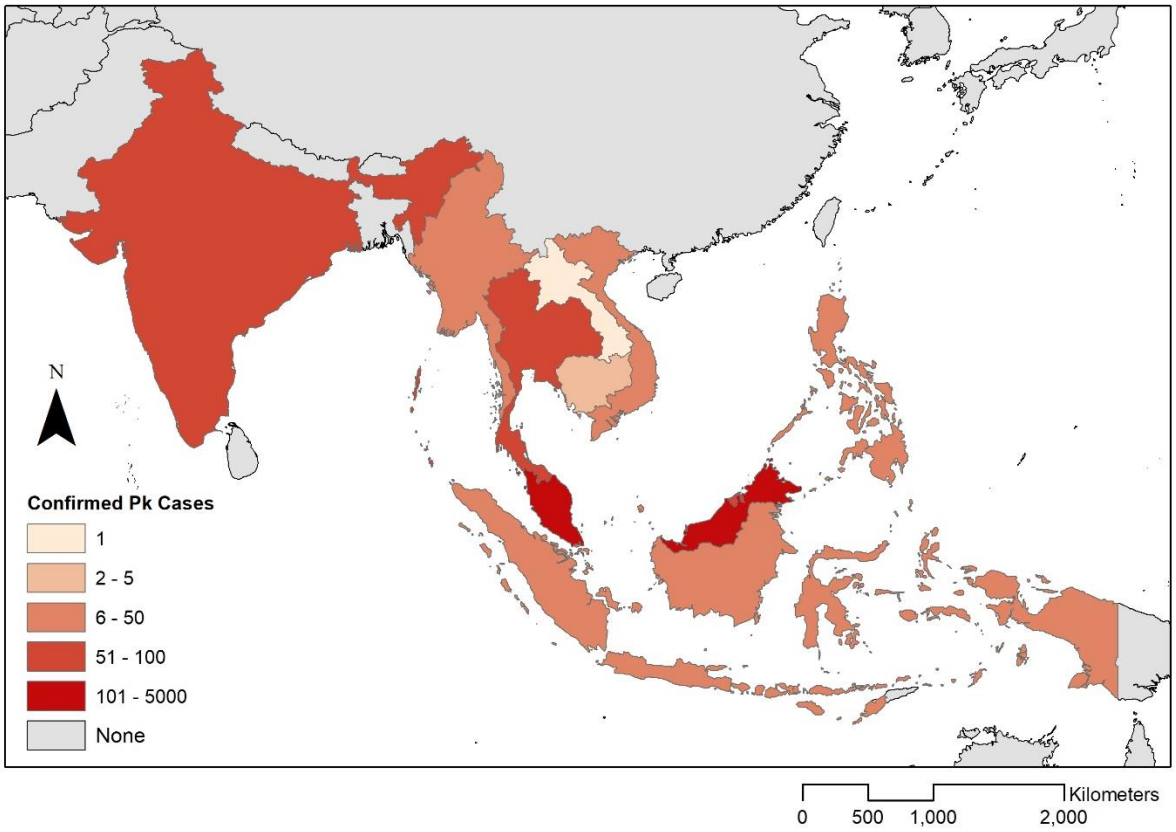
Increasing numbers of human *P. knowlesi* cases have been identified throughout Southeast Asia, posing a threat to current malaria control and elimination programmes in this region [9, 11, 13]. Although the majority of *P. knowlesi* cases respond to treatment, infection with *P. knowlesi* can cause severe and fatal disease in people [16-20]. Designing effective control programmes requires detailed information on the spatial epidemiology of *P. knowlesi* as well as an understanding of how environmental changes are likely to impact this distribution.

Molecular studies indicate *P. knowlesi* is not a newly emergent malaria species and likely predates human settlement in Southeast Asia [15]. *P. knowlesi* was first described in macaques in the 1930s

and the first naturally acquired human case was reported in 1965 in peninsular Malaysia [21, 22]. Subsequent epidemiological investigations of people residing within the area where this individual was infected did not identify any additional *P. knowlesi* cases in people, although *P. knowlesi* was detected in two out of the four long-tailed macaques screened [23]. No further natural infections were reported until 2004, when the application of molecular diagnostic tools identified a large focus of *P. knowlesi* infections in the Kapit division of Sarawak in Malaysian Borneo [11]. Retrospective studies have since detected *P. knowlesi* infections from archival blood films collected in the mid-1990s in Malaysia and Thailand, suggesting that previous human infections were misdiagnosed as other species by microscopy [24, 25].

Since 2004, human *P. knowlesi* infections have been reported from a number of Southeast Asian countries [9, 26]. While sporadic cases have been reported from countries including the Philippines, Thailand, Vietnam, Cambodia, Myanmar, Singapore, Brunei, India, Indonesia and China, *P. knowlesi* is now the most common cause of human malaria in areas of Malaysian Borneo [12, 13, 27-40] (Figure 2.1). In the Malaysian state of Sabah, despite an overall decrease in malaria notifications following successful malaria control measures, the percentage of suspected *P. knowlesi* notifications has increased from 2% (59/2741) of all malaria notifications in 2004 to 62% (996/1606) of reported malaria cases in 2013 [12, 13]. Similar trends have been reported in the neighbouring state of Sarawak, where suspected *P. knowlesi* now accounts for the majority of reported malaria cases [41].

Figure 2.1 Reports of PCR confirmed *P. knowlesi* cases by country from 1996 – 2014 (data obtained from [26] and aggregated by country)



However, the true burden of *P. knowlesi* remains poorly understood due to frequent misidentification as other human malaria species by microscopy and the limited availability of *P. knowlesi*- specific molecular diagnostic capabilities. *P. knowlesi* appears microscopically similar to the human malaria species *P. malariae* but also can be misdiagnosed as *P. falciparum* and *P. vivax* [14, 42, 43]. Rapid diagnostic tests (RDTs) developed for other malaria species are also insufficiently sensitive to detect *P. knowlesi* and can lead to misdiagnosis as other species [44-46]. While it is difficult to determine whether there is a genuine increase in human cases rather than improved detection, evidence of a significant increase in numbers and proportions of *P. knowlesi* malaria suggests transmission of *P. knowlesi* is rising in areas of Malaysian Borneo [12, 13, 47, 48].

2.2 Environmental change and P. knowlesi

Land use changes, such as deforestation and agricultural expansion, have been proposed as the main drivers of the apparent emergence of human *P. knowlesi* [15, 49]. Anthropogenic land use changes can affect infectious disease transmission by altering the abundance, distribution and behaviour of

people, wildlife reservoirs and disease vectors (Table 2.1). These environmental changes are particularly pronounced in tropical areas, where more than 80% of new agricultural land came from tropical rainforests between 1980 and 2000 and an estimated 2100 km^2 of forest were lost per year between 2000 and 2012 [50, 51] (Figure 2.2). The Malaysian state of Sabah has been named a global hotspot of forest loss and degradation due the rapid conversion of land for oil palm plantations and other agricultural activities [52, 53].

Figure 2.2. Forest loss from 2000 – 2016 in Southeast Asia (forest cover defined as greater than 50% canopy cover, data obtained from [51])

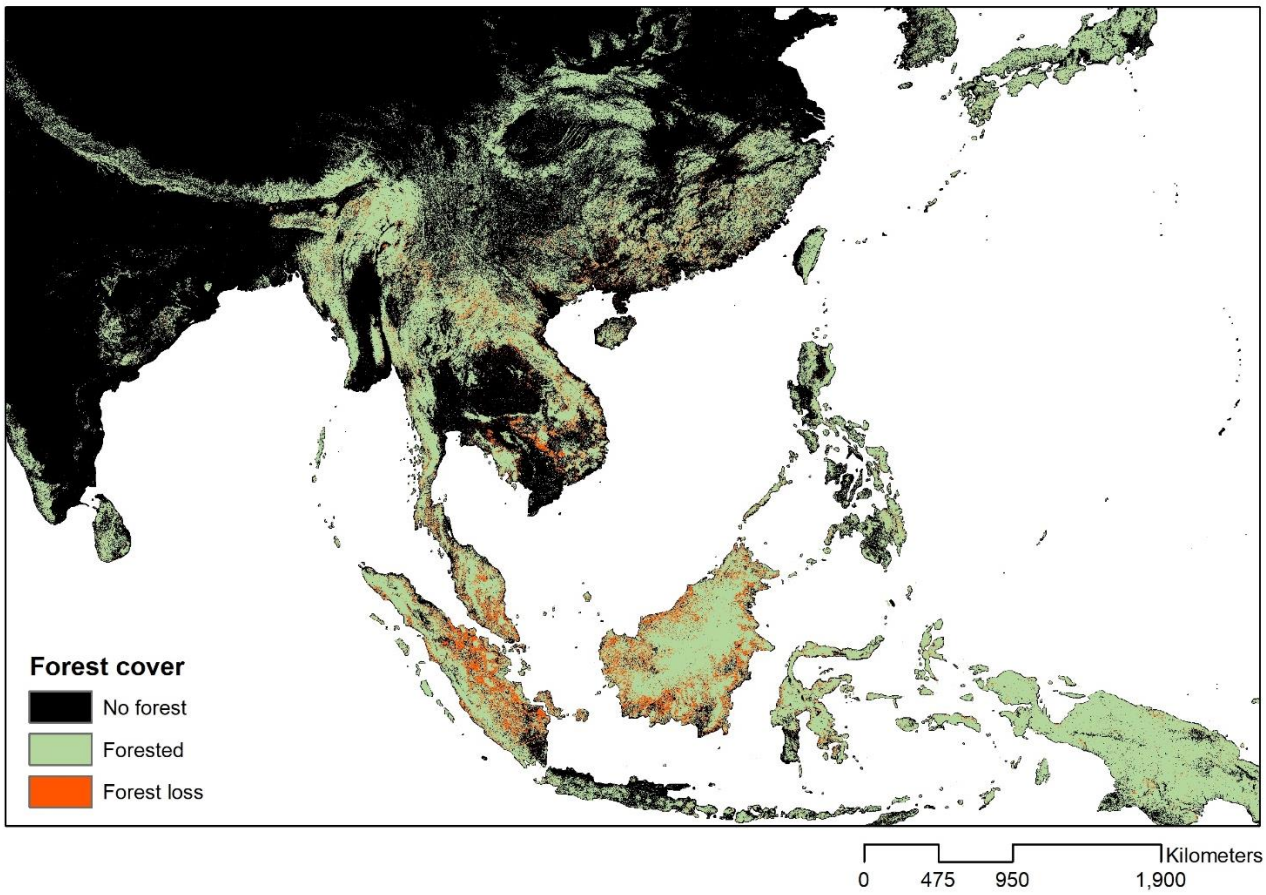


Table 2.1. Examples of effects of land use change on malaria risks

Environmental changes		References
Deforestation	Increases in anopheline larval breeding sites in response to forest clearing in the Amazon	[54]
	Initial decreases in vector densities followed by colonisation by more efficient malaria vectors	[6, 55]
	Changes in vector habitat suitability linked with forest disturbance	[56, 57]
	Changes in ecological structure and biodiversity increasing or decreasing vector densities, availability of blood meals and resulting disease risks	[58-60]
Agricultural expansion	Effects of irrigation systems	[61, 62]
	Expansion of rubber and rice paddies associated with increases in anopheline densities	[8, 63]
Socio-demographic changes		
Population at risk	Influx of susceptible populations into endemic areas in response to increased economic opportunity	[64, 65]
	Increase and movement of migrant worker populations in the Amazon and Southeast Asia	[66, 67]
	Occupational changes, such as forestry and extraction activities bringing people into vector habitats	[68, 69]
Socioeconomic status	Increased income following agricultural development leading to decrease in malaria risk	[70]
	Improved housing structure due to development reducing malaria risks	[71, 72]
Wildlife reservoirs		
Origin of malaria	*P. falciparum* originated from non-human primates	[73]
Spatial overlap with wildlife hosts	Increased contact between people and non-human primates hypothesised as main driver of human infections with *P. knowlesi* and *P. cynomolgi* in Asia and *P. simium* and *P. brasilianum* in South America	[74-77]
Maintenance of malaria infections	Human malaria species circulating in great apes and gorillas in West and Central Africa	[78, 79]

2.2.1 Mosquito habitats

Physical changes in the environment, such as changes in vegetation, microclimate and soil, can affect the species composition and abundance of vector populations [8]. For example, deforestation has been shown to create environmental conditions favourable for larval breeding sites of malaria vectors in the Peruvian Amazon rainforest [54]. Within Southeast Asia, agriculture such as rubber plantations and rice paddies have been associated with increased anopheline densities as well as increases in malaria incidence [8]. In some cases, deforestation has been reported to cause initial depletion of forest dwelling vectors followed by colonisation of the area by more efficient vector species and overall increases in malaria transmission [6, 55].

Previous studies have implicated the primarily exophagic *Anopheles leucosphyrus* group of mosquitoes as the main vector of *P. knowlesi* and found relatively high biting rates in both farm edge and forest areas [80, 81]. Entomological studies in Kapit, Sarawak incriminated *An. latens* as the main vector of *P. knowlesi* in this area and observed this species is both attracted to macaques in the canopy and humans on the ground [81, 82]. Studies within Sabah have implicated *An. balabacensis* as the primary vector of *P. knowlesi* in Northwestern Sabah; this species was also historically the main vector of human malaria species within the region and has been experimentally shown to be able to transmit *P. knowlesi* [83-85]. A study of *P. knowlesi* vectors across a forest disturbance gradient in Sabah found abundance of *An. balabacensis* was higher in previously logged forests compared with primary forests and this vector were present at both ground and canopy levels, suggesting the potential for this mosquito to transmit *P. knowlesi* between canopy-dwelling primates and people at ground level [86]. Additional investigations of anopheline mosquito densities within villages reporting knowlesi cases found higher densities of *An. balabacensis* in environments around cases households and identified vectors infected with simian malarias in peridomestic settings [87]. A systematic analysis of *P. knowlesi* mosquito vectors also suggested the leucosphyrus mosquito habitat to primarily consist of disturbed forest, although this analysis was limited by the crude environmental data used [10]. However, together these data indicate that changing land use patterns are affecting the distribution and behaviour of mosquito vectors and that conversion of previously intact forests to agricultural land may increase the abundance of these vectors.

2.2.2 Macaque habitats

Changing habitats are likely to have similar effects on the distribution and density of the macaque hosts. *P. knowlesi* has been frequently identified in both long-tailed and pig-tailed macaques (*Macaca fascicularis* and *M. nemestrina*) and likely infects *M. leonina*, a close relative of the pig-tailed macaque recently classified as a separate species [10, 88]. Previous studies have indicated a relatively high proportion of macaques are infected with *P. knowlesi* and genetic studies of human infections have identified two distinct parasite populations associated with each macaque reservoir species [15, 89]. While the possibility of human to human transmission of *P. knowlesi* has been demonstrated experimentally, the high parasite diversity found within macaques suggests transmission remains primarily zoonotic [89-91]. Macaques are also frequently found in close contact to human settlements and in highly disturbed environments, where loss of natural habitat may lead to increased dependence on anthropogenic food sources and closer contact with people

[10, 92, 93]. Habitat disturbance can also influence primate disease transmission by altering ranging patterns, increasing density or crowding within forest patches or weakening immunity through exposure to other pathogens e.g. [93-98]. A mathematical modelling study of *P. knowlesi* highlighted how transmission might increase with higher contact rates between people and macaques around habitat edges [75].

2.2.3 Human use of the environment

Environmental changes also can also bring humans into closer contact with disease reservoirs and vectors as new areas are settled and agricultural and forest activities are undertaken. Individual human movement patterns influence the exposure to disease vectors and wildlife reservoirs within different environments [99, 100]. These movements occur on different spatial and temporal scales, from long term migrations to daily movements in areas surrounding households [64, 101]. Regional migration can lead to immunologically naïve individuals moving to areas of disease transmission or introduce infected individuals to areas previously free from disease [102]. On a local scale, deforestation has been associated with higher levels of human activities in forest areas and increased exposure to anopheline malaria vectors, previously described as frontier malaria [65, 69]. Other forest-related occupational activities, such as logging, rubber tapping and mining, have also been linked to higher malaria risks [7, 66-68].

In Kapit, Sarawak, the majority (83%; 93/107) of molecularly confirmed *P. knowlesi* patients reported some type of forest exposure [19]. Similarly, a study of patients from a referral hospital in Sabah found that most *P. knowlesi* cases (92%, 119/130) had spent time in a forest or plantation the previous month [18]. However, reports from a district hospital in the largely deforested area of Kudat in North-western Sabah describe a wide age distribution of *P. knowlesi* cases (0.7 to 89 years) and lack of association with forest activities [37]. A subsequent case control study conducted within Kudat and Kota Marudu districts found a similarly wide age distribution (3 – 85 years), with 9% of cases occurring in children under the age of 15 [103]. This study also identified plantation work, forest activities and sleeping outside overnight as associated with increased *P. knowlesi* risk, however, a minority of cases, including children, had no reported farm or forest activities, suggesting the possibility of peri-domestic transmission. The protective effects of indoor residual spraying (IRS) and associations between risk and vegetation surrounding the household provides additional evidence for transmission around the household. Further research is needed to understand the

importance of peri-domestic transmission and how this relates to disruption of macaque and mosquito habitats.

2.3 Current knowledge gaps and future research needs

Despite these apparent linkages between land use change and *P. knowlesi* transmission, environmental risk factors for human infections with *P. knowlesi* are unknown. This is limited by the availability of population-level epidemiological information on *P. knowlesi* infection and exposure as well as detailed land use and land cover (LULC) data within areas of knowlesi transmission. Few studies have explored the spatial distribution of *P. knowlesi* or examined association between environmental factors and *P. knowlesi* risk [9, 10, 26, 104]. These studies have all relied on passive hospital reporting at regional levels and have only included limited environmental data (such as classification of areas as forest or other). Further research is needed to understand the fine-spatial distribution of *P. knowlesi* in the community and associations with landscape factors.

2.3.1 Spatial distribution of exposure and infection

Despite increasing amounts of data available for symptomatic *P. knowlesi* cases presenting at hospital facilities, little is known about the patterns of infection and exposure in the community. It is unknown how many asymptomatic malaria cases have occurred or how many symptomatic cases were resolved before medical treatment was required or did not seek treatment.

With the improved sensitivity of malaria diagnostics, low level asymptomatic malaria parasite carriage is increasingly being described (e.g. [105-109]). Population level epidemiological surveys have reported a substantial proportion of individuals with low density infections that are not detectable by conventional malaria microscopy [110]. These infections have been estimated to account for 50% of malaria infections, with higher proportions of submicroscopic infections described in low transmission settings (settings with a community parasite prevalence of less than 10%) [111]. In regions of low transmission intensity after successful malaria control initiatives, submicroscopic infections can be particularly prevalent, with reports of over 80% of *P. falciparum* infections in these settings being submicroscopic [110].

While submicroscopic parasitemia has been described for human malaria species, particularly *Plasmodium falciparum* and *P. vivax*, limited data are available on submicroscopic carriage of zoonotic malaria species. Asymptomatic *P. knowlesi* infection had previously only been identified by

few studies and submicroscopic parasitemia had not been described [31, 112]. During two large scale cross-sectional malariometric surveys in Vietnam, three individuals were identified as *P. knowlesi* positive using molecular techniques [31]. A more recent study in Indonesia also identified an asymptomatic *P. knowlesi* case through a reactive case detection programme screening individuals residing in the same village as symptomatic malaria cases [113]. However, no community-based surveys have been conducted in areas reporting relatively high knowlesi transmission and cross-sectional surveys are required to understand the prevalence of these infections in Sabah and understand spatial patterns of disease infection.

The distribution of community-level exposure can also be assessed by prevalence of antibodies to species-specific malaria antibodies, reflecting previous exposure to malaria [114-116]. Age-specific prevalence of these antibodies can be used to calculate seroconversion rates and evaluate changes in transmission over time; this measure has been shown to be closely correlated with other indicators of malaria transmission intensity such as parasite prevalence or entomological inoculation rates [117]. Serological indices of exposure also have increased utility in low transmission settings where the probability of detecting infections is very low [114, 118]. While numerous antigens have been described for *P. falciparum* and *P. vivax*, species-specific antigens for *P. knowlesi* have only recently been developed [119]. Sero-epidemiological approaches had not previously been applied to *P. knowlesi*.

2.3.2 Risk factors for *P. knowlesi* infection and exposure

Although small scale spatial variations in malaria risk have been reported for other malaria species and variability of *P. knowlesi* risk has been reported on regional scales, fine-scale spatial patterns of *P. knowlesi* have not been described and population-level risk factors remain unknown [9, 104]. Human behaviour and contact with macaques, mosquitoes and the habitats these populations are associated with are likely to influence risk; characterising these requires spatially and temporally accurate data on the environment and distributions of people, animals and disease vectors from individual to local and regional levels.

Numerous studies have highlighted how individual movement into different environments may affect *P. knowlesi* risk through occupational and forest activities however these studies have all relied on self-reported data (e.g. [11, 18, 103]). Quantifying and mapping these local movement patterns can provide further insights on how environment determines individual infection risk [101]. For example, contact with households likely to be infested with infected mosquitoes has been

demonstrated to drive infection risk for dengue in an urban area of Peru [120]. Increasing availability of GPS technology and mobile phones have been used to map human mobility patterns; these studies have revealed marked heterogeneity in individual movement and risk behaviours [102, 121-127]. Individual GPS trackers and detailed behavioural interviews and questionnaires have also been used to study movements on a finer spatial scale to contact with vector habitats for studies of dengue, schistosomiasis and malaria [100, 120, 128-132]. These GPS trackers have been used to monitor spatial interactions and potential for disease transmission between livestock and wildlife but have not yet been applied to studying zoonotic disease transmission between people and wild animal populations [133]. Further applications of these technologies can be used to evaluate the complex ways local land use affects *P. knowlesi* transmission.

Data on the role of local environments can be used to design population-based cross-sectional surveys to understand environmental risk factors across a wider ecological gradient. Geostatistical techniques are widely used to quantify the associations between infectious diseases and spatially correlated environmental data and develop predictive spatial models of disease occurrence and transmission [134]. Resulting risk maps have been widely used to predict the geographical distributions of infectious diseases and inform disease control and surveillance programmes e.g. [135-139]. Further research is needed to assess the relationship between ecological factors and the spatial patterns of *P. knowlesi* in the community in order to understand the geographical variation in *P. knowlesi* risk and predict the impacts of future land use changes.

3 Study design

This thesis will explore associations between environmental change and the risk of *P. knowlesi* infection in Sabah, Malaysia. This thesis will focus on describing patterns of *P. knowlesi* infection and exposure at a variety of spatial scales and evaluating associations between individual-, household- and village-level risk and a range of environmental and behavioural factors. Novel technologies and techniques will be employed to characterise the environment and contact between people, macaque and mosquito vector populations.

3.1 Aim

To describe the spatial epidemiology of *Plasmodium knowlesi* in Sabah, Malaysia and characterise how environmental changes and human behaviour affect the risk of human exposure and infection at a range of spatial scales

3.2 Objectives

- *To describe historical patterns of P. knowlesi incidence and association with environmental factors in Northern Sabah*
  Initial analysis on the historical distribution of *P. knowlesi* within Northern Sabah will be conducted to characterise *P. knowlesi* transmission within the area and identify case study areas for intensive study. Additionally, this analysis will examine the relationship between village level incidence and environmental characteristics to explore the hypothesis that forest loss is associated with increased *P. knowlesi* incidence.

- *To assess the probability of human exposure to P. knowlesi based on individual movement patterns and identify characteristics of individuals and places associated with exposure*
  GPS tracking of individuals in case study communities will be conducted to characterise local movement patterns during peak mosquito biting periods, in order to 1.) analyse individual movement patterns and develop predictive maps of human space use relative to spatial and environmental factors and 2.) assess relative exposure risks for *P. knowlesi* at individual and community levels based on modelled mosquito biting rates.

- *To detect potential asymptomatic P. knowlesi cases by screening individuals residing in the household and villages of symptomatic cases*

This analysis attempted to detect whether asymptomatic *P. knowlesi* infections were present in the community to explore the hypothesis that, similar to other human malaria species, asymptomatic and submicroscopic human *P. knowlesi* infections would be present in areas with on-going knowlesi transmission.

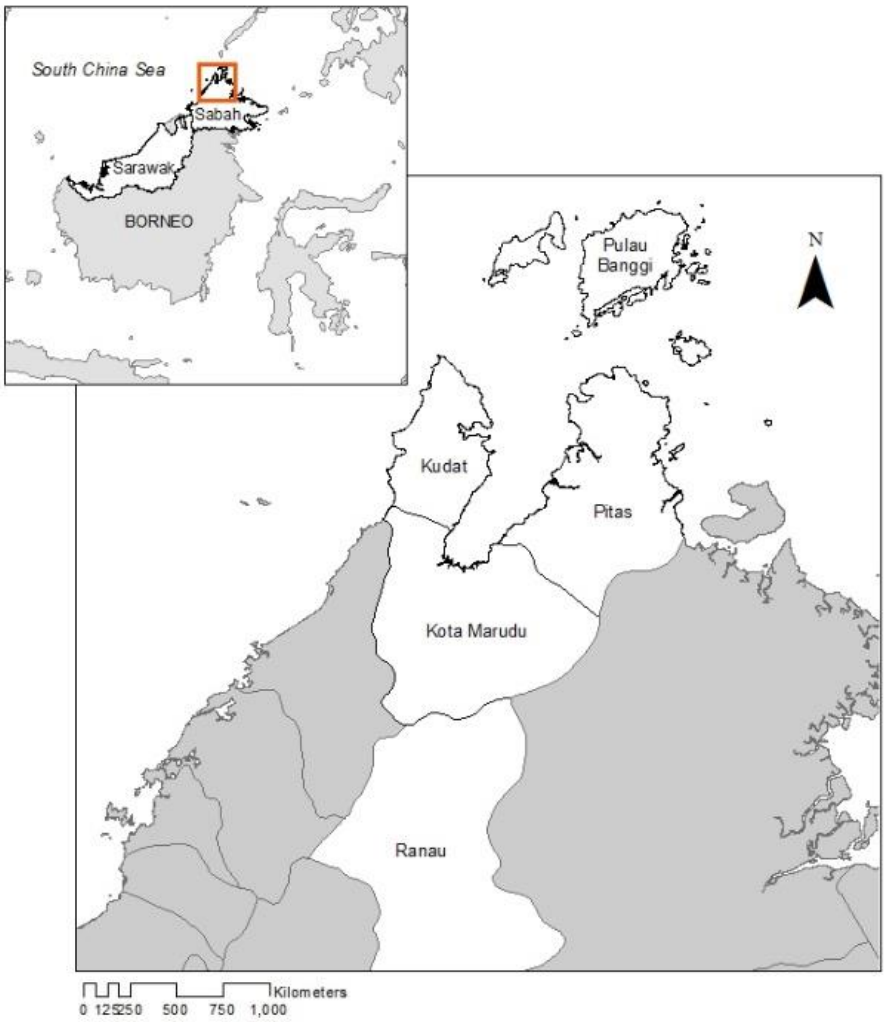
- *To characterise detailed spatial patterns of infection and exposure in case study communities*
  This study aimed to describe infection and exposure (as defined by species-specific serological methods) within case study communities. The main objective of this study was to estimate sero-prevalence to knowlesi and other malaria species and determine associated risk factors. As a secondary objective, this study attempted to assess the extent of asymptomatic infections. Initial results from this study were used to guide the design of a larger cross-sectional survey.

- *To identify population-level spatial and environmental risk factors for human infection and exposure in Northern, Sabah*
  The primary objective of this study was to estimate transmission intensity of *P. knowlesi* and other malaria parasites, as measured by prevalence of antibodies to specific malaria antigens. In order to estimate sero-prevalence across Northern Sabah, an environmentally-stratified two- stage cross-sectional survey was conducted across four districts. Design of this survey and data collected was informed by identified environmental risk factors from analyses of historical data and intensive study areas. Results from this survey were analysed with detailed land use classification and other environmental data to identify population-level environmental risk factors for exposure.

3.3 Study site

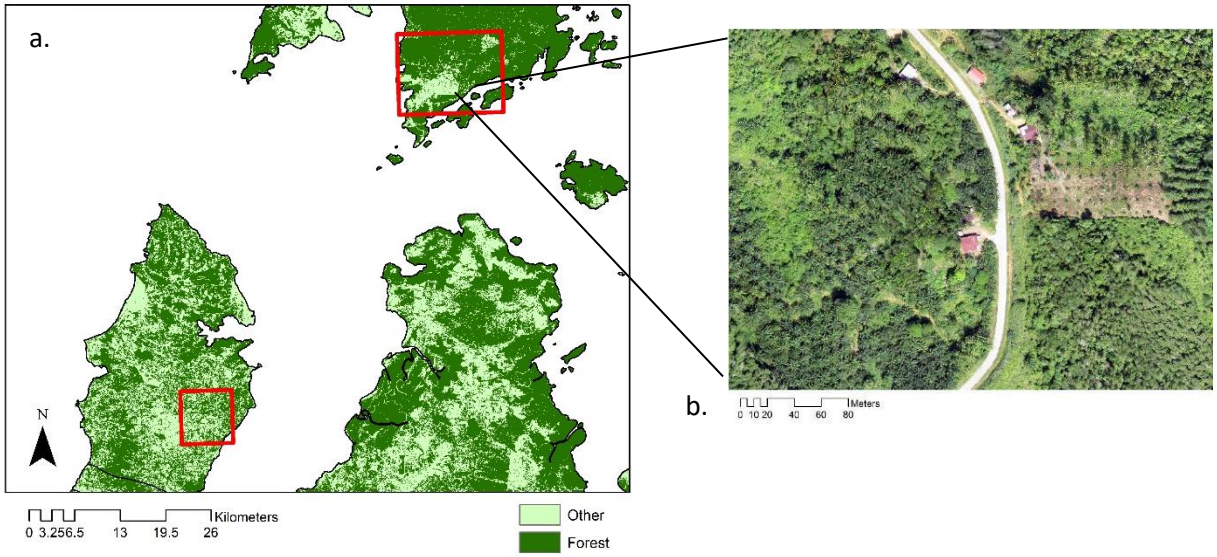
These studies will be conducted in four districts in Northern Sabah in Malaysian Borneo, an area with a population of approximately 285,000 people, predominantly of Rungus and Dusun ethnicities [140] (Figure 3.1). The climate is tropical, with no dry season and higher rainfall from November through March. This area includes coastal and inland regions, with elevations ranging from sea level to over 1,000 MSL. These districts are currently undergoing substantial environmental change due to conversion of land for oil palm plantations and other agricultural activities [52, 53]. All districts have central referral hospitals serving defined catchment areas. Malaria is a notifiable disease and patients have access to diagnosis and treatment free of charge. All clinics refer cases to the central district hospital where hospitalisation is mandatory for all malaria patients until a negative blood smear for malaria parasites has been obtained.

Figure 3.1. Kudat, Kota Marudu, Pitas and Ranau districts, Northern Sabah, Malaysia



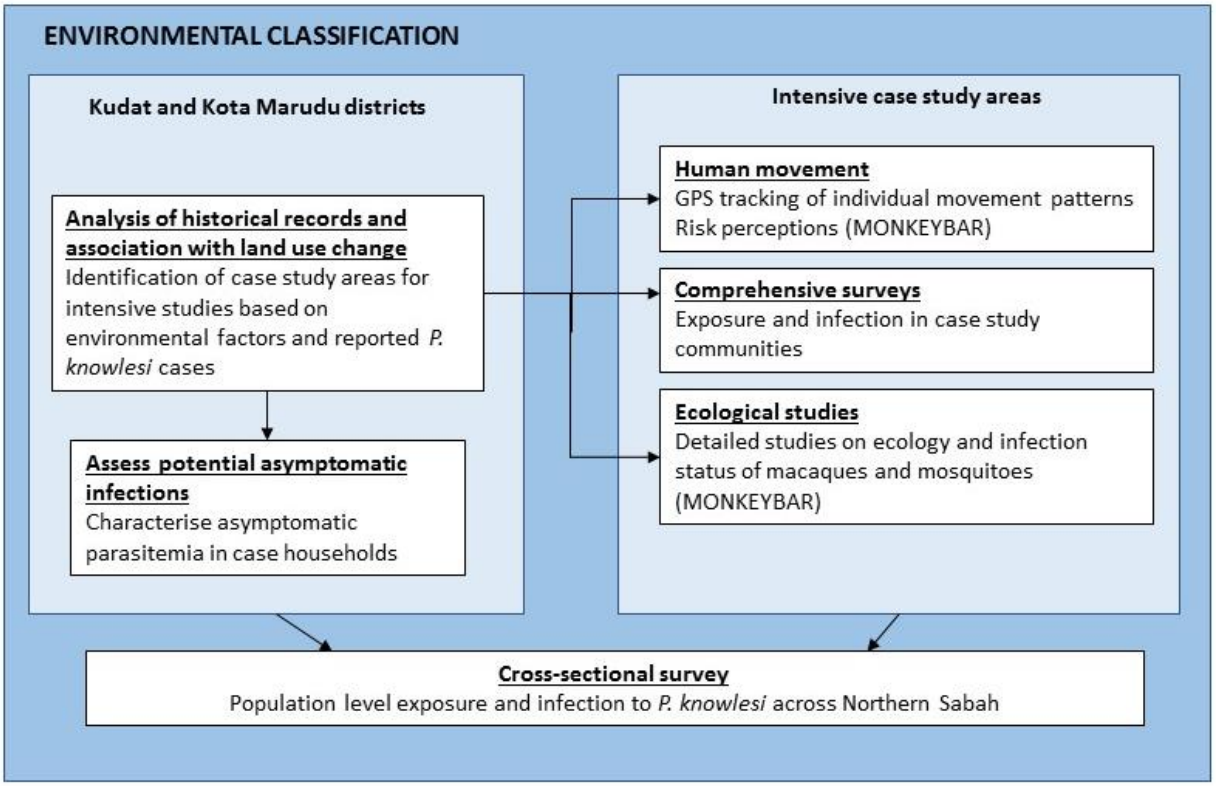
Due to high number of reported *P. knowlesi* cases, Kudat district was selected as the focus of a large multidisciplinary project on biomedical, environmental and social risk factors for human infection with zoonotic malaria (MONKEYBAR; http://malaria.lshtm.ac.uk/MONKEYBAR). From 2004 to 2014, the proportion of suspected *P. knowlesi* cases reported in Kudat district hospital increased from 2.6% (8/305) to 51.0% (185/363), with knowlesi comprising 57.8% (412/713) of all malaria cases reported from Kudat between 2013 and 2015 [13, 103].

Figure 3.2. Intensive study sites and forest cover in Kudat and Pulau Banggi; b. high resolution aerial image of area within intensive study site



Initially, activities to characterise fine scale patterns of exposure and infection will be conducted in two case study areas with on-going *P. knowlesi* transmission: a highly fragmented and largely deforested area in mainland Kudat and an area with remaining intact secondary forest undergoing large scale clearing in Pulau Banggi (Figure 3.2). These will include high resolution mapping of environmental changes, tracking of human movement patterns and community wide comprehensive surveys on exposure and infection to *P. knowlesi*. Additional activities will be conducted in Kudat and Kota Marudu districts to evaluate the presence of asymptomatic *P.* knowlesi carriage and historical patterns of *P. knowlesi* transmission. The results of these activities will be used to design more extensive activities as part of a cross sectional survey covering, Kudat, Kota Marudu, Pitas and Northern Ranau (Figure 3.3).

Figure 3.3. Diagram of activities and study sites



** MONKEYBAR activities conducted by wider research programme outside of this thesis

4 Methods of environmental classification

Human behaviour and contact with macaques, mosquitoes and the habitats these populations are associated with are likely to influence risk; characterising these requires spatially and temporally accurate data on the environment. Geospatial data can be acquired from different sources, such as satellite-based remote sensing, aerial surveys and ground-based Global Positioning System (GPS) mapping (e.g. [141-146]). Obtaining temporally accurate environmental data of sufficient resolution is particularly challenging in Malaysian Borneo, an area characterised by high cloud cover and rapid environmental change [53, 147]. To address these issues, a range of remote-sensing and ground-based methods were utilised during these studies. This chapter will discuss the data sources used and the methodological approaches to combining different data sources, including a published paper on the uses and limitations of unmanned aerial vehicles (UAV) for epidemiology and how data collected by UAV was combined with satellite-based remote sensing and field data collection.

4.1 Approach to land classification

Land use/ land cover change analysis techniques are used to monitor physical changes to the environment (land cover) and human management and modification of terrestrial surfaces (land use) [148]. These approaches frequently classify data obtained from air and space-borne sensors into discrete land cover classes relevant to the ecological processes of interest [134, 149, 150]. Supervised classification techniques utilise representative training data of pre-defined land cover classes to classify land cover and develop thematic maps while unsupervised classification techniques identify clusters based only on characteristics of imagery, which can later be grouped into land classes [143].

For this study, supervised classification techniques were utilised to identify land cover classes hypothesised to be important for parasite transmission. Although forest loss had been suggested as a key driver of *P. knowlesi* emergence, limited data were available on densities of macaques and mosquito vectors within degraded habitats and no spatial data were available on where individuals are infected. Because of this, land cover classes were identified iteratively. Initial exploratory analyses on the effects of deforestation were conducted using tree cover data classifying land cover as forest or other [51]. Fine-scale aerial mapping of study sites where entomological, primatology and human movement studies were conducted to identify potentially important habitat types; this was supplemented by questionnaire data on human movement and interactions with macaques

[151-153]. These data were used to refine the final land classification system used for extensive surveys across a wider ecological gradient in Northern Sabah. As mosquito densities and macaque roosting locations were observed to vary between different agriculture types [87, 152, 154], initial classification of areas as agriculture or forest was expanded to include specific plantation types. This was augmented by inclusion of land types found in the study area which were not present in the intensive study areas (e.g. intact forest, wet rice paddies, pulpwood plantations). A hierarchical classification system was used for final land classification, as described by Table 4.1.

Table 4.1: Land cover classification for extensive cross-sectional survey

Level 1 classification	Level 2 classification	
Forest	Intact closed canopy forest	Intact virgin forest, closed canopy (protected forest reserves), old growth secondary forest with over 90% canopy cover and area of over 0.5 ha [155]
	Secondary forest	Secondary forest, closed canopy cover*
	Mangrove forest	Mangroves
Cropland	Oil palm	Predominantly oil palm
	Rubber	Predominantly rubber trees
	Pulpwood	Predominantly pulpwood plantations
	Rice paddy	Wet rice paddy, irrigated fields
	Mixed agriculture and other crops	Other crops and gardens
Cleared	Shrubland, sparse vegetation	Cleared land or areas with limited vegetation consisting of shrubs, grasses and young forest, open canopy cover*
	Built environment	Roads, houses and other buildings
Water bodies	Water bodies	Oceans, rivers, lakes and other water bodies

* Canopy cover is defined as closed (more than 60% cover), open (10-60% cover) and sparse (1-10% cover) [156]

Land classification utilised a range of remote sensing and geographical information system datasets, characterised by varying spatial, temporal and spectral resolutions (**Error! Reference source not found.**). As freely available satellite data (such as Google Earth) was out of date and did not reflect the current land cover in study area, initial mapping was performed using a UAV (Paper 1). While this allowed the collection of highly spatially and temporally accurate data, a typical one hour flight could only cover a 120ha area, making coverage of the entire intensive study site (38,560ha) or larger extensive survey site unrealistic. Moreover, aerial data of the site could only provide an instantaneous view of land cover and could not be used to explore historical forest loss. Instead, a model-based approach was applied using non-contiguous aerial data and GPS data collected during field surveys as training data to classify satellite data and generate spatially contiguous land classification maps for multiple time periods and the larger study area. Additional training data was obtained from publicly available datasets on forest cover [157, 158]; due to the high rates of change in these areas, the accuracy of these datasets was verified by ground-truthing and consultations with the Sabah Forestry Department.

Table 4.2. Geospatial data used for mapping land cover

Parameter	Description	Resolution	Source
Elevation	Elevation (metres above sea level)	30 m	ASTER Global Digital Elevation Map [159]
Slope and aspect	Slope and aspect (degrees)	30 m	Calculated from ASTER Global Digital Elevation Map
Landsat 8 bands	Reflectance (Wavelength 0.45-2.29 µm)	30 m	Landsat8 reflectance data calculated using ENVI [160]
NDVI	Normalised differential vegetation index	30 m	Calculated from NIR and Red Landsat8 bands
Forest cover	Annual forest cover and forest loss (forest defined as over 50% canopy cover)	30 m	Derived from Landsat images, calculated by [51]
Intact forests	Undisturbed forest with minimum area of 50,000ha, minimum width of 10km, minimum corridor width of 2km	30 m	Obtained from Intact Forest Landscapes initiative, processed from Landsat [158]
Forest/ non-forest map	Forest/ non- forest map with forest classified as over 90% canopy cover with a minimum area of 0.5 ha	25 m	2015 global PALSAR-2/ PALSAR Forest/Non-Forest Map derived from Synthetic Aperture Radar (SAR) [161]
Global distribution of mangroves	Global mangrove distribution	30 m	Obtained from UNEP, processed from 2011 GLS and Landsat data [157]
Location of roads and houses	GPS coordinates, UTM (Zone 50 North)	5 m	Mapped during GPS field surveys
Distance to roads and houses	Distance from nearest road and house	5 m	Calculated from GPS coordinates
High resolution satellite imagery	Natural colour imagery of selected areas	6.5 m	RapidEye satellite data (imagery date July 2015) [162]
Aerial images	Aerial images of study site	0.1 m	Collected using Sensefly UAV as described by [151]
Surface models of study site	Digital surface models of study site	0.1 m	Calculated from UAV data using Postflight Terra 3D [151]

4.2 Aerial mapping

The justification and methodology for using UAVs for epidemiological studies is discussed in Paper 1. This paper outlines how UAVs were used for this study and the potential applications for the use of UAVs across other epidemiological contexts.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Kimberly Fornace
Principal Supervisor	Prof. Chris Drakeley
Thesis Title	Spatial epidemiology of P. knowlesi in Northern Sabah, Malaysia

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Trends in Parasitology		
When was the work published?	November 2014		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conducted the literature re[illegible]s, analysed all data and wrote the paper. All authors commented on the manuscript.

Student Signature: _____ **Date:** 9 August 2018

Supervisor Signature: _____ **Date:** 09/08/18

4.2.1 Mapping infectious disease landscapes: unmanned aerial vehicles and epidemiology (Paper 1)

Keywords: GIS, unmanned aerial vehicle (UAV), drone, spatial epidemiology, malaria, Plasmodium knowlesi

Authors:

Kimberly M Fornace[1], Chris J Drakeley[1], Timothy William[2,3,4], Fe Espino[5], Jonathan Cox[1]

1. Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK
2. Infectious Diseases Society Sabah- Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Sabah, Malaysia
3. Infectious Diseases Unit, Clinical Research Centre, Queen Elizabeth Hospital, Kota Kinabalu, Sabah, Malaysia
4. Sabah Department of Health, Kota Kinabalu, Sabah, Malaysia
5. Research Institute for Tropical Medicine, Department of Health, Filinvest, Alabang, Muntinlupa City, Philippines

Corresponding author:

Fornace, K. M. (Kimberly.Fornace@lshtm.ac.uk)

Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine

Keppel Street, London, WC1E 7HT, United Kingdom

4.2.2 Abstract

The potential applications of UAVs (unmanned aerial vehicles, or drones) have generated intense interest across many fields. UAVs offer the potential to collect detailed spatial information in real time at relatively low cost and are being used increasingly in conservation and ecological research. Within infectious disease epidemiology and public health research UAVs can provide spatially and temporally accurate data critical to understanding the linkages between disease transmission and environmental factors. Using UAVs avoids many of the limitations associated with satellite data (e.g. long repeat times, cloud contamination, low spatial resolution). However, the practicalities of using UAVs for field research limit their use to specific applications and settings. UAVs fill a niche but do not replace existing remote sensing methods.

4.2.3 Applications of unmanned aerial vehicles

Increasing attention has been focused on the potential usages of unmanned aerial vehicles (UAVs, or drones). UAVs have been used for a variety of civilian purposes ranging from law enforcement, fire fighting and parcel delivery to wildlife population monitoring [146, 163, 164]. UAVs offer the potential to collect detailed geospatial information in real time at relatively low cost. UAVs can also be an effective method of monitoring situations too dangerous or costly for traditional aerial surveys, such as for mapping forest fires and ice floes in the Arctic or conducting anti-poaching patrols [165, 166]. These advantages have led to the application of UAVs for ecological research studies evaluating land use and cover change and conducting aerial surveys of large wild animals, such as dugongs, rhinoceros and orang-utans [166-170]. Additionally, UAVs have been used in agriculture to monitor vegetation levels, crop growth and distribution of weeds [171, 172].

There are also numerous potential applications for UAVs in public health. UAVs can be used to locate people and monitor human population movements of nomadic and migrant groups to allow targeting of surveillance and public health interventions [173]. UAVs have also been used to facilitate access and sample collection from remote locations. For example, a UAV was developed to allow the transportation of test samples from remote rural clinics to national laboratories in South Africa [174]. UAVs can also be used for disaster management and emergency relief operations to monitor situations as well as to deliver medical supplies to inaccessible or dangerous locations. During the aftermath of the Typhoon Haiyan in the Philippines, UAVs were used by aid organisations to assess the extent of the typhoon damage and plan relief measures and reconstruction [175]. Aid

organisations have also started piloting the use of UAVs to deliver medical supplies to areas inaccessible by road in Haiti, the Dominican Republic and Lesotho [176].

UAVs can also be used to collect other types of environmental data of public health relevance. Environmental factors such as radiation and air pollution vary spatially with important consequences for human health. Monitoring equipment has been fitted to UAVs to measure levels of environmental toxins and pollutants [177, 178]. Further applications could include mapping health infrastructure, such as water and sanitation systems and locations of health facilities.

Within infectious disease epidemiology, UAVs provide a new alternative to collect detailed geo-referenced information on environmental and other spatial variables influencing the transmission of infectious diseases. Land use change, for example through deforestation or agricultural expansion, has been widely documented as a major driver of infectious disease emergence and spread [179-183]. Anthropogenic environmental changes can modify the transmission of zoonotic and vector-borne diseases by disrupting existing ecosystems and altering the geographic spread of human populations, animal reservoirs and vector species [184, 185]. For example, the emergence of malaria in new areas of South America and Southeast Asia has been associated with the clearing of tropical forests resulting in changes in anopheline mosquito densities and contact with people [186]. Changes in forest cover affect the life cycle and distribution of disease vectors by altering microclimates, availability of breeding sites and ecological community structures [187]. Simultaneously, deforestation is associated with higher levels of human activity within forest environments, leading to increased exposure to forest-breeding vectors [188]. Understanding rapidly changing patterns of human settlements and vector distributions in this context is vital to predicting disease risks and effectively targeting disease control measures.

4.2.4 Satellite data versus aerial data

Epidemiologists rely on accurate spatial and environmental data to describe variations in vector-borne and zoonotic disease risk, establish early warning systems, model disease transmission and estimate disease burden [189]. These data can include detailed information on land cover, climatic variables and distributions of human and animal populations. Geospatial data can be obtained from a range of sources, such as satellite-based remote sensing, aerial surveys and ground-based GPS surveys.

Satellite remote sensing is increasingly being used to obtain environmental data on land cover, vegetation, soil type, surface water and rainfall for infectious disease research [141]. Satellite data are characterised by varying spatial, temporal and spectral resolutions. Temporal resolution relates to the frequency with which a satellite returns to a specific location, while spectral resolution is defined by the wavelength interval size on the electromagnetic spectrum and the number of intervals the measured by the satellite's sensor. Higher spectral resolution allows image classification or transformation (such as for vegetation indices) using information beyond the visible range of the electromagnetic spectrum. A new generation of sensors such as QuickBird, IKONOS and GeoEye (http://www.digitalglobe.com) provide imagery with very high spatial resolution (<1 m) but are limited by relatively low temporal and spectral resolutions [190]. Cloud cover, a frequent issue in tropical areas, may also limit the usefulness of the data, particularly if an area is visited infrequently [191]. Additionally, obtaining high-resolution data can be prohibitively expensive. If data are needed for specific time points, this may require paying a premium to specifically task sensors to collect data for defined areas of interest. Effective analysis and application of satellite data also requires suitably trained personnel as well as specialised software.

More accessible to most public health programmes, satellite data from sensors such as Landsat (http://landsat.gsfc.nasa.gov/) and Moderate Resolution Imaging Spectroradiometer (MODIS; http://modis.gsfc.nasa.gov/) are freely available in the public domain. These sensors produce very detailed spectral data but, compared to sensors such as QuickBird, have much coarser spatial resolutions (15-60 m for Landsat, 250-1000 m for MODIS, depending on wavelength). Data from these sensors are well suited to studies of infectious diseases or disease vectors at regional, national or sub-national level that incorporate information on either land cover or other environmental variables, e.g. [192-200]. They are less well suited, however, to studies that require either very detailed environmental mapping or frequent monitoring of land use and habitat.
Alternatively, due to these limitations, many ecological studies rely on the use of aerial surveys conducted by light aircraft to monitor land cover and conduct wildlife population estimates. Aerial surveys are a standard method of estimating population sizes of large animals and can also be used to collect aerial photography for habitat assessments [201-203]. Aerial surveys can also use light detection and ranging (LiDAR) systems, a technology which measures distance by reflected light from targets illuminated by lasers, to create high resolution maps of land cover and measure canopy heights [204, 205]. Studies may also utilise ground-based GPS surveys to map the distribution of human settlements and wildlife populations. In order to identify malaria cases and evaluate risk factors in forested areas of Vietnam, GPS surveys were conducted to identify locations of villages

and nomadic groups[206]. Both aerial and ground-based surveys can provide highly accurate information but are extremely resource intensive and may not always be feasible or affordable.

The use of UAVs can supplement other remote sensing data used for infectious disease epidemiology. UAVs allow the mapping of small geographical areas at user defined time points and spatial resolutions. UAVs can be used to produce high-resolution aerial photography as well as collect data on other variables such as elevation. Epidemiologists can respond quickly to changing disease reports to map areas immediately and as frequently as required. However, despite these advantages, the use of UAVs is not always an appropriate technology.

4.2.5 Case study: mapping environmental risk factors for zoonotic malaria

Between December 2013 and May 2014, we conducted 158 flights with a UAV to collect data for an epidemiological study. The flights were carried out in two study sites in Sabah, Malaysia and one site in Palawan, the Philippines. These activities were completed as part of a larger multi-disciplinary study to characterise biomedical, environmental and social risk factors for human infection with the zoonotic malaria *Plasmodium knowlesi* (http://malaria.lshtm.ac.uk/research/projects/malaria-research-epidemiology-20). Maintained by long- and pig-tailed macaques, *P. knowlesi* is an emerging pathogen likely to be affected by deforestation and changing patterns of land use resulting in increased contact between people, mosquito vectors and wildlife reservoirs [207, 208]. The study requires detailed spatial information to integrate human and macaque movement and vector bionomics to understand the epidemiology of infection.

Figure 4.1. Use of the Sensefly eBee to map land cover in Malaysia. A) Setting up the Sensefly eBee before a flight B) Launching the Sensefly eBee

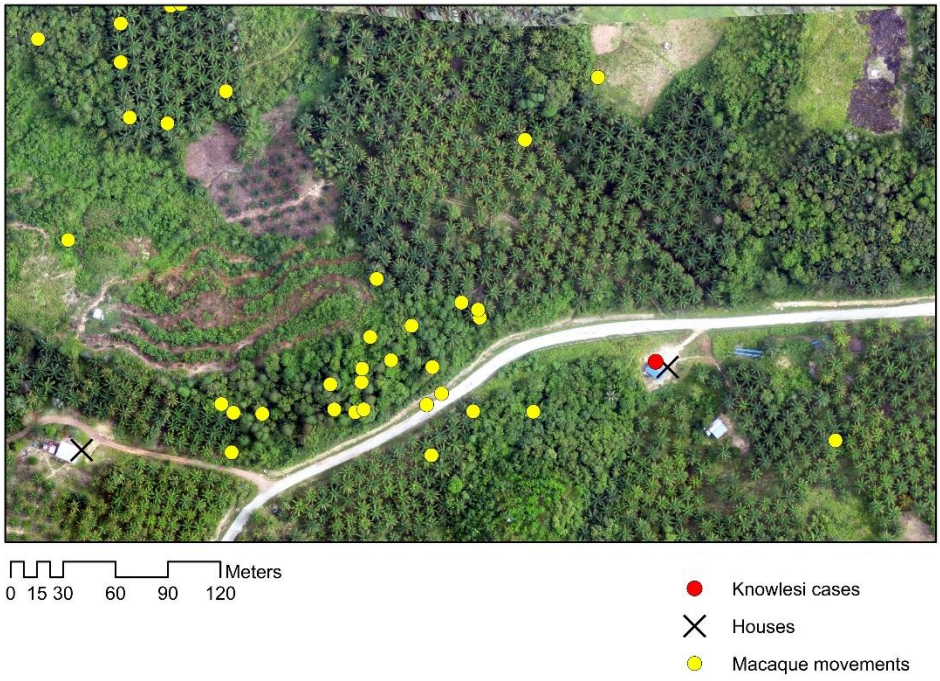
The commercially available Sensefly eBee UAV was used for all mapping exercises (Sensefly, Cheseaux-Lausanne, Switzerland; Figure 4.1). The eBee can fly up to 50 minutes and uses a 16 megapixel digital camera to record aerial imagery which can be used to produce maps and digital surface models. All UAV flight plans were programmed and monitored using the eMotion2 software (Sensefly, Cheseaux-Lausanne, Switzerland) and post flight image processing was completed using Postflight Terra 3D (Pix4D SA, Lausanne, Switzerland). ArcGIS (ESRI, Redlands, USA) was used for data analysis and generation of three-dimensional models (Figure 4.2). Previews of aerial photography and digital surface models were generated in real time while full data processing took several hours.

Figure 4.2. Three-dimensional model of study site in Sabah, Malaysia

Within the study sites, the eBee was flown at approximately 350 to 400 meters above the take-off point. Publicly available digital elevation data from ASTER Global Digital Elevation Model (ASTER GDEM) was used to develop flight plans [209]. Out of 158 flights, 127 (80%) flights generated usable data. The most frequent reasons for failed flights were high winds, rain and battery failure. Out of these flights, six flights (5%) were obscured by low clouds and needed to be repeated. The average area covered by a single flight was 124 hectares or 1.24 km^2 with an image overlap between 80-90% and average resolution of 11.22 cm per pixel. Mapping exercises were conducted on 26 days between December 2013 and May 2014, with repeated flights over areas identified as having high rates of land use change (Figure 4.3). The resulting maps were overlaid with GPS data on locations of households and malaria cases and used to characterise land use types and create a spatial sampling frame for further sampling (Figure 4.4)[210].

Figure 4.3. Mapping changes to land cover in study site in Sabah, Malaysia. A) Study site in February 2014. B) Same study site in May 2014 after the start of clearing to create a rubber plantation.

Figure 4.4. Macaque movements around human *Plasmodium knowlesi* case household



Benefits of UAV mapping

For this project spatially and temporally detailed data on the dynamics of land use and land cover are required in order to explore interactions between environmental factors, disease vectors, and human and primate hosts in the light of increasing disease transmission. As is commonly the case in tropical settings, it proved impossible to obtain recent cloud-free satellite data for our field sites via the archives of satellite data providers. Imagery on Google Earth (http://www.earth.google.com), which uses data from the same archives, was out of date and was inadequate for characterising the

study site (for example, large-scale clearings at one study site appeared as intact forest in Google Earth imagery). The paucity of available satellite data and lack of certainty in the ability to obtain data for key time points moving forward were the principal reasons for electing to carry out land cover/land use mapping using a UAV.

Data from UAVs share many of the characteristics of high resolution satellite data, although in the case of UAVs the user has more control over the spatial resolution of the resulting imagery (depending on flight altitude, imagery from UAVs typically has a spatial resolution of 4-20 cm; currently the highest spatial resolution available from commercial high resolution satellite sensors is 41 cm). As with satellite data, UAVs can produce "stereo" imagery which, using standard photogrammetric tools, can be used for Digital Elevation Model (DEM) generation, 3D visualization, and feature extraction.

One of the main benefits of using UAVs is the ability to obtain data in real time and to repeatedly map areas of interest as frequently as required. In one of our sites in Sabah, Malaysia, development began on clearing secondary forest to establish a rubber plantation. As the clearing occurred within a limited geographical area, progress of the clearing and the resulting land changes could be mapped quickly and updated routinely. This ability to map changes as they occur is critical for understanding how land use change affects the distribution of human populations and disease vectors.

Depending on the size of the area to be covered and the resolution of data needed, the costs of purchasing and operating a UAV can compare favourably with purchasing high resolution satellite data over repeated time points. A wide range of UAV models are available, with low cost options like the Conservation Drone available for several hundred dollars to high-end specialised drones costing hundred thousands of dollars [167]. For our purposes, we chose a commercially available fixed wing UAV costing approximately USD 25,000 for the UAV and associated software. While there are less expensive models of UAVs available, this UAV could be easily used without significant training or technical knowledge, allowing multiple members of the project team to be trained on operating the UAV. Various models of UAVs are commercially available with different specifications. The choice of a UAV model should depend on the financial and technical resources available, anticipated spatial and temporal scales of the mapping project and types of data required.

Limitations of UAVs

Although UAVs represent a new source of data for epidemiological investigations, there are still significant potential limitations in their use. Similar to light aircraft, small UAVs cannot fly in all weather conditions. The ability to withstand certain weather conditions is determined by the size and specifications of the UAV used. The model we used could not be used during rain or with winds over 45 km per hour (12 m per second). We also found that high temperatures at our study sites (frequently in excess of 35 °C) could cause the UAV to overheat after multiple continuous flights. While not as much of an issue as with satellite data, low cloud cover can also limit the visibility of data collected at certain times of day or in areas with poor visibility. The variability of weather conditions can make it difficult to plan exact flight times ahead of time and even when conditions appear suitable, areas frequently need to be remapped to get sufficient data. Some land types, such as forest, are more difficult to map due to the difficulty matching overlapping images and may need to be mapped repeatedly or at higher resolutions.

Additionally, mapping exercises using a UAV require adequate resourcing. While small areas can be mapped quickly, mapping larger areas can require significant amounts of field personnel time. The amount of time needed to map an area is highly dependent on local weather conditions and the image resolution required. If higher resolutions of data are needed, UAVs need to be flown at lower heights and can cover shorter distances per flight. The number of flights conducted per day may also be constrained by the availability of electricity and ability to recharge UAV batteries. High levels of usage can necessitate the purchase of additional equipment and spare parts as well as lead to higher maintenance costs. Further, processing and analysis of UAV data can be computationally intensive, requiring computers with high specifications that may not always be available in the field. While data can be processed at later dates, the immediate processing of collected images allows rapid assessment of data quality and better planning of further fieldwork. UAVs have also been limited by the lack of multispectral data although this recently UAVs have been modified to record other data of interest; for example, UAVs have been fitted with near infrared (NIR) cameras to measure biomass of forest areas [211, 212]. Currently, the spectral resolution of most UAVs is limited in comparison to available satellite data; however, this is a rapidly developing technology and may change in the near future.

Challenges can also be encountered while applying for official permissions for carrying out UAV mapping. As the use of UAVs is relatively uncommon, there is frequently no clear regulatory

framework for applying for permissions. For our research, we were required to apply for permissions from multiple agencies, ranging from Ministries of Defence, Civil Aviation Authorities, conservation and development councils and land use planning authorities. While guidelines are in place for traditional aerial surveys, these guidelines were not always appropriate for relatively short, low altitude UAV flights. For example, some regulations required the submission of detailed flight plans to allow redirection of other aircraft within the area despite the differences in flight heights between a small UAV and larger airplanes. It is also worth noting that insurance associated with the use of UAVs is potentially restrictive. Although we encountered no safety issues with using a UAV, all project staff needed to be instructed on safe handling of equipment.

4.2.6 Concluding remarks and future perspectives

Detailed investigations of environmental factors influencing the transmission of infectious diseases are vital to effectively target surveillance and control programmes. UAVs present a new opportunity to obtain high-resolution geo-referenced data in real time. These data can be used to better understand how land use changes affect the emergence and spread of infectious diseases by monitoring distribution of human populations and changes to the habitats of disease vectors and wildlife reservoirs. We demonstrated the utility of this method by using a UAV to obtain environmental data for an epidemiological investigation on risk factors for zoonotic malaria.

The use of UAVs is most appropriate when detailed maps of relatively small geographical areas are needed in areas where high-resolution satellite data is not readily available. UAVs may not be appropriate for large-scale data collection due to the time and resources required to operate them. Also, despite the modification of some UAVs to record data at different wavelengths, UAVs do not have the spectral resolution of most satellite data. Within smaller areas, UAVs can be used to generate high resolution data on land cover, vegetation and elevation and can be used to monitor changes in habitats for vectors and wildlife reservoirs on a fine spatial scale. Additionally, UAVs can provide a valuable alternative to other data sources when data are needed either in real time or at very frequent time points.

4.2.7 Acknowledgements

We would like to thank Gaim James Lunkapis (Universiti Sabah Malaysia), Tommy Rowel Abidin (Infectious Disease Society Sabah), Albert M Lim (Infectious Disease Society Sabah) and Judy Dorothy Marcos (Research Institute for Tropical Medicine) for their help with UAV mapping exercises.

Additional thanks to Beth Downe (LSHTM) and the teams in Sabah, Malaysia and Palawan, the Philippines for their support with this project. We acknowledge the Medical Research Council, Natural Environment Research Council, Economic and Social Research Council, and Biotechnology and Biosciences Research Council for the funding received for this project through the Environmental and Social Ecology of Human Infectious Diseases Initiative (ESEI). Grant Number: G1100796

4.3 Training and validation data generation

In addition to mapping intensively studied areas repeatedly, UAV imagery was used to generate training data for classification of LANDSAT data. Supervised classification requires collection of training and validation data to define land cover classes; the accuracy of this classification can be substantially influenced by the amount and representativeness of training data as well as the number of classes defined and dimensionality of the data used [213, 214]. Training samples should be representative of the classes and of sufficient quantity to account for variability across the study site [215]. Classification accuracy has been shown to improve with larger training data sizes [216, 217], with required sample sizes determined by data dimensionality and class numbers. Ideally, training data would be collected using a randomly distributed sampling strategy [214, 216]; however, depending on the terrain and method of collecting data, this may not always be feasible. Stratified sampling methods have also been used to ensure rare land classes are included; these have been shown to improve classification accuracy where randomly selected data may under-represent or exclude rare land types [218].

Frequently used guidelines suggest the use of a minimum of $30p$ cases per class for training, where p is the number of variables (bands) used for classification [143]. As the lack of sufficiently representative training data for all classes has been shown to negatively impact classification, a stratified approach to training data collection was used, collecting approximately equal numbers (minimum 240) of cases for each class [143, 215, 218-220]. The study area was defined as land area within 5 km of selected kampungs and a subset of randomly selected kampungs stratified by forest cover were mapped by UAV. Additional areas with land cover classes not captured by initial mapping exercises were identified by local knowledge and mapped. In total, 177 usable drone flights were completed across all four districts with over 200km^2 of aerial imagery generated. Although training data would have ideally been randomly selected, this was not possible due to logistics and difficult terrain. Instead, areas representative of specific land classes across the study area were identified from aerial imagery and manually digitised.

Due to the difficulties accessing forested and mangrove areas by UAV (insufficient landing areas and high winds), additional data on the extent of undisturbed forests and mangrove forests was obtained from the ALOS-PALSAR Forest-Non-Forest Maps, Intact Forest Landscapes project and United Nations Environment Programme [155, 157, 158]. To cross-validate these data, we obtained three high resolution natural colour RapidEye satellite images acquired in July 2015 and manually identified representative training data from this imagery [162]. Additionally, we obtained data on

the extent of industrial pulpwood plantations (primarily *Acacia* species) from Gaveau et. al [53]. As small-scale pulpwood plantations are not present in this region and the data on industrial plantations could be verified by local forestry officials, we masked these areas from the data to be classified and used these spatial boundaries for the final thematic map. Training data for all other classes was rasterised to 30m resolution and values extracted. From this data, we sampled points a minimum of 60m apart and roughly proportional to the expected proportion of land types to maximise classification accuracy [214]. The final data set included 70,648 points with 55,648 points used as training data and 15,000 points withheld for independent validation. Final categories for classification included forest, mangrove, rubber, mixed agriculture, oil palm, irrigated fields, built environment and shrubland.

4.4 Random forest classification

Land cover maps were derived using a random forest classifier, an ensemble classifier creating multiple decision trees using randomly selected subsets of training samples [221]. This approach is widely used in remote sensing due to the ability to handle large datasets with high levels of collinearity [222] and has been utilised for land cover classification from multispectral, hyperspectral and radar data (e.g. [223-227]). As well, random forests have been used to integrate data from multiple sources, such as the inclusion of elevation and soil data with satellite imagery to improve classification accuracy [225, 228]. These models can also be run iteratively, with least predictive variables excluded at each run [228, 229]. For this thesis, all analyses were completed in R using the randomForest and RRF packages using LANDSAT data with elevation, slope and aspect data derived from ASTER digital elevation models and infrastructure data obtained from aerial and ground-based GPS surveys (**Error! Reference source not found.**) [160, 209, 229]. For the land use classification for the extensive study site, a cloud free composite LANDSAT image for 2015 was obtained from [51]. Water bodies were masked using a water mask derived from [161]. All data were resampled to 30m per pixel.

A regularised random forest was used for feature selection. Within this regularization framework, feature selection is penalised if the new feature does not add substantially new predictive information, limiting model overfitting and increasing performance accuracy [230]. For selected features, the model was tuned to determine the optimum number of variables per split (mtry) and analyses were run with high numbers of decision trees (over 1000) to ensure stability. The final classification was derived by averaging the class probabilities from all decision trees [221]. Trees were grown with different bootstrapped samples of two-thirds of the training data, with the

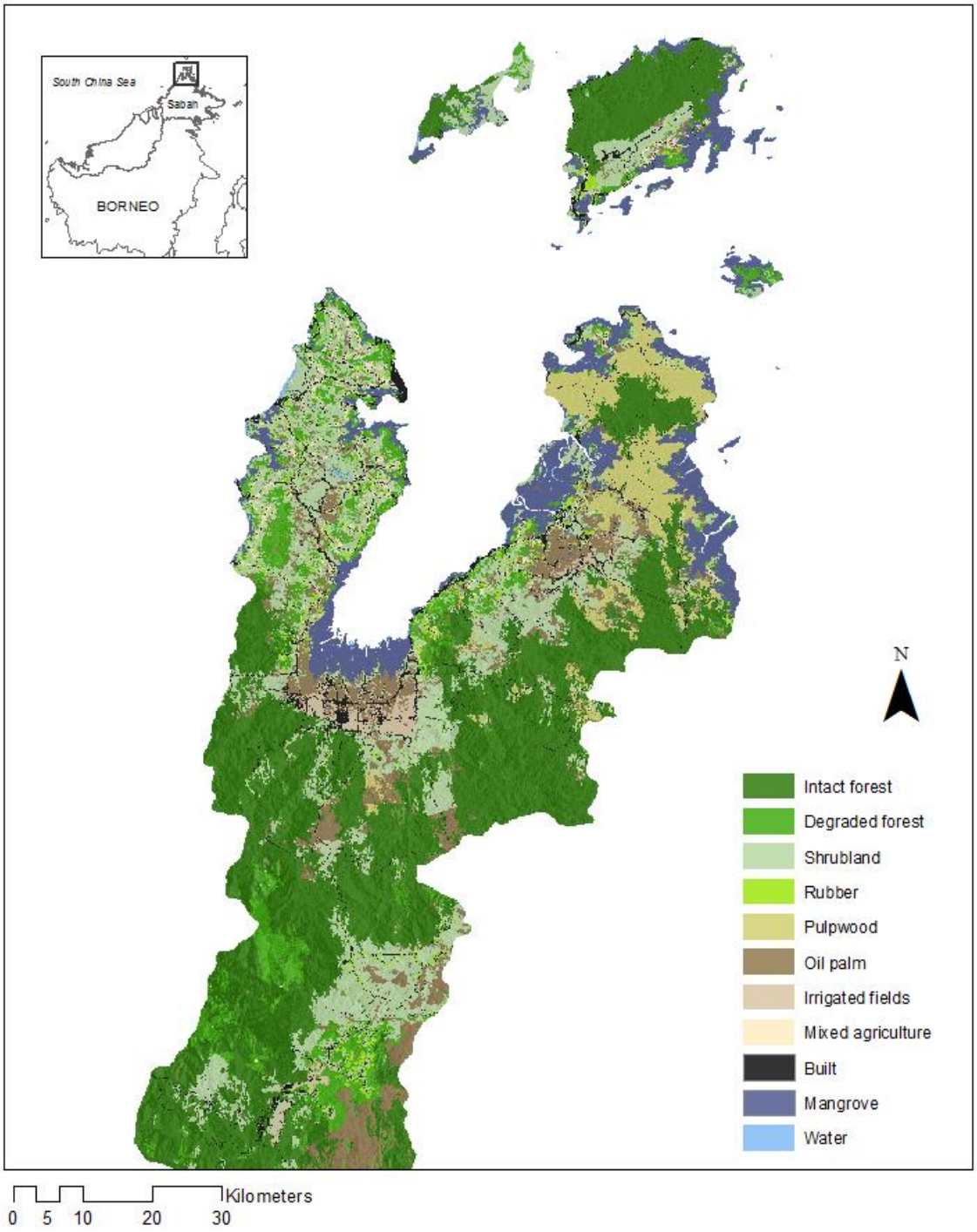
remaining third of the data used in an internal cross validation procedure to derive an “out-of-bag” (OOB) error [222]. Resulting predictions were exported as a 30m resolution raster file.

4.5 Post-classification processing and accuracy assessment

As the data was classified on a per pixel basis, post-classification processing was completed to reduce the classification noise. Due to the complexity of biophysical environments (i.e. a single pixel containing multiple habitat or vegetation types), the resulting classification contained misclassified isolated pixels, creating a “salt and pepper” appearance [213]. A post-classification workflow was implemented in ArcGIS: first, the Majority filter tool was applied to remove isolated pixels, next, class boundaries were smoothed using the Boundary Clean tool, and finally, small isolated regions (less than 90m x 90m) were generalised to the nearest class. As the incorporation of ancillary GIS data can increase classification accuracy, mapped road networks and locations of pulpwood plantations were rasterised and merged with the classified data [231]. Additionally, data classified as forest cover was divided into two sub-classes (disturbed and intact forest) based on spatial overlap with JAXA forest maps [161]. These maps were derived from 2015 synthetic aperture radar (SAR) data capable of penetrating cloud cover and have a reported overall accuracy of 95% [155].

The initial OOB error for random forest classification was estimated at 3.68%. To assess the classification accuracy of the final processed land use classification, classified values were extracted for the withheld validation points. Final classification accuracy was highly accurate (Kappa score: 0.948) (Figure 4.5).

Figure 4.5. Final land use classification for extensive study site in Northern Sabah, Malaysia



4.6 Spatial scale and configuration

The resulting land cover maps were analysed to identify associations between disease risks and environmental factors. These associations are modulated by the spatial scale of analysis; ecological processes affecting the distribution of people, disease vectors and wildlife hosts may occur at highly local to larger regional scales. For example, *P. knowlesi* transmission depends on the scale of human

movements, macaque home ranges as well as local environmental conditions influencing vector breeding sites [75]. For other vector-borne zoonotic diseases, variations in host richness and ecological community structure have been shown to be important at a fine spatial scale while changes in climate and other abiotic factors are more important across larger scales [232]. Data-driven approaches can be used to explore the strength of associations between environmental variables at different spatial scales and disease risks [233, 234]. These have been applied to estimated landscape characteristics at different buffer radii around sites, identifying the distance with the best explanatory power [233]. This approach has been expanded to include landscape variables at multiple spatial scales [235] and demonstrated inclusion of environmental predictors at different spatial scales can improve model fit [234].

In addition to the proportion of different land types at varying spatial scales, increasing evidence illustrates landscape configuration can also modify disease transmission [184]. Fragmentation occurs when habitats become spatially separated due to land changes, creating a mosaic of land types. This increases the area of edges between different habitats and create ecotones, where previously separated populations are brought into closer contact and can increase disease incidence. For example, primate populations in highly fragmented forest areas have higher parasite infection levels, likely due to closer contact with other non-human primate groups and people [93, 97, 236]. Additionally, fragmentation may result in small patches with increased crowding and higher contact rates between hosts and populations. The increased density of birds within smaller forest patches has been associated with higher levels of tick infestation, likely due to the higher probability of a tick finding a host [237]. These changes to landscape also lead to changes in microclimates within fragments and have been shown to impact mosquito densities [238, 239].

Numerous metrics have been used to describe levels of fragmentation, measuring the patch density, edge lengths and ratio between edge and interior areas [240, 241]. These analyses can also be applied at varying spatial scales to explore associations between landscape configuration at different distances and disease risks [234]. Within this thesis, proportions and fragmentation of forest and other land types were calculated at different buffer radii surrounding houses and kampungs with identified *P. knowlesi* infection and exposure; final models included the landscape factors explaining the most variation in disease risks. This analysis was expanded to evaluate different land classes and data sets, from the impact of forest around villages in reported incidence to more complex landscapes around households included in cross sectional surveys.

5 Historical patterns of *P. knowlesi* incidence and association with landscape factors

5.1 Background and implications for future research

Although the emergence of *P. knowlesi* in Sabah had been widely hypothesised to be driven by deforestation, no studies had examined associations between disease and environmental factors. Similarly, fine scale spatial patterns of *P. knowlesi* had not been described, with previously published reports aggregating data by district level [12, 13]. This study aimed to explore the associations between village level incidence of *P. knowlesi* and forest cover, configuration and loss as classified by Hansen et. al [51]. Specifically, this study examined the importance of spatial scale by evaluating associations between incidence and forest cover within a range of buffer areas around village centroids, results of which were then inform the design of the extensive cross-sectional survey and generate future hypotheses on the importance of specific environmental characteristics.

During the initial development of this model, multiple approaches were trialled, including fitting models with zero inflation and without including spatial autocorrelation. As variograms did not identify any clear patterns of spatial autocorrelation and, Moran's I, although significant, was close to 0, the initial model included a random effect for village but did not include spatial autocorrelation. In response to reviewer comments, we expanded this model to include spatial autocorrelation, although the results were not substantially different. Reviewers additionally suggested the use of higher resolution data on NDVI as differences in vegetation levels surrounding households have previously been correlated with malaria risks in Southeast Asia. We did explore the possibility of including this data; however, after systematically reviewing 198 available Landsat images for this period, 93% were significantly affected by striping and 55% were affected by cloud cover. Additionally, as only village centroids were available, it was not possible to examine household level risks or effects of fine-scale land use changes.

Inherent limitations were present within this study, such as the reliance on hospital-based reporting of microscopy results and the limited environmental data used. Hansen et. al classify forest based on tree canopy cover, with areas of over 50% canopy cover classified as forest and the disturbance or removal of tree canopy cover to levels below 25% as forest loss [51]. This definition substantially misclassifies agroforestry plantations such as palm oil and rubber common in tropical regions,

including Malaysian Borneo [242]. Substantial differences in vectoral capacity between different forest and plantation types have previously been reported within Sabah, suggesting these distinctions may be important for *P. knowlesi* transmission [56, 86, 152]. Subsequent studies were designed to address these limitations: assessing the importance of asymptomatic infections and the distribution within the community and further refining land classification definitions used.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Kimberly Fornace
Principal Supervisor	Prof Chris Drakeley
Thesis Title	Spatial epidemiology of P. knowlesi in Northern Sabah, Malaysia

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Emerging Infectious Diseases		
When was the work published?	February 2016		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the experiment, performed all analysis and wrote the paper. Other authors assisted with collecting data and suggesting analysis methods.

Student Signature: _____ **Date:** 9 August 2018

Supervisor Signature: _____ **Date:** 09/08/18

5.2 Association between landscape factors and spatial patterns of *Plasmodium knowlesi* infections in Sabah, Malaysia (Paper 2)

Authors: Kimberly M Fornace[1], Tommy Rowel Abidin[2], Neal Alexander[1], Paddy Brock[3], Matthew J Grigg[2, 4], Amanda Murphy[4], Timothy William[2, 5, 6], Jayaram Menon[5], Chris J Drakeley[1], Jonathan Cox[1]

Keywords: *Plasmodium knowlesi*, deforestation, malaria

Affliations:

1. London School of Hygiene and Tropical Medicine, London, UK
2. Infectious Diseases Society Sabah- Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Sabah, Malaysia
3. University of Glasgow, Glasgow, UK
4. Menzies School of Health Research, Darwin, Australia
5. Clinical Research Centre, Sabah Department of Health, Kota Kinabalu, Sabah, Malaysia
6. Jesselton Medical Centre, Kota Kinabalu, Sabah, Malaysia

5.2.1 Abstract

The zoonotic malaria species, *Plasmodium knowlesi*, is now the main cause of human malaria in Malaysian Borneo. It has been hypothesized that deforestation and associated environmental and population changes are the main drivers of this apparent emergence. We assembled village-level data for *P. knowlesi* incidence for two districts, Kudat and Kota Marudu, in Sabah, Malaysia for 2008-2012. Malaria records from routine reporting systems were adjusted to reflect the diagnostic uncertainty of microscopy for *P. knowlesi*. Negative binomial spatial auto-regressive models were developed to assess potential associations between *P. knowlesi* incidence and environmental variables derived from remotely-sensed data. Marked spatial heterogeneity in *P. knowlesi* incidence was observed and village-level counts of *P. knowlesi* cases were positively associated with both forest cover and historical forest loss in surrounding areas. These results suggest that deforestation and associated environmental changes are likely to represent key drivers for *P. knowlesi* transmission in these areas.

5.2.2 Introduction

Since the initial description of a large cluster of human infections with the zoonotic malaria *Plasmodium knowlesi* in Malaysian Borneo in 2004, increasing numbers of *P. knowlesi* cases have been identified throughout Southeast Asia [9, 11]. Although most *P. knowlesi* cases respond to treatment, infection can cause severe and fatal disease [14]. Understanding the distribution of *P. knowlesi* and risk factors associated with this disease is critical for designing appropriate public health interventions.

Maintained by long and pig-tailed macaques (*Macaca fascicularis* and *M. nemestrina*), the geographical range of *P. knowlesi* is limited by the distribution of the mosquito vectors and simian hosts [9]. Within this range, the risk of *P. knowlesi* human infection is highly variable. While sporadic *P. knowlesi* cases have been reported in a number of Southeast Asian countries, *P. knowlesi* is now the most common cause of human malaria in Malaysian Borneo [9, 11]. In the Malaysian state of Sabah suspected *P. knowlesi* notifications, as a proportion of all malaria notifications, increased from 2% (59/2741) in 2004 to 62% (996/1606) in 2013 [12, 13].

Molecular studies indicate that zoonotic *P. knowlesi* is not a newly emergent malaria species and is likely to predate human settlement in Southeast Asia [15]. *P. knowlesi* was first described in

macaques in the 1930s and the first naturally acquired human case was reported in 1965 in peninsular Malaysia [21, 22]. However, the true burden of *P. knowlesi* remains poorly understood due to frequent misidentification as other human malaria species by microscopy and the limited availability of *P. knowlesi*- specific molecular diagnostic capabilities. *P. knowlesi* appears microscopically similar to the human malaria species *P. malariae* but can be misdiagnosed as *P. falciparum* or *P. vivax* [14]. While it is difficult to determine the potential contribution of improved detection to recent increases in human case numbers, the rise in *P. knowlesi* relative to other malaria species is strongly suggestive of increasing *P. knowlesi* transmission in Malaysian Borneo [12, 13].

Land use changes, such as deforestation and agricultural expansion, have been proposed as the main drivers of this apparent emergence [14]. Deforestation and related forest activities have been associated with changes in malaria vector populations and associated disease incidence globally [6]. Changes in vegetation, microclimate and soil composition can affect the species composition and abundance of mosquito populations [8]. In Malaysia, previous studies have implicated the primarily exophagic *Anopheles leucosphyrus* group of mosquitoes as the main vector of *P. knowlesi* and found relatively high biting rates in farm edge and forest areas [80, 81, 84, 243] . Human disturbed environments have been associated with changes in primate behaviour and increased contact with people [93]. Fragmentation of existing habitats can also increase the frequency of disease transmission by creating transition areas with increased spatial overlap between human, mosquito and wildlife populations or by altering vector ecology [239, 244]. These edge effects have been described for malaria and other vector-borne zoonotic diseases (e.g. [2, 245]) but have not been evaluated for *P. knowlesi*. A previous mathematical modelling study also highlighted the potential for increased transmission resulting from increased spatial overlap between people, macaques and mosquitoes at forest edges [75].

Despite these apparent linkages between land use and *P. knowlesi* transmission, detailed environmental risk factors for human infections with *P. knowlesi* are unknown. Although variability of *P. knowlesi* risk has been reported at a regional scale, patterns of *P. knowlesi* transmission have not been described at a sub-district spatial scale [9, 104]. Further, to our knowledge, no studies have formally evaluated associations between *P. knowlesi* and environment and landscape characteristics. This study examined the changing incidence of *P. knowlesi* in Kudat and Kota Marudu districts in North-western Sabah, an area of relatively high *P. knowlesi* transmission [37]. The aim was to describe the spatial and temporal patterns of *P. knowlesi* incidence within these districts and explore

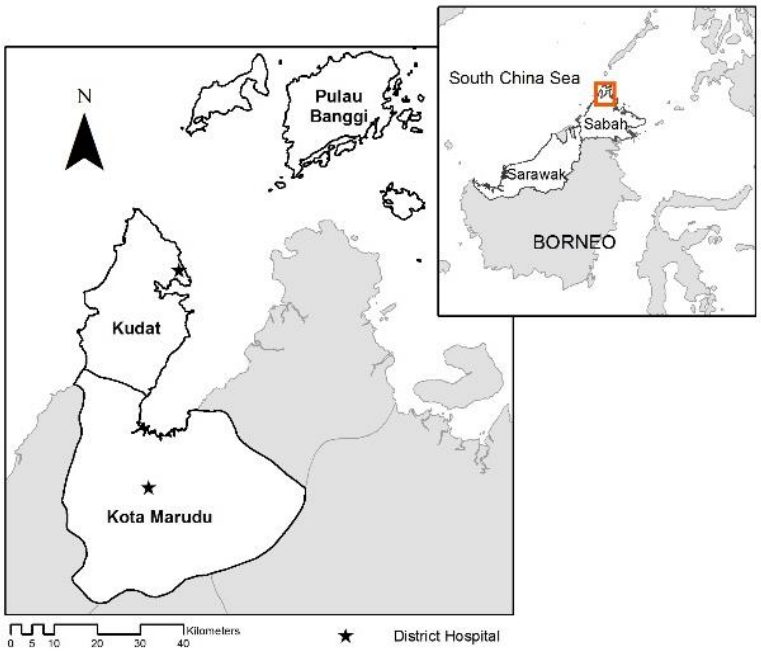
potential associations between village-level incidence and deforestation and other environmental factors. Understanding these relationships is vital to predicting and responding to future disease outbreaks as well as to understanding the underlying mechanisms of *P. knowlesi* emergence.

5.2.3 Methods

Study site and population:

This study was conducted in Kudat and Kota Marudu districts in North-western Sabah, Malaysia (07.38-06.19°N, 116.62-117.46°E), an area of 3,204 km^2 with a population of approximately 120,000 people, predominantly of Rungus and Dusun ethnicities [140] (Figure 5.1). The climate is tropical, with no dry season and higher rainfall from November through March. This area includes coastal and inland regions, with elevations ranging from sea level to 1,000 m above sea level, undergoing substantial environmental change due to conversion of land for oil palm plantations and other agricultural activities [52]. Both Kudat and Kota Marudu districts have central referral hospitals serving defined catchment areas where patients have access to diagnosis and treatment free of charge. All clinics refer cases to the central district hospital where hospitalisation is mandatory for all malaria patients until a negative blood smear for malaria parasites has been obtained.

Figure 5.1. Kudat and Kota Marudu districts in Sabah, Malaysia



Geo-location of cases and villages:

A retrospective review was conducted of malaria cases reported at Kudat and Kota Marudu district hospitals from 2008 to 2012. Data on diagnosis, date, demographic information and address of all malaria patients were obtained from hospital laboratory microscopy records. Villages and populations were identified from the 2010 Malaysian National Census, the most recent census conducted in this area, and geo-located as part of a larger interdisciplinary study (MONKEYBAR ESEI project (http://malaria.lshtm.ac.uk/research/projects/malaria-research-epidemiology-20) [140]. Village populations were updated using published population growth rates for Malaysia [246]. All locations were confirmed using imagery available through Google Earth (https://www.google.com/earth/) or other freely available satellite data. Patient addresses were matched to the census data or the nearest reported village from the census data. Administrative boundaries were used to define the extent of urban areas, within which village data were combined.

Calculating the proportion of knowlesi cases

Plasmodium knowlesi is microscopically similar to *P. malariae* but can also be misdiagnosed as *P. falciparum* or other human malaria species. In hospital microscopy records, the species of malaria is recorded as determined by morphology, with no separate listing for *P. knowlesi*. Uncertainty in diagnosis resulted in some infections being recorded as *P. malariae or P. malariae/ knowlesi*. To

estimate the true proportion of *P. knowlesi*, sensitivity and specificity of microscopy diagnosis as *P. malariae* was calculated for a subset of 539 malaria patients for whom both microscopy and molecularly confirmed results were available, including all cases from Kudat and Kota Marudu hospitals referred to a tertiary care hospital during this period and all cases recruited at these hospitals between 2013 and 2014 [37, 247]. The proportion of malaria cases reported as *P. knowlesi* per village per year was adjusted for this sensitivity and specificity using a Bayesian estimation of true incidence from apparent incidence obtained by testing individual samples (Equation 1) [248].

Equation 1:

$$Pos_i \sim Binomial\ (p_i, n_i)$$
$$p_i = \pi * SE + (1 - \pi) * (1 - SP)$$

Models were fitted in R through the prevalence and rjags packages interfacing with JAGS, using two chains containing 1,000 burn-in samples and 5,000 retained samples (R, Vienna, Austria, http://www.R-project.org, JAGS v3.4.0). Sensitivity and specificity were parameterised using beta-PERT distributions of the minimum, maximum and most likely values of sensitivity and specificity.

Environmental data:

Topography and land use data were extracted from a variety of datasets derived from remote-sensing data. These data were evaluated for buffer areas with a radius of 1, 2 and 5 km from the centre of each village, chosen to explore a range of spatial scales at which environment may be relevant based on the typical distribution of households, farming land and local movements. Elevation data with a spatial resolution of 30 m were obtained from the ASTER Digital Global Elevation Model [209]. The average annual normalised difference vegetation index (NDVI) was calculated from Moderate Resolution Imaging Spectroradiometer (MODIS) 16-day composites at 250 m resolution [249]. NDVI quantifies the greenness of vegetation and is influenced by climatic factors such as rainfall and temperature. It has been used extensively to predict malaria incidence and develop early warning systems in other contexts (e.g. [250, 251]).

Tree cover data, derived from classified Landsat imagery at a 30 m resolution, were obtained from Hansen *et al.*, 2013 [51]. Annual forest cover maps for the study districts were produced, with forest defined as ≥50% tree crown cover density; while this represents forested areas, this cannot distinguish between types of forest or agro-forestry such as rubber or oil palm. The proportion of forest coverage, the proportion of forest loss during the year for which incidence was estimated and the proportion of cumulative forest loss for the previous five years were calculated for each buffer

radius for each village and time point. The effect of forest configuration was evaluated as the number of forest patches per each radius, a standard metric representing landscape fragmentation, using the Landscape Ecology Statistics plugin for Quantum GIS [252]. Distributions of these variables were examined and quartiles were used to categorise variables.

Additionally, as this analysis relied on passive reporting of malaria, travel time to the nearest clinic where cases would seek treatment for a febrile illness was included as a measure of access to care. Travel time to the clinic from each village was estimated using reported travel times from community interviews and cases recruited as part of a population-based case control study [247].

Statistical analysis:

Annual *P. knowlesi* incidence for each village was mapped and smoothed maps were produced to visualise the data using a kernel density estimation method, a standard method for interpolating point location data. As the data were skewed relative to Poisson distribution, potential associations between environmental factors and reported *P. knowlesi* cases at village level were assessed using general linearized mixed models with a negative binomial distribution and an offset for population in R [253]. To account for correlation between repeat measurements on the same village, village was included as a random effect. Bivariable analysis was carried out for each covariate; variables for which $p<0.2$ were included in multivariable models. Significance of single variables was assessed using likelihood ratio tests and final model selection was based on Akaike Information Criterion (AIC). For correlated variables (such as elevation at different buffer ranges) a single variable was selected for inclusion based on marginal AIC values. Potential bias due to residual spatial autocorrelation in the model was explored for island and mainland areas through Moran's I. Based on this statistic, the negative binomial model was fit with a spatial correlation component that was estimated with a distance-based Matern correlation function.

Ethics:

This study was approved by the Medical Review and Ethics Committee of the Ministry of Health, Malaysia.

5.2.4 Results

Malaria in Kudat and Kota Marudu districts from 2008 – 2012

A total of 405 villages were mapped in Kudat and Kota Marudu districts with a median population of 168 people (IQR: 80, 313) and 44 households (IQR: 20, 78) per village. A total of 2,006 malaria cases were reported between 2008-2012, with 833 malaria cases in Kota Marudu district, 1,014 in Kudat district and 159 reported from outside these districts. Standard reporting forms did not include specific age data but classified cases as adults (65.7%; 1318/2006) or children (32.8%; 657/2006), with a small number of records (1.5%; 31/2006) missing this information. The majority of malaria cases were male (66%; 1330/2006). Most villages (60%, 245/405) reported at least one malaria case during this period. The number of malaria cases reported varied annually, with marked seasonal variations in both case numbers and rainfall. Almost half (47%, 878/1847) of reported malaria cases were diagnosed as suspected *P. knowlesi*/ *P. malariae* by microscopy. An additional 27% (512/1847) and 25% (457/1847) were diagnosed as *P. falciparum* and *P. vivax* respectively.

Out of 346 samples diagnosed as *P. knowlesi*/ *P. malariae* by microscopy (including mixed infections) and sent for molecular confirmation, 90% (313/346) were confirmed as *P. knowlesi* using PCR. Sensitivity and specificity of microscopy diagnosis for *P. knowlesi* were 95% (95% CI: 0.92, 0.97) and 84% (95% CI: 0.79, 0.89) respectively (Table 5.1). Although previous studies have reported frequent misdiagnosis, very few samples (3%, 16/539) were incorrectly identified as other species by microscopy. As PCR results were only available for confirmed malaria patients, the probability of detecting sub-microscopic infections could not be estimated.

Table 5.1. Comparison of PCR and microscopy results

	P. knowlesi PCR positive	*P. knowlesi* PCR negative	**Total**
P. malariae positive by microscopy	313	33	346
P. malariae negative by microscopy	16	177	193
Total	329	210	539

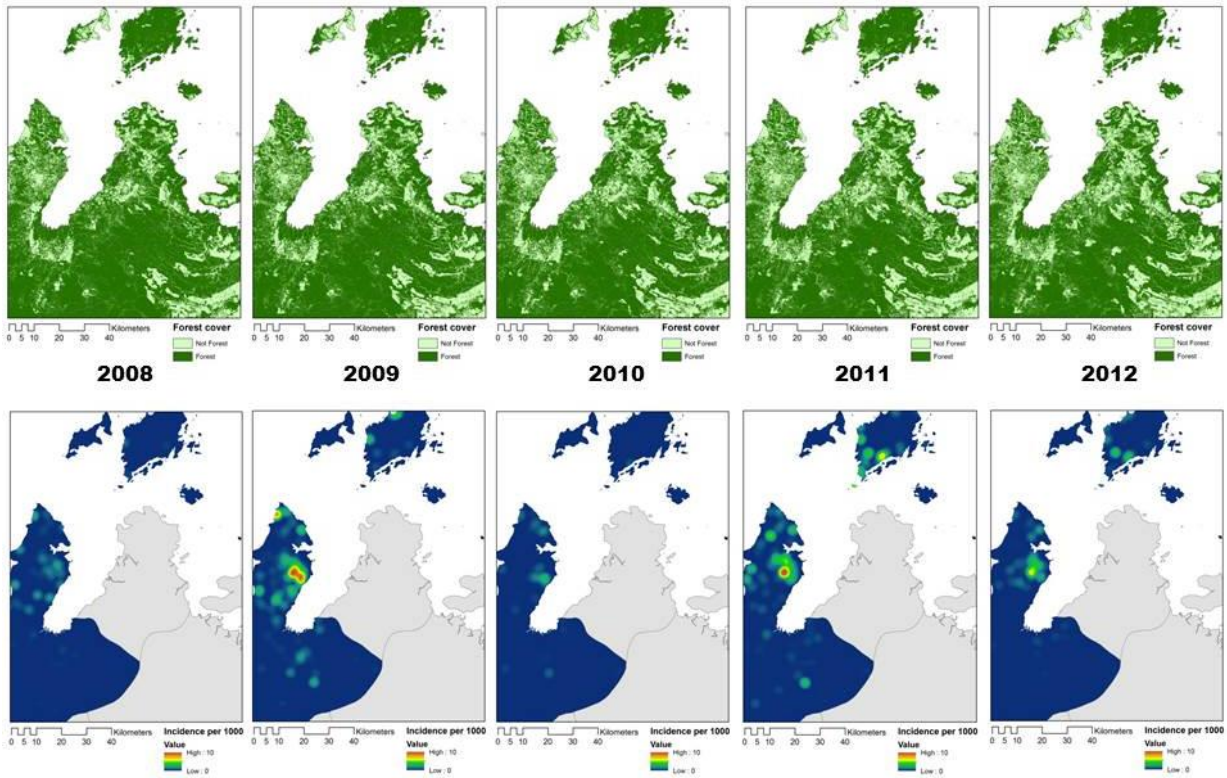
Based on these values, the true numbers of *P. knowlesi* cases were estimated as 739 (95% CI: 664, 794) for Kudat and Kota Marudu districts for this time period. The estimated annual parasite incidence (API) for *P. knowlesi* calculated for each village ranged from 0 to 102 cases per 1,000

people per year, with an overall mean API of 1.84. Out of the 245 villages reporting malaria cases, 24% (59/245) had an estimated mean API for *P. knowlesi* of less than 1, 44% (108/245) had an API of 1-5, 11% (26/245) had an API of 5-10, 6% (15/245) had an API of 10-20 and 2 had an API of >20 (Figure 3).

Association with environmental variables:

Based on estimates from remote sensing data [51], forest cover declined by 4.8% in Kudat and Kota Marudu between 2008 and 2012 (Figure 5.2). While in many instances this loss was highly localized, large tracts of forest were cleared in the interior of Pulau Banggi. Overall, substantial environmental change was observed, with 39% (157/405) of villages losing >10% of the forest cover within a 1 km radius between 2008 and 2012, 44% (179/405) of villages losing >10% within a 2 km radius and 51% (206/405) of villages losing >10% of forest cover within a 5 km radius.

Figure 5.2. Forest cover and estimated P. knowlesi incidence from 2008 - 2012



Lower elevation, more than 65% forest cover (within a 2 km radius) and higher historical forest loss were associated with greater incidence *P. knowlesi*. Travel times to clinic, NDVI and patch numbers were not significant at any radius in the bivariate analysis.

Table 5.2. Land use characteristics and bivariable statistics

Variable	Bivariable	
	IRR (95% CI)**	P Value
Proportion of forest remaining (%)		
Within 1 km radius		
Less than 65%	-	0.59
65% or more	1.08 (0.82, 1.43)	
Within 2 km radius		
Less than 65%	-	0.06
65% or more	1.30 (0.99, 1.71)	
Within 5 km radius		
Less than 65%	-	0.77
65% or more	1.05 (0.78, 1.40)	
Proportion of forest lost in current year (%)		
Within 1 km radius		
Less than 1%	-	
1 – 2%	0.94 (0.74, 1.19)	0.11
Over 2%	1.22 (0.99, 1.51)	
Within 2 km radius		
Less than 1%	-	
1 – 2%	1.50 (1.20, 1.87)	0.001
Over 2%	1.36 (1.09, 1.71)	
Within 5 km radius		
Less than 1%	-	
1 – 2%	0.97 (0.75, 1.27)	0.01
Over 2%	1.57 (1.19, 2.06)	
Proportion of forest lost in past five years (%)		
Within 1 km radius		
Less than 8%	-	0.60
8 – 14%	1.01 (0.76, 1.33)	
Over 14%	1.14 (0.85, 1.53)	
Within 2 km radius		
Less than 8%	-	
8 – 14%	1.66 (1.25, 2.20)	< 0.001
Over 14%	2.03 (1.46, 2.80)	
Within 5 km radius		
Less than 8%	-	
8 – 14%	0.93 (0.69, 1.25)	0.32
Over 14%	1.13 (0.81, 1.58)	
Mean elevation (per 10 meters above sea level)		
Within 1km radius	0.99 (0.97, 1.00)	0.12
Within 2km radius	0.98 (0.97, 1.00)	0.06
Within 5km radius	0.97 (0.96, 0.99)	0.01
Travel time to clinic (minutes)	1.00 (0.99, 1.00)	0.001

** Incident rate ratio

Table 5.3. Multivariable negative binomial regression of *P. knowlesi* incidence (including population offset)

Variable	IRR (95% CI)**	P value
Proportion forest remaining within 2 km radius		
Less than 65%	-	0.0004
65% or greater	1.51 (1.42, 1.99)	
Proportion forest lost in the past 5 years with 2km		
Less than 8%	-	
8 – 14%	1.68 (1.27, 2.22)	< 0.0001
Over 14%	2.22 (1.53, 2.93)	
Mean elevation (per 10 metres above sea level) within 5km radius	0.98 (0.96, 0.99)	0.001

** Incident rate ratio

Spatial patterns:

Maps of interpolated *P. knowlesi* incidence show distinctive spatial patterns that vary from year to year, with some areas of high incidence persisting over time (Figure 2). To assess the degree of residual spatial correlation after fitting the negative binomial regression, the residuals of the best fitting model were mapped and Moran's I was calculated (mainland: Moran's I: 0.07, $p < 0.0001$; islands: Moran's I: 0.07, $p = 0.07$). As Moran's I was significant, the final model was adjusted for spatial autocorrelation (Tables 5.2 and 5.3).

5.2.5 Discussion

This study aimed to describe spatial and temporal patterns of *P. knowlesi* infection within Northern Sabah and evaluate potential associations between village-level *P. knowlesi* incidence and key environmental factors. Although land use changes can impact the emergence of infectious diseases, this has not previously been evaluated for *P. knowlesi.* This study shows an important link between deforestation, environmental characteristics and the reported incidence of *P. knowlesi* malaria in this area of Sabah, Malaysia.

Village-level malaria data reveal marked spatial and temporal heterogeneity in *P. knowlesi* incidence in Kudat and Kota Marudu districts. Adjusting for the sensitivity and specificity of microscopy, *P. knowlesi* was the most common cause of human malaria cases, which is consistent with other studies within this geographic area [12, 13, 37]. Quantification of annual forest loss indicates that substantial environmental changes occurred between 2008 and 2012, with many villages losing substantial proportions of surrounding forest cover.

The proportion of forest surrounding the village was significantly associated with *P. knowlesi* incidence in the final model potentially reflecting the fact that forest environments are the habitats of macaques and the mosquito vectors. Studies have reported higher vectorial capacity and sporozoite rates for this species in forest environments relative to agricultural and settled lands [80, 83]. Long-tailed macaques have also been reported in a variety of environments, including degraded secondary forest areas [92]. The significance of both forest and forest loss likely reflects that forested areas undergoing substantial change are where transmission is occurring, similar to previously described frontier malaria [69].

Notably, higher incidence of *P. knowlesi* was associated with higher proportions of forest loss surrounding a village during the previous five years. This association could be due to changes in macaque or mosquito habitats as well as higher levels of human activity. Density of long-tailed macaques has been reported to increase in response to deforestation; loss of previous habitats can result in crowding within forest patches with potential implications for disease transmission [92]. Land use changes have also been shown to affect abundance and community composition of potential vectors [56, 239]. Deforestation and associated agricultural development are also associated with changes in human risk, due both to changes in the distribution and behaviour of people as more employment opportunities become available and to changes to individual movement patterns due to forest clearing and agricultural activities [69, 245]. Although the initial clearing of forests may deplete vector populations, initially reducing malaria transmission, this may be followed by colonisation of cleared areas by more efficient vector species and subsequent increases in transmission [6]. The fact that historical forest loss was more significantly associated with *P. knowlesi* incidence than forest loss in the current year suggests that increases in transmission are related to longer term changes in vector, host or human populations involved but additional longitudinal studies would be required to investigate this hypothesis.

The effect of habitat may also be reflected in associations with elevation. Elevation was negatively correlated with *P. knowlesi* incidence, although the range of elevations within the study site was limited, with the majority of villages at elevations below 100m. Both macaques and vectors are more frequently described in low elevation areas but have been reported at higher elevations [92, 254].

Neither forest configuration nor vegetation indices appeared to be strongly associated with incidence. As NDVI is a measure of vegetation greenness, it may not differentiate between forest and other types of agriculture (e.g. oil palm), as it tends to become saturated in tropical environments. Studies of other zoonotic diseases have found fragmentation of specific land cover classes, in addition to forest, and their interactions influence disease transmission more than overall landscape configuration [2, 245]. Other fragmentation metrics could be included to evaluate the effects of forest patch size and shape on *P. knowlesi* transmission.

The buffer sizes evaluated represent a range of potential scales at which these variables may be important for *P. knowlesi* transmission. Human infections are a product of many interacting factors occurring across different spatial scales, with signals across a range of scales relating to different contributions of drivers. As this study covers an extensive area, behaviours, village distributions and ecology vary within these regions. Villages are typically within small spatial areas but farming practices range from small scale swidden farming to large scale plantations, affecting the scale of human interactions with the environment. Future analyses could include more detailed spatial data on household locations and human movement patterns.

The main limitation of this study is the reliance on records of malaria cases presenting to the hospital; these cases may represent only a proportion of malaria cases in the community. While malaria is a notifiable disease and distance to the hospital was included as a measure of access to care, this does not account for any asymptomatic malaria cases that may have occurred, or symptomatic cases that were resolved without treatment. Previous studies of other malaria species in similar transmission settings have described a large proportion of asymptomatic carriage within communities; however this has not yet been evaluated for *P. knowlesi*. A cross-sectional survey of 2,019 people in central Vietnam identified three people positive for *P. knowlesi*, all of whom were asymptomatic at the time of the survey and for the subsequent six months, demonstrating that asymptomatic carriage of *P. knowlesi* can occur [31].

This study was also limited by the environmental data used. The forest cover and forest loss data used were aggregated by year and finer scale temporal associations between land cover and incidence could not be explored. This dataset was limited by the definition of forest by canopy cover; this does not allow the differentiation between types of forest or crops and did not allow differentiation between patches of different types of land. Additionally, analysis was limited by the spatial resolution of remote sensing data, the use of a centroid point to represent village location

and the use of circular buffers rather than polygons reflecting actual village shape. Exploratory spatial analysis did suggest spatial heterogeneity among the village-level data and these spatial effects were included within the model. Further work with more spatially specific outcome and environmental data could investigate these spatial patterns in more detail.

Despite inherent limitations in the outcome and covariate data used in this study, results are strongly suggestive of a link between environmental change and reported incidence of emerging *P. knowlesi* in northern Sabah. Spatial analysis of environmental factors affecting disease emergence can be used to target surveillance and public health activities to areas expected to have higher disease risk.

5.2.6 Conclusion

While further population-based studies are needed to define environmental risk factors, this study indicates deforestation is associated with human cases of *P. knowlesi* within North-western Sabah, Malaysia.

5.2.7 Acknowledgements

We would like to thank Albert M Lim, Lina Marlina Gulam (Infectious Disease Society Sabah), Beth Downe (LSHTM) and the MONKEYBAR team in Sabah with their help with this study. N. Alexander receives salary support from the Medical Research Council and Department for International Development (MR/K012126/1). We acknowledge the Medical Research Council, Natural Environment Research Council, Economic and Social Research Council, and Biotechnology and Biosciences Research Council for the funding received for this project through the Environmental and Social Ecology of Human Infectious Diseases Initiative (ESEI). Grant Number: G1100796.

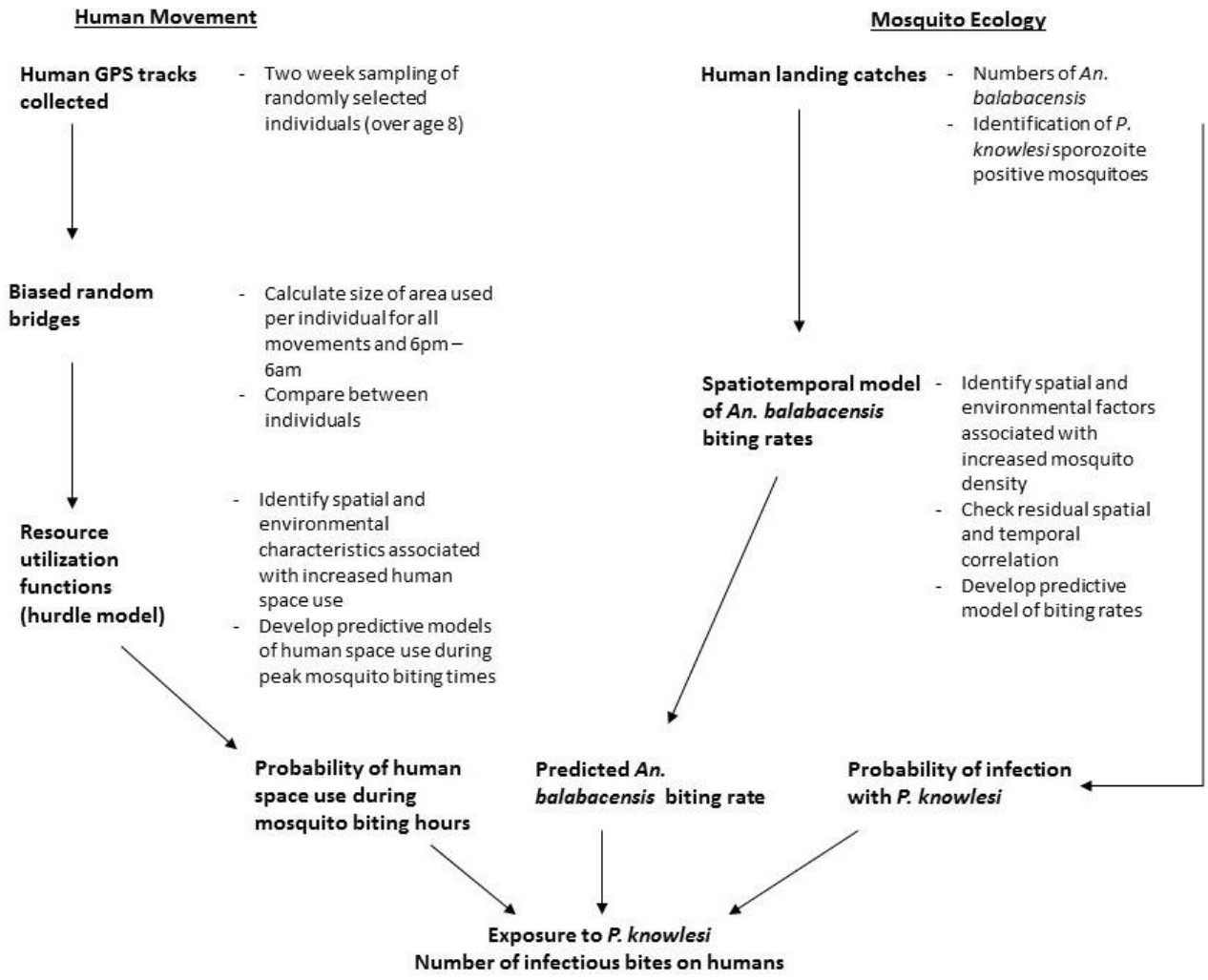
6 Human movement and exposure to *P. knowlesi*

6.1 Background and aims

Out of necessity, approaches to assess environmental risk factors for disease typically rely on landscape data surrounding individual households. While an individual may spend the majority of their time within the vicinity of their residence, this area does not necessarily represent where an individual is most likely to be exposed to a disease. This is also supported by varying associations between disease occurrence and landscape variables at different spatial scales surrounding households, likely due partially to human movement into these areas [234, 235, 255]. Numerous studies have identified travel to nearby plantations and forest areas as a risk factor for *P. knowlesi* infection (e.g. [11, 18, 103]); however, the spatial range and frequency at which these movements occur remains unknown. Further, little is known about differences in movement patterns in different demographic groups and whether infections reported in women and young children are likely to arise from exposure to forest and plantation environments [37]. The presence of both vectors and macaques around households suggests some individuals may be infected within these areas; however, detailed quantitative data on individual mobility patterns relative to mosquito habitats is required to assess the probability of peri-domestic transmission and understand how local movement impacts disease risks [87].

To characterise these local mobility patterns, we conducted a GPS tracking study in two case study communities in Northern Sabah, Malaysia and combined these data with models of mosquito biting to assess exposure risks. GPS technology has previously been utilised to map fine-scale movements and quantify differences in individual risk behaviours for epidemiological studies [101, 102, 122, 131]. We aimed to characterise local movement patterns and identify individuals and locations associated with increased *knowlesi* exposure risks by: 1. analysing individual movement patterns and developing predictive maps of human space use relative to spatial and environmental factors, 2. modelling biting rates of the main *knowlesi* vector *An. balabacensis*, and 3. assessing the relative exposure risks for *P. knowlesi* (Figure 6.1).

Figure 6.1: Conceptual framework of analysis of human movement, mosquito biting rates and estimation of exposure



6.2 Methods

6.2.1 Study site and spatial data

This study was conducted in the two case study communities (Matunggong in mainland Kudat, population and Limbuak in Pulau Banggi) in Northern Sabah described in Chapters 3 & 8. Demographic data and GPS locations of primary residences were collected for all individuals residing within these areas during the time of the study [256].

Potential spatial and environmental covariates for these sites during the study period were assembled from a range of ground-based and remote-sensing data sources summarised in Table 6.1.

Land use was categorised as forest, agriculture, cleared areas (no vegetation) and water bodies. The enhanced vegetation index (EVI) was used to capture temporal changes in vegetation levels; this index captures photosynthetic activity and has higher sensitivity in high biomass areas compared to the normalised difference vegetation index (NDVI) frequently used. Due to the high cloud cover within this area, EVI at a high spatial resolution could not be obtained for all time periods. Instead, EVI data at a lower spatial but higher temporal resolution was used and monthly averages were calculated from all available cloud-free data and resampled to 30m per pixel [257]. Rainfall data was obtained from the Tropical Rainfall Measuring Mission (TRMM), a satellite using radar to measure the intensity and distribution of rainfall in tropical and subtropical areas [258]. Additional topographic measures derived from the ASTER global digital elevation model and Euclidean distances to houses, roads and forests were calculated using R. Data on temperature were obtained from the Malaysian Meteorological Department station in Kudat; as this area is close to the equator and these data exhibited limited monthly variation, temperature was not included in the final model.

Table 6.1. data sources for potential covariates

Covariate	Description	Spatial resolution	Source
Land use type	Classified land use type as forest, agriculture, clearing and water	30m	Derived from Landsat [160], described in [256]
EVI	Monthly mean enhanced vegetation index (0-1)	250m	MODIS Vegetation Indices [257]
TWI	Topographic wetness index	30m	Calculated from ASTER Digital Elevation Model [159]
Elevation	Meters above sea level (m)	30m	ASTER Digital Elevation Model [159]
Slope	Degrees of slope incline (degree)	30m	Calculated from ASTER Digital Elevation Model [159]
Aspect	Compass direction slope is facing (degree)	30m	Calculated from ASTER Digital Elevation Model [159]
Distance to houses	Euclidean distance from nearest household, calculated from household GPS points (m)	1m	Mapped during study
Distance to roads	Euclidean distance from nearest road, calculated from GPS tracks of roads (m)	1m	Mapped during study
Population density	Estimated population per 100m for 2015 (adjusted for UN estimates)	100m	WorldPop [259]
Precipitation	Average monthly rainfall (mm/ day)	0.25°	Tropical Rainfall Measuring Mission [258]
Temperature	Average monthly temperature (°C)	-	Malaysian Meteorology Department

6.2.2 GPS tracking

This study utilised a rolling cross-sectional design to characterise local movements within these communities, targeting a minimum of 50 individuals per each study site. During pre-defined two week intervals, participants were randomly selected and asked to carry a QStarz BT-QT13000XT GPS tracking device (QStarz, Taipei, Taiwan). Individuals were excluded if they were not currently primarily residing in the study area, under 8 years old or did not consent. GPS devices were programmed to record coordinates continuously at one minute intervals. Participants were asked to wear the GPS device at all times outside the house and instructed on how to use the device and change batteries.

Trained fieldworkers visited the participant every two days to confirm the device was functioning and replace batteries (questionnaires and participant information in Annex 2). During these visits, structured questionnaires were administered on locations visited, vector preventative practices, macaque encounters and use of the GPS device. Fieldworkers also recorded whether the device was working and if the individual was observed carrying the device to assess compliance levels. Individuals were excluded from the analysis if insufficient GPS data was collected (less than 33% of sampling period included) or individuals were observed not using the device for 2 or more visits.

6.2.3 Human utilisation distributions (biased random bridges)

Biased random bridges (BRBs) were used to calculate individual utilisation distributions (UDs), the probability of an individual being in a location in space within a particular time. UDs are widely used in movement ecology to estimate the home range of an individual or species and can be related to landscape characteristics [260]. UDs are typically calculated using kernel smoothing methods, smoothing observed locations of individuals or species to estimate average probabilities of use per grid cell to create a probability density function. BRB improves on previous kernel density estimates by estimating the utilisation distribution as a time-ordered series of points, taking advantage of the autocorrelated nature of GPS tracks to bias movement predictions towards subsequent locations in a time series [260]. This allows for interpolation of missing values and adjustment for spatial error to estimate utilisation distributions reflecting both the intensity (mean residence time per visit) and frequency of individual visits to specific locations [260]. Within this study, large proportions of GPS fixes were missed due to technical issues with batteries and GPS signal; BRBs were used to interpolate between known locations and adjust for missing data.

To fit the BRBs, the maximum time threshold between points before they were considered uncorrelated (T_{max}) was set as 3 hours, the minimum distance between relocations (L_{min}) was set as 10 metres and the minimum smoothing parameter (h_{min}) was set as 30 metres to account for relocation uncertainty and the resolution of habitat data. UDs were calculated separately for each individual for all movement and night-time only movements (6pm – 6am). Individuals were excluded if less than 200 GPS points were available between 6pm and 6am. Estimates of the core utilisation area (home range) were based on the 99^{th} percentile, representing the area with a 99% cumulative probability distribution of use by the sampled individual.

6.2.4 Resource utilisation functions

To assess relationships between space use and environmental factors and develop predictive maps of community space use, we fit resource utilisation functions, regression models in which the UDs are used as the response variable with spatial and environmental predictors. Resource utilisation functions can improve on models using raw GPS count points as the response when there is location uncertainty and missing data [261]. The probability density function (UD) per individual was rasterised to 30m^2 grid cells and environmental and spatial covariates extracted for each grid cell. Potential environmental covariates included distance to the individual's own house, distance to closest house, distance to roads, land use class (forest, agriculture, cleared or water), distance to forest edge, elevation and slope. Models were limited to land areas within the study area, defined as within 5km of households. Multicollinearity between selected variables was assessed through Pearson correlation and variance inflation factors and binomial and gamma general linearised models were fit using the Bayesian Information Criteria (BIC) to select the most parsimonious model. Separate models were fit for each of the two sites.

Resource utilisation was modelled as a Bayesian semi-continuous (hurdle) model with two functionally independent components, a Bernoulli distribution for presence/absence of individual *j* in a specific grid cell *i* (z_i) and a gamma distribution for the UD in grid cells visited (y_i) [262, 263]. We defined absences as grid cells with a UD of less than 0.00001 and randomly sampled equal numbers of absences within the study site for each presence point for each individual [264]. Occurrence was defined as:

$$z_{ij} = \begin{Bmatrix} 1 & \text{if } y_{ij} > 0 \\ 0 & \text{otherwise} \end{Bmatrix}$$

The utilisation distribution for grid cells visited was defined as:

$$y_{ij} = \begin{Bmatrix} \text{NA} & \text{if } y_{ij} = 0 \\ y_{ij} & \text{otherwise} \end{Bmatrix}$$

The full model was specified as:

$$z_{ij} \sim \text{Bernoulli}(\phi_{ij})$$

$$y_{ij} \sim \text{Gamma}\left(\frac{\mu_{ij}^2}{\sigma^2}, \frac{\mu_{ij}}{\sigma^2}\right)$$

With the linear predictor for the Bernoulli model specified as:

$$\text{logit}(\pi_{ij}) = \beta_0 + X_{ij}^T \beta_i + \gamma_j$$

Where β_0 represents the intercept, $X_{ij}^T\beta_i$ represents a vector of covariate effects and γ_j represents the additive terms of random effects for individual. For the Gamma component, σ^2 is the variance and the linear predictor μ_{ij} is specified as:

$$\log(\mu_{ij}) = \alpha_0 + X_{ij}^T\alpha + \gamma_j$$

With α_0 representing the intercept, $X_{ij}^T\alpha$ representing a vector of coordinates and γ_j representing the random effects. Weakly informative normal priors specified as *Normal*(0,0.01) were used for all intercepts and coefficients. Bayesian inference was implemented using integrated nested Laplace approximation (INLA), as described by [265]. This approach uses a deterministic algorithm for Bayesian inference, increasing computational efficiency relative to Markov chain Monte Carlo and other simulation-based approaches [262]. We did not explicitly include spatial autocorrelation as several distance-based covariates were included (e.g. distance from own house) and multicollinearity between covariates and spatial structure has been shown to decrease performance of resource utilisation functions [261]. Predictive performance was assessed using deviance information criterion (DIC), area under the receiver operating curve (AUC) and root mean square error (RMSE).

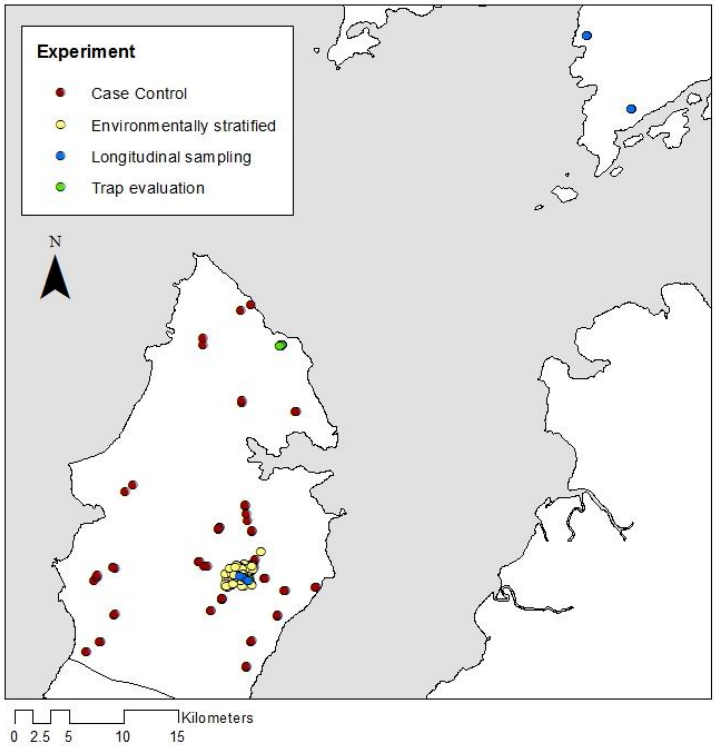
6.2.5 Mosquito human landing catch data

To estimate vector biting rates, we assembled data from entomological surveillance activities using outdoor human landing catches conducted while GPS tracking was on-going, including: monthly longitudinal surveillance [152], investigations surrounding the households of cases and controls [87], trap comparison experiments conducted in Kudat [266], and environmentally stratified catches conducted within the Kudat study site (Ng, unpublished) (Table 2, Figure 2). We limited this data to counts of *An. balabacensis*, the primary *knowlesi* vector which comprises over 95% of *Anopheles* caught in this region. Plausible environmental covariates were assembled, including land use type, slope, aspect, elevation, topographic wetness index, enhanced vegetation index (EVI), population density and average monthly temperature and rainfall.

Table 6.2. Human landing catch data used

Experiment	Study Design	Sampling dates	Time period	Data points
Longitudinal sampling	Longitudinal monthly sampling at sentinel sites in Kudat and Pulau Banggi, described by [152]	August 2013-December 2014	12 hour catches, 6pm – 6am	106
Case control	Sampling around households of knowlesi cases and matched controls, described by [87]	February 2014-July 2014	12 hour catches, 6pm – 6am	56
Environmentally stratified sampling	Monthly sampling in randomly selected 100m^2 grid cells, stratified by land use type (Ng et. al, unpublished data)	February 2015 – December 2015	6 hour catches, 6pm – 12am	213
Trap evaluation	Comparison of human landing catches and electric net traps in Latin square design, only data from human landing catches included (Hawkes et. al, unpublished data)	November 2013 – February 2014	12 hour catches, 6pm – 6am	40

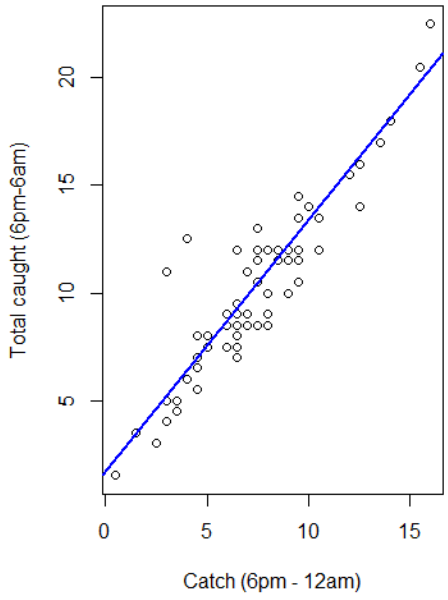
Figure 6.2. Human landing catch locations



As only two locations were surveyed in Pulau Banggi, we excluded all islands and limited the analysis of mosquito data to mainland Kudat. All data included was collected by trained catchers following the same protocol for catching and identifying mosquitoes. However, while three experiments conducted 12 hour catches (6pm – 6am), one experiment only collected mosquitoes for 6 hours

(6pm – 12am), during the peak mosquito biting times. To estimate the total numbers of *An. balabacensis* which would have been caught over a 12 hour period, a linear model was fit for all data from collected during 12 hour catches using the numbers of *An. balabacensis* caught between 6pm – 12am as a predictor (Figure 3). This model was used to predict the total number of *An. balabacensis* which would have been caught over 12 hours for the 6 hour experiments.

Figure 6.3. Total *An. balabacensis* from 6 hour catches vs. 12 hour catches and fitted linear model (R^2= 0.85)



6.2.6 Mosquito biting rate model development

To identify variables to include, Pearson correlation analysis was used to assess multicollinearity between selected environmental variables (Table 3). As slope and TWI had a strong negative correlation, only TWI was included in analysis. A negative binomial generalised linear regression model was implemented in R, with the number of person-nights for each data point included as an offset. Latitude and longitude were initially included as fixed effects within models. Bivariable analysis was conducted for each variable and variables for which $p < 0.2$ were included in the multivariable model. Likelihood ratio tests and the Akaike Information Criterion (AIC) were used for final model selection. Results from the final model are shown in Table S4. Potential residual spatial autocorrelation was assessed through Moran's I.

Table 6.3. Pearson correlation of selected covariates

	TWI	**Elevation**	**Population**	**Aspect**	**Slope**	**Rainfall**	**EVI**	**Forest distance**
TWI	1.00	-0.35	0.09	0.09	-0.80	0.01	0.05	-0.15
Elevation	-0.35	1.00	-0.01	-0.08	0.21	-0.10	-0.11	0.08
Population	0.09	-0.01	1.00	-0.01	-0.04	-0.05	-0.05	-0.01
Aspect	0.09	-0.08	-0.01	1.00	0.03	0.01	0.02	-0.06
Slope	-0.80	0.21	-0.04	0.03	1.00	0.02	-0.05	0.14
Rainfall	0.01	-0.10	-0.05	0.01	0.02	1.00	0.11	-0.07
EVI	0.05	-0.11	-0.05	0.02	-0.05	0.11	1.00	-0.08
Forest distance	-0.15	0.08	-0.01	-0.06	0.14	-0.07	-0.08	1.00

Table 6.4. Non-spatial negative binomial regression with log person-nights as offset

Covariate	Coefficient	Standard Error	P value
Population density	-0.030	0.012	0.012
EVI	1.785	0.628	0.005
Distance to forest (100m)	0.064	0.033	0.052

To explore potential temporal autocorrelation, biting rates per person per night and the residuals of the final non-spatial, non-temporal regression model were plotted by date (Figure 6.4). The Dickey-Fuller test was used to test for data stationarity (Dickey-Fuller = -4.3266, Lag order = 10, p value < 0.01) and the autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to correlation between time lags (Figure 6.5). No significant time lags were identified and the Box-Ljung Test on model residuals did not detect significant remaining temporal autocorrelation after accounting for spatial autocorrelation (X-squared=2.73, p value = 0.987), suggesting that it would not be appropriate to use autoregressive models for time series analysis for these data.

Figure 6.4. a. *An. balabacensis* biting rates per night per person (mean monthly biting rate in black); b. Biting rates by time and residuals of spatial, non-temporal model by time

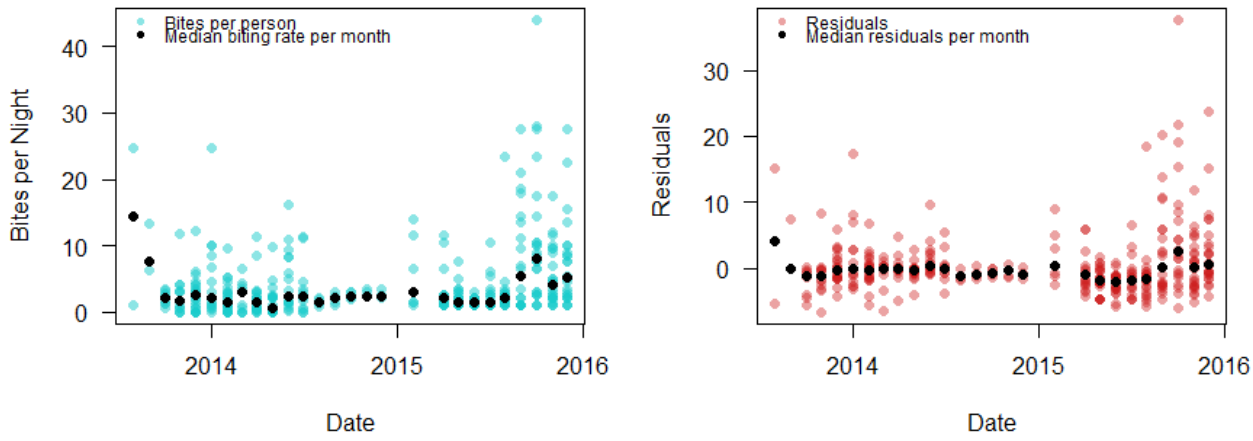
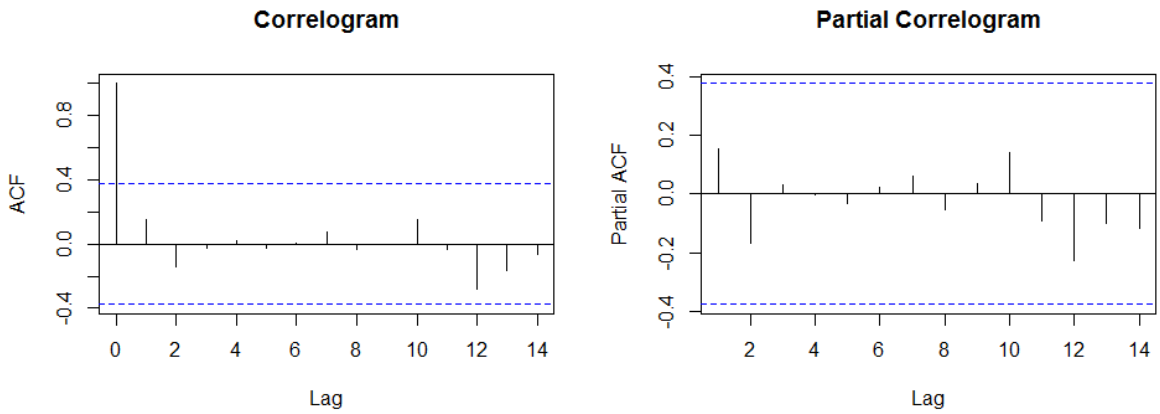


Figure 6.5. Autocorrelation and partial autocorrelation functions



A Bayesian hierarchical spatial model was implemented using integrated nested Laplace approximation (INLA) for latent Gaussian models as described by [265]. The outcome for this model was counts of *An. balabacensis* bites, denoted as Y_{ij}; i = 1…n; t = 1…n; where i indexes location and t indexes month. The number of person-nights per catch was included as an offset to adjust for different numbers of catchers conducting HLCs during different experiments. As the data was over dispersed, a negative binomial distribution was used to model y_{ij}:

$$[y_{it}|\lambda,\kappa] = \frac{\Gamma(y_{it}+\kappa)}{\Gamma(\kappa)y_{it}!}\left(\frac{\kappa}{\kappa+\lambda}\right)^{\kappa}\left(\frac{\lambda}{\kappa+\lambda}\right)^{\lambda}$$

where λ is equal to the mean (μ_{it}) and κ is the dispersion parameter, equal to μ_{ij}^2/ variance(μ_{ij})^2- μ_{ij}. The linear predictor η_{it} was specified as:

$$\eta_{it} = \log(\mu_{it}) = \log(N_{it}) + \beta_0 + X_{it}^T\beta + \mathrm{w_i} + e_t$$

Where N_{it} represents the offset of person-nights for each HLC catch, β_0 represents the intercept, $X_{it}^T\beta$ represents a vector of covariates, w_i is the spatial effect and e_t is the temporal effect. The spatial effect w_i was modelled as a Matern covariance function between locations s_j and s_k:

$$W \sim Multivariate\ Normal\ (0, \Sigma)$$

$$\Sigma_{jk} = Cov(\xi(s_j), \xi(s_k)) = Cov(\xi_j, \xi_k) = \frac{\sigma^2}{\Gamma(\lambda)2^{\lambda-1}}(\kappa||s_i - s_k||)^{\lambda}K_{\lambda}\big(\kappa||s_j - s_k||\big)$$

Where $||s_j - s_k||$ denotes the Euclidean distance between locations s_j and s_k, σ^2 is the spatial process variance and K_λ is a modified Bessel function of the second kind and order $\lambda > 0$. κ is a scaling parameter related to r, the distance at which spatial correlation becomes negligible, by $r = \sqrt{8\lambda}/\kappa$. A stochastic partial differential equations (SPDE) approach was used, representing the spatial process by Gaussian Markov random fields (GMRF) by partitioning the study area into non-intersecting triangles [267]. This approach represents the covariance matrix Σ by the inverse of the precision matrix Q of the GMRF [262, 267].

Prior distributions were specified on fixed effects and hyperparameters. A vague normal prior distribution was used for the intercept. Informative priors were used for coefficients based on previous studies examining relationships between *An. balabacensis* presence and EVI, forest cover and population density with priors specified as $\beta_{EVI} \sim N(1,0.01)$, $\beta_{Forest} \sim N(1, 0.01)$ and $\beta_{pop} \sim N(1,0.01)$ [10, 152]. Priors for spatial hyperparameters were specified as range $r \sim N(10,0.01)$ and standard deviation $\sigma \sim N(0.1, 0.01)$ parameterised as described by Lindgren and Rue [268].

Although clear time lags were not identified, monthly variation in catch numbers was observed. As no seasonal patterns were observed in either mosquito numbers or rainfall patterns, four plausible models were identified (without month or with month included as a linear effect, random effect or random walk model of order 1). Models were evaluated based on deviance information criterion (DIC), marginal likelihood and root mean square error (RMSE) using 20% of the data withheld for independent validation (Table 6.5). The predictive performance of the model was assessed using the mean logarithmic score, calculated as – mean(log(CPO_{Yij})), where CPO_{Yij} is the conditional predictive ordinate of the observed value y_{ij} calculated using leave-one out cross validatory posterior predictive distribution evaluated at the observed value y_{ij}, with lower values indicating better predictive performance [269]. Based on these statistics, model 4 was used for subsequent results with month specified as a random walk order 1 as:

$$\Delta x_t = x_t + x_{t+1} \sim Normal(0, \tau^{-1})$$

$$\theta = \log \tau$$

With a weakly informative penalised complexity prior used for theta [270].

Table 6.5. Model selection statistics

Model		DIC*	Marginal Likelihood	Model complexity*	RMSE*	Mean log-score
M1	Spatial effect only	2191.38	-1138.23	11.95	4.78	3.51
M2	Spatial effect + month as linear effect	2195.43	-1127.61	5.64	4.75	3.51
M3	Spatial effect + month as random effect	2064.97	-1100.18	31.29	4.74	3.31
M4	Spatial effect + month as random walk	2057.64	-1118.46	25.83	4.73	3.30
M5	Spatial effect + month as random walk + space-time interaction term	1772.98	-1080.80	248.30	3.78	11.21

* DIC: deviance information criterion, RMSE: root mean square error (bites per person), Effective number of parameters used as a measure of model complexity

6.2.7 Exposure to P. knowlesi

Borrowing from environmental epidemiology and using an approach outlined by Stoddard et. al [100], we calculate relative exposure risks as a derived quantity from human resource utilisation and mosquito biting rate models. First, to explore individual variation in risk, we adapt a simple exposure assessment model where an individual's risk of exposure is the sum of the probability of exposure to an infected vector across all locations visited. For the individuals tracked, we define the relative risk of individual *i* being exposed to an infected vector in month *t* as:

$$r_{it} = \sum_{j=1}^{J} y_{ij} a_{jt} Z_j$$

Where *j* indexes the grid cells visited, y_{ij} is the utilisation distribution, a_{jt} is the number of bites per individual and Z_j is the proportion of infectious mosquitoes. Based on reported sporozoite rates for each location, we parameterise *Z* as *Beta*(10,875) for Kudat and *Beta*(28,974) for Pulau Banggi [152]. As these vectors are rarely reported indoors and human landing catch data was not available for inside houses, we excluded areas within houses. Second, to evaluate places important for exposure for the entire community, we calculate the number of infectious bites per grid cell for each month as:

$$r_{jt} = \sum_{i=1}^{I} y_{ij} a_{jt} Z_j$$

Where y_{ij} is the predicted utilisation distribution for all individuals within the community per grid cell *j*.

6.3 Results

6.3.1 GPS Tracking

Between February 2014 and May 2016, 285 individuals participated in the GPS tracking study with 243 included in the final analysis, including 109 in Banggi and 134 in Kudat (example Figure 6.6). Participants included 121 men and 122 women, with ages ranging from 8 to 85 years (median 29.5). The most commonly reported occupation was farm or plantation work (n=74) with individuals also reporting occupations including fishing, office work and student. 80 individuals were housewives, retired or reported no occupation.

A total of 3,424,913 GPS points were collected, representing 6,319,885 person-minutes of sampling time. Median sampling duration was 16.27 days (IQR 13.72 – 19.97), with points recorded for a median of 59.1% (IQR: 46.9% - 71.1%) of the sampling duration. Distances travelled ranged from no travel outside the house to 116km, with a median distance travelled of 1.8km. While a median of 1.4% of all GPS locations were recorded in secondary forest (IQR: 0.003 – 13.22%), few points were recorded in the forest during peak mosquito biting times at night (median 0%, IQR: 0 – 10.1%). Home range sizes varied by gender and occupation with larger distances covered by individuals at the more rural Pulau Banggi site (Table 6.6; Figure 6.7). Farming activities were primarily conducted within the immediate vicinity of the households. Although substantial differences were reported in all movements between seasons, no seasonal differences were observed between 6pm – 6am.

Figure 6.6. a. individual wearing a GPS tracking device; b. example of GPS tracking data collected from a single individual in Matunggung, Kudat overlaid with aerial imagery

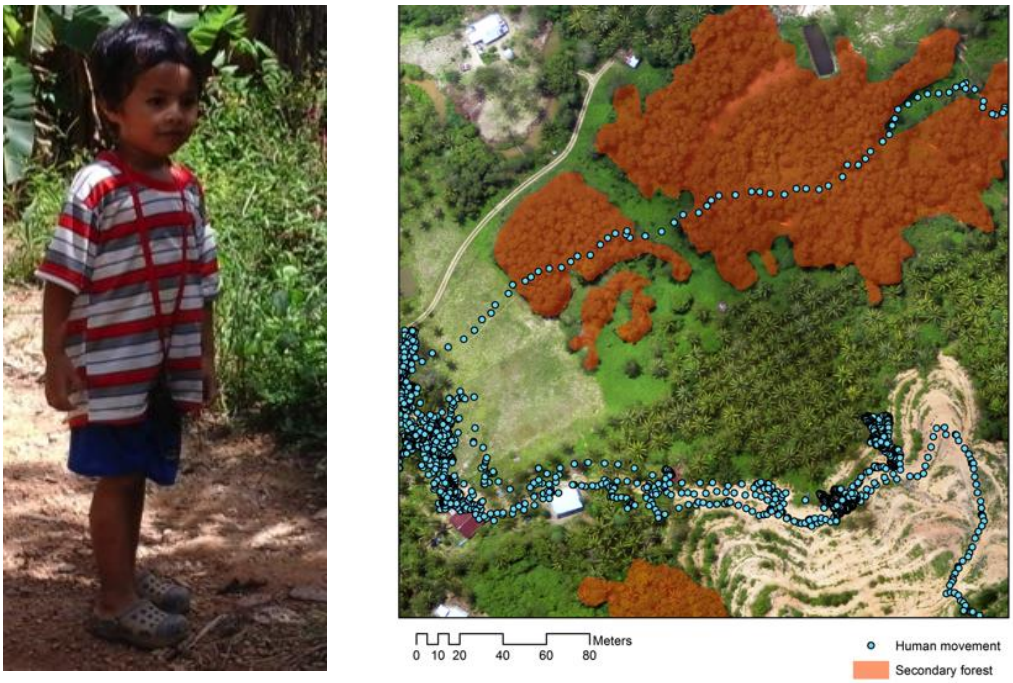
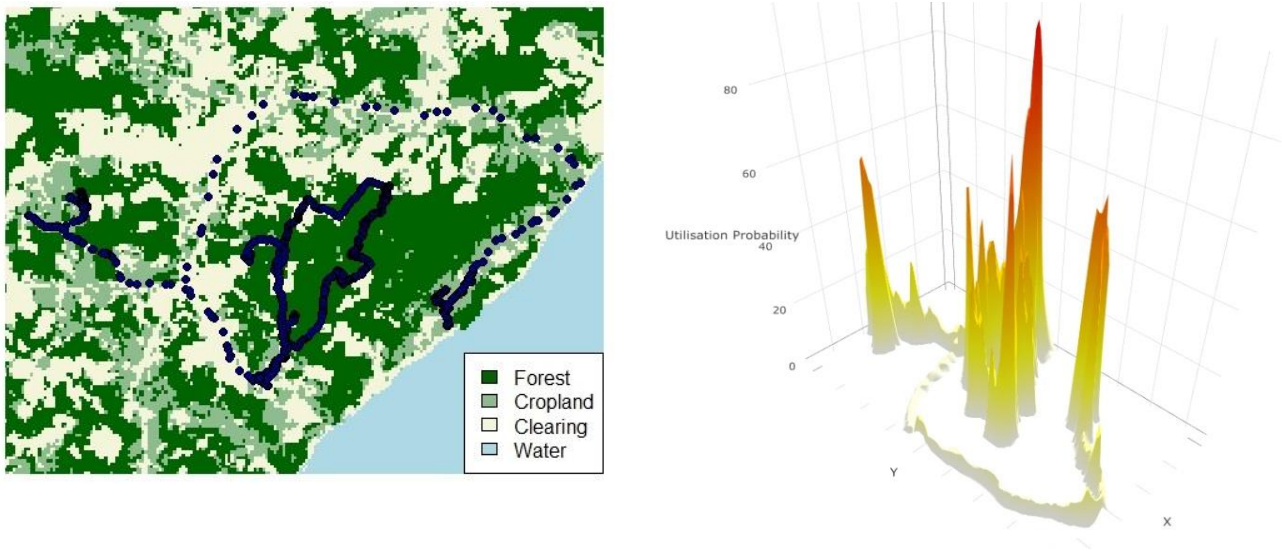


Table 6.6. Home range estimations by demographic groups

	Area of 99% UD for all movement (hectares) Median (IQR)	Area of 99% UD from 6pm – 6am (hectares) Median (IQR)
Demographic group		
Men	32.09 (7.07, 148.93)	4.50 (2.79, 19.53)
Women	74.25 (12.24, 320.74)	6.08 (2.79, 24.17)
Children (under 15)	26.01 (6.39, 151.94)	3.83 (2.79, 8.73)
Occupation		
Farming	29.34 (8.15, 324.38)	6.75 (2.79, 19.80)
Plantation work	49.14 (9.72, 201.33)	4.59 (2.79, 27.72)
Fishing	442.49 (40.07, 1189.00)	227.16 (4.05, 465.14)
Office work	96.80 (63.61, 256.75)	13.63 (2.88, 20.14)
Other	19.98 (6.30, 26.82)	2.97 (2.61, 18.27)
No employment/ housewife	43.38 (11.97, 157.59)	3.60 (2.79, 19.12)
Site		
Pulau Banggi	99.99 (24.57, 387.54)	7.74 (2.88, 58.05)
Kudat	12.02 (3.94, 85.55)	2.97 (2.70, 11.77)
Season		
Dry (February – July)	28.62 (5.45, 252.45)	4.19 (2.79, 19.60)
Wet (August – January)	54.90 (17.23, 160.99)	4.64 (2.79, 19.35)

Figure 6.7. a. GPS tracks from single individual over classified land use map and, b. utilitisation distribution calculated for this individual using BRB



6.3.2 Resource utilisation functions

Human space use during peak mosquito biting hours was mostly predictable and negatively correlated with distance from the individual's house, other houses and roads (Table 6.7). The AUC for the presence/ absence models was 0.936 for Kudat and 0.938 for Pulau Banggi and RMSE for the overall model was 0.0073 and 0.0043 for Kudat and Pulau Banggi respectively. Despite marked differences between different demographic groups and seasons observed during all movements, these factors did not improve the predictive power of the model for movements within the land areas of the study site between 6pm and 6am.

Table 6.7. Resource utilisation function fixed effects from final model

	Kudat			Banggi		
	Mean	SD	95% CI	Mean	SD	95%CI
Probability of presence/ absence						
Intercept	3.383	0.839	3.218, 3.547	3.571	0.104	3.368, 3.775
Distance from own house (km)	-0.954	0.006	-0.966, -0.942	-0.543	0.003	-0.548, -0.539
Distance from forest (km)	5.997	0.177	-5.650, 6.344	-1.845	0.050	-1.944, -1.746
Distance from road (km)	-5.552	0.057	-5.663, -5.441	-3.656	0.019	-3.694, -3.618
Distance from houses (km)	-0.504	0.030	-0.563, -0.444	0.176	0.007	0.162, 0.189
Elevation (100 MSL)	-0.710	0.025	-0.759, -0.662	-1.268	0.037	-1.340, -1.197
Slope (degrees)	-0.0244	0.002	-0.028, -0.021	-0.009	0.001	-0.012, -0.006
Utilisation distributions for locations present						
Intercept	-6.846	0.866	-8.549, -5.147	-5.676	1.017	-7.673, -3.681
Distance from own house (km)	-0.583	0.004	-0.590, -0.576	-0.308	0.002	-0.311, -0.305
Distance from forest (km)	12.012	0.199	11.621, 12.403	-1.771	0.049	-1.868, -1.675
Distance from road (km)	-0.833	0.054	-0.939, -0.728	-1.532	0.011	-1.554, -1.511
Distance from houses (km)	-0.819	0.023	-0.864, -0.773	-0.239	0.006	-0.249, -0.228
Elevation (100 MSL)	0.664	0.027	0.610, 0.718	-0.297	0.003	-0.303, -0.297
Slope (degrees)	-0.021	0.002	-0.024, -0.018	-0.034	0.001	-0.036, -0.031

6.3.3 Vector density

Between August 2013 and December 2015, 6061 *An. balabacensis* were caught from 415 sampling nights from 183 unique locations. The median biting rate was 2 bites per night per individual, however this biting rate ranged from 0 – 44 bites per person per night. While results showed strong spatial autocorrelation and some monthly variation, no significant time lags or seasonal patterns were identified. Preliminary regression identified positive relationships between EVI and distance to forest with *An. balabacensis* density and a negative relationship with human population density (SI). There was no association between land use type and vector density. The final posterior rate ratio estimates after including spatial and temporal autocorrelation are presented in Table 6.8. Despite considerable uncertainty for some of the study area, the RMSE calculated from an independent sample of data was 3.51 for the best fitting model (Table 6.5).

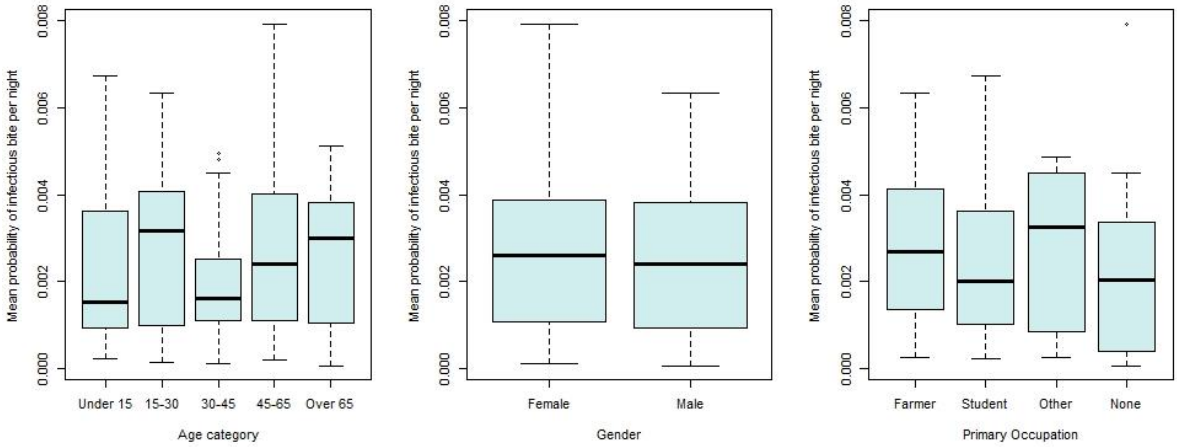
Table 6.8. Posterior rate ratio estimates and 95% Bayesian credible interval (BCI) for model 4

Covariate	95% BCI Rate Ratio			
	Mean	2.5%	50%	97.5%
Population density	0.967	0.918	0.970	1.003
EVI	1.484	0.472	1.300	3.548
Distance to forest (km)	1.190	0.377	1.061	2.746
Spatial range (km)	15.146	4.791	13.235	36.491

6.3.4 Individual exposure

For individuals included in the GPS tracking study in the Kudat intensive site, estimated exposure to infected mosquitoes varied markedly between individuals, with mean probabilities over the study period ranging from 6.9 x 10^{-5} to 0.008 bites from infected mosquitoes per night. Within the Kudat site, there were no clear differences in exposure between genders and no clear age patterns (Figure 6.8). Individuals with the highest predicted exposure primarily included farmers, although also included some students reporting spending time outside in the village during the evening. Overall, higher mean exposure rates were predicted in farmers and other occupations compared with students and unemployed individuals. Exposure was not correlated with the size of the home range or distance travelled.

Figure 6.8. Mean probability of infectious bites per night over the study period by demographic group



6.3.5 Community level exposure

Resource utilisation functions were fit for all members of the community in Matunggong, Kudat and used to develop predictive maps of community-wide space use (Figure 6.9). By overlaying these space use predictions on predicting vector biting rates and sporozoite rates, maps of exposure to infected vectors were generated (Figure 6.10). Despite substantial spatial and temporal heterogeneity, the highest exposure risks were found near forest edges and in close proximity to households (Figure 6.11).

Figure 6.9. Estimated total person-nights per grid cell per night for the entire community in Matunggong intensive site

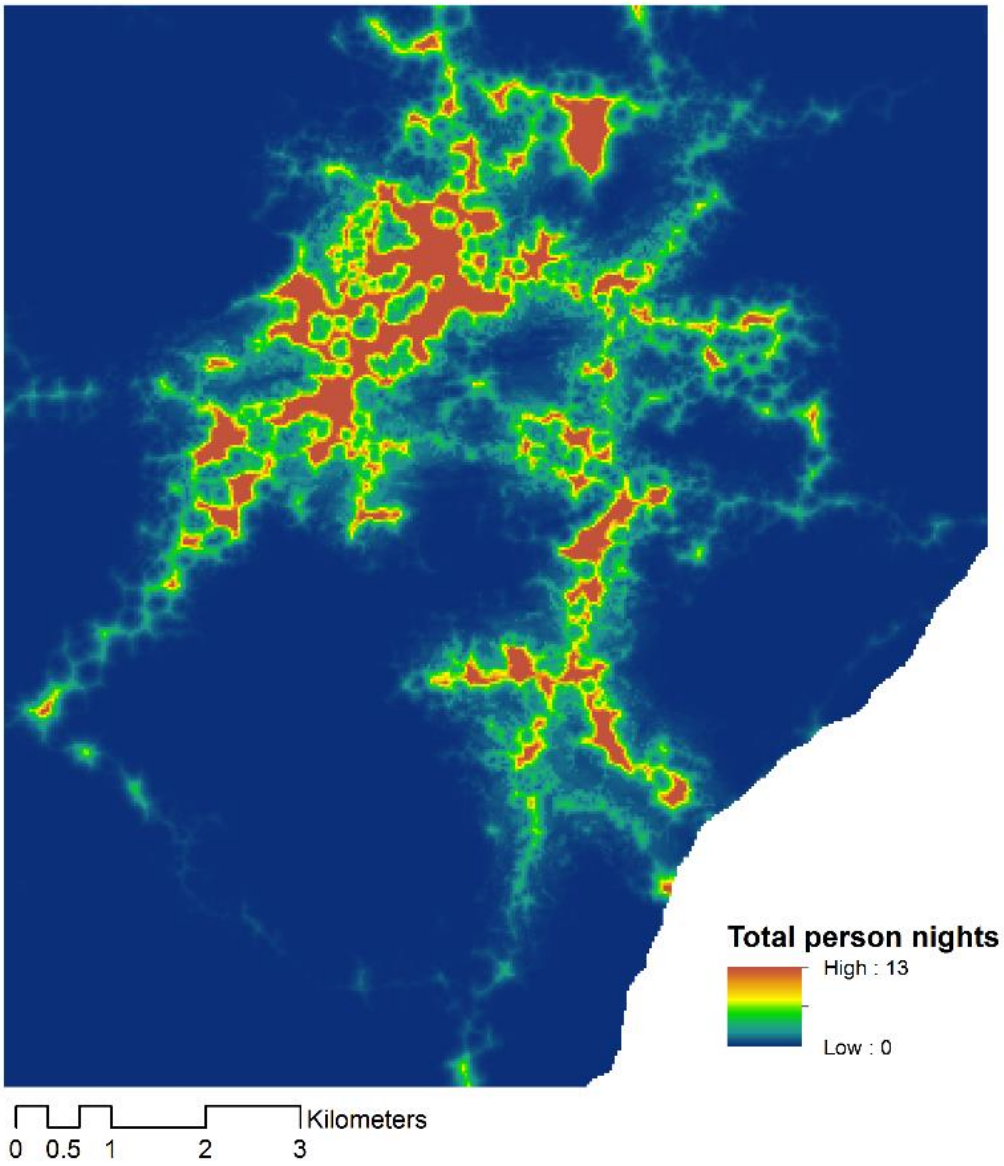


Figure 6.10. Median number of bites from infected *An. balabacensis* received by community members per grid cell per night over the study period

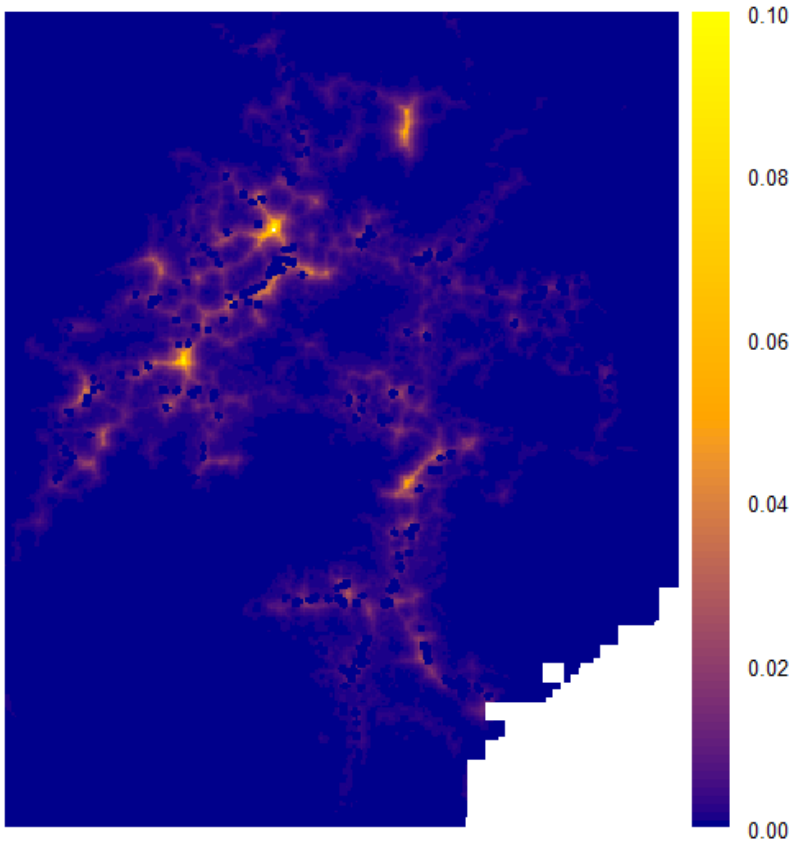
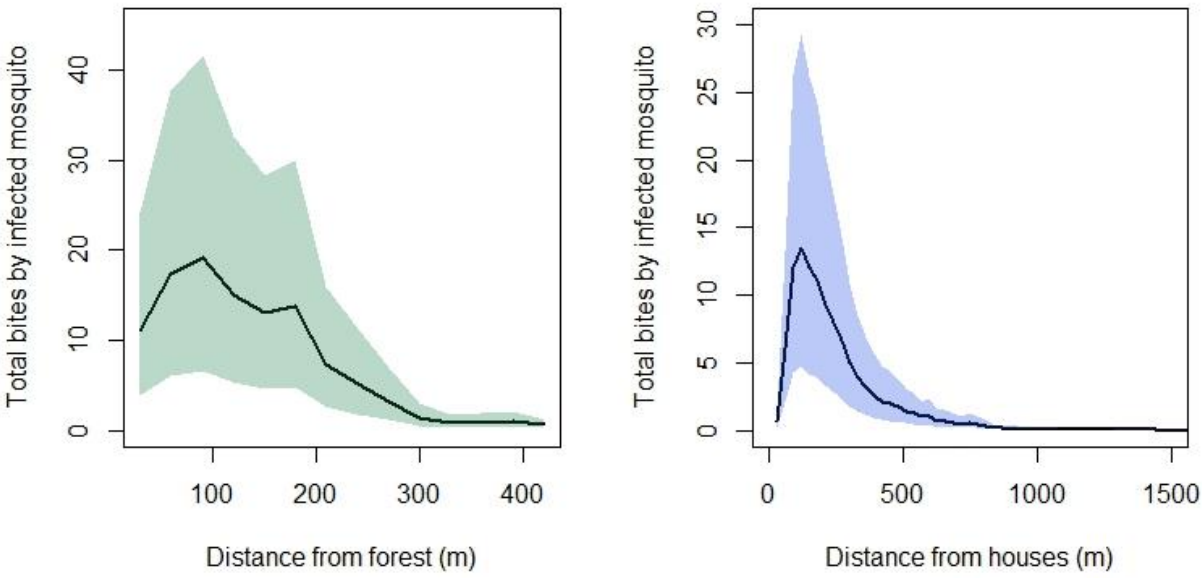


Figure 6.11. Total numbers of expected bites per night by distance from forest and houses (shaded areas represent 95% BCI)



6.4 Discussion and implications for future research

Results from this study highlight the importance of individual movement patterns in determining exposure to zoonotic and vector-borne diseases such as *P. knowlesi*. Although *P. knowlesi* has previously been associated with forest exposure (e.g. [103]) and higher biting rates have been reported in forest interiors ([152]), increased probability of people utilising areas surrounding households suggests the possibility of peri-domestic transmission and emphasising the need to incorporate human movement into disease models. This study additionally demonstrates the utility of ecological methods to understand human movement and identify geographical areas associated with higher contact with disease vectors.

Within these communities, movement patterns during peak vector biting times (6pm – 6am) were highly predictable, with spatial and environmental factors predicting close to 95% of the grid cells visited in both sites. However, despite this finding, substantial variation in exposure was predicted between individuals, suggesting some individual level variation during these times. Within the Matunggong site, there were no significant differences in exposure between men and women and individuals with high exposure risks were identified across all occupations and age groups. While this finding differs from clinical reports, a comprehensive survey in this community identified equal proportions of sero-positive men and women and data on asymptomatic infections suggests higher numbers of non-clinical infections in women [256, 271]. As this study was only designed to evaluate routine movements, the impact of less routine activities (such as infrequent hunting trips) cannot be determined and may contribute to the difference in clinical cases.

This study was additionally limited by the availability of vector data; as human landing catch data was assembled opportunistically, there was not uniform spatial and temporal coverage of the study site, potentially biasing results. This likely contributed to the lack of temporal autocorrelation detected and could be improved with routine human landing catches collected at regular time intervals in representative areas across the site. An additional limitation to estimating mosquito biting rates was the difficulty obtaining spatially and temporally resolute remote sensing data for predictors due to high cloud cover [272]. The limited vector data combined with non-stationarity of data on outlying islands precluded the development of mosquito models for Pulau Banggi. Other dynamic modelling approaches could be explored to refine biting rate estimates and develop predictions for other sites [273, 274].

As limited data was available and few sporozoite positive mosquitoes were identified, uniform estimates of sporozoite rates were used across the entire study site. These estimates could be improved to reflect spatial and temporal heterogeneity of infection levels by extending models to incorporate estimates of macaque density, infection levels and biting rates in different habitats [275]. Additionally, estimates of human risk could be refined to incorporate vector control measures and other factors which determine whether exposure to sporozoite positive mosquitoes leads to development of infection. Together, these data highlight the need for spatially resolute movement and vector data to incorporate heterogenous contact and refine malaria risk models.

Despite these limitations, this is the first large scale GPS tracking study to utilise ecological methods to predict resource utilisation during peak mosquito biting periods. This study highlights the importance of incorporating heterogeneous patterns of human space use into disease models, as the majority of human exposure may occur in areas with lower vector biting rates but much greater probabilities of human use.

7 Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia

7.1 Background and implications for future research

A key knowledge gap in understanding the epidemiology of *P. knowlesi* is the extent of asymptomatic parasitemia and what proportion of infected individuals are captured by hospital-based surveillance systems. In order to assess whether asymptomatic knowlesi infections were present in the study areas, individuals residing in the same household or kampung as symptomatic cases were screened for malaria during a population-based case control study [247]. While this population is not representative of the wider population within the study area and results could not be used to estimate prevalence within the community, it was assumed the probability of detecting asymptomatic infections within this population would be higher as they were residing in an area at a time when transmission was documented. The presence of asymptomatic infections, particularly in women and groups comprising a small proportion of clinical cases, highlights the need for population-based community surveys to assess the prevalence of knowlesi infection and exposure. Additionally, the identification of household clusters of infections suggests peri-domestic transmission as well as the possibility of human-to-human transmission. Although the current dataset could not be used to investigate the role of peri-domestic transmission, subsequent entomological studies and individual risk factors identified by the case control study suggested the environment around the household could influence infection risk and should be explored in future cross-sectional surveys [87, 103].

A critical limitation in available diagnostic tools became apparent during the analysis of these samples. All individuals included were screened by microscopy and a genus-specific *Plasmodium* PCR assay [276]. While initially 18% (206/1147) individuals were found to be *Plasmodium* positive, only 20 (1.7%) were identified as positive for *P. knowlesi*, the most common malaria species within this region [277]. An additional three molecular assays, including two real time PCR assays, were used to screen for *P. knowlesi* [278-282]. Although some clear positives were identified (64 *P. knowlesi* positive by at least two assays), molecular results were not always in agreement. This could have been due to the limited amount of DNA used, crude extraction method (by Chelex) or extremely low parasite densities (only 1 individual was positive by microscopy). Latent class analysis was used to estimate the proportion of samples positive for *P. knowlesi* and the sensitivity and specificity of

assays. Results highlight differences between assays as well as the need for improved knowlesi-specific molecular diagnostics. As all *P. knowlesi* infections were detected by the *Plasmodium* genus-level PCR, we chose to apply this method for subsequent surveys. Additional work was undertaken to identify *P. knowlesi* specific molecular diagnostics for low density infections.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Kimberly Fornace
Principal Supervisor	Prof Chris Drakeley
Thesis Title	Spatial epidemiology of P. knowlesi in Northern Sabah, Malaysia

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Journal of Infectious Diseases		
When was the work published?	March 2016		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*		Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I led and conducted all data analysis and wrote the paper. Other authors (Betson, Nuin, Ying) conducted all laboratory analysis.

Student Signature: _____ **Date:** 9 August 2018

Supervisor Signature: _____ **Date:** 09/08/18

7.2 Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia (Paper 3)

Authors: Kimberly M Fornace[1*], Nor Afizah Nuin[2], Martha Betson[3,$], Matthew J Grigg[4,5], Timothy William[4,6,7], Nicholas M Anstey[4,5], Tsin W Yeo[4,5], Jonathan Cox[1], Lau Tiek Ying[2] and Chris J Drakeley[1]

Affiliations:

1. London School of Hygiene and Tropical Medicine, London, UK
2. Biotechnology Research Institute, Universiti Sabah Malaysia, Kota Kinabalu, Malaysia
3. Royal Veterinary College, London, UK
4. Infectious Diseases Society Sabah – Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Malaysia
5. Menzies School of Health Research, Darwin, Australia
6. Clinical Research Centre, Sabah Department of Health, Kota Kinabalu, Malaysia
7. Jesselton Medical Centre, Kota Kinabalu, Malaysia

[$] Current address: University of Surrey, Guilford, UK

[*] Corresponding author

7.2.1 Abstract

Although asymptomatic carriage of human malaria species has been widely reported, the extent of asymptomatic, submicroscopic parasitemia for *Plasmodium knowlesi* is unknown. This study sampled individuals residing in households or villages of symptomatic malaria cases with the aim of detecting submicroscopic *P. knowlesi* in this population. Four published molecular assays were used to confirm the presence of *P. knowlesi.* Using latent class analysis the estimated proportion of asymptomatic individuals was 6.9% (5.6 -8.4%). This study confirms the presence of a substantial number of asymptomatic mono-infections across all age groups; further work is needed to estimate prevalence in the wider community.

7.2.2 Introduction

Malaria epidemiological surveys have reported a substantial proportion of individuals with low density infections that are not detectable by conventional microscopy [110]. Meta-analysis of studies that used molecular amplification techniques suggest these submicroscopic infections represent on average 50% of malaria infections. This proportion can be as high as 80% in low malaria transmission settings where community parasite prevalence by microscopy is less than 10% [111].

Although described for human malaria species, particularly *Plasmodium falciparum* and *P. vivax*, limited data are available on submicroscopic carriage of zoonotic malaria species. *P. knowlesi*, a zoonotic malaria parasite maintained by macaques, has been described throughout Southeast Asia and is now the most common cause of human malaria in Malaysian Borneo [14]. To our knowledge, limited asymptomatic *P. knowlesi* infection has only been identified in two studies in Vietnam and submicroscopic parasitemia in people has not been widely reported. During two cross-sectional malariometric surveys in Vietnam, three individuals were identified as *P. knowlesi* positive using molecular techniques [31]. These infections were found in two young children (ages 2 and 3 years) and an adult man (age 27), all of whom were asymptomatic at the time and for six months following the survey. Similarly, multiple co-infections with *P. knowlesi* were detected in younger age groups through active case detection [112].

Despite rarely causing clinical disease, submicroscopic malaria infections can contribute to malaria transmission. Experimental evidence has demonstrated that individuals with submicroscopic infections are capable of infecting mosquitoes; while these individuals may infect fewer mosquitoes

than individuals with higher parasite counts, the high numbers of individuals with low-density infections may result in them contributing substantially to malaria transmission [283]. Understanding the prevalence of these infections and the extent to which they contribute to malaria transmission is critical for designing effective malaria control programmes.

Data on submicroscopic parasitemia are also needed to better understand disease progression. Both parasite and host factors will influence whether infections remain asymptomatic or become symptomatic and potentially life threatening. The data on *P. knowlesi* infection dynamics in exposed populations are very limited. The parasite has a distinct 24 hour asexual development cycle with common severe disease and case fatality rates similar to those recorded for *P. falciparum* [18]. This potential for rapid disease progression makes improving our understanding of *P. knowlesi* particularly important.

This study aimed to detect potential asymptomatic *P. knowlesi* cases by screening individuals residing in the same households and villages of clinical *P. knowlesi* cases recruited during a population-based case control study in an area of known *P. knowlesi* transmission in Northwestern Sabah, Malaysia [247]. *P. knowlesi* is the main cause of clinical human malaria in this region and clustering of cases at household level has previously been reported [37]. Individuals were screened by microscopy and multiple molecular methods to determine whether asymptomatic *P. knowlesi* carriage is present in this population and estimate the proportion of infected individuals.

7.2.3 Methods

Study design

The study sites in Kudat and Kota Marudu districts, Sabah, Malaysia have been described elsewhere [247]. The area is served by two district hospitals and numerous referral clinics. Malaria is a notifiable disease in Malaysia and all malaria patients have access to treatment free of charge.

As part of the case control study, consenting clinical cases positive for any species of malaria by microscopy were recruited at district hospitals and visited at their homes within two weeks of initial infection detection. Community controls were randomly selected afebrile individuals residing in the same village as malaria cases for the previous three weeks as described by [247]. Blood samples were also collected from all consenting individuals residing in the same household as both cases and controls; these included a blood smear for detection of malaria parasites by microscopy and whole

blood stored on filter paper (3MM; Whatman, UK) and in a 500 µl EDTA tube (Becton-Dickinson, NJ, USA). Demographic details were recorded for all individuals residing in the same households or villages as cases and individuals were asked about their history of fever. Malaysian health policy mandates that all cases of malaria are referred to the district hospital for treatment, and the case-control study prospective surveillance system [247] enabled detection of subsequent clinical disease among asymptomatically infected controls.

Ethics

This study was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health, the Health Research Ethics Committee of Menzies School of Health Research, Australia and the Research Ethics Committee of London School of Hygiene and Tropical Medicine (NMRR-12-537-12568). Written informed consent was obtained from all participants in this study.

Detection of malaria infection

Thick and thin blood smears were examined by a trained malaria microscopist. DNA was extracted from 10 µl red blood cell pellets using the Chelex-100 boiling method and nested PCR assay targeting the *Plasmodium* small subunit ribosomal RNA (ssRNA) gene was performed to identify *Plasmodium* genus positive samples as described by [276] and in Supplementary Information. Primers targeting a region of the nest 1 ssRNA product were then used to detect *P. knowlesi* [277]. Positive controls of confirmed clinical cases of *P. knowlesi* and other species were used for all PCR reactions.

Due to difficulties in determining species in some of the genus positive samples, additional methods were used on a subset of 374 samples. A nested PCR assay targeting the cytochrome B gene of *P. knowlesi* was used to identify *P. knowlesi* positives [279]. Samples were also run using two real-time PCR assays, one targeting the *P. knowlesi* ssRNA gene and strain H chromosome 13 plasmepsin [278, 280], as described in Supplementary Information.

Statistical analysis

Data were analysed using R statistical software version 3.1.1 (The R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org). In the absence of a gold standard for *P. knowlesi* species specific diagnosis, latent class analysis was used to estimate the proportion of infected individuals and the sensitivity and specificity of these tests using the randomLCA package in

R (v 1.0.2) [284]. Test type was included as a random effect to account for conditional dependence between assays. Competing models with one to three latent classes, representing possible diagnostic classes, were developed and model selection was based on Bayes information criterion (BIC). Posterior probabilities for each latent class were estimated, using parametric bootstrap methods to estimate confidence intervals. Individuals were assigned to infected or uninfected classes based on the predicted probabilities.

7.2.4 Results

A total of 1147 blood samples were collected from December 2012 until May 2014. Only one individual was microscopy positive but did not report a history of fever. Initial genus PCR results using the ssRNA primers as described [276] found 18% (206/1147) of these individuals positive for *Plasmodium* species. From this sample set, 1.7% (20/1147; 95% CI: 1.0% - 2.5%) were confirmed to be *P. knowlesi* positive using the *P. knowlesi* specific ssRNA primers [277]. As *P. knowlesi* is the predominant cause of human malaria in this area and many *Plasmodium* genus positive samples could not be identified, alternate assays were run on all of the genus positive samples (n=206) and a subset of genus negative samples (n=168) [278-280].

The genus-specific nested PCR was highly sensitive, detecting all of the *P. knowlesi* infections identified by every method (sensitivity 100%, 95% CI: 95%- 100%). A total of 9.8% (112/1147) samples had at least one positive *P. knowlesi* test result and 3.0% (34/1147) had three or more positive test results. The diagnostic sensitivity and specificity of each *P. knowlesi* specific assay for this population was estimated using latent class analysis (Table 1), demonstrating that the standard nested PCR used for *P. knowlesi* identification had lower sensitivity to detect submicroscopic infections (estimated sensitivity 15%; 95% CI: 8-23%) than the other molecular assays used. Using these estimates of sensitivity and specificity, the infection prevalence in this population was estimated at 6.9% (95% CI: 5.6 – 8.4%).

Table 7.1. Prevalence of latent classes and conditional probabilities for latent class model[1]

Assay	Result	N [2]	Latent classes	
			Infected	Non-infected
			Conditional probability (%)	
ssRNA Nested PCR	Positive	20	15.0	2.7
	Negative	1127	85.0	97.3
CytB Nested PCR	Positive	67	59.1	6.8
	Negative	305	40.9	93.2
ssRNA Real Time PCR	Positive	71	87.9	2.4
	Negative	264	12.1	97.6
Plasmepsin Real Time	Positive	60	81.3	0.9
PCR	Negative	229	18.7	99.1

1. Latent classes were assigned using the results from all available assays
2. All assays were not performed on all samples due to insufficient DNA availability

The majority (73/84; 87%) of *P.knowlesi* infected individuals predicted by the latent class analysis did not report a history of a fever (Table 2), similar to the non-infected population (947/1063; 89%), and none of these individuals reported to hospital with clinical malaria following this survey. Infected individuals were present in both households of cases and controls and in all age groups (Table 2). Households with multiple infected individuals were identified, including 11 households with two asymptomatic infected individuals and two households with three infected individuals. All but one of these households also reported a symptomatic case. The majority of infections with *P. knowlesi* were mono-infections; 5% (4/84) were co-infections with *P. falciparum* and 2% (2/84) with *P. malariae*.

Table 7.2. Population characteristics and infection prevalence based on probabilities of latent classes

Demographic characteristics	Study population % (n = 1147)	*P. knowlesi* infected[1] % infected (95% CI)
District		
Kota Marudu	39.8% (457)	9.0% (6.4-11.6)
Kudat	60.1% (690)	6.2% (4.4 – 8.0)
Residence		
Case household	44.6% (512)	8.4% (6.0-10.8)
Control household	55.4% (635)	6.5% (4.6-8.4)
Gender		
Men	46.6% (535)	6.2% (4.2-8.2)
Women	53.4% (612)	8.3% (6.1-10.5)
Age		
Under 15	28.2% (323)	9.6% (6.4-12.8)
15 to 45	42.9% (492)	6.7% (4.5-8.9)
46 to 60	16.9% (194)	6.7% (8.3-17.7)
Over 60	12.0% (138)	5.1% (1.4-8.8)
Self-reported fever in past month		
Fever	11.1% (127)	8.7% (3.8-13.6)
No fever	87.3% (1001)	7.3% (5.7-8.9)
Don't know	1.6% (19)	0%
Microscopy positive		
Positive	0.1% (1)	100%
Negative	99.9% (1147)	7.2% (5.7-8.7)

[1] Based on probability of class membership (uninfected median p= 0.998; infected, median p = 0.997)

7.2.5 Discussion

This is the first study to describe a high level of submicroscopic, asymptomatic *P. knowlesi* carriage in an exposed human population. Although this is not a true prevalence survey and many of the infections detected were from individuals residing in the same household as symptomatic cases, infections were also found in unrelated individuals residing in the same village during that time. Moreover, a high proportion of infections were detected in children under the age of 15 and women, groups not previously considered to be at high risk for *P. knowlesi* infections. Although a small proportion of infections reported history of fever, none of these infections were reported as developing into symptomatic malaria. However, longitudinal studies are needed to fully understand the disease progression and potential development of acute disease. Given the on-going case recruitment in these health facilities [247], mandatory referral policy and the previous history of these households reporting to the clinic, it is unlikely that these infections developed into clinical malaria, suggesting there is a substantial number of asymptomatic *P. knowlesi* infections in the community. Further, the distribution of these infections in different demographic groups may not be captured by the passive health surveillance systems.

The majority of infections were submicroscopic and could not be detected by conventional malaria microscopy. Unlike previously reported submicroscopic infections, the majority (93%) were not co-infected with other species. Additionally, the inconsistent results obtained from multiple published and validated molecular assays demonstrate that these infections may be at or below the level of detection for assays developed for use primarily on human clinical samples. Using this relatively small blood volume with very low density infections means these samples may be missed even in repeated PCR assays. Understanding the wider community prevalence will require better optimised molecular assays as well as population level surveys.

It remains unknown if and how much humans contribute to the infectious reservoir for *P. knowlesi* transmission. While the submicroscopic parasite densities of other malaria species have been shown to be capable of infecting mosquitoes, all experimental infections with *P. knowlesi* have been from clinical malaria cases [14]. The identification of multiple human infections in different demographic groups within limited geographical areas including households suggests the possibility of peri-domestic transmission. Although this cannot be directly attributed to human to human transmission, it is probably indicative of exposure to the same infected vectors. Further studies could use molecular typing techniques to identify whether strains of *P. knowlesi* identified within the same areas are likely to have come from a common source. Entomological and primatological studies are also needed to evaluate the presence of potential *P. knowlesi* vectors in these village environments and the proximity of infected reservoirs.

As this study included only individuals in the same household or village as a symptomatic case, these data cannot be used to generalise about the community prevalence of *P. knowlesi* or to further understand spatial and temporal patterns of *P. knowlesi* infection. Similarly, as samples were collected at a single time point for each individual, the duration and fluctuations in parasite densities over time cannot be determined. Longitudinal data as well as data on treatment seeking behaviours for different demographic groups is required to determine factors contributing to whether specific groups are under-represented by hospital surveillance systems.

Despite these limitations, this study does illustrate the presence of asymptomatic *P. knowlesi* mono-infections within communities located in *P. knowlesi* endemic areas and highlights the need for further studies to evaluate population wide prevalence. Current molecular tools are still limited for detection of very low parasite densities and these need further evaluation and optimisation.

Additional community-based surveys are currently planned to evaluate the prevalence of these infections in the wider community.

7.2.6 Supplementary information:

Molecular detection of Plasmodium:

Multiple molecular methods were used to identify *P. knowlesi* infections. DNA was extracted from 10µl red blood cell pellets using the Chelex-100 boiling method. Initial screening for *Plasmodium* infections was done using a nested PCR assay targeting the ssRNA [276]. This nested PCR assay used the genus-specific primers rPLU1 (5'-TCA AAG ATT AAG CCA TGC AAG TGA-3') and rPLU5 (5'-CCT GTT GTT GCC TTA AAC TTC-3') for nest 1 using 4 µl of DNA and rPLU3 (5'-TTT TTA TAA GGA TAA CTA CGG AAA AGC TGT-3') and rPLU4 (5'-TAC CCG TCA TAG CCA TGT TAG GCC AAT ACC-3') for nest 2 using 2 µl of template [276]. Thermal cycling conditions for primary and nested PCRs were 35 cycles at 94°C, 60°C and 72°C. Products were visualised on a 2% agarose gel. The sensitivity of this assay as a screening tool was evaluated by screening all genus positive samples and a subset of 180 genus negative samples.

Detection of Plasmodium species:

Species specific primers targeting the ssRNA gene were used as described by [276], using the same conditions as nest 1. As cross reactivity between *P. vivax* and *P. knowlesi* has been reported, primers targeting an alternate region of the nest 1 ssRNA product were used for *P. knowlesi*; these comprised PkF1140 (5'-GATTCATCTATTAAAAATTTGCTTC-3') and PkR1150 (5'-TCTTTTCTCTCCGGAGATTAGAACTC-3') [277]. Conditions for nest 2 of this PCR were 35 cycles at 50°C, 72°C and 94°C using 2µl of template DNA and all results were visualised on a 2% agarose gel. Positive controls for confirmed clinical cases of *P. knowlesi* and other species were used for all PCR reactions.

Due to difficulties in identifying some of the genus positive samples to species, additional methods were used on a subset of 374 samples, including 168 genus negative samples and 206 genus positive samples. Previously reported minimum detection thresholds are included in Supplementary Table 1.

Supplementary Table 1: Minimum detection thresholds of *P. knowlesi* assays

Assay	Study population
ssRNA Nested PCR	1 to 10 parasite genomes per reaction [277]
CytB Nested PCR	Not reported [279]
ssRNA Real Time PCR	10 gene copies per reaction [280]
Plasmepsin Real Time PCR	10 to 100 gene copies per µl [278], 1-6 parasites per µl [285]

A nested PCR assay targeting the cytochrome B gene of *P. knowlesi* was used to identify *P. knowlesi* positives [279]. The primers PCBF (5'-ATGCTTTATTATGGATTGGATGTC-3') and PCBRed (5'-ACATAATTATAACCTTACGGTCTG-3') were used for the first nest and PkCB (5'-TATTCTTCTTTAGTGGATTATTTA-3') and PkCBed (5'-GTATTGTTCTAATCAGTGTA-3') were used for the second nest [281, 282]. Thermocycler conditions were 95°C, 50°C, 72°C, with 35 cycles for the first nest and 25 cycles for the second nest.

Samples were also run using two real-time PCR assays, one targeting the *P. knowlesi* ssRNA gene and one targeting strain H chromosome 13 plasmepsin [278, 280]. For the ssRNA real-time assay, the primers Plasmo 1 [5'-GTTAAGGGAGTGAAGACGATCAGA-3'] and Plasmo 2 [5'AACCCAAAGACTTTGATTTC TCATAA-3'], and a FAM-labelled Pk probe [5'-FAM-CTCTCCGGAGATTAGAACTCTTAGATTGCT -BHQ1-3'] were employed [6]. Real-time PCR reactions contained 200 nM of each primer, 80 nM probe, SensiFAST™ Probe No-ROX mastermix (Bioline, London, UK) and 2 l of DNA template in a total volume of 20 l. Reactions were performed using the CFX96 real-time PCR detection system (Biorad, Hemel Hempstead, UK) under the following cycling conditions: initial denaturation at 95°C for 5 min followed by 45 cycles of 95°C for 10 sec and 60°C for 30 sec. For the plasmepsin real-time assay, the forward and reverse primer sequences were 5'- TAACATGGTAATCATACATAAGG-3' and 5'-TAAGGAAATGCCAACTCTTG-3', respectively, and the probe was FAM -TCAGCCAACAACACTTACAG-BHQ1 [7]. Reaction conditions were as for the ssRNA assay except the probe was used at 200 nM. The cycling conditions were: initial denaturation at 95°C for 5 min followed by 40 cycles of 95°C for 10 sec and 55°C for 30 sec. Ct values of 45 and 40 were used for the ssRNA and plasmepsin assays respectively.

8 Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines

8.1 Background and implications for future research

The identification of asymptomatic *P. knowlesi* infections highlighted the need for community surveys to characterise patterns of *P. knowlesi* infection and exposure. While most clinical cases were adult men, a relatively high proportion of asymptomatic *P. knowlesi* infections were in women and children, suggesting passive surveillance systems may not fully reflect exposure in the community [103, 286]. Assessing community-level exposure is further complicated by the low *P. knowlesi* incidence and poor sensitivity of diagnostics. Serological methods, measuring species-specific antibodies to malaria parasites, have been used to characterise malaria epidemiology in settings where the probability of detecting infections is low [114]. The recent development of species-specific recombinant antigens for *P. knowlesi* [119] allowed the application of these techniques within these study sites.

This paper was submitted as a dual submission with a paper outlining the development of *P. knowlesi*- specific antigens [119] (included in Annex 1). During the submission process, a key concern of reviewers was the sensitivity and specificity of these antigens to identify *P. knowlesi* exposed individuals. To illustrate the potential utility of these antigens for sero-epidemiological studies, we developed predictive models classifying individual seropositivity based on these results. While individual immune responses are highly heterogenous for a range of factors, these models correctly classified 88% of individuals (IQR: 84-90%). We then expanded this analysis to evaluate contributions of all four antigens to model predictions and identify the candidate marker of serological exposure used by this study [119]. To further address this issue, we included additional details on the antigen used within this manuscript and also added supplementary information showing the magnitude of individual responses to all antigens by age group.

Results from these surveys demonstrate the utility of these serological methods and illustrate differences in knowlesi transmission from communities within the same geographic region. These data also highlight differences in transmission and risk factors between non-zoonotic malaria and *P. knowlesi*. However, due to the limited sampling design (comprehensive sampling of selected case

study communities), results cannot be used to generalise about the wider population. Households within the same communities were in close proximity, with limited variation in potential spatial or environmental risk factors. Additionally, the use of seroreactivity as a primary endpoint complicates analysis of environmental risk factors; due to the longevity of antibody responses, exposure may be related to past rather than current land use. These initial results informed the design of a larger population-based cross-sectional survey sampling across different ecological gradients. Full questionnaires for both surveys are included in Annex 2. Estimates of exposure and infection rates from case study communities in Sabah were used to calculate sample sizes and land use classification was expanded to encompass the range of ecologies within this wider geographical area. Further analysis of serological data, classifying exposure as recent or past exposure (e.g. [287]), could be explored to understand how current environmental conditions contribute to disease risks.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Kimberly Fornace
Principal Supervisor	Prof Chris Drakeley
Thesis Title	Spatial epidemiology of P. knowlesi in Sabah, Malaysia

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	PLoS Neglected Tropical Diseases		
When was the work published?	June 2018		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I designed the study, performed all data analysis and wrote the paper. Other authors performed all laboratory analyses and revised the manuscript.

Student Signature: _____ **Date:** 9 August 2018

Supervisor Signature: _____ **Date:** 09/08/18

8.2 Exposure and infection to *Plasmodium knowlesi* in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines (Paper 4)

Authors: Kimberly M Fornace[1*], Lou S Herman[1], Tommy R Abidin[2], Tock Hing Chua[3], Sylvia Daim[3], Pauline J Lorenzo[4], Lynn Grignard[1], Nor Afizah Nuin[5], Lau Tiek Ying[5], Matthew J Grigg[2,6], Timothy William[2,7], Fe Espino[4], Jonathan Cox[1], Kevin K A Tetteh[1], Chris J Drakeley[1]

1. Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK
2. Infectious Diseases Society Kota Kinabalu – Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Malaysia
3. Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia
4. Research Institute of Tropical Medicine, Department of Health, Muntinlupa City, Philippines
5. Biotechnology Research Institute, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia
6. Menzies School of Health Research and Charles Darwin University, Darwin, Australia
7. Jesselton Medical Centre, Kota Kinabalu, Malaysia

* Corresponding author

8.2.1 Abstract

Background

Primarily impacting poor, rural populations, the zoonotic malaria *Plasmodium knowlesi* is now the main cause of human malaria within Malaysian Borneo. While data is increasingly available on symptomatic cases, little is known about community-level patterns of exposure and infection. Understanding the true burden of disease and associated risk factors within endemic communities is critical for informing evidence-based control measures.

Methodology/ Principal findings

We conducted comprehensive surveys in three areas where *P. knowlesi* transmission is reported: Limbuak, Pulau Banggi and Matunggung, Kudat, Sabah, Malaysia and Bacungan, Palawan, the Philippines. Infection prevalence was low with parasites detected by PCR in only 0.2% (4/2503) of the population. *P. knowlesi* PkSERA3 ag2 antibody responses were detected in 7.1% (95% CI: 6.2 – 8.2%) of the population, compared with 16.1% (14.6 – 17.7%) and 12.6% (11.2 – 14.1%) for *P. falciparum* and *P. vivax* as measured by species-specific AMA1 and MSP1 antibody responses. Sero-prevalence was low in individuals <10 years old for *P. falciparum* and *P. vivax* consistent with decreased transmission of non-zoonotic malaria species. Results indicated marked heterogeneity in transmission intensity between sites and *P. knowlesi* exposure was associated with agricultural work (OR 1.63; 95% CI 1.07-2.48) and higher levels of forest cover (OR 2.40; 95% CI 1.29-4.46) and clearing (OR 2.14; 95% CI 1.35-3.40) around houses. Spatial patterns of *P. knowlesi* exposure differed from exposure to non-zoonotic malaria and *P. knowlesi* exposed individuals were younger on average than individuals exposed to non-zoonotic malaria.

This is the first study to describe serological exposure to *P. knowlesi* and associated risk factors within endemic communities. Results indicate community –level patterns of infection and exposure differ markedly from demographics of reported cases, with higher levels of exposure among women and children. Further work is needed to understand these variations in risk across a wider population and spatial scale.

8.2.2 Author summary

Plasmodium knowlesi is a species of malaria parasite found in wild macaque populations which is now the main cause of human malaria in Malaysian Borneo. Spread from macaques to people through infected mosquitoes, human *P. knowlesi* malaria cases have primarily been reported in adult

men working in forests or plantations. However, little data is available on the extent of asymptomatic infections or people exposed to *P. knowlesi* not reporting to clinics. We conducted comprehensive surveys of three case study communities in Malaysian Borneo and Palawan, the Philippines with varying numbers of *P. knowlesi* cases reported. In addition to testing for infection, we measured species-specific antibody responses to *P. knowlesi* and other malaria species to identify exposed individuals. Few asymptomatic infections were detected and varying levels of *P. knowlesi* exposure was detected between sites. *P. knowlesi* exposure was identified in both men and women and associated with farm work and forest and clearing around the house. Spatial patterns and risk factors for *P. knowlesi* differed from other malaria species, highlighting the need for *knowlesi* specific disease control measures. Results suggest more people are exposed to *P. knowlesi* than are identified at clinics and exposure to *P. knowlesi* may occur in different demographic groups and geographic areas than previously reported.

8.2.3 Introduction

After the initial recognition of a large number of human cases of the zoonotic malaria *Plasmodium knowlesi* in 2004 and advent of routine diagnosis of malaria cases by molecular methods, increasing numbers of human *P. knowlesi* cases have been reported in Southeast Asia and *P. knowlesi* is now the most common cause of human malaria in Malaysian Borneo [11, 13, 41]. Although regional control programmes have reduced the incidence of other malaria species in Malaysia and the Philippines, such as *P. falciparum* and *vivax,* the emergence of *P. knowlesi* presents a challenge to malaria elimination programmes. Despite increasing amounts of data available for symptomatic malaria cases presenting at hospital facilities, little is known about patterns of *P. knowlesi* exposure and infection at a community level [26].

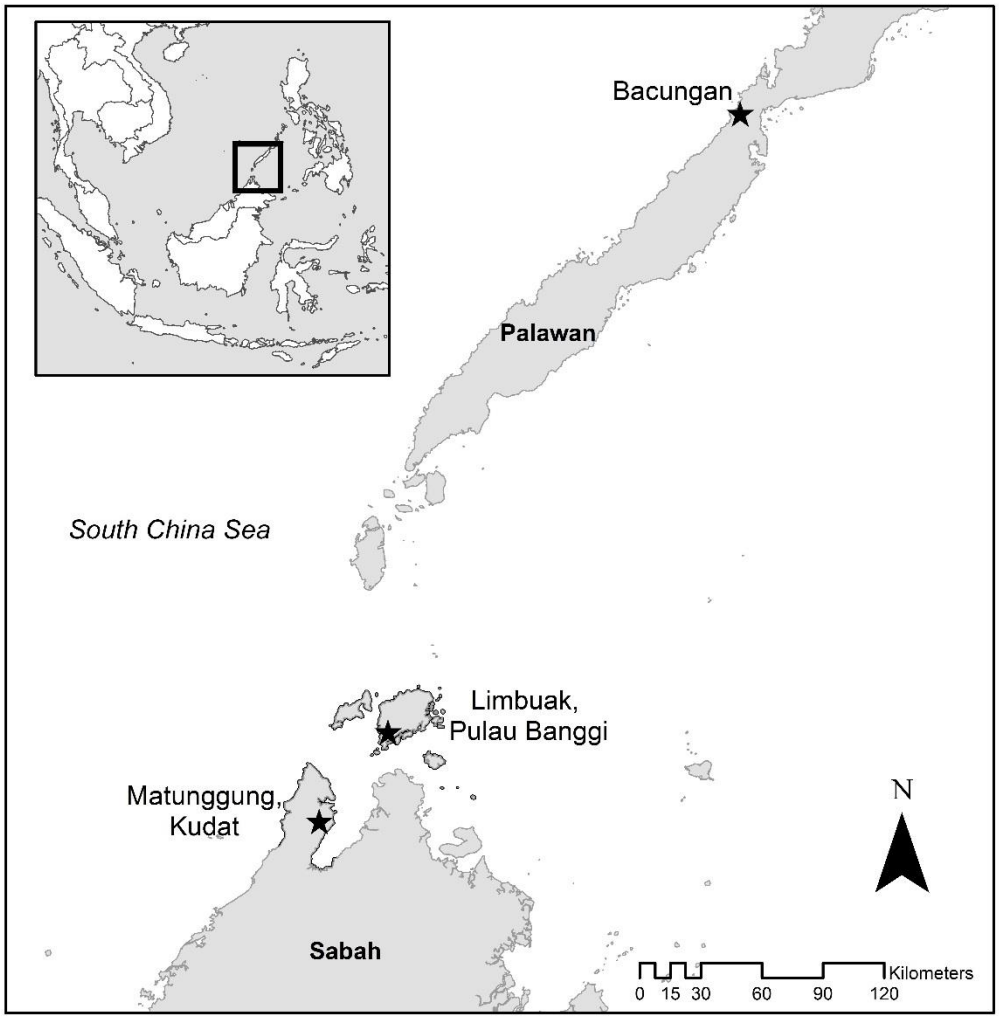
Effectively targeting resources to identify and control *P. knowlesi* requires a detailed understanding of environmental and social risk factors. Carried by long and pig-tailed macaques (*Macaca fasicularus* and *M. nemestrina*), environmental changes affecting contact between people, mosquito vectors and simian hosts are believed to contribute to this apparent emergence of *P. knowlesi* in people [14, 75]. *Anopheles balabacensis,* the main knowlesi vector, has been associated with forest environments but is also found in peridomestic and agricultural areas [87, 152]. Associations between deforestation and increases in village-level incidence have been shown for clinical cases but this may not fully reflect exposure in the wider community [255]. Additionally, multiple studies have reported asymptomatic *P. knowlesi* infections, including in women and children, demographic groups comprising a minority of cases reported to facilities [31, 112, 113, 286, 288].

Patterns of community-level exposure can be assessed by the prevalence of specific antibodies against malaria parasites; these antibodies reflect exposure to previous infection and can be used to characterise the level of transmission and identify areas or groups with higher transmission [115]. These serological markers may be particularly useful in low transmission settings, where the probability of detecting infections is low [114]. Seroconversion rates derived from age specific sero-prevalence have also been shown to be closely correlated with more traditional measures of malaria transmission intensity, such as entomological inoculation rates or parasite prevalence, and can be used to identify differences in spatial patterns in transmission [116, 117]. Further, as these antibody responses represent exposure over time, longer term transmission patterns and temporal changes in transmission can be evaluated [289]. There are an increasing

number of reagents for serological studies available for both *P .falciparum* and *P. vivax* e.g. [117, 118, 287]but antigens specific for *P.knowlesi* have only recently been described[119].

This study aimed to characterise these community level patterns of serological exposure to and prevalence of asymptomatic parasitemia of *P. knowlesi* and other malaria species in three case study communities where *P. knowlesi* transmission has been reported; a largely deforested and highly fragmented site at Matunggong, Kudat, an area with large patches of secondary forest bordering large scale clearing for an oil palm plantation in Limbuak, Pulau Banggi in Sabah, Malaysia and an area with intact secondary forest and some remaining primary forest in Bacungan, Palawan, The Philippines (Figure 8.1). These areas were selected as areas representative of locations were *P. knowlesi* transmission is occurring based on district hospital reports and were the sites of integrated entomology, primatology and social science studies within a wider research programme on risk factors for *P. knowlesi* (http://malaria.lshtm.ac.uk/MONKEYBAR). *P. knowlesi* is the main cause of reported human malaria in both the Matunggong and Limbuak sites while only few sporadic *P. knowlesi* cases have been reported from Bacungan [27, 37, 247]. Based on reporting of symptomatic cases to the national malaria programmes, the annual parasite incidence per 1000 people for *P. knowlesi* in 2014 was 12 for Matunggong, 2 for Limbuak and 0 for Bacungan.

Figure 8.1. Study site locations in Matunggong, Kudat and Limbuak, Pulau Banggi in Sabah, Malaysia and Bacungan, Palawan, Philippines



8.2.4 Methods

Ethics approval and informed consent

This study was approved by the Medical Research Sub-Committee of the Malaysian Ministry of Health (NMRR-14-713-21117), the Institutional Review Board of the Research Institute for Tropical Medicine, Philippines and the Research Ethics Committee of the London School of Hygiene and Tropical Medicine (8340). Written informed consent was obtained from all participants or parents or guardians and assent obtained from children under 18 in this study and all methods were performed in accordance with relevant guidelines and regulations.

Sampling methods

This study involved comprehensive sampling of all individuals residing within the study areas. Study sites were selected based on the locations of previously reported clinical *P. knowlesi* cases and all households within these communities were enumerated and geo-located. All individuals were asked to participate in the study and consenting individuals were interviewed on demographic characteristics, movement patterns, malaria prevention methods and land use practices. Individuals were excluded if they were less than 3 months old, had not primarily resided in the area for the past month or could not be reached after three attempts to contact them, including during evenings and weekends. Finger-prick blood samples were collected to test for malaria infection and exposure; these included blood smears to detect malaria parasites by microscopy and approximately 200µl whole-blood specimens collected in a tube containing EDTA (Becton-Dickinson, Franklin Lakes, New Jersey) and three 20µl spots stored on filter paper (3MM, Whatman, Maidstone, United Kingdom). Filter paper was dried and stored with desiccant at 4°C.

Detection of malaria infection

All blood smears were examined by trained malaria microscopists. DNA was extracted from filter paper or 10 µl blood pellets using the Chelex-100 boiling method and a nested polymerase chain reaction (PCR) method targeting the *Plasmodium* small subunit ribosomal RNA (ssRNA) was used to identify malaria infected individuals, as described by [276, 286]. This assay used the genus-specific primers rPLU1 (5'-TCA AAG ATT AAG CCA TGC AAG TGA-3') and rPLU5 (5'-CCT GTT GTT GCC TTA AAC TTC-3') for nest 1 and rPLU3 (5'-TTT TTA TAA GGA TAA CTA CGG AAA AGC TGT-3') and rPLU4 (5'-TAC CCG TCA TAG CCA TGT TAG GCC AAT ACC-3') for nest 2. Thermal cycling conditions for primary and nested PCRs were 35 cycles at 94°C, 60°C and 72°C. Samples positive for the *Plasmodium* genus were then screened using species specific primers targeting the ssRNA region; for *P. knowlesi* these included PkF1140 (5'-GATTCATCTATTAAAAATTTGCTTC-3') and PkR1150 (5' GAGTTCTAATCTCCGGAGAGAAAAGA 3') for 35 cycles at 50°C, 72°C and 94°C. All products were visualised on a 2% agarose gel. PCR for malaria infection was performed at laboratories at the Universiti Sabah Malaysia in Malaysia and Research Institute for Tropical Medicine in the Philippines, with PCR validation of a subset of samples at the London School of Hygiene and Tropical Medicine in the UK.

Serological detection of exposure

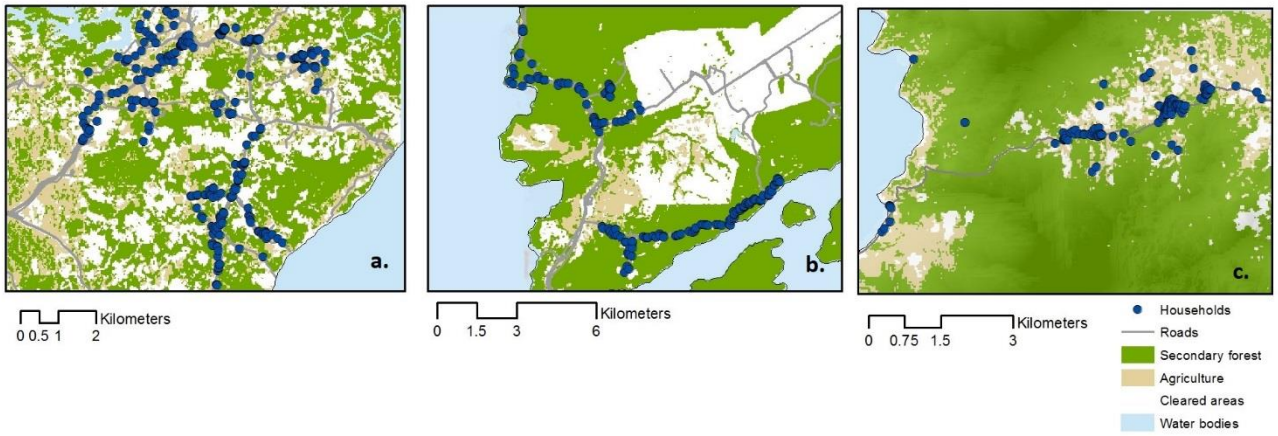
Enzyme-linked immunosorbent assays (ELISA) were performed as previously described [290]. Briefly, 3 mm disc was excised from each dried blood spot and incubated in reconstitution buffer (PBS/tween with sodium azide) overnight at 4^0C. Antibodies were eluted from the blood spots equivalent to a 1:100 dilution of whole blood or a 1:200 dilution of serum [16]. Antibody responses were measured against apical membrane antigen-1 or the 19 kDa fragment of merozoite surface protein-1 for *P. vivax* (PvAMA-1 and PvMSP-1_{19}, respectively), *P. falciparum* (PfAMA-1 [291, 292] and PfMSP-1_{19} [293] and *P.knowlesi* SERA3 antigen 2[119]. The *Pk* serine repeat antigen (SERA) 3 antigen 2 (PKNH_0413400; chromosome 4) is a novel recombinant protein, N-terminally located between positions 826-998 aa, inclusive. SERA3 (1079 aa) belongs to a multigene family whose members encode a papain-like cysteine protease domain [294] . In *P falciparum*, the N-terminal domain of SERA 5 is showing promise as a potential vaccine candidate [295, 296]. The recombinant protein was expressed in *Escherichia coli* and affinity purified by a GST tag. Knowlesi -exposed hospital clinical case control samples showed antigen specific reactivity to the SERA3 antigen 2 recombinant when compared to responses from European malaria naïve and Ethiopian vivax-exposed serum samples (Herman et al. submitted) Eluates were tested in duplicate at a final concentration of 1:1000 for all antigens except 1:2000 for PfAMA-1. In addition, blank wells and a dilution series of the appropriate positive plasma pool were added per plate. Positive controls based on a hyper-immune endemic adult Tanzanian pool [117], a lyophilised anti-malaria patient sample (NIBSC, UK; 72/96) and pooled *Pk*-exposed hospital serum samples were used to assay for *P. falciparum*, *P. vivax* and *P. knowlesi* antigens, respectively. Polyclonal rabbit anti-human IgG-HRP (Dako, Denmark) was used at 1/15,000 dilution and plates were developed using TMB (One component HRP microwell substrate, Tebu-bio). Optical density (OD) values were measured at 490 nm with a microplate reader. Values in excess of 1.5 CV between duplicates were considered fails and re-ran. OD values were corrected by subtracting the background of the blank well per plate. For *P. falciparum* and *P. vivax* OD readings, values were normalised between plates using a standardised control. Normalisation was not done for *P. knowlesi* results due to the lack of standard control. All serological analysis was performed at the Universiti Malaysia Sabah and the London School of Hygiene and Tropical Medicine.

Environmental classification

All households and roads within these areas were geo-located using a hand-held GPS (global positioning system). Land cover maps were derived from LANDSAT 8 30m resolution satellite images

[160] and supervised classification was performed using random forests [221, 225]. In order to generate training data, high- resolution aerial images of areas within study sites were produced using the Sensefly eBee unmanned aerial vehicle flown at 400 metres above ground level (UAV or drone; Sensefly, Cheseaux-sur-Lausanne, Switzerland) and processed using Postflight Terra 3D (Pix4D SA, Lausanne, Switzerland) as described by [151]. These data were manually digitised and classified as forest, agricultural land (including cropland and agroforestry such as rubber and palm oil), open areas and water bodies. Additional data on elevation, aspect and slope was extracted from the ASTER global digital elevation model [159]. All data were resampled to 30m per pixel and datasets including topographic variables, distance from roads and houses, normalised differential vegetation indices (NDVI) and Landsat satellite data were included in the initial model. Random forest models were run using 10,000 trees to ensure stability and were run iteratively with least predictive variables excluded at every run [229]. A random subset of the training data for each site was withheld to independently validate the classification; estimated classification accuracy was 88%, 97% and 85% for Matunggung, Limbuak and Palawan respectively (Figure 8.2).

Figure 8.2. Land use classification of study sites**.** a. Highly deforested and fragmented site at Matunggong, Kudat, Sabah, Malaysia; b. Some forested area bordering large scale clearing in Limbuak, Pulau Banggi, Sabah, Malaysia; c. Mostly intact forest in Bacungan, Palawan, Philippines



These classified land cover maps were used to calculate distance from the household to the forest edge. The proportions of different land types surrounding all households were evaluated for 100m, 500m and 1000m buffer radii. Additionally, the level of forest fragmentation was assessed within 500m and 1000m of each household; this was represented as the ratio of forest perimeter to forest area as described by [240]. All geographic data was stored and visualised in a Geographic Information System using ArcGIS (ArcGIS, Redlands, USA) and all other analysis was performed using R statistical software (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org).

Statistical analysis and data management

Questionnaire data was collected electronically using Pendragon Forms VI (Pendragon Software Corporation, Chicago, USA) and analysed using R statistical software. To define sero-positive individuals, mixture models were fit for normalised optical densities (ODs), with the distribution of ODs modelled as two Gaussian distributions. Cut off values to define sero-prevalence for each antigen were defined as the mean OD of the sero-negative population plus 3 standard deviations for *P. falciparum* and *P. vivax* as described by [114]. For the *P.knowlesi* antigen a more parsimonious cutoff value was defined as the mean OD plus 5 standard deviations due to a lack of prior data. Because the assays were run in different laboratories, cut off values were defined separately for each antigen, malaria species and location (Palawan and Sabah). For *P. falciparum* and *P. vivax*, individuals were considered positive if they were positive for either MSP-1 and/or AMA-1. Reversible catalytic models were fit to age sero-prevalence data using maximum likelihood methods; these models were then used to generate age sero-prevalence curves and estimate the seroconversion rate (SCR) [117]. Evidence of historical changes in transmission was explored by using profile likelihood plots. Models with two SCR were assessed by likelihood ratio tests and used if the fit was significantly better ($p < 0.05$) than models with a constant seroconversion rate [289]. Models were fit separately for each parasite species and site.

Risk factors associated with *P. knowlesi* sero-positivity were evaluated using multivariate logistic regression, with household included as a random effect to account for correlation between individuals from the same household. An additional model was developed to compare individuals sero-positive for *P. knowlesi* with those sero-positive for non-zoonotic malaria species. Explanatory variables included age, gender, site, individual and household level farming activities, residence in the area, elevation and distance to forest. Additionally, the proportions and configuration of different land types were extracted for each household at 100m, 500m and 1000m radii and categorised as greater or less than 30% coverage within a specific radius in the final model. Univariate analysis was conducted for all explanatory variables and variables with $p < 0.2$ were included in multivariate analyses. For highly correlated variables (such as land cover proportions at different radii), single variables were selected based on marginal increases in Akaike Information Criterion (AIC). The final adjusted models were developed by retaining all variables significant at a 0.05 level and variables were added in a forward stepwise fashion to check for interactions.

Potential residual spatial autocorrelation of exposure to *P. knowlesi* was assessed separately for all sites using Moran's I.

Correlation between spatial patterns of exposure to *P. knowlesi* and nonzoonotic malaria species was explored through correlograms, plots of spatial autocorrelation with lag distances. First, ODs were log-transformed and adjusted for age by linear regression as described by [116]. For each site, cross-correlograms of antibody responses to *P.knowlesi* and each other antigen were plotted. Correlation ranges were determined by significance values ($p < 0.05$) of individual bins of lag distances of 500m. Pairwise correlation between antibody responses was determined using a simple Mantel test to test the significance of associations [297, 298].

8.2.5 Results

The total populations resident in the sites were 1260 in Matunggong, 1009 in Limbuak and 686 in Bacungan. Surveys were conducted from October 2014 to January 2015 in Limbuak (n=795) and Matunggong (n=1162) sites in Sabah and in September 2014 in Bacungan, Palawan (n=546). During this time, no clinical *P. knowlesi* cases were reported from the Bacungan study site while one *P. knowlesi* case was reported in the Limbuak site and two cases were reported in Matunggong site. The median age of participants was 24 years (age range 3 months – 99 years) and similar proportions of men and women were sampled in all study sites. While only 22% (538/2503) of individuals reported their primary occupation as farming or plantation work, the majority of individuals (74%; 1846/2503) reported their household engaged in some agricultural activities (Table 1). The proportion of forest cover within 1km of the houses in each site ranged from 39% in Matunggong, 55% in Bacungan to 82% in Limbuak (Fig 2). The Matunggong site was the most highly fragmented, with a forest perimeter to area ratio of 0.03 compared to 0.01 in Bacungan and 0.005 in Limbuak.

Table 8.1. Demographic and environmental characteristics of included participants

	Limbuak, Pulau Banggi (n = 795)	Matunggong, Kudat (n = 1162)	Bacungan, Palawan (n = 546)
Demographic factors			
Gender			
Female, % (n)	52.5 (417)	51.8 (602)	43.6 (238)
Male, % (n)	47.5 (378)	48.2 (560)	56.4 (308)
Age in years, median (IQR)	22 (9 – 44)	25 (10 – 47)	25 (11 – 44)
Farming or plantation work as main occupation, % (n)	14.2 (113)	30.6 (356)	12.6 (69)
Household farm activities, % (n)	68.1 (542)	88.6 (1030)	50.1 (274)
Stay overnight outside village, % (n)	8.2 (65)	13.6 (161)	29.5 (161)
Environmental factors			
Elevation (metres above sea level), median (IQR)	11 (8 – 15)	50 (35 – 75)	84 (77 – 114)
Distance to forest edge (metres), median (IQR)	30 (30 – 60)	95 (68 – 120)	67 (30 – 108)
Proportion of cleared areas around house (%), median (IQR)			
Within 100m	0.43 (0.21 – 0.65)	0.63 (0.46 – 0.74)	0.39 (0.26 – 0.61)
Within 500m	0.14 (0.10 – 0.24)	0.38 (0.28 – 0.47)	0.22 (0.16 – 0.26)
Within 1000m	0.14 (0.09 – 0.17)	0.37 (0.31 – 0.39)	0.18 (0.16 – 0.20)
Proportion of agriculture around house (%), median (IQR)			
Within 100m	0.14 (0.03 – 0.31)	0.33 (0.23 – 0.49)	0.43 (0.32 – 0.60)
Within 500m	0.06 (0.05 – 0.14)	0.38 (0.28 – 0.48)	0.39 (0.36 – 0.42)
Within 1000m	0.05 (0.03 – 0.10)	0.31 (0.24 – 0.36)	0.37 (0.29 – 0.39)
Proportion of forest around house (%), median (IQR)			
Within 100m	0.31 (0.12 – 0.50)	0.03 (0 – 0.08)	0.08 (0 – 0.20)
Within 500m	0.71 (0.59 – 0.81)	0.22 (0.13 – 0.34)	0.37 (0.32 – 0.43)
Within 1000m	0.79 (0.76 – 0.86)	0.33 (0.27 – 0.39)	0.44 (0.40 – 0.52)
Forest area to perimeter ratio around house, median (IQR)			
Within 500m	0.02 (0.01 – 0.02)	0.05 (0.04 – 0.06)	0.04 (0.03 – 0.04)
Within 1000m	0.01 (0.01 – 0.01)	0.03 (0.03 – 0.04)	0.04 (0.03 – 0.04)

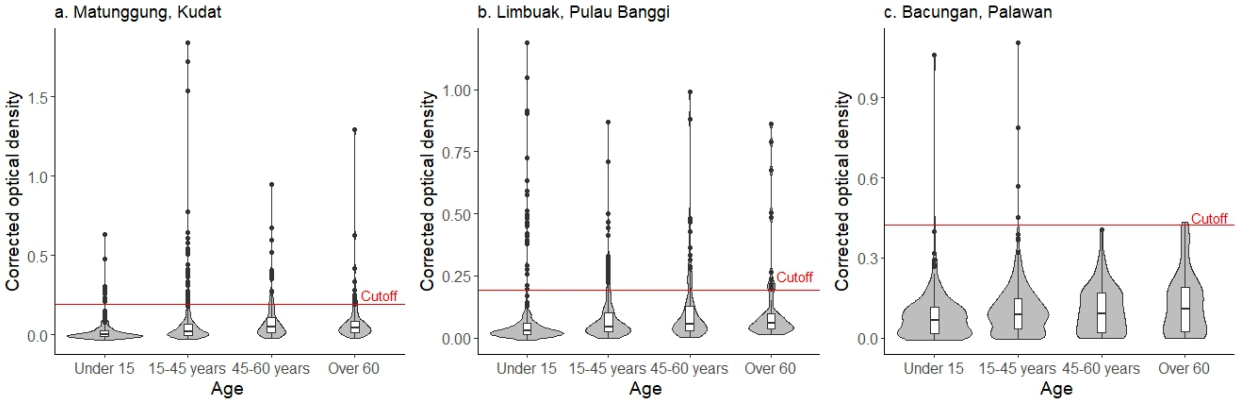
Infection with malaria

Two microscopy positive individuals were identified from the Matunggong, Kudat site; these were both subsequently identified as *P. knowlesi* mono-infections by PCR. All PCR infections were re-confirmed at the laboratory in the U.K. Both of these individuals were male plantation workers (ages 21 and 25) residing in the same household. An additional two individuals in Matunggong were microscopy negative but identified as *P. knowlesi* infected by molecular methods; these included a three year old girl and 33 year old woman residing in different villages within the study site. Only one out of these four infected individuals identified self-reported history of fever. None of the survey participants in either the Limbuak or Bacungan sites were positive by microscopy or PCR and no infections with any other malaria species were identified in any participants.

Serological assessment of exposure to P. knowlesi

Overall, 7.1% (178/2503) of the population surveyed was seropositive to *P. knowlesi* (Figure 8.3). Exposure varied substantially between study sites, with the highest *P. knowlesi* antibody prevalence detected in Limbuak, Pulau Banggi (11.7%; 93/795) followed by 6.8% (79/1162) in Matunggong Kudat. Bacungan, Palawan had the lowest sero-prevalence (1.1%; 6/546). Similar reactivity to *P. knowlesi* was observed in men (optical density (OD): med: 0.035, IQR: 0.006 – 0.094) and women (OD: median: 0.035, IQR: 0.007 – 0.089) and gender was not significantly associated with *P. knowlesi* sero-positivity (OR: 0.99, 95% CI: 0.71- 1.37, p=0.95).

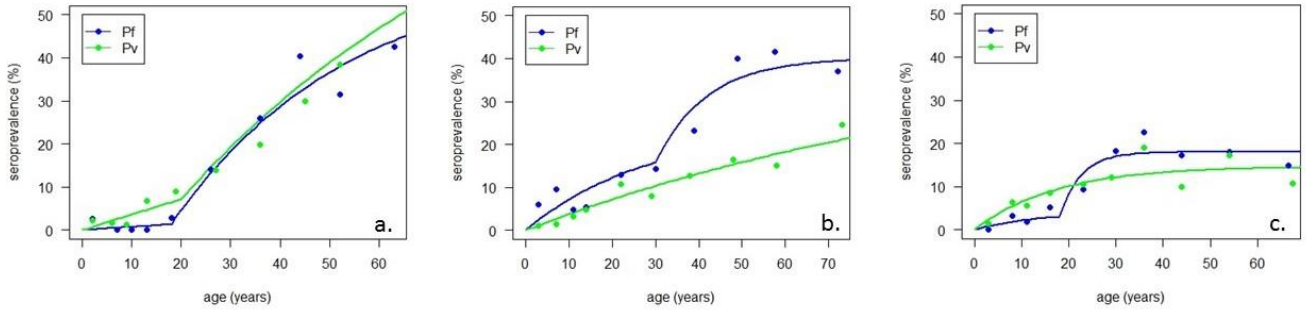
Figure 8.3. Violin plots of *P. knowlesi* antibody density by age group: a. Matunggong, Kudat, b. Limbuak, Pulau Banggi, c. Bacungan Palawan



Antibody prevalences to *P. falciparum* and *P. vivax* were higher in all sites, with 16.1% (364/2266) of individuals sero-positive to one or both *P. falciparum* antigens and 12.6% (270/2141) positive for one or more *P. vivax* antigens. Sero-prevalence to *P. falciparum* was 16.9% (196/1162) in Matunggong, 13.5% (107/795) in Limbuak and 10.4% (61/587) in Bacungan. In contrast, reactivity to *P. vivax* was highest in Limbuak (16.7%; 133/795) with sero-prevalences of 6.9% (80/1162) and 9.7% (57/587) in Matunggong and Bacungan respectively. Due to insufficient samples and non-systematic errors in labelling, results for all antigens were not available for all individuals. Out of individuals with complete test results for all antigens, 25.7% (499/1941) of participants were sero-positive to at least one species of malaria and 7.9% (54/1941) were sero-positive for two or more malaria species. Of individuals exposed to *P. knowlesi*, 29.7% (53/ 178) were also positive for *P. falciparum* or *P. vivax* antigens. There was no evidence of correlation between *P. knowlesi* and other antigens tested (Supplementary Information, Fig S1).

Sero-prevalence was positively associated with increases in age for all antigens tested. However, despite this, seroreactivity, including individuals with high antibody titres was still detected in the youngest age groups and 4.2% (39/921) individuals under 15 years had antibodies to *P. knowlesi* (Fig 3), 3.5% (29/821) had antibodies to *P. falciparum* and 2.9% (23/792) to *P. vivax* (Supplementary Information, Fig. S2). Changes in age sero-prevalence were more pronounced for *P. falciparum* and *P. vivax*, with 32.9% (78/237) and 28.1% (64/228) reactivity to *P. falciparum* and *P. vivax* in individuals over the age of 60 years. In contrast, antibodies for *P. knowlesi* were detected in 9.4% (25/265) of individuals over 60 years old and the highest sero-prevalence was detected in adults from 45 – 60 years old (11.6%; 43/370). As reactivity to *P. knowlesi* was low and not evenly distributed through the population, seroconversion rates (SCR) for *P. knowlesi* could not be calculated. Historical changes in falciparum transmission intensity were apparent in all sites and SCR models fitted with two forces of infection suggest substantial reductions in *P. falciparum* transmission occurred 18 – 30 years ago ($p < 0.05$) (Fig 4). Strong evidence of decreased transmission intensity for *P. vivax* was only seen in Limbuak, where transmission decreased over 25-fold in the past 20 years.

Figure 8.4. Seroprevalence curves for each location used to calculate SCRS (λ) a. Pulau Banggi, Sabah (Pf λ1: 0.0196 (0.0078 – 0.0492); Pf λ2: 0.0008 (0.0002 – 0.0031); Pv λ1: 0.0185 (0.0136 – 0.0248)) b. Matunggong, Kudat ((Pf λ1: 0.0588 (0.0198 – 0.1746); Pf λ2: 0.0085 (0.0063 – 0.0116); Pv λ: 0.0039 (0.0024 – 0.0064)) c. Bacungan, Palawan (Pf λ1: 0.1441 (0.0175 – 1.1892); Pf λ2: 0.0031 (0.0012 – 0.0086); Pv λ: 0.0086 (0.0044 – 0.0166))



Factors associated with P. knowlesi *sero-positivity*

Demographic and environmental characteristics of survey participants are summarised in (Table 1). In addition to age and site, reporting farm or plantation work as a primary occupation was positively associated with *P. knowlesi* sero-positivity (Table 8.2). Higher proportions of forest cover within 1km of the household and cleared areas within 500m of the house were both associated with increased odds of *P. knowlesi* positivity. While forest fragmentation, elevation and agricultural land around the house were significant within the univariate analysis, none of these variables were

significant in the final multivariate model (Supplementary information). Similar proportions of men and women reacted to *P. knowlesi* in all sites and gender was not associated with sero-positivity.

Table 8.2. Multivariate analysis of risk factors for *P. knowlesi* seropositivity (comparison of *P. knowlesi* exposed individuals with non-exposed individuals)

	Adjusted OR (95% CI)	P value
Age		
Under 15 years	-	< 0.001
15 – 45 years	2.05 (1.30-3.22)	
45 – 60 years	2.94 (1.70-5.11)	
Over 60 years	2.46 (1.32-4.58)	
Site		
Palawan	-	< 0.001
Mainland Kudat	4.30 (1.66, 11.15)	
Pulau Banggi	10.83 (4.50, 26.10)	
Main occupation farm or plantation work		
No	-	0.025
Yes	1.63 (1.07, 2.48)	
Forest cover within 1km		0.004
Less than 30%	-	
Over 30% forest cover	2.40 (1.29, 4.46)	
Proportion of cleared/ open area within 500m of house		0.001
Less than 30%	-	
Over 30% cleared	2.14 (1.35, 3.40)	

Individuals reacting to *P. knowlesi* were more likely to be younger than individuals sero-positive for only non-zoonotic malaria species (Table 8.3). Forest cover was not associated with exposure to non-zoonotic malaria and malaria positive individuals residing in areas with high forest cover around the house had 4.86 (95% CI: 2.30 – 11.37) the odds of being positive for *P. knowlesi*. Similarly, cleared areas around the house were also positively associated with *P. knowlesi* cases compared to other malaria species.

Table 8.3. Multivariate analysis of risk factors for *P. knowlesi* seropositivity in malaria exposed individuals. (comparison of *P. knowlesi* exposed individuals with individuals exposed to other non-zoonotic malaria species)

	Adjusted OR (95% CI)	P value
Age		
Under 15 years	-	0.05
15 – 45 years	0.72 (0.37-1.39)	
45 – 60 years	0.53 (0.26-1.06)	
Over 60 years	0.38 (0.18-0.82)	
Site		
Palawan	-	< 0.001
Mainland Kudat	3.79 (1.50, 11.00)	
Pulau Banggi	6.55 (2.88, 17.68)	
Proportion of forest within 1km of house		
Less than 30%	-	< 0.001
Over 30% cleared	4.86 (2.30, 11.37)	
Proportion of cleared/ open area within 500m of house		
Less than 30%	-	0.001
Over 30% cleared	2.70 (1.60, 4.66)	

Based on Moran's I, there was no evidence of residual spatial autocorrelation for *P. knowlesi* antibody responses (Moran's I $p > 0.2$ for all sites). There was no significant spatial correlation detected between age-adjusted antibody responses for *P. knowlesi* and other malaria species for either Matunggong or Limbuak ($p > 0.30$ for all pairwise comparisons). Comparisons between *P. knowlesi* and other malaria species could not be evaluated for Bacungan due to the low prevalence of *P. knowlesi* sero-positivity.

8.2.6 Discussion

This is the first study to describe exposure to *P. knowlesi* through antigen specific antibody responses and associated risk factors and is one of few studies to assess *P. knowlesi* carriage prevalence at a community level. Results indicate spatial and temporal patterns of *P. knowlesi* transmission differ markedly from other non-zoonotic malaria species within the region. Although *P. knowlesi* sero-positivity was associated with some landscape attributes within these communities, extensive cross sectional surveys are needed to identify ecological risk factors across a broader geographic scale.

Sero-prevalence data indicate distinct heterogeneities in *P. knowlesi* transmission intensity between sites. Although formal comparisons between *P. knowlesi* infection and exposure could not be undertaken due to the low prevalence of parasite carriage, these geographical differences in transmission mirror hospital-based reporting rates in the study sites at Kudat, Pulau Banggi and Palawan [27, 37, 247]. These results also highlight the utility of serological techniques to identify differences in transmission intensity in settings where the sensitivity of parasite prevalence surveys is limited by the scarcity of infected individuals and suboptimal diagnostics. This is the first time these knowlesi-specific antigens have been used at a population level to assess species-specific exposure to malaria. Although high levels of homology between *P. knowlesi* and *P. vivax* indicate the possibility of cross reactivity between antigens, relatively low numbers of individuals were identified as sero-positive for both knowlesi and vivax (2%; 43/2102 individuals with results for both assays) and individuals could have been plausibly exposed to both species due to the co-endemicity of these species within this region. Additional work has been done to characterise response to *P. knowlesi* in vivax-exposed individuals and validate these antigens for population-based studies[119].

Changes in seroconversion rates can also reflect temporal changes in malaria transmission. In Sabah, state-wide malaria notification records describe dramatic decreases in clinical *P. falciparum* and *P. vivax* cases within the past 20 years following the scale up of malaria control and elimination programmes [13]. The Philippines has also reported a substantial decline in the number of malaria cases reported in the past few decades, most notably for *P. falciparum*[299]. These changes are evident in seroconversion rates to non-zoonotic malaria species from the 3 sites with over 5-fold difference between current and previous SCRs. *P. knowlesi* exposure was identified in children under 5 in all sites, suggesting recent or on-going transmission albeit at a low level. Further work is needed to refine *P. knowlesi* serological analysis to allow for antigenic variation, identify further antigenic targets and assess the differential responsiveness of individuals and longevity of antibody responses [89].

Despite these similarities between existing case data and community-level exposure to *P. knowlesi*, levels of exposure between different demographic groups varied markedly from clinical case reports. While clinical *P. knowlesi* has been commonly reported in adult men, men and women had similar antibody reactivity to *P. knowlesi* antigens in all sites [18, 247]. Within Kudat district, wide age distributions and family clusters of knowlesi cases have previously been described; however, from 2012-2015, 73% (84/115) and 77% (27/35) of all clinical cases reported from Kudat and Pulau Banggi respectively were men [247]. Asymptomatic knowlesi carriage has been detected

in higher proportions of women by this study and other studies; however these results are extremely limited by sampling design and the small numbers of infected individuals detected [31, 286]. As forest and agricultural activities have been identified as risk factors for clinical *P. knowlesi* infection, more men could develop clinical infections due to higher exposure or number of bites; however, this requires further research to assess [103]. Larger scale population-based cross sectional surveys are needed to determine if these patterns occur in the wider community and if *P. knowlesi* affects groups which may be underrepresented by current passive surveillance systems.

P. knowlesi exposure was also associated with landscape factors. Both the proportion of forest cover and cleared areas around the household were positively associated with knowlesi sero-positivity, potentially reflecting the higher vectorial capacity and sporozoite rates reported in secondary forest within these study sites [152]. Although plantation or farm work as a primary occupation was associated with increased exposure and previous reports have described associations between *P. knowlesi* and forest activities, data on movement into different environments was not available for all survey participants [18, 19]. Instead, to explore the potential range of spatial interactions between people and mosquito vectors, proportions of habitat were evaluated at different buffer distances around houses. The significance of both clearing and forest areas at different radii suggests the importance of edge effects, transition areas between habitats where increased overlap of human, macaque and mosquito populations may occur [10, 82, 255]. Despite this, no associations were identified between metrics of fragmentation or distance to forest edges; this may reflect the limited environmental variation within these small spatial scales. Future studies could assess the importance of these variables across different ecotypes as well as collect more detailed data on the human movement into different environments, particularly during peak mosquito biting times.

These spatial patterns differed markedly from exposure to other non-zoonotic malaria species. Individuals with antibodies to *P. knowlesi* were more likely to reside in areas with higher proportions of forest cover; this may reflect differences in disease dynamics between species or temporal changes in transmission. Because of the longevity of antibody responses and the rapid rates of land use change within these areas, seroreactivity to non-zoonotic species is probably more likely to be associated with past rather than current environmental factors. The main mosquito vector species of *P. knowlesi*, *Anopheles balabacensis*, was historically incriminated as the main vector of other human malaria species within these same areas [243, 300]. While these vectors have been primarily associated with forest habitats, high vector densities have also been reported in small

scale farms and other habitat types [81, 82, 152]. Deforestation and increased application of vector control measures may have triggered changes in vector composition and biting preferences; similarly, habitat changes and encroachment of human settlements into forest areas may have also led to changes in macaque population densities and closer contact between macaques, people and mosquito vectors [6, 75].

The main limitations of this study are the non-randomised population sampling approach and limited geographical scale. While this study describes fine scale patterns of malaria exposure and infection within these three case study communities, these results cannot be generalised to extrapolate *P. knowlesi* risks across wider populations. As this study surveyed three relatively homogenous populations, there was minimal variation in environment, ethnicity, socioeconomic status and access to healthcare within each site. Identifying environmental and population-level risk factors will require randomised sampling across a wider ecological gradient; community level data on presence and absence of exposure and infection are required to understand spatial heterogeneity of disease transmission and develop and refine predictions of disease risk [235]. Additionally, extensive surveys of parasite prevalence may allow the application of genetic approaches to track parasite diversity and transmission and explore the roles of host and parasite genetic factors.

Despite these limitations, this study describes *P. knowlesi* infection and exposure within these communities and illustrates how serologic markers can be used to describe differences in transmission intensity between malaria species in low transmission settings. Results from these surveys indicate patterns of *P. knowlesi* exposure and infection within the community may be substantially different from cases detected by passive surveillance systems. Cross sectional surveys across a broader geographical scale are needed to describe spatial variation in transmission intensity and identify associated environmental and population-based risk factors. Integration of serology into these surveys would provide vital information on rare infections for control programmes [301].

8.2.7 Acknowledgements

The authors would like to thank Albert M Lim (Infectious Diseases Society Sabah), Judy Dorothy Madarcos, Jo Anne Bibit (Research Institute for Tropical Medicine) and the teams in Sabah and Palawan for their support for this project.

8.2.8 Supporting information

Fig S1. Comparison of normalised optical densities for *P. knowlesi* and other antigens. a. optical densities and cut offs for Sabah, Malaysia; b. optical densities and cut offs for Palawan, Philippines

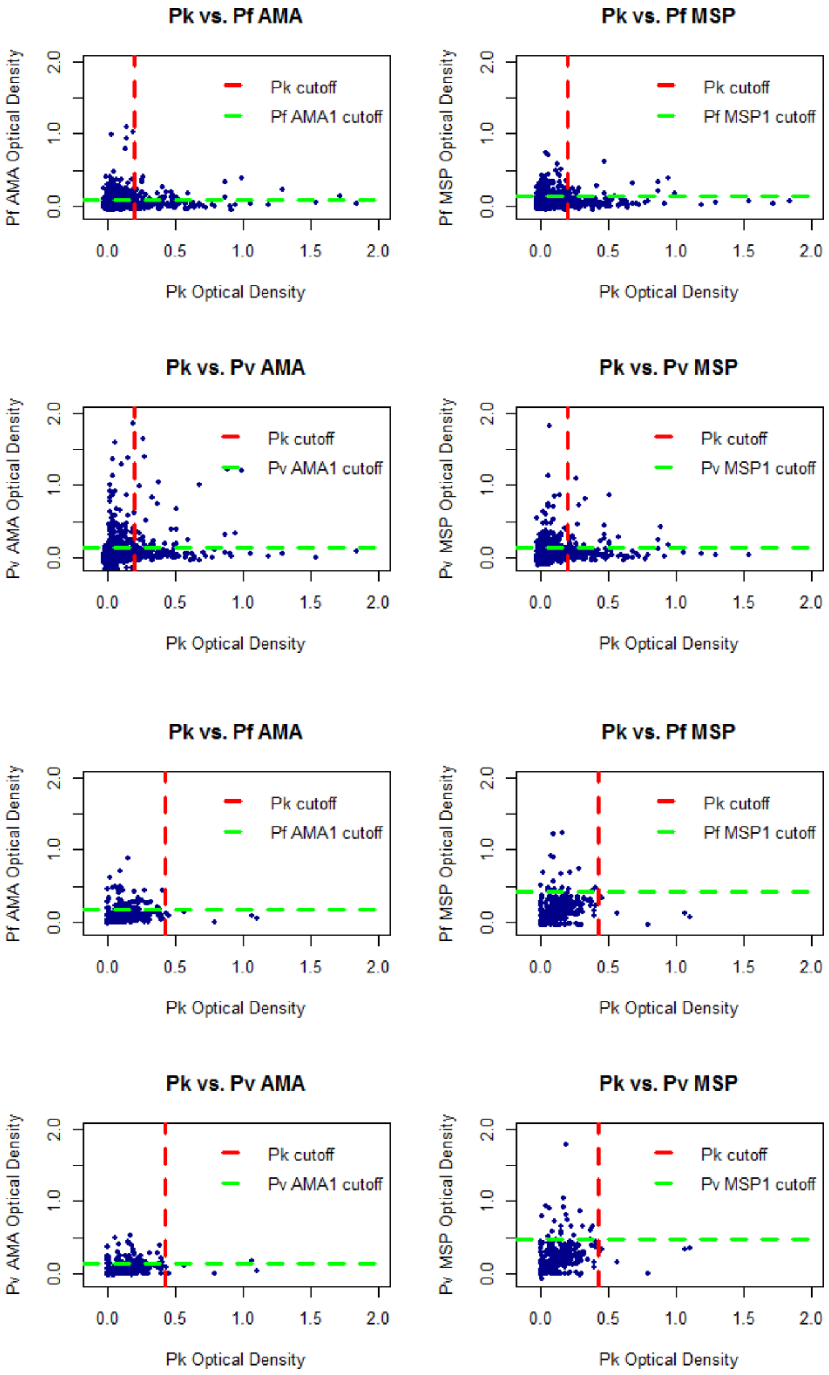
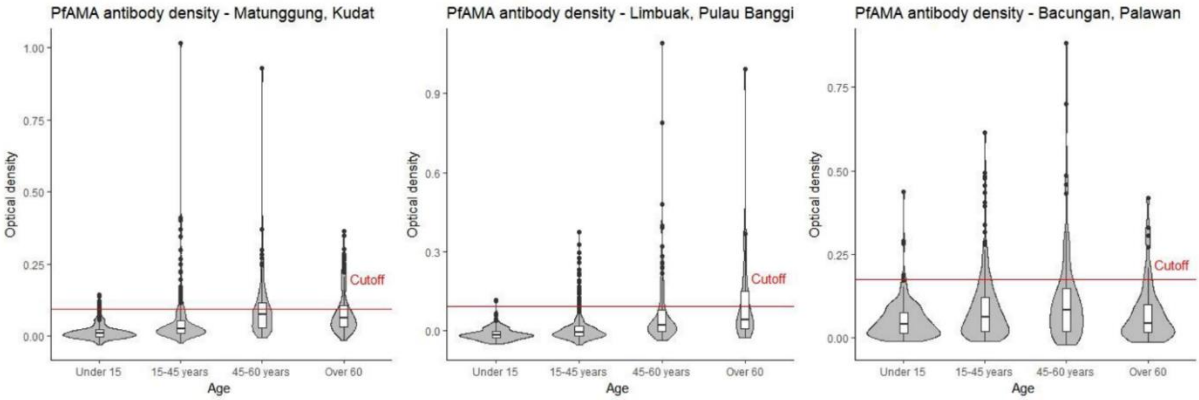
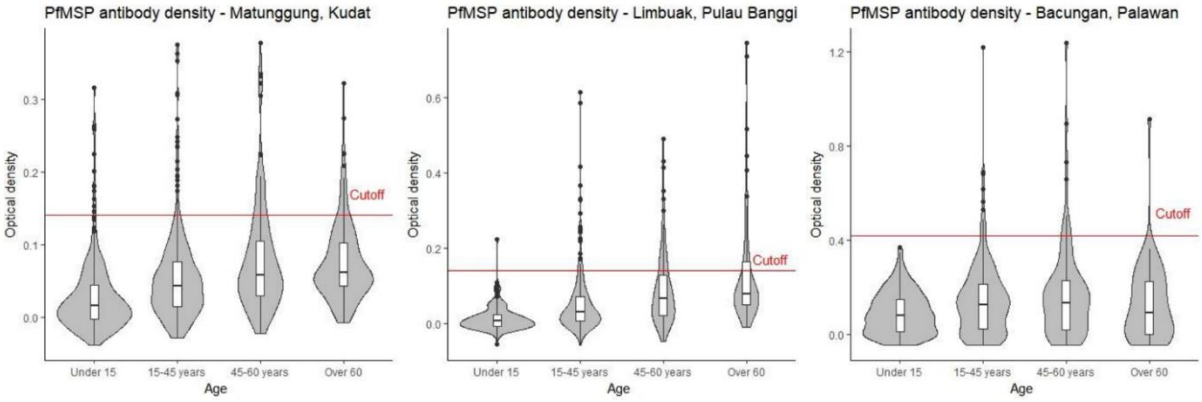


Fig S2. Violin plots of antibody density for *P. falciparum* and *P. vivax* by age group. a. *P. falciparum* AMA-1 antibody density; b. *P. falciparum* MSP-1 antibody density; c. *P. vivax* AMA-1 antibody density; d. *P. vivax* MSP-1 antibody density

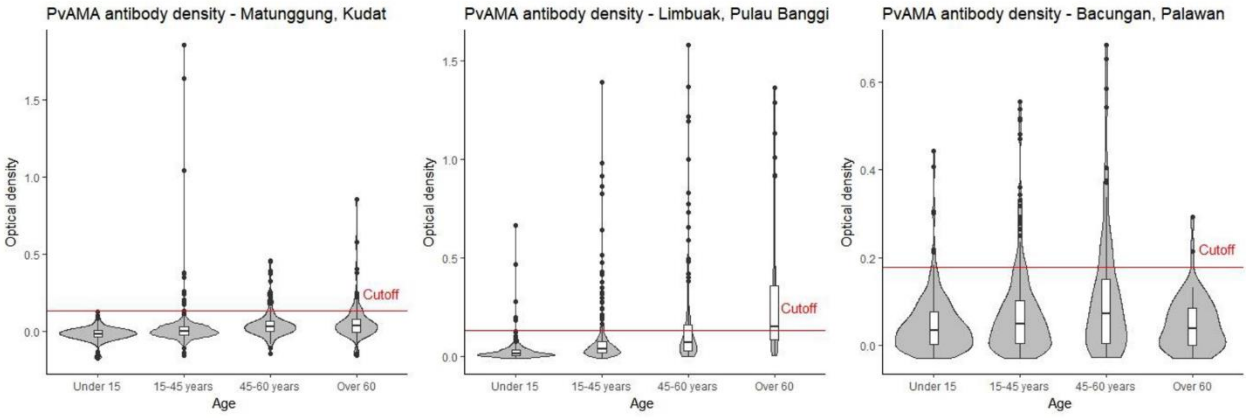
a. *P. falciparum* AMA-1 antibody density



b. *P. falciparum* MSP-1 antibody density



a. *P. vivax* AMA-1 antibody density



b. *P. vivax* MSP-1 antibody density

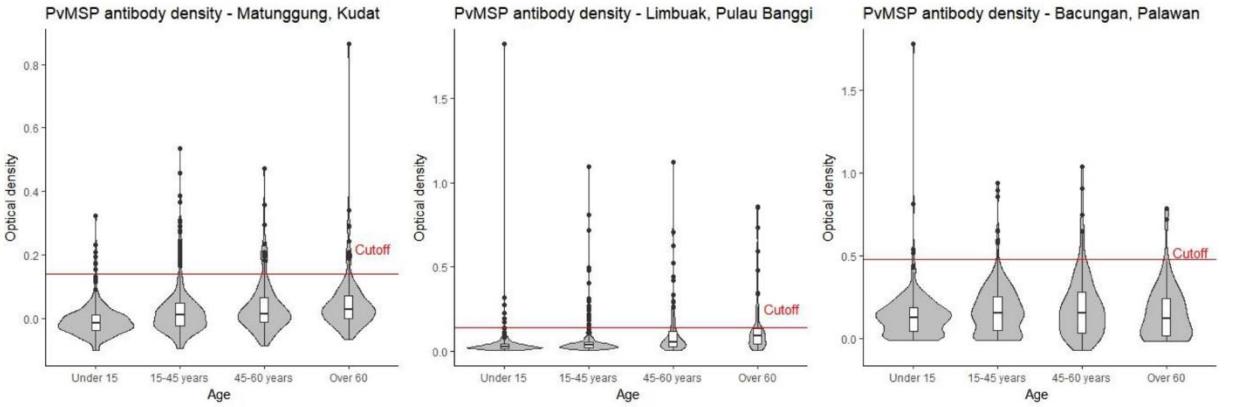


Table S1. Proportions and univariate analysis of risk factors for *P. knowlesi* seropositivity

	All		Pk Seropositive		Unadjusted OR (95%CI)*	
	N	%	N	%		P value
Total samples	2503		178	7.1%		
Site						
Palawan	546	21.8%	6	1.1%	-	< 0.001
Mainland Kudat	1162	46.4%	79	6.8%	7.02 (2.93, 16.86)	
Banggi	795	31.8%	93	11.7%	12.80 (5.31, 30.89)	
Gender						
Female	1257	50.2%	90	7.2%	-	0.95
Male	1246	49.8%	88	7.1%	0.99 (0.71, 1.37)	
Age						
Under 15 years	921	36.8%	39	4.2%	-	< 0.001
15 – 45 years	947	37.8%	71	7.5%	2.19 (1.42, 3.38)	
45 – 60 years	370	14.8%	43	11.6%	4.28 (2.61, 7.03)	
Over 60 years	265	10.6%	25	9.4%	3.40 (1.93, 6.00)	
Primary occupation						
Other occupation	1965	78.5%	118	6.0%	-	< 0.001
Farmer/ plantation worker	538	21.5%	60	11.2%	2.23 (1.52, 3.25)	
Farm or plantation activities						
No	657	26.2%	47	7.2%	-	0.51
Yes	1846	73.8%	131	7.1%	1.17 (0.734, 1.86)	
Stay overnight outside kampung						
No	2116	88.4%	161	7.6%	-	0.02
Yes	387	11.6%	17	4.4%	0.53 (0.30, 0.95)	
Elevation						
Under 50 MSL	1396	55.7%	130	9.3%	-	< 0.001
50-100 MSL	730	29.2%	30	4.1%	0.41 (0.25, 0.67)	
Over 100 MSL	377	15.1%	18	4.7%	0.46 (0.25, 0.86)	
Forest within 50m of house						
Forest within 50m of house	852	34.0%	76	8.9%	-	0.11
Forest over 50m from house	1651	66.0%	102	6.2%	0.71 (0.47, 1.07)	
Proportion of forest within 1km of house						
Less than 30%	421	16.8%	17	4.0%	-	< 0.001
30-40% forest cover	608	24.3%	43	7.1%	1.52 (1.04, 2.20)	
40 – 75% forest cover	842	33.6%	45	5.3%	1.31 (0.67, 2.53)	
Over 75% forest cover	632	25.3%	73	11.6%	3.16 (1.64, 6.07)	
Proportion of agriculture within 1km						
Less than 10%	658	26.3%	83	12.6%	-	< 0.001
10 – 25% agricultural land	533	21.3%	39	7.3%	0.54 (0.33, 0.90)	
25 – 35% agricultural land	656	26.2%	35	5.3%	0.37 (0.22, 0.62)	
Over 35% agricultural land	656	26.2%	21	3.2%	0.22 (0.12, 0.39)	
Proportion of cleared/ open area within 500m of house						
Less than 15%	530	21.2%	40	7.5%	-	0.01
15 – 30% cleared	890	35.6%	40	4.5%	0.56 (0.32, 0.99)	
30 – 40% cleared	579	23.1%	49	8.5%	1.16 (0.65, 2.05)	
Over 40% cleared	504	20.1%	49	9.7%	1.29 (0.72, 2.30)	

9 Environmental risk factors and exposure to *Plasmodium knowlesi* across Northern Sabah: a cross-sectional survey

9.1 Background and aims

Understanding patterns of *P. knowlesi* across a range of ecological conditions is critical to effectively target disease control resources as well as to identify areas that may be at risk of future outbreaks. Although previous work has identified individual and environmental risk factors for *P. knowlesi* (**Chapters 5 and 8**; [103]), all of these studies relied on passively collected clinical data or community-level data from a limited geographical area. Initial data on asymptomatic infections and serological exposure indicate *P. knowlesi* infections may occur across a wider demographic range than represented by clinical data and spatial patterns of these infections remain unknown. Further, previous analyses across larger spatial scales are limited by relatively crude environmental and spatial data available.

To address this need, we conducted an environmentally-stratified cross-sectional survey across four districts in Northern Sabah encompassing a broad range of ecologies. The primary objective of this survey was to estimate transmission intensity of *P. knowlesi* and other malaria parasites as measured by prevalence of antibodies to specific malaria antigens and to characterise population-level risk factors for sero-positivity. As a secondary objective, we measured the prevalence of asymptomatic parasitemia to *P. knowlesi* and other malaria species within these communities.

9.2 Methods

9.2.1 Ethics approval and informed consent

The Medical Research Sub-Committee of the Malaysian Ministry of Health (NMRR-14-713-21117) and the Research Ethics Committee of the London School of Hygiene and Tropical Medicine (8340) approved this study. Written informed consent was obtained from all study participants and procedures were performed according to relevant local guidelines.

9.2.2 Study site and population

This study was conducted in 2015 across Kudat, Kota Marudu, Pitas and Ranau districts in Northern Sabah in Malaysian Borneo and was the site of integrated entomology, primatology and social science studies as part of the MONKEYBAR project (**Chapter 3**). These districts have a combined, predominantly rural population of around 280,000 and contain a wide variety of land cover and ecologies [302]. This area includes mainland Borneo as well as outlying islands and elevation ranges from sea-level to over 4000 metres above sea level (MSL) at Mt. Kinabalu. The climate is tropical and rainfall varies monthly; however, widespread droughts and high smoke pollution due to El Niño were described in Borneo during the year of the survey [303]. While *P. knowlesi* cases have been reported from all districts, the relative importance of different malaria species varies by district, suggesting *P. knowlesi* epidemiology varies at this spatial scale [13]. Malaysia is targeting malaria elimination by 2020 and there is an active malaria control programme operating within this region.

9.2.3 Sampling design and survey methodology

To estimate sero-prevalence within this region, we used a non-self-weighting two stage sampling design. First, all villages within the study area were geo-located, with urban areas excluded. Based on previous associations between *P. knowlesi* village-level incidence and forest cover within a 2km radius, we stratified villages into three strata based on percentage of forest cover in 2014 within a 2km circular buffer of the village centroid, with forest defined as over 50% canopy cover [51]. As *knowlesi* sero-prevalence was estimated as 11.7% and 6.8% within two case study areas within this region, we expected sero-prevalence to be around 10% within the general population [256]. To adequately define sero-prevalence within each strata, we aimed for 95% confidence with 80% power and assumed a design effect of 2. Assuming an average household size of four and a 15% non-response rate, we calculated a sample size of 883 households per strata with an overall sample size of 2650 households. At the first stage, equal numbers of villages were randomly selected from each strata. Second, we enumerated and geo-located all households within selected villages and 20 households were randomly selected per village. For villages with fewer than 20 households, all households were sampled and additional villages within the same strata were randomly selected until the target sample size was met.

All individuals residing in selected households for the previous month were asked to participate in the survey. Individuals were excluded if they were less than 3 months old or could not be reached after 3 attempts to contact them. Individual and household-level data from consenting individuals

was collected electronically using Pendragon Forms VI (Pendragon Software Corporation, Chicago, USA), questionnaire included in **Annex 2**. Finger prick blood sampling was used to prepare blood smears to detect malaria parasites by microscopy, whole-blood specimens in tubes containing EDTA (Becton-Dickinson, Franklin Lakes, USA) and blood spots on filter paper (3MM, Whatman, Maidstone, United Kingdom).

9.2.4 Laboratory methods (serology and PCR)

For DNA extraction, whole blood samples were pooled into 10 x 10 matrices with 40µl of each sample loaded on one vertical and one horizontal pool. The 400µl pools were extracted on a QIAsymphony SP/AS instrument (Qiagen, UK) using QIAsymphony DNA Midi Kit (Qiagen, UK) and eluted in 200µl of elution buffer provided with the kit. Extracted DNA pools were amplified by genus-specific 18S ribosomal DNA nested PCR using methods described by [276]. Nested PCR products were analysed on 1.5% agarose gels. Genus-positive sample pools were de-pooled and reamplified. Positive samples were speciated using methods described by [276, 288] and visualised on agarose gels.

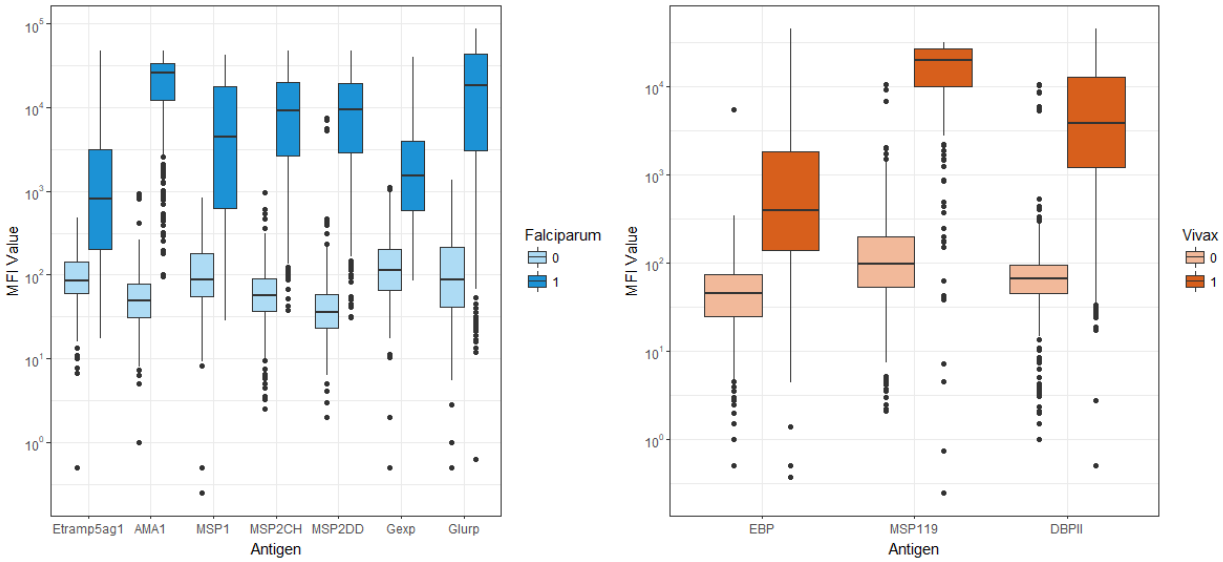
9.2.5 Classification of seropositivity

To identify sero-positive individuals, we utilised an ensemble approach for binary classification using the Super Learner algorithm [304]. Super Learner is a data adaptive meta-learning algorithm which estimates the optimal combination of base learning algorithms for prediction and has previously been used to classify serological responses [287, 305]. This algorithm obtains better predictive performance than could be obtained from individual base learners and allows for complex, non-linear relationships between magnitude of antibody responses and exposure [306]. Additionally, as all steps of the modelling can be fully automated, bias in model selection or antibody importance is avoided.

As supervised classification algorithms were used to identify exposure status, training datasets of known sero-positive and sero-negative samples were assembled for each malaria species. Ideally, training data would be from individuals within this population with known exposure status; however, due to the continued transmission of all malaria species assayed, it was not possible to identify unexposed individuals within the study site. Instead, we utilised samples from malaria-unexposed populations. For *P. knowlesi*, we additionally included samples from malaria-endemic areas in Africa and South America as described by [119].

Sero-positive training data for *P. falciparum* included all available molecularly confirmed *P. falciparum* cases from Northern Sabah followed up from Day 0 to 1 year after diagnosis (n=47) [103] and longitudinal samples from individuals over the age of 5 in a previously hyper-endemic area experiencing massive reductions in transmission following an intervention (Ssewanyana, in preparation). These samples were selected to represent both recent and historical *P. falciparum* exposure. Similarly for *P. vivax* classification, sero-positives included individuals in Northern Sabah with molecularly confirmed *P. vivax* infections (n=99) [103], confirmed *P. vivax* exposed individuals from other endemic areas (Ethiopia and Brazil, as described in [119] and positive *P. vivax* controls n=371). Samples from UK residents with no history of travel was used as a negative reference population for both species (n=510) (NIBSC, UK; 72/96). Responses to all available antigens were used for classification, with only *P. vivax* AMA1 omitted due to the high level of homology with *P. knowlesi* AMA1 [119] (Figure 9.1).

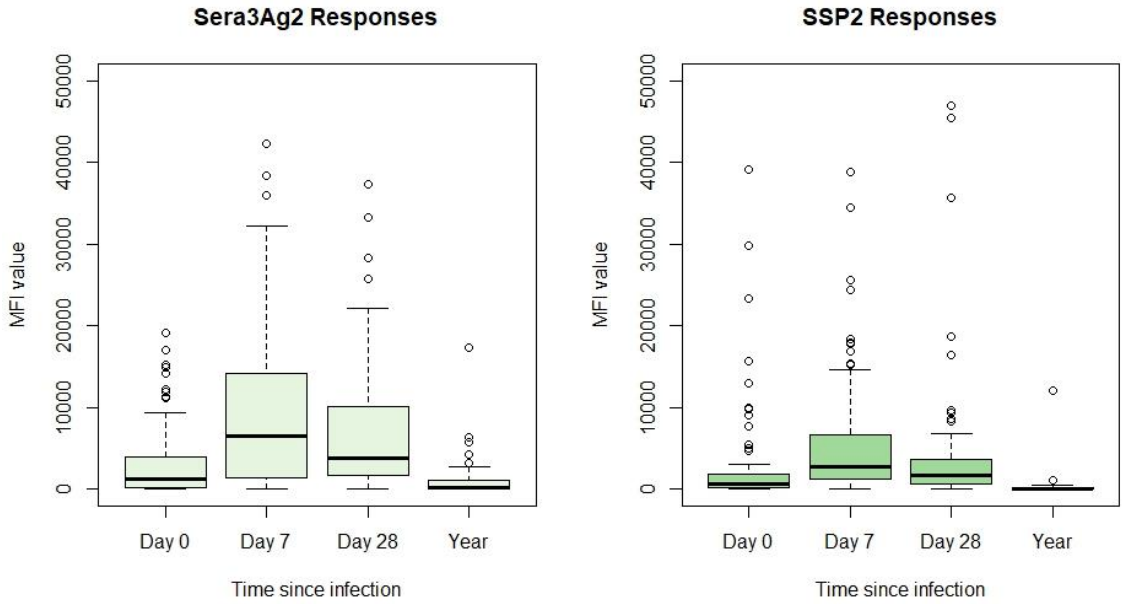
Figure 9.1. Median Fluorescence Intensity (MFI) antibody responses of training data used for classification in known positive and negative individuals for a. *P. falciparum*; b. *P. vivax*



In contrast to *P. falciparum* and *P. vivax*, species-specific antigens have only recently been developed for *P. knowlesi* and limited data is available on the longevity or individual variation of antibody responses. Using three knowlesi- specific antigens from a panel developed by Herman et. al [119], we first evaluated temporal changes in magnitude of antibody responses from a cohort of molecularly confirmed *P. knowlesi* cases in Northern Sabah followed up at different time points from diagnosis, including day 0 (n=126), day 7 (n=76), day 28 (n=79) and 1 year (n=40) [103]. Results suggest antibody responses were relatively short-lived, peaking at day 7 and becoming undetectable

after 1 year (Figure 9.2). Although further studies are required to fully assess temporal changes in responses, we chose to assemble a sero-positive training dataset from day 7 and day 28 antibody responses to identify recent *P. knowlesi* exposure. While high responses were observed to knowlesi AMA1, this antigen was excluded from the final model due to the high levels of correlation between *P. vivax* and *P. knowlesi* AMA1 responses. We additionally included vivax-exposed individuals from areas without *P. knowlesi* transmission in the negative training data for *P. knowlesi*.

Figure 9.2. Temporal changes in antibody responses in *P. knowlesi* cases: a. *P. knowlesi* Sera3Ag2, b. *P. knowlesi* SSP2



Using the training data with a binary response for species-specific exposure, we fit the Super Learner algorithm as a weighted combination of five component models: random forests [221], boosted regression trees [307], support vector machines [308], K-nearest neighbour [309] and Lasso classification [310]. Weights for each base learner were calculated using the Nelder-Mead method to maximise Area Under the ROC Curve (AUC) [311]. To avoid overfitting, we used a random 70% of the dataset to build the model with the remaining data used for independent validation. The full dataset with 10-fold cross validation was used to make predictions. Multiple imputation by chained equations was used to estimate missing values for antibody responses in test data [312]. Evidence of historical changes in classified *P. falciparum* and *P. vivax* data were explored using reverse catalytic models with a constant or two sero-conversion rates assessed by likelihood ratio tests as described in [256, 289].

9.2.6 Individual risk factor analysis

Individual level risk factors for knowlesi sero-positivity were evaluated using a multivariate binomial generalised mixed model including household and cluster as a random effect. A socioeconomic status index was constructed using principal component analysis incorporating assets, household structure and household head education [313]. Data on elevation, aspect and slope was obtained from the ASTER Global Digital Elevation Model and extracted for each household location [159]. To evaluate access to healthcare, cost surface rasters with 30m resolution were created using an estimated speed of 60 km/hr for highways, 20 km/hr for other roads, 15 km/hr for boats and 5 km/hr for areas with no road or water access. Travel times were calculated as least cost estimates of travel times from each household to the nearest clinic and the nearest hospital [314].

9.2.7 Environmental risk factors

Next, to identify household-level environmental risks, we extracted land cover data for households. Detailed land cover classification for the study area was derived from satellite and aerial-based remote sensing sources, as described in (**Chapter 4**). As the strength of the association between different environmental variables and disease can vary by spatial scale (eg.[233, 234]), the proportion of the 11 land classes were extracted at 8 circular buffer radii around households: 100m, 200m, 300m, 500m, 1000m, 2000m, 3000m and 5000m. Additionally, to assess the importance of landscape configuration, fragmentation indices were extracted for each land class at each buffer radii. These included perimeter area ratio, shape index (the patch perimeter divided by the minimum possible perimeter of the patch area) and fractal dimension index (a measure of shape complexity) [241].

To identify important environmental factors, we applied a data-mining pathway utilising random forests to identify optimal variable subsets at different spatial scales based on classification performance [234]. As 90% of households had only one knowlesi sero-positive individual and only 1.5% had more than two sero-positive individuals, we used a binary classification with households classed as sero-positive if one or more household members were exposed. Initially, to explore the importance of variables at different spatial scales, we fit random forest models from 100 datasets of randomly sampled households using 80% of available data with the remaining data withheld for independent validation. Relative variable importance (RVI), a measure of importance based on a mean decrease in prediction accuracy after permuting the selected predictor, was used to explore the contribution of different variables and predictive ability was evaluated using AUC [315].

We then applied the Boruta algorithm, a feature selection algorithm designed to reduce data dimensionality and identify important features [316]. This algorithm compares the variable importance of the predictor values with shadow variables, permuted variables with no association with classification; based on the statistical significance of the importance between predictors and shadow variables over multiple random forest iterations, predictor variables are declared important or unimportant [315]. From the list of important variables, we removed highly correlated variables (Pearson's correlation coefficient > 0.8) and resulting variables were assessed for inclusion into final model.

9.2.8 Bayesian hierarchical modelling

Univariate analysis was conducted for all potential explanatory variables, including questionnaire data, landscape and spatial variables. Variables with p < 0.2 were assessed for inclusion in a multivariate binomial general linearised model and added in a forward stepwise manner to check for interactions. Final variable selection included all variables with p < 0.05. Prevalence estimates of sero-positivity to each malaria species were weighted to represent the entire population within the study area, with sampling weights calculated from the total population residing in ecotypes corresponding to each strata. Residual spatial autocorrelation was assessed using Moran's I.

The final model was developed as a Bayesian hierarchical model implemented in INLA, incorporating two levels for individual and household level effects. Individual seropositivity was denoted as $y_{ij}\ i = 1 \ldots n$; $j = 1 \ldots m$, where i is the individual and j is the household. The full model was specified as:

$$y_{ij} \sim \text{Binomial}(\pi_{ij}, n_{ij})$$

With the linear predictor for the Bernoulli model specified as:

$$n_{ij} = \text{logit}(\pi_{ij}) = \beta_0 + X_{ij}\,\beta_i + \alpha_j \gamma_j$$

Where β_0 represents the intercept, $X_{ij}\,\beta_i$ represents a vector of individual covariate effects and $\alpha_j \gamma_j$ represents the additive terms of random effects for household with a vector of household level coefficients α_j.

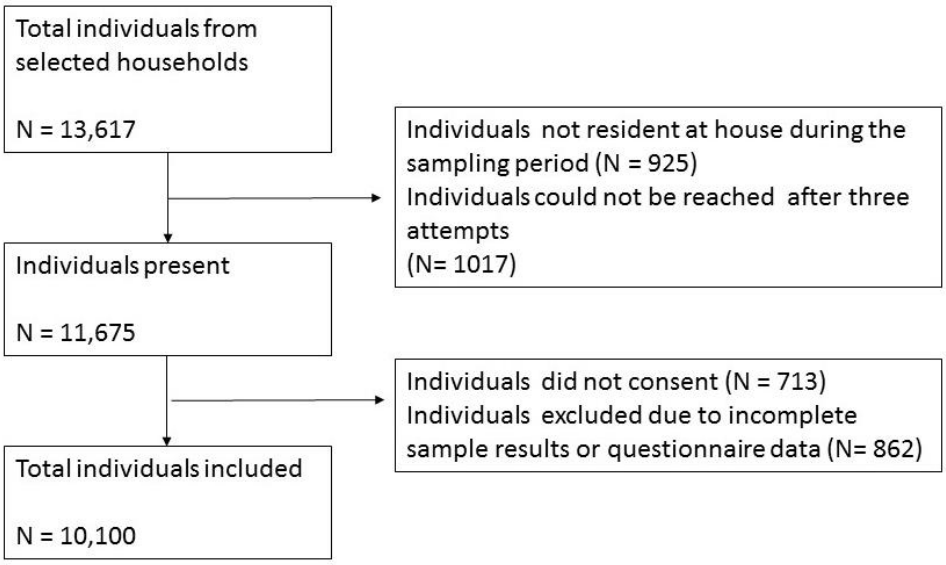
All landscape covariates were mean-centred and scaled so regression coefficents represent effects per standard deviation. Weakly informative priors of *N(0, 0.01)* were used for intercepts and fixed effect coefficients and penalised complexity priors were used for the spatial effect as described by [270]. The default parameter of logGamma (1, 0.00005) was used for the precision of the random

effect (τ_γ). Predictive performance was assessed using deviance information criteria (DIC) and AUC. All analyses were performed in R version 3.5 [317].

9.3 Results

Between September and December 2015, 10,100 individuals were sampled from 2849 households in 180 villages (Figure 9.3). The sample population included 47% men and 53% women with a median age of 25 years (range: 3 months – 105 years). Use of malaria prevention measures was high, with 78.5% (7930/10100) and 46% (4645/10100) of participants reporting using bed nets and insecticide respectively. Roughly half (46%; 4622/10100) of participants reported contact with monkeys, with similar contact rates reported between men and women. While 3% of the population self-reported having a fever, no symptomatic microscopy positive malaria cases were identified during the survey.

Figure 9.3. Recruitment flow chart for survey



9.3.1 Infection to malaria

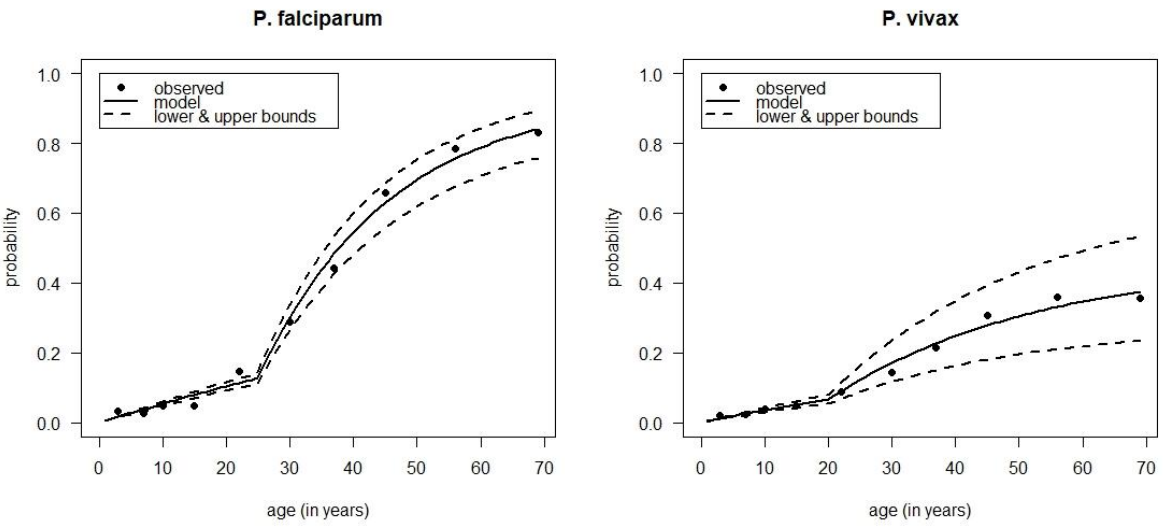
During the survey, no individuals were identified as microscopy positive for malaria. From the samples pooled for DNA extraction and PCR, a total of 9 samples were identified as *Plasmodium* positive including two positive mono-infections with *P. knowlesi* and one mixed infection with *P. knowlesi* and *P. vivax*. The remaining infections identified included three *P. malariae* infections, one *P. vivax* infection, one mixed *P. vivax* and *P. malariae* infection and one infection which could not be speciated. As none of the infections identified were in the same households or geographic areas, a

subset of 498 samples were extracted and assayed individually to assess whether PCR sensitivity was affected by pooling. Of these samples a further 58 were identified as *Plasmodium* positive, including an additional 40 samples identified as *P. falciparum* and 10 samples which could not speciated. As limited *P. falciparum* transmission has been reported in these areas, further work is being conducted to confirm these results and speciate the remaining samples.

9.3.2 Serological responses to malaria

Models for falciparum and vivax identified exposed individuals highly accurately (cross-validated AUCs: 0.977- 1 and 0.980 – 1 respectively). Overall exposure prevalence was estimated at 32.4% (95% CI: 31.4-33.4%) for *P. falciparum* and 16.4% (95% CI: 15.6-17.1%) for *P. vivax* and strongly positively associated with age for both species. While there was limited evidence of exposure in children under 10 (less than 5%), estimated sero-prevalence was highest in individuals above 70 (falciparum 83.1%, 95% CI 80.6-85.4%; vivax 35.5%, 95% CI 32.5- 38.6%). Historical changes in the force of infection were apparent for both species with the time of change estimated as 25 years ago for *P. falciparum* and 20 years for *P. vivax* ($p < 0.01$, Figure 9.4). Seroconversion rates (λ) were estimated as λ1 0.047 (95% CI: 0.042 – 0.053) and λ2 0.006 (0.005-0.006) for *P. falciparum* λ1 0.017 (0.012 – 0.023) and λ2 0.004 (0.008 – 0.022) for *P. vivax*.

Figure 9.4. Reverse catalytic models for *P. falciparum* and *P. vivax*



As limited antibody response and training data was available for *P. knowlesi*, models were less accurate although still correctly classified the majority of knowlesi exposure (cross-validated AUC: 0.841 – 0.997). Sero-prevalence to knowlesi was 5.1% (95% CI: 4.8-5.4%) in the study population,

with high responders identified across all age groups. Similar sero-prevalences were estimated in all strata, with a slightly higher sero-prevalence in villages with low forest cover in 2014 (Table 9.1).

Table 9.1. Estimated *P. knowlesi* sero-prevalence by strata

Strata	Total population	Number sampled	Sero-prevalence (95% CI)
Low forest cover	59438	3352	5.79% (5.28 – 6.30%)
Medium forest cover	39720	3339	4.64% (4.18 – 5.10%)
High forest cover	41576	3409	4.87% (4.27 – 5.47%)

9.3.3 Individual-level risk factors for *P. knowlesi*

The characteristics of included individuals and results of the univariate analysis are included in Supplementary information, Table 9.3. As well as age and gender, travel to forest areas and contact with macaques were both associated with increased odds of *P. knowlesi* seropositivity. While bed net use and other malaria prevention methods were not associated with knowlesi exposure, the use of insecticides was associated with decreased odds of sero-positivity. Additionally, individuals at higher elevations and individuals residing in houses less than 1m from the ground had lower risks of knowlesi exposure. Although occupational activities such as farm and plantation work were significant in the univariate analysis, these did not improve the final model after adjusting for age and gender.

9.3.4 Environmental risk factors for *P. knowlesi*

Out of a total of 352 environmental variables, 157 were identified as potentially important predictor variables. Marked differences in relative variable importance were identified at different spatial scales, with some variables showing clear peaks at single radii and others showing peaks at multiple distances (Figure 9.5). Unsurprisingly, some landscape variables were highly correlated (Figure 9.6) and a further 83 variables were excluded with a Pearson's correlation coefficient > 0.8.

Figure 9.5. Relative variable importance of 100 models by radius distance for a. Proportion of pulpwood, b. Intact forest perimeter- area ratio, c. Fractal dimension index of irrigated farmland (rice paddies), d. Oil palm perimeter area ratio

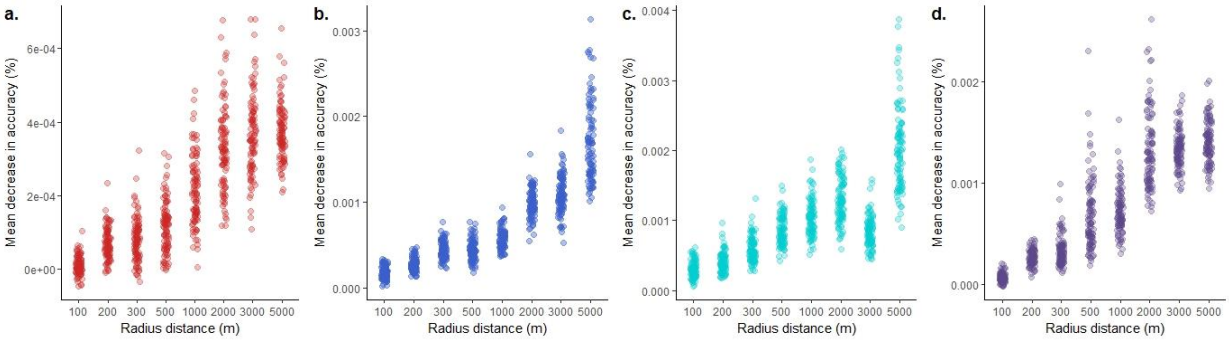
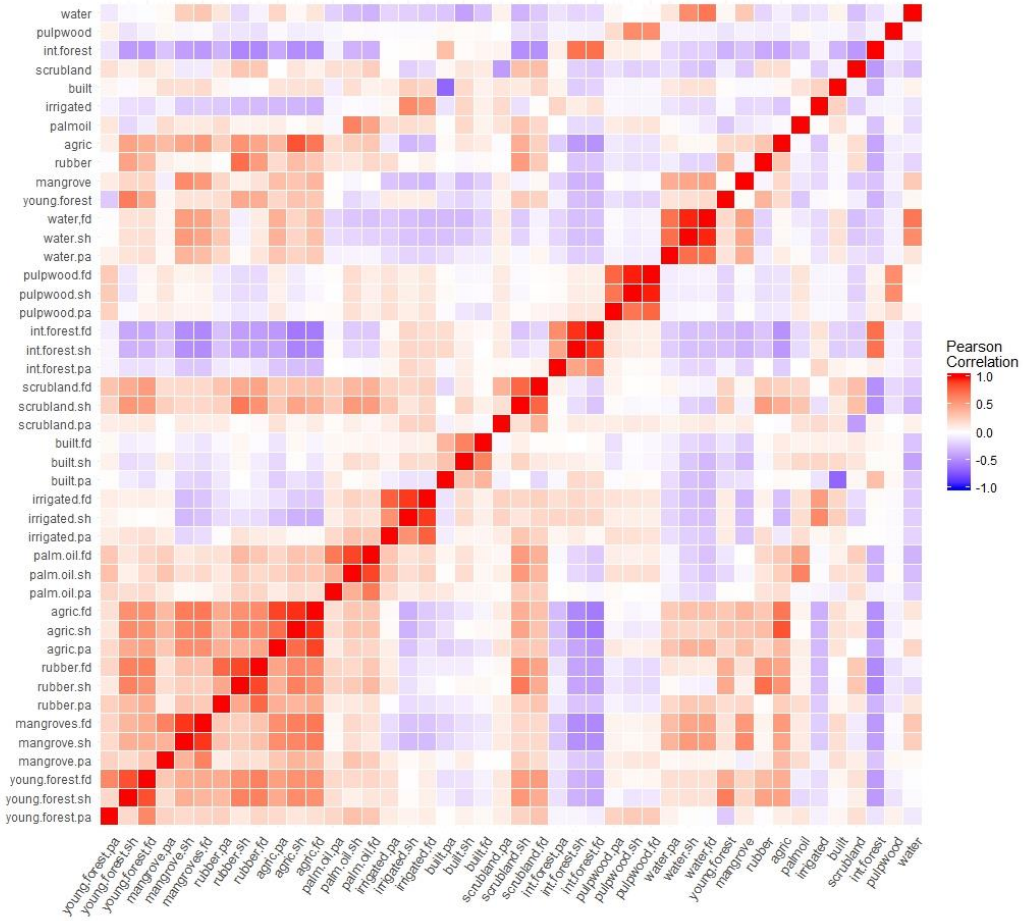


Figure 9.6. Pearson correlation for all land cover variables within a 1 km radius



9.3.5 Bayesian hierarchical model

Results of the final hierarchical model are presented in Table 9.2. The final model included four landscape variables at varying spatial scales. As Moran's I detected small but statistically significant residual spatial autocorrelation (Moran's I: 0.022, p = 0.001), we additionally explored including a spatial effect modelled as a Matern covariance function. As the spatial model did not substantially

improve predictive performance (AUC: 0.776 for spatial model vs 0.767 for non-spatial model), we reported results from the most parsimonious non-spatial model.

Table 9.2. Posterior estimates of coefficients for fixed effects for final hierarchical model of *P. knowlesi* exposure risk

Variable	Estimated	2.5% BCI	97.5% BCI
Individual level effects			
Age (per 10 years)	0.286	0.245	0.328
Male gender	0.216	0.030	0.403
Reported contact with macaques	0.338	0.137	0.542
Reported forest activities	0.621	0.356	0.881
Use insecticides	-0.269	-0.461	-0.078
Household level effects			
House at ground level	-0.277	-0.467	-0.087
Elevation (per 1000m)	-0.762	-1.311	-0.249
Intact forest perimeter- area ratio (5000m radius)*	-0.156	-0.290	-0.036
Irrigated farming fractal dimension (300m radius)*	0.156	0.061	0.249
Proportion of pulpwood plantations (3000m radius)*	0.141	0.064	0.213
Oil palm perimeter area ratio (3000m radius)*	0.096	0.003	0.186

* Variables scaled and mean-centred; increase per standard deviation

9.4 Discussion

This is the first large-scale cross-sectional survey to characterise population-level exposure and infection to *P. knowlesi* in an endemic area. Consistent with previous findings in case study communities [256], results reveal marked historical decreases in transmission intensity of *P. falciparum* and *P. vivax* while recent *P. knowlesi* exposure was identified across all demographic and age groups. Further, results suggest that both agricultural expansion and fragmentation are associated with increased risks of knowlesi exposure.

Similar to a previous case control study [103] and in contrast to other community studies [256, 271], men had higher risks of *P. knowlesi* exposure, although women and children were also identified as knowlesi-exposed. Exposure to all malaria species was strongly correlated with age, likely due to historical changes in transmission as well as increased exposure in older age groups. Clinical knowlesi cases have been most commonly identified in older age groups [11, 18, 47, 103] and, similarly, seropositivity has also been associated with increased age [256]. *P. knowlesi* exposure was also associated with contact with macaques and forest activities but exposure was not associated with duration or frequency of forest activities or overnight travel outside the house in the past month;

this may be due to the longer duration of serological positivity compared to infection data. While most malaria prevention methods, including bed net use, had no association with knowlesi seropositivity, insecticide use had a protective effect, highlighting the utility of this control method. This finding is supported by entomological data suggesting most bites occur outdoors in early evening (6pm – 10pm) before people are likely to be using bed nets [87, 152].

Elevation was negatively associated with *P. knowlesi* exposure, consistent with previous studies in this area [255] and other studies finding decreased malaria risks at higher altitudes (e.g. [117, 318, 319]). Associations between land cover metrics and exposure risks provide further evidence of linkages between habitat and knowlesi transmission. Larger patches of intact forest within a 5km range, represented by lower perimeter-area ratios, were associated with higher knowlesi transmission. Although macaques and the main mosquito vector have been reported in a range of habitat types [56, 87, 92], higher sporozoite rates have been reported from interior forest areas [152] and these larger forest patches may be important for maintaining transmission. Conversely, higher fragmentation of oil palm plantations was associated with increased knowlesi exposure, indicating fragmented landscapes and edge effects may also contribute to human risks. Oil palm plantation work has previously been identified as a risk factor for clinical *P. knowlesi* infection [103]. Configuration of irrigated land, predominantly rice paddies, in close proximity to the house was also associated with higher risks; rice paddies have been associated with increased knowlesi risk in Sabah as well as malaria risk across Southeast Asia [62, 103, 320-322]. The proportion of pulpwood plantations within 3km of households was additionally associated with higher knowlesi exposure. Little data is available on macaques or vectors within this land type and the importance of these agro-forestry systems should be explored further.

Understanding the linkages between land cover and disease risk necessitates characterising the complex interactions between human land use and movement and environmental characteristics [185]. For example, adult men may be more likely to reside near plantations and local movement patterns and vector prevention practices of these groups may determine whether these land types are important for transmission. As well, adjusting for self-reported forest activities may explain why other forest types previously associated with knowlesi risks and high vector densities were not included in the final model. *P. knowlesi* transmission is influenced by the distribution of people, macaques and mosquitoes across the environment, all of which likely occur at different spatial scales. To account for this, we utilised a data mining approach to identify which factors were important at different distances [234]. Future work could explore how spatially explicit land use

predictions can be derived from GPS tracking and questionnaire data to better understand how environment shapes disease risks.

This study was limited by diagnostic methods utilised. The infection data is likely substantially underestimated due to the poor sensitivity of molecular techniques for *P. knowlesi*, particularly as the samples were pooled. Although the development of knowlesi specific antigens presents a valuable alternative for characterising transmission, further longitudinal data on the duration and magnitude of responses to these antibodies would allow better identification of recent exposure. Despite attempts to select antigens not cross reactive between *P. knowlesi* and *P. vivax* [119], there is still remaining correlation between the responses to both species and the strong correlation with age limits the ability to assess relationships between different malaria species. Future work should explicitly include the sensitivity and specificity of these methods into the hierarchical modelling framework.

Despite these limitations, this is the first large scale population-based survey to characterise *P. knowlesi* transmission. Identification of knowlesi exposed individuals across different demographic groups illustrates that the extent of exposure within the community is likely not captured by passive surveillance systems. Clear relationships between land cover and knowlesi exposure were identified, highlighting the role of land use change in the spread of this disease. As well, the methodology utilised to characterise land use and land cover and to identify significant variables within a large, complex dataset could be applied to other zoonotic and vector-borne diseases. Further longitudinal and modelling studies are needed to fully understand how changing landscapes affect future disease risks.

9.5 Supplementary information

Table 9.3. Univariate analysis and summary statistics

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
INDIVIDUAL CHARACTERISTICS				
Age category				
Under 5	1026	4	Ref	
15- 15	2672	49	4.85 (1.74 - 13.52)	
15-30	1970	92	13.05 (4.76 - 35.78)	
30-55	2842	204	21.42 (7.90 - 58.07)	
Over 55	1590	166	33.13 (12.16 - 90.26)	< 0.001
Gender				
Female	5324	236	Ref	
Male	4776	279	1.35 (1.14 - 1.62)	0.0014
Ethnicity				
Bajau	884	47	Ref	
Dusun	5074	228	0.81 (0.57 - 1.16)	
Rungus	2682	155	1.09 (0.75 - 1.57)	
Sungoi	410	24	1.07 (0.61 - 1.87)	
Other	1050	61	1.08 (0.71 - 1.66)	0.1067
Self-reported previous malaria diagnosis				
No	8771	391	Ref	
Yes	1329	124	2.28 (1.82 - 2.86)	< 0.001
Report taking anti-malaria medication				
No	10026	513	Ref	
Yes	74	2	0.50 (0.12 - 2.13)	0.3022
Treatment- seeking behaviour during fever: obtain medicines from clinic				
No	7085	351	Ref	
Yes	3015	164	1.11 (0.90 - 1.36)	0.3454
Treatment- seeking behaviour during fever: take traditional medicines				
No	9672	485	Ref	
Yes	428	30	1.40 (0.92 - 2.12)	0.1247
Treatment- seeking behaviour during fever: go to the hospital				
No	4820	248	Ref	
Yes	5280	267	0.98 (0.81 - 1.19)	0.8429
Treatment- seeking behaviour during fever: don’t seek treatment				
No	8775	457	Ref	
Yes	1325	58	0.84 (0.62 - 1.13)	0.2491
Treatment- seeking behaviour during fever: don’t seek treatment				
No	10086	514	Ref	
Yes	14	1	1.32 (0.16 - 11.14)	0.8052
Occupation				
Other	296	21	Ref	
Fishing	180	7	0.49 (0.20 - 1.24)	
Office/shop	371	28	1.08 (0.58 - 2.02)	< 0.001

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
Rubber	254	24	1.42 (0.74 - 2.73)	
Palm oil plantation	96	6	0.90 (0.34 - 2.42)	
student	2740	54	0.25 (0.15 - 0.43)	
farmer	1412	147	1.55 (0.94 - 2.57)	
none	4751	228	0.64 (0.40 - 1.05)	
Farm work				
No	8075	335	Ref	
Yes	2025	180	2.37 (1.93 - 2.90)	< 0.001
Occupation place				
In village	2586	164	Ref	
In district	1864	74	0.58 (0.43 - 0.78)	
Around the house	5538	273	0.75 (0.61 - 0.93)	
Different district	112	4	0.54 (0.19 - 1.54)	0.002
Travel to or from work or school between 11pm and 6am				
No	9211	477	Ref	
Yes	889	38	0.79 (0.55 - 1.13)	0.184
Travel to or from work or school between 5pm and 10pm				
No	8448	407	Ref	
Yes	1652	108	1.43 (1.13 - 1.80)	0.003
Walk to work or school				
No	7490	340	Ref	
Yes	2610	175	1.53 (1.26 - 1.87)	< 0.001
Walk to work or school through forest				
No	9146	436	Ref	
Yes	954	79	1.82 (1.40 - 2.38)	< 0.001
Go to forest				
No	9345	428	Ref	
Yes	755	87	2.91 (2.23 - 3.80)	< 0.001
Go to forest between 11pm and 6am				
No	10016	507	Ref	
Yes	84	8	2.09 (0.96 - 4.57)	0.086
Go to forest between 5pm and 10pm				
No	9664	461	Ref	
Yes	436	54	2.99 (2.15 - 4.14)	< 0.001
Go to forest at night (5pm – 6am)				
No	9622	457	Ref	
Yes	478	58	2.95 (2.15 - 4.04)	< 0.001
Hunting in forest				
No	9900	494	Ref	
Yes	200	21	2.31 (1.41 - 3.79)	0.002
Collect wood in forest				
No	9870	480	Ref	
Yes	230	35	3.82 (2.54 - 5.74)	< 0.001
Cleared land in the past year				
No	8185	349	Ref	< 0.001

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
Yes	1915	166	2.28 (1.85 - 2.81)	
Involved in construction in the past year				
No	9945	505	Ref	
Yes	155	10	1.33 (0.68 - 2.63)	0.422
Other activities in evenings				
Sport	611	30	Ref	
Other	491	44	1.99 (1.20 - 3.32)	
None	7332	339	0.95 (0.63 - 1.41)	
Visiting outside house	1403	84	1.24 (0.79 - 1.95)	
Fishing	263	18	1.49 (0.79 - 2.81)	0.001
Any early morning activities outside the house				
No	9768	503	Ref	
Yes	332	12	0.72 (0.39 - 1.33)	0.270
Any evening activities outside the house				
No	7587	382	Ref	
Yes	2513	133	1.05 (0.84 - 1.30)	0.673
Usually bathe outside				
No	7510	354	Ref	
Yes	2590	161	1.37 (1.11 - 1.69)	0.004
Usually bathe at river				
No	9565	491	Ref	
Yes	535	24	0.86 (0.55 - 1.36)	0.517
Usually bathe outside at night				
No	7571	374	Ref	
Yes	2529	141	1.14 (0.92 - 1.42)	0.227
Typical amount of time spent outside of the house at night				
1 - 3 hours	1643	96	Ref	
Less than 1 hour	6318	320	0.84 (0.66 - 1.09)	
More than 3 hours	359	20	0.93 (0.55 - 1.59)	
Don't know	1780	79	0.72 (0.52 - 0.99)	0.254
Have stayed outside the village in the past month				
No	9678	486	Ref	
Yes	422	29	1.39 (0.92 - 2.10)	0.134
Have slept outside walls (out of houses) in the past month				
No	10050	511	Ref	
Yes	50	4	1.54 (0.51 - 4.59)	0.464
Sleep under a bednet				
No	2170	87	Ref	
Yes	7930	428	1.37 (1.06 - 1.76)	0.013
Use insecticide				
No	5455	305	Ref	
Yes	4645	210	0.78 (0.64 - 0.95)	0.012
Use a fan to prevent mosquitoes				
No	8124	438	Ref	
Yes	1976	77	0.69 (0.53 - 0.91)	0.006

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
Use smoke to prevent mosquitoes				
No	9242	464	Ref	
Yes	858	51	1.22 (0.88 - 1.69)	0.236
Use window screens to prevent mosquitoes				
No	10044	513	Ref	
Yes	56	2	0.70 (0.16 - 3.04)	0.622
Don't use any mosquito prevention				
No	9852	504	Ref	
Yes	248	11	0.93 (0.48 - 1.79)	0.822
Contact with monkeys				
No	5478	207	Ref	
Yes	4622	308	1.92 (1.58 - 2.34)	< 0.001
Contact with long-tailed macaques				
No	5674	222	Ref	
Yes	4426	293	1.83 (1.51 - 2.23)	< 0.001
Contact with pig-tailed macaques				
No	9271	464	Ref	
Yes	829	51	1.25 (0.91 - 1.72)	0.183
Monkeys seen around the house				
No	8816	425	Ref	
Yes	1284	90	1.53 (1.18 - 1.99)	0.002
Monkeys seen around the village				
No	7433	339	Ref	
Yes	2667	176	1.52 (1.24 - 1.86)	< 0.001
Monkeys seen around the farm or plantation				
No	9013	431	Ref	
Yes	1087	84	1.72 (1.32 - 2.23)	< 0.001
Frequency of monkey sightings				
Never	5449	207	Ref	
Monthly	1774	112	1.80 (1.39 - 2.32)	
Yearly	617	33	1.52 (1.02 - 2.27)	
Weekly	1080	81	2.19 (1.64 - 2.93)	
Daily	1180	82	1.99 (1.49 - 2.66)	< 0.001
HOUSEHOLD CHARACTERISTICS				
Socioeconomic status				
Quartile 1	2251	155	Ref	
Quartile 2	2526	125	0.69 (0.53 - 0.91)	
Quartile 3	2626	123	0.65 (0.49 - 0.85)	
Quartile 4	2697	112	0.57 (0.43 - 0.75)	< 0.001
Length of time resident at current house				
1 to 5 years	1635	90	Ref	
Over 5 years	8059	404	0.91 (0.71 - 1.19)	
Less than 1 year	385	20	0.95 (0.55 - 1.64)	
Unknown	21	1	0.86 (0.10 - 7.51)	0.927
Age of house				

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
1 to 5 years	1704	98	Ref	
Over 5 years	7993	396	0.85 (0.66 - 1.10)	
Less than 1 year	368	20	0.95 (0.55 - 1.63)	
Unknown	35	1	0.47 (0.06 - 3.86)	0.576
Household head education				
None	2224	164	Ref	
Primary	4190	196	0.60 (0.47 - 0.76)	
Secondary	3686	155	0.53 (0.41 - 0.68)	< 0.001
Corrugated iron roof				
No	1136	54	Ref	
Yes	8964	461	1.12 (0.81 - 1.54)	0.494
Concrete or tile floor				
No	6854	371	Ref	
Yes	3246	144	0.79 (0.63 - 0.98)	0.031
Wood or bamboo walls				
No	1957	90	Ref	
Yes	8143	425	1.17 (0.90 - 1.51)	0.229
House height				
Ground level	3478	161	Ref	
Less than 1m	2094	79	0.81 (0.60 - 1.09)	
Over 1m	4416	273	1.39 (1.11 - 1.73)	
Over water	112	2	0.36 (0.08 - 1.56)	< 0.001
Gaps in eaves of the house				
No	5561	264	Ref	
Yes	4539	251	1.19 (0.98 - 1.45)	0.077
Number of windows in the house that can close				
None	1088	57	Ref	
Some	4546	241	1.02 (0.73 - 1.41)	
All	4466	217	0.91 (0.66 - 1.27)	0.567
Insect screens observed in house				
No	9410	480	Ref	
Yes	690	35	1.01 (0.68 - 1.49)	0.968
Kitchen outside of house				
No	9661	495	Ref	
Yes	439	20	0.88 (0.53 - 1.44)	0.600
House has a toilet				
No	1279	83	Ref	
Yes	8821	432	0.73 (0.56 - 0.96)	0.028
Toilet is inside the house				
No	5787	307	Ref	
Yes	4313	208	0.90 (0.74 - 1.10)	0.289
Piped water inside the house				
No	4296	246	Ref	
Yes	5804	269	0.78 (0.64 - 0.95)	0.015
Household owns cattle				

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
No	9907	496	Ref	
Yes	193	19	2.19 (1.25 - 3.83)	0.010
Household owns buffalo				
No	9985	506	Ref	
Yes	115	9	1.61 (0.74 - 3.51)	0.254
Household owns goats				
No	10011	512	Ref	
Yes	89	3	0.64 (0.19 - 2.21)	0.465
Household owns pigs				
No	9871	495	Ref	
Yes	229	20	1.94 (1.14 - 3.31)	0.021
Household has pet monkey				
No	9863	499	Ref	
Yes	237	16	1.37 (0.76 - 2.45)	0.308
Other household in village has a pet monkey				
No	8267	428	Ref	
Yes	1833	87	0.90 (0.69 - 1.17)	0.423
Monkeys observed raiding household crops				
No	9006	445	Ref	
Yes	1094	70	1.35 (1.01 - 1.80)	0.050
River observed near house				
No	4911	264	Ref	
Yes	5189	251	0.87 (0.72 - 1.07)	0.191
Pond observed near house				
No	8157	413	Ref	
Yes	1943	102	1.02 (0.80 - 1.30)	0.888
Well observed near house				
No	8435	422	Ref	
Yes	1665	93	1.14 (0.88 - 1.47)	0.329
Water-filled plastic containers observed near house				
No	7473	398	Ref	
Yes	2627	117	0.81 (0.64 - 1.02)	0.076
Lake observed near house				
No	10018	509	Ref	
Yes	82	6	1.48 (0.58 - 3.78)	0.434
House near sea				
No	9262	474	Ref	
Yes	838	41	0.96 (0.67 - 1.38)	0.830
Household farms fruit				
No	8527	422	Ref	
Yes	1573	93	1.22 (0.95 - 1.59)	0.130
Household farms rubber				
No	6564	312	Ref	
Yes	3536	203	1.24 (1.01 - 1.51)	0.040
Household farms corn				

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
No	9854	493	Ref	
Yes	246	22	1.87 (1.12 - 3.12)	0.023
Household keeps livestock				
No	10082	514	Ref	
Yes	18	1	1.07 (0.12 - 9.67)	0.95059
Household farms vegetables				
No	8454	421	Ref	
Yes	1646	94	1.17 (0.91 - 1.51)	0.236
Household has rice paddies				
No	8791	430	Ref	
Yes	1309	85	1.38 (1.05 - 1.81)	0.022
Household farms oil palm				
No	9106	460	Ref	
Yes	994	55	1.07 (0.78 - 1.48)	0.681
Distance of farming land from the house				
Near the house	2110	135	Ref	
Same village	3810	189	0.75 (0.58 - 0.97)	
No farmland	3747	166	0.66 (0.51 - 0.86)	
Outside the village	433	25	0.89 (0.54 - 1.45)	0.018
Use pesticides for farming				
No	7162	356	Ref	
Yes	2938	159	1.08 (0.88 - 1.34)	0.456
Swidden farming				
No	5338	255	Ref	
Yes	4762	260	1.16 (0.95 - 1.41)	0.136
Household collects wood from forest				
No	7716	372	Ref	
Yes	2384	143	1.26 (1.01 - 1.57)	0.043
Household collects food from forest				
No	8074	399	Ref	
Yes	2026	116	1.17 (0.92 - 1.48)	0.206
Household collects medicine from forest				
No	8627	430	Ref	
Yes	1473	85	1.14 (0.87 - 1.49)	0.343
Travel time to nearest clinic				
Quartile 1	2495	102	Ref	
Quartile 2	2592	147	1.44 (1.08 - 1.91)	
Quartile 3	2420	127	1.31 (0.98 - 1.76)	
Quartile 4	2593	139	1.35 (1.01 - 1.80)	0.065
Travel time to nearest hospital				
Quartile 1	2467	110	Ref	
Quartile 2	2599	109	0.94 (0.70 - 1.26)	
Quartile 3	2533	132	1.19 (0.90 - 1.59)	
Quartile 4	2501	164	1.55 (1.18 - 2.05)	0.001
Elevation (metres above sea level)				

Variable	Total number	Knowlesi exposed	Crude Odds Ratio (95% CI)	P value
Under 50 MSL	5701	322	Ref	
50-250 MSL	2248	111	0.86 (0.67 - 1.09)	
250-500 MSL	1227	63	0.89 (0.65 - 1.21)	
Over 500 MSL	924	19	0.34 (0.21 - 0.55)	< 0.001

10 Discussion

This thesis aimed to characterise how environmental change effects human *P. knowlesi* risks by 1.) describing the spatial epidemiology of knowlesi infection and exposure on a range of spatial scales and 2.) evaluating associations with environmental and behavioural factors. This chapter highlights the main research findings from each chapter and describes the strengths, weaknesses, and conclusions of this work. The implications of these findings and future research directions for *P. knowlesi* and other vector-borne zoonotic diseases are discussed.

10.1 Main research findings

10.1.1 Historical patterns of *P. knowlesi* incidence

Although the initial cluster of human *P. knowlesi* cases in Borneo was identified ten years prior to the start of this thesis, limited data was available on the distribution of *P. knowlesi* cases at fine spatial scales. To assemble baseline data to guide further research plans, we assembled data on reported cases in Kudat and Kota Marudu from 2008 – 2013 (**CHAPTER 5**). A key limitation of data collected during this time period was the misdiagnosis of *P. knowlesi* as *P. malariae* by microscopy and inconsistent application of molecular techniques able to differentiate between these species. To overcome this issue, we estimated sensitivity and specificity of microscopy for a subset of 346 molecularly confirmed samples and calculated adjusted village level knowlesi incidence. Results revealed marked fine-scale spatial heterogeneity varying annually. Importantly, this analysis also identified positive associations between adjusted village level incidence and both forest cover and forest loss in surrounding areas, showing the first link between deforestation and *P. knowlesi* incidence. Additionally, as this data was only collected through clinical reports, no conclusions could be drawn about the spatial distribution of *P. knowlesi* infection in the wider population as data was not available on the proportion of asymptomatic infections or individuals who did not seek treatment.

10.1.2 Community-level *P. knowlesi* infection and exposure

An important limitation of this retrospective analysis was the reliance on passive surveillance systems. Despite having strong government reporting systems offering free treatment for malaria, these programmes only capture symptomatic cases attending health facilities and the levels of asymptomatic knowlesi infections remained unknown. Initially, to identify whether asymptomatic

infections were present in this population, we screened individuals residing in the same households or villages as symptomatic *P. knowlesi* cases, a population assumed to be most likely to have been exposed to *P. knowlesi* (**CHAPTER 7**). Within this population, asymptomatic infections were identified in all age groups and in both men and women, with an estimated prevalence of 6.9% (5.6-8.4%). This data suggests infection with *P. knowlesi* is more common than previously thought and may affect a wider demographic group than the primarily adult men reporting as symptomatic cases. Critically, this work also highlighted the limitations of molecular diagnostics to correctly identify low density infections; sensitivity of the frequently used species-specific ssRNA nested PCR [277] was only estimated at 15% and the Real Time PCR assays used only had sensitivity ranging from 81.3-87.9%. The only highly sensitive assay able to identify asymptomatic infections was the genus level ssRNA nested PCR described by [276]; however, this assay could not differentiate between species and had poor specificity.

Alternatively, species-specific antibody responses have been used to characterise levels of exposure within the community; these serological techniques may be particularly useful in low transmission settings with low probabilities of infection. Following the development of species-specific antigens for *P. knowlesi*, we conducted comprehensive surveys of three case study areas, including two areas of relatively high *P. knowlesi* transmission identified during retrospective analyses and an area with few sporadic *P. knowlesi* cases reported in the Philippines (**CHAPTER 8**). As in the study of asymptomatic infections, similar sero-prevalence was estimated in men and women and across all age groups. While agricultural work and some environmental risk factors for *P. knowlesi* were identified, population-level risk factors needed to be assessed across a wider population rather than in case study communities. Both *P. falciparum* and *P. vivax* showed evidence of historical decreases in transmission in all sites and spatial patterns of exposure were not correlated with *P. knowlesi*.

To address these limitations, an environmentally stratified two stage cross-sectional survey was conducted across four districts in Northern Sabah. The initial association between village level knowlesi incidence and forest cover within a 2km radius was used to group villages into three strata and over 10,000 individuals were sampled within a 3 month period (**CHAPTER 9**). The ability to identify asymptomatic infections by PCR remained limited, with sensitivity even further reduced by the need to pool samples for DNA extraction and PCR due to the large number of samples. Due to this limitation, sero-positivity was used as the main outcome for all risk factor analysis. To identify more recent exposure corresponding to the year the environmental and demographic data was collected, we utilised a supervised classification approach with a positive training dataset assembled

from recently infected individuals and a negative training dataset including both individuals from malaria unexposed areas (i.e. the United Kingdom) and individuals exposed to *P. falciparum* and *P. vivax* in regions outside Southeast Asia. Using this methodology, we estimated *knowlesi* sero-prevalence as 5.1% (4.8-5.4%) within this region and identified population-level risk factors for recent exposure, including age, forest activities, contact with macaques and housing structure. Contrary to prior studies, there were increased odds for knowlesi exposure in men compared to women, however; sero-positive individuals were identified in all demographic groups. Clear decreases in transmission were observed for *P. falciparum* and *P. vivax*, with few sero-positive individuals identified in younger age groups.

10.1.3 Environmental risk factors

A core focus of this thesis is utilising remote sensing technologies to explore environmental risk factors for *P. knowlesi*. While initial retrospective studies demonstrated a link between deforestation and *P. knowlesi* incidence, this study was limited by the coarse spatial data utilised, in terms of aggregating incidence at a village level represented by a point centroid as well as the limited forest cover used, classifying an area as forest or other based on a 50% canopy threshold [51]. As subsequent epidemiological studies were conducted for this thesis, populations were mapped down to individual household and movements and land cover was classified into more descriptive categories at higher resolutions (**CHAPTER 4**).

Environmental, land use and demographic factors interact on a range of spatial and temporal scales, making identifying significant factors particularly challenging. Initial reports of *P. knowlesi* cases were primarily in adult men reporting frequent trips into forested environments and mosquito vectors are known to be primarily exophagic; if transmission primarily occurs in areas far from households, there would be limited utility in identifying environmental risk factors in the immediate vicinity of the house. To explore how human movement contributes to transmission risk, we conducted a GPS tracking study, mapping movements of over 300 individuals on a fine-spatial scale (**CHAPTER 6**). This study, along with questionnaire data on reported movements, revealed most individuals spent the majority of their time within several hundred meters of the house and rarely travelled more than 5km during peak mosquito biting periods.

As the detailed distribution of macaques and mosquito vectors across these environments remained unknown, we employed a data mining approach to identify what environmental variables at what spatial scales were most predictive of knowlesi risk (as defined by household sero-exposure).

Limiting these variables to within the 5km range predicted by human movement models, we utilised a machine learning feature selection approach to identify the most predictive variables, identifying landscape features and individual factors associated with knowlesi risk. Habitat characteristics such as less fragmented intact forests, fragmentation of plantation areas and irrigation were associated with increased knowlesi risks, illustrating the importance of land use and land cover for transmission of this parasite.

10.2 Research in context

During the course of this work, *P. knowlesi* was increasingly recognised as a significant public health problem and a potential threat to malaria elimination within Southeast Asia. Within the past five years, research has further characterised clinical and epidemiological aspects of *P. knowlesi*, leading to changes in public health policy. Notably within Malaysia, policy responses included implementing molecular diagnostics for all malaria cases and modifying national treatment guidelines for *P. knowlesi* [323, 324]. At the same time, Malaysia achieved massive reductions in nonzoonotic malaria cases, with reported cases decreasing from over 12,000 in 2001 to 723 in 2016 [13, 325]. *P. knowlesi* accounted for 69% of malaria cases in Malaysia in 2016, with a general upward trend in *P. knowlesi* cases reported despite interannual variations apparently due to climatic factors [325]. A World Health Organisation (WHO) expert consultation on *Plasmodium knowlesi* conducted in Malaysia in 2017 concluded there was no evidence of widespread human to human transmission, suggesting *P. knowlesi* remains primarily a zoonosis [325]. While the inclusion of *P. knowlesi* in malaria elimination programmes is still debated by WHO, Malaysia targets elimination of nonzoonotic malaria species by 2020, with the majority of remaining public health costs due to zoonotic species. As *P. knowlesi* has now been identified in multiple countries in the region, including over 450 cases reported in Indonesia, and guidelines for *P. knowlesi* are largely based on research from Malaysia, further research is needed on the extent and epidemiology of knowlesi regionally.

With increased awareness and improved surveillance, notable advances have been made towards understanding the true burden and distribution of *P. knowlesi*. Initial risk maps based on isolated reported cases have been further refined to account for the distribution of macaque hosts and mosquito vectors [9, 10, 26]. Simultaneously, asymptomatic infections have been identified in both Sabah and Sarawak, highlighting that reported cases may not fully reflect infection within the community [271, 326]. Application of molecular diagnostics also identified other human cases with the primate malaria *P. cynomolgi*, demonstrating the spillover potential for other primate *Plasmodium* species [74]. As further data is available on human infections, new evidence has

emerged on the spatial and environmental factors contributing to transmission. Within Sabah, the main *knowlesi* mosquito vector was incriminated and identified in peri-domestic areas surrounding households [87, 152]. Entomological data, along with the identification of indoor residual spraying as a protective risk factor, support the continued use of vector control measures around households [103]. The possibility of peri-domestic transmission is further supported by the relationship between disease risks and land use change and fragmentation in areas immediately surrounding households [235, 255]. As the body of research around *P. knowlesi* continues to increase, further advances may be made in guiding control programmes for knowlesi and other zoonotic diseases.

10.3 Limitations and further research questions

10.3.1 Diagnostic challenges

Despite these gains, further research is needed to fully understand the mechanisms driving the emergence of *P. knowlesi* and refine control strategies. While this thesis has aimed to characterise the epidemiology of *P. knowlesi* in people, a key limitation has been available diagnostic techniques. Microscopy lacks specificity to distinguish between *P. knowlesi* and other malaria species and, despite the ability to identify *knowlesi* in symptomatic cases, commonly used molecular techniques lack the sensitivity to identify low density asymptomatic infections [42, 271]. Real Time PCR methods and sequencing approaches have much higher sensitivity but these methods are prohibitively expensive to apply at a population level. Similarly, the sensitivity of molecular methods for identifying other potentially zoonotic malaria species has not been well characterised. Within this thesis, different molecular techniques frequently yielded inconsistent results. In the cross-sectional survey, despite extensive optimisation of the PCR assay and sensitivity testing, surprisingly few PCR positives were identified from pooled PCR samples. The identification of further *Plasmodium* infections on un-pooled samples suggests poor sensitivity for this method and need for further development.

The development of species-specific antigens for *P. knowlesi* infection has allowed further characterisation of *P. knowlesi* transmission intensity, a particularly valuable methodology when the probability of detecting infections is low due to both infrequent infections and poor sensitivity of diagnostics [114]. While duration of these antibody responses appears short-lived and we have attempted to optimise analysis methods to detect recent exposure, these methods could be further expanded by development of a larger panel of antigens or validation against a larger longitudinal dataset of antibody responses. Further development of a panel of antigens could allow for selection

of markers of recent exposure, as has been developed for *P. falciparum* [287]. Additionally, methods for analysing multiple serological responses need further refinement; within this thesis, machine learning methods were applied to analyse multi-antigen responses, but these would benefit from more extensive training data and further validation.

Alternatively, as more data is available on the limitations of different diagnostics methods, analysis methods could be expanded to address this uncertainty. Bayesian hierarchical models have been used to account for sensitivity and specificity of diagnostic methods [327, 328]. While a similar approach was initially used to assess sensitivity and specificity of microscopy (**CHAPTER 5**), future models of disease risk should explicitly incorporate these diagnostic limitations. Additionally, models could be further refined to identify geographic areas with high transmission by using both serological and molecular results with other available measures of transmission intensity such as the entomological inoculation rate [115, 117].

10.3.2 Transmission dynamics

Even with these limitations, there is increasing epidemiological evidence describing infections and exposure in people. In contrast, little data exists about the prevalence or infectiousness in macaque hosts. Identification of infections in macaques is limited by the same diagnostic challenges as in human samples, further complicated by the difficulty of obtaining macaque blood samples or isolating *Plasmodium* from faecal samples [154, 329]. Collecting further data on macaque populations is complicated by both logistical difficulties in catching, sampling and detecting macaques as well as strict ethical guidelines on the sampling of non-human primates, both from local and international guidelines (e.g. CITES, the Convention on International Trade in Endangered Species of Wild Fauna and Flora). Although little is known about the pathology of knowlesi infections in macaques, the identification of large proportions of macaque populations as infected suggests that infections are likely to be long term and present in low densities [15]. Further, there is limited data on the basic population dynamics and distribution of macaques and how these populations respond to habitat changes. Within the Matunggong, Kudat study site, a case study report of a GPS-collared macaque troop suggests that macaque ranging behaviour is disturbed by deforestation events, with macaques moving in close proximity to houses where symptomatic cases were detected [153]. Within this same geographic area, phylogenetic analysis provided further evidence of common *P. knowlesi* isolates between macaques, human cases and *An. balabacensis* mosquitoes in this region [330].

While these data anecdotally support the theory that transmission remains primarily zoonotic and driven by increased spatial overlap between people, macaques and mosquitoes in response to land use change, the possibility of human to human transmission cannot be ruled out. Experimental studies demonstrated *P. knowlesi* can be transmitted from infected individuals to other people via the mosquito vector [91]. Mathematical modelling studies suggest that while human to human transmission is unlikely, and humans are probably do not play a major role in parasite maintenance, plausible scenarios of human to human transmission were identified [75, 235]. High densities of the main mosquito vector identified around populated peri-domestic areas further highlight the plausibility of transmission between people [87].

To better understand transmission, further research is needed to characterise fine-scale habitat preferences and densities of vectors and primate hosts. Currently available data is either too geographically limited to allow for extrapolation to other areas (e.g. [152, 153]) or too broadly aggregated at larger spatial scales to examine effects of changing movement patterns or habitat selection in response to land use change [10]. More detailed data on the densities of these populations would additionally allow exploration of vector biting preferences based on host availability in different environments. Incorporating detailed spatial data on host, vector and human populations would enable further insights into the mechanistic links between knowlesi incidence and land use change as well as quantification of the contribution of different populations to transmission and design of targeted control strategies.

10.3.3 Long term ecological changes

One of the key limitations of this thesis is the limited temporal resolution. Land use change is a dynamic process; initial impacts on disease transmission from disruption of existing ecosystems may change over time as transmission reaches new equilibrium states. Following deforestation, subsequent stages of forest succession and agricultural development may create new habitats for disease vectors and hosts while simultaneously leading to expanding human settlements and changes in occupational activities [6, 69]. Commonly used geostatistical modelling approaches quantifying associations between disease metrics and spatially dependent environmental variables assume static relationships between disease and the environment and fail to account for temporal changes in transmission, which may over or under-estimate associations between land use change and disease [134, 331].

Within this thesis, all datasets collected were cross sectional or of limited duration (less than five years) and unable to fully capture the breadth of these longer-term changes. In some cases, spatial data is used in lieu of longitudinal data, for example, classifying forest cover by succession stage rather than following up an area since time of disturbance. The strong association found between historical deforestation and current disease incidence suggest temporal fluctuations in *P. knowlesi* incidence occur several years after forest clearing (**CHAPTER 5**); however, further research is needed to elucidate the relative contributions of deforestation, climate and social factors. If longitudinal datasets could be assembled on disease incidence and environmental characteristics, epidemiological time-series methods could be utilised to explore temporal associations between disease incidence and environmental factors and examine scales of variability (e.g. through wavelet analysis). These approaches have been widely used to study forcing from climate change but have not be applied to deforestation [332]. Future research is needed to explore the significance and interactions between factors operating at different spatial and temporal scales, such as seasonal disease dynamics and longer term climate and ecological changes, including forest loss and fragmentation at a range of spatial scales [235, 333, 334]. These methods could additionally be extended to explore potential feedback loops; although deforestation may be linked to increased knowlesi transmission, increased economic development and risk behaviours may mitigate these risks (e.g. [335]).

10.4 Implications for infectious disease emergence and re-emergence

P. knowlesi exemplifies an emerging public health issue driven by landscape changes and human encroachment into wildlife habitats. The potentially devastating human health consequences of these anthropogenic environmental changes on infectious disease transmission have been illustrated by recent outbreaks of zoonotic diseases such as Ebola and Zika [336]. Although these impacts of land changes can be challenging to predict, with disruptions to ecological systems leading to either increases or decreases in human disease risks depending on a range of context-specific environmental, biological and social factors, understanding the drivers for disease emergence and identifying high risk populations and areas to establish surveillance is critical for preventing the next outbreak [337-339].

Methodologies within this thesis demonstrate iterative approaches to identifying people and places associated with higher risks using increasingly accessible remote sensing data. Spatially and temporally accurate data on both disease and land cover is needed to determine risk factors and

monitor transmission in changing landscapes. This requires first characterising the spatial epidemiology of disease and examining whether community-level patterns of infection and exposure are captured by clinical case reports through passive surveillance systems [340]. Integrating this data with detailed data on land cover, use and configuration allows identification of landscape factors influencing disease risks and can refine predictive models and guide design of spatially-targeted interventions. Data driven analysis pathways, such as those described in Chapter 9, can be used to interrogate highly correlated, high dimensional environmental datasets to uncover significant associations and narrow the field of plausible mechanisms connecting land use change to disease [234].

The most direct implications of the work conducted within this thesis are for other emerging and vector-borne diseases within Borneo. This region of Southeast Asia has been identified as a hotspot for disease emergence due to the high levels of biodiversity, rapid environmental change in forested areas and global connectedness through human travel and trade [341]. Within the past decades, the first large outbreak of Nipah virus was reported from West Malaysia and traced to land use changes bringing bats and pigs into closer proximity; this disease has subsequently been reported in other Asian countries, with high human mortality rates described in both India and Bangledesh [342, 343]. High prevalence of lethal zoonotic viruses such as Macacine herpesvirus 1 (B virus) have also been described in wild macaques in Malaysia and the increased proximity between people and macaques due to habitat changes may also lead to elevated infection risks [344].

Changes to vector biology resulting from habitat modification can further undermine vector-borne disease control programmes in this region. Malaysia is moving towards malaria elimination; in addition to the emergence of *P. knowlesi* and reports of *P. cynomolgi*, changing mosquito bionomics and outdoor biting patterns may challenge elimination efforts [152, 325]. Land use and land cover changes may also impact *Aedes* mosquito populations, with high *Aedes* densities reported in rubber plantations and in peri-urban areas of Sabah [63, 345]. In contrast to many other settings, a large proportion of dengue cases in Malaysia are reported from rural areas, with associations reported with land cover factors [346, 347]. Similarly, Zika virus has been reported in Sabah, with recent cases identified in individuals with no history of travel or European tourists visiting Sabah [348-351]. Both Zika and dengue have been identified in free-ranging primates in Malaysian Borneo, highlighting the relevance of habitat changes for disease transmission for pathogens other than *P. knowlesi* [352, 353]. Developing appropriate surveillance systems for emerging diseases requires further characterising links between environmental change and infectious disease transmission.

10.5 Conclusions

This thesis has attempted to characterise the spatial epidemiology of *P. knowlesi* in Sabah, Malaysia and identify linkages between human *P. knowlesi* and environmental change. Despite identifying key knowledge gaps and areas where further research is needed, this thesis has contributed to understanding of the baseline community-level epidemiology of *P. knowlesi* and identified associations with environmental change.

Further, the methods utilised within this thesis are widely applicable to other zoonotic and vector-borne diseases with suspected linkages to land use change. As people continue to modify natural landscapes, understanding how these changes influence infectious disease transmission and emergence is critical to guide future disease control programmes and fully understand the implications of land use and management.

11 References

1. Patz, J.A., et al., *Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence.* Environ Health Perspect, 2004. **112**(10): p. 1092-8.
2. Lambin, E.F., et al., *Pathogenic landscapes: interactions between land, people, disease vectors, and their animal hosts.* Int J Health Geogr, 2010. **9**: p. 54.
3. Jones, K.E., et al., *Global trends in emerging infectious diseases.* Nature, 2008. **451**(7181): p. 990-3.
4. Gottdenker, N.L., et al., *Anthropogenic land use change and infectious diseases: a review of the evidence.* Ecohealth, 2014. **11**(4): p. 619-32.
5. Kilpatrick, A.M. and S.E. Randolph, *Drivers, dynamics, and control of emerging vector-borne zoonotic diseases.* Lancet, 2012. **380**(9857): p. 1946-55.
6. Guerra, C.A., R.W. Snow, and S.I. Hay, *A global assessment of closed forests, deforestation and malaria risk.* Ann Trop Med Parasitol, 2006. **100**(3): p. 189-204.
7. Hahn, M.B., et al., *Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon.* PLoS One, 2014. **9**(1): p. e85725.
8. Yasuoka, J. and R. Levins, *Impact of deforestation and agricultural development on anopheline ecology and malaria epidemiology.* Am J Trop Med Hyg, 2007. **76**(3): p. 450-60.
9. Moyes, C.L., et al., *Defining the geographical range of the Plasmodium knowlesi reservoir.* PLoS Negl Trop Dis, 2014. **8**(3): p. e2780.
10. Moyes, C.L., et al., *Predicting the geographical distributions of the macaque hosts and mosquito vectors of Plasmodium knowlesi malaria in forested and non-forested areas.* Parasit Vectors, 2016. **9**(1): p. 242.
11. Singh, B., et al., *A large focus of naturally acquired Plasmodium knowlesi infections in human beings.* Lancet, 2004. **363**(9414): p. 1017-24.
12. William, T., et al., *Changing epidemiology of malaria in Sabah, Malaysia: increasing incidence of Plasmodium knowlesi.* Malar J, 2014. **13**: p. 390.
13. William, T., et al., *Increasing incidence of Plasmodium knowlesi malaria following control of P. falciparum and P. vivax Malaria in Sabah, Malaysia.* PLoS Negl Trop Dis, 2013. **7**(1): p. e2026.
14. Singh, B. and C. Daneshvar, *Human infections and detection of Plasmodium knowlesi.* Clin Microbiol Rev, 2013. **26**(2): p. 165-84.
15. Lee, K.S., et al., *Plasmodium knowlesi: reservoir hosts and tracking the emergence in humans and macaques.* PLoS Pathog, 2011. **7**(4): p. e1002015.
16. William, T., et al., *Severe Plasmodium knowlesi malaria in a tertiary care hospital, Sabah, Malaysia.* Emerg Infect Dis, 2011. **17**(7): p. 1248-55.
17. Rajahram, G.S., et al., *Deaths due to Plasmodium knowlesi malaria in Sabah, Malaysia: association with reporting as Plasmodium malariae and delayed parenteral artesunate.* Malar J, 2012. **11**: p. 284.
18. Barber, B.E., et al., *A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy.* Clin Infect Dis, 2013. **56**(3): p. 383-97.
19. Daneshvar, C., et al., *Clinical and laboratory features of human Plasmodium knowlesi infection.* Clin Infect Dis, 2009. **49**(6): p. 852-60.
20. Grigg, M.J., et al., *Artesunate-mefloquine versus chloroquine for treatment of uncomplicated Plasmodium knowlesi malaria in Malaysia (ACT KNOW): an open-label, randomised controlled trial.* Lancet Infect Dis, 2016. **16**(2): p. 180-8.
21. Chin, W., et al., *A Naturally Acquited Quotidian-Type Malaria in Man Transferable to Monkeys.* Science, 1965. **149**(3686): p. 865.

22. Knowles, R. and B.M. Das Gupta, *A study of monkey-malaria and its experimental transmission to man.* Indian Medical Gazette, 1932. **67**: p. 301-320.
23. Warren, M., et al., *Cycles of jungle malaria in West Malaysia.* Am J Trop Med Hyg, 1970. **19**(3): p. 383-93.
24. Lee, K.S., et al., *Plasmodium knowlesi from archival blood films: further evidence that human infections are widely distributed and not newly emergent in Malaysian Borneo.* Int J Parasitol, 2009. **39**(10): p. 1125-8.
25. Jongwutiwes, S., et al., *Plasmodium knowlesi Malaria in humans and macaques, Thailand.* Emerg Infect Dis, 2011. **17**(10): p. 1799-806.
26. Shearer, F.M., et al., *Estimating Geographical Variation in the Risk of Zoonotic Plasmodium knowlesi Infection in Countries Eliminating Malaria.* PLoS Negl Trop Dis, 2016. **10**(8): p. e0004915.
27. Luchavez, J., et al., *Human Infections with Plasmodium knowlesi, the Philippines.* Emerg Infect Dis, 2008. **14**(5): p. 811-3.
28. Putaporntip, C., et al., *Differential prevalence of Plasmodium infections and cryptic Plasmodium knowlesi malaria in humans in Thailand.* J Infect Dis, 2009. **199**(8): p. 1143-50.
29. Sermwittayawong, N., et al., *Human Plasmodium knowlesi infection in Ranong province, southwestern border of Thailand.* Malar J, 2012. **11**: p. 36.
30. Jongwutiwes, S., et al., *Naturally acquired Plasmodium knowlesi malaria in human, Thailand.* Emerg Infect Dis, 2004. **10**(12): p. 2211-3.
31. Van den Eede, P., et al., *Human Plasmodium knowlesi infections in young children in central Vietnam.* Malar J, 2009. **8**: p. 249.
32. Khim, N., et al., *Plasmodium knowlesi infection in humans, Cambodia, 2007-2010.* Emerg Infect Dis, 2011. **17**(10): p. 1900-2.
33. Jiang, N., et al., *Co-infections with Plasmodium knowlesi and other malaria parasites, Myanmar.* Emerg Infect Dis, 2010. **16**(9): p. 1476-8.
34. Lee, Y.C., et al., *Epidemiological characteristics of imported and locally-acquired malaria in Singapore.* Ann Acad Med Singapore, 2009. **38**(10): p. 840-9.
35. Zhou, X., et al., *A molecular survey of febrile cases in malaria-endemic areas along China-Myanmar border in Yunnan province, People's Republic of China.* Parasite, 2014. **21**: p. 27.
36. Cox-Singh, J., et al., *Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening.* Clin Infect Dis, 2008. **46**(2): p. 165-71.
37. Barber, B.E., et al., *Epidemiology of Plasmodium knowlesi malaria in north-east Sabah, Malaysia: family clusters and wide age distribution.* Malar J, 2012. **11**: p. 401.
38. Ninan, T., et al., *Plasmodium knowlesi malaria infection in human.* Brunei Int Med J, 2012. **8**(6): p. 358-361.
39. Tyagi, R.K., et al., *Discordance in drug resistance-associated mutation patterns in marker genes of Plasmodium falciparum and Plasmodium knowlesi during coinfections.* J Antimicrob Chemother, 2013. **68**(5): p. 1081-8.
40. Figtree, M., et al., *Plasmodium knowlesi in human, Indonesian Borneo.* Emerg Infect Dis, 2010. **16**(4): p. 672-4.
41. Singh, B. and C. Daneshvar, *Plasmodium knowlesi malaria in Malaysia.* Med J Malaysia, 2010. **65**(3): p. 166-72.
42. Barber, B.E., et al., *Limitations of microscopy to differentiate Plasmodium species in a region co-endemic for Plasmodium falciparum, Plasmodium vivax and Plasmodium knowlesi.* Malar J, 2013. **12**: p. 8.
43. Lee, K.S., J. Cox-Singh, and B. Singh, *Morphological features and differential counts of Plasmodium knowlesi parasites in naturally acquired human infections.* Malar J, 2009. **8**: p. 73.
44. Barber, B.E., et al., *Evaluation of the sensitivity of a pLDH-based and an aldolase-based rapid diagnostic test for diagnosis of uncomplicated and severe malaria caused by PCR-confirmed*

Plasmodium knowlesi, Plasmodium falciparum, and Plasmodium vivax. J Clin Microbiol, 2013. **51**(4): p. 1118-23.
45. Foster, D., et al., *Evaluation of three rapid diagnostic tests for the detection of human infections with Plasmodium knowlesi.* Malar J, 2014. **13**: p. 60.
46. Grigg, M.J., et al., *Combining parasite lactate dehydrogenase-based and histidine-rich protein 2-based rapid tests to improve specificity for diagnosis of malaria Due to Plasmodium knowlesi and other Plasmodium species in Sabah, Malaysia.* J Clin Microbiol, 2014. **52**(6): p. 2053-60.
47. Yusof, R., et al., *High proportion of knowlesi malaria in recent malaria cases in Malaysia.* Malar J, 2014. **13**: p. 168.
48. Joveen-Neoh, W.F., et al., *Incidence of malaria in the interior division of sabah, malaysian borneo, based on nested PCR.* J Parasitol Res, 2011. **2011**: p. 104284.
49. Cox-Singh, J. and B. Singh, *Knowlesi malaria: newly emergent and of public health importance?* Trends Parasitol, 2008. **24**(9): p. 406-10.
50. Gibbs, H.K., et al., *Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s.* Proc Natl Acad Sci U S A, 2010. **107**(38): p. 16732-7.
51. Hansen, M.C., et al., *High-resolution global maps of 21st-century forest cover change.* Science, 2013. **342**(6160): p. 850-3.
52. Bryan, J.E., et al., *Extreme differences in forest degradation in Borneo: comparing practices in Sarawak, Sabah, and Brunei.* PLoS One, 2013. **8**(7): p. e69679.
53. Gaveau, D.L., et al., *Rapid conversions and avoided deforestation: examining four decades of industrial plantation expansion in Borneo.* Sci Rep, 2016. **6**: p. 32017.
54. Vittor, A.Y., et al., *Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, Anopheles darlingi.* Am J Trop Med Hyg, 2009. **81**(1): p. 5-12.
55. Durnez, L., et al., *Outdoor malaria transmission in forested villages of Cambodia.* Malar J, 2013. **12**: p. 329.
56. Brant, H.L., *Changes in abundance, diversity and community composition of mosquitoes based on different land use in Sabah, Malaysia*. 2011, Imperial College London: London.
57. Loaiza, J.R., et al., *Disturbance and mosquito diversity in the lowland tropical rainforest of central Panama.* Sci Rep, 2017. **7**(1): p. 7248.
58. Randolph, S.E. and A.D. Dobson, *Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm.* Parasitology, 2012. **139**(7): p. 847-63.
59. Laporta, G.Z., et al., *Biodiversity can help prevent malaria outbreaks in tropical forests.* PLoS Negl Trop Dis, 2013. **7**(3): p. e2139.
60. Nah, K., Y. Kim, and J.M. Lee, *The dilution effect of the domestic animal population on the transmission of P. vivax malaria.* J Theor Biol, 2010. **266**(2): p. 299-306.
61. Leonardo, L.R., et al., *A study of the environmental determinants of malaria and schistosomiasis in the Philippines using Remote Sensing and Geographic Information Systems.* Parassitologia, 2005. **47**(1): p. 105-14.
62. Keiser, J., et al., *Effect of irrigation and large dams on the burden of malaria on a global and regional scale.* Am J Trop Med Hyg, 2005. **72**(4): p. 392-406.
63. Tangena, J.A., et al., *Risk and Control of Mosquito-Borne Diseases in Southeast Asian Rubber Plantations.* Trends Parasitol, 2016. **32**(5): p. 402-15.
64. Pindolia, D.K., et al., *Quantifying cross-border movements and migrations for guiding the strategic planning of malaria control and elimination.* Malar J, 2014. **13**: p. 169.
65. Barbieri, A.F., i.O. Sawyer, and B.S. Soares-Filho, *Population and Land Use Effects on Malaria Prevalence in the Southern Brazilian Amazon.* Human Ecology, 2005. **33**(6): p. 847-874.
66. Wai, K.T., et al., *Spatial distribution, work patterns, and perception towards malaria interventions among temporary mobile/migrant workers in artemisinin resistance containment zone.* BMC Public Health, 2014. **14**: p. 463.

67. Satitvipawee, P., et al., *Predictors of malaria-association with rubber plantations in Thailand.* BMC Public Health, 2012. **12**: p. 1115.
68. Basurko, C., et al., *Deforestation, agriculture and farm jobs: a good recipe for Plasmodium vivax in French Guiana.* Malar J, 2013. **12**: p. 90.
69. de Castro, M.C., et al., *Malaria risk on the Amazon frontier.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2452-7.
70. Tusting, L.S., et al., *Socioeconomic development as an intervention against malaria: a systematic review and meta-analysis.* Lancet, 2013. **382**(9896): p. 963-72.
71. Tusting, L.S., et al., *Housing Improvements and Malaria Risk in Sub-Saharan Africa: A Multi-Country Analysis of Survey Data.* PLoS Med, 2017. **14**(2): p. e1002234.
72. Tusting, L.S., et al., *The evidence for improving housing to reduce malaria: a systematic review and meta-analysis.* Malar J, 2015. **14**: p. 209.
73. Liu, W., et al., *Origin of the human malaria parasite Plasmodium falciparum in gorillas.* Nature, 2010. **467**(7314): p. 420-5.
74. Ta, T.H., et al., *First case of a naturally acquired human infection with Plasmodium cynomolgi.* Malar J, 2014. **13**: p. 68.
75. Imai, N., et al., *Transmission and control of Plasmodium knowlesi: a mathematical modelling study.* PLoS Negl Trop Dis, 2014. **8**(7): p. e2978.
76. Brasil, P., et al., *Outbreak of human malaria caused by Plasmodium simium in the Atlantic Forest in Rio de Janeiro: a molecular epidemiological investigation.* Lancet Glob Health, 2017. **5**(10): p. e1038-e1046.
77. Guimaraes, L.O., et al., *The genetic diversity of Plasmodium malariae and Plasmodium brasilianum from human, simian and mosquito hosts in Brazil.* Acta Trop, 2012. **124**(1): p. 27-32.
78. Boundenga, L., et al., *Diversity of malaria parasites in great apes in Gabon.* Malar J, 2015. **14**: p. 111.
79. Sundararaman, S.A., et al., *Plasmodium falciparum-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria.* Proc Natl Acad Sci U S A, 2013. **110**(17): p. 7020-5.
80. Vythilingam, I., et al., *Plasmodium knowlesi in humans, macaques and mosquitoes in peninsular Malaysia.* Parasit Vectors, 2008. **1**(1): p. 26.
81. Tan, C.H., et al., *Bionomics of Anopheles latens in Kapit, Sarawak, Malaysian Borneo in relation to the transmission of zoonotic simian malaria parasite Plasmodium knowlesi.* Malar J, 2008. **7**: p. 52.
82. Vythilingam, I., et al., *Natural transmission of Plasmodium knowlesi to humans by Anopheles latens in Sarawak, Malaysia.* Trans R Soc Trop Med Hyg, 2006. **100**(11): p. 1087-8.
83. Wong, M.L., et al., *Dynamics of Anopheles balabacensis on Banggi Island and Kudat, Sabah, Malaysia in relation to Plasmodium knowlesi.* 2015.
84. Hii, J., *Anopheles malaria vector in Malaysia with reference to Sabah*, in *Problems of malaria in endemic countries*, R. Harinasuta, Editor. 1985, Southeast Asian Medical Information Centre: Tokyo. p. 71-81.
85. Collins, W.E., P.G. Contacos, and E.G. Guinn, *Studies on the transmission of simian malarias. II. Transmission of the H strain of Plasmodium knowlesi by Anopheles balabacensis balabacensis.* J Parasitol, 1967. **53**(4): p. 841-4.
86. Brant, H.L., et al., *Vertical stratification of adult mosquitoes (Diptera: Culicidae) within a tropical rainforest in Sabah, Malaysia.* Malar J, 2016. **15**(1): p. 370.
87. Manin, B.O., et al., *Investigating the Contribution of Peri-domestic Transmission to Risk of Zoonotic Malaria Infection in Humans.* PLoS Negl Trop Dis, 2016. **10**(10): p. e0005064.
88. Malaivijitnond, S., et al., *Boundary zone between northern and southern pig-tailed macaques and their morphological differences.* Primates, 2012. **53**(4): p. 377-89.

89. Divis, P.C., et al., *Admixture in Humans of Two Divergent Plasmodium knowlesi Populations Associated with Different Macaque Host Species.* PLoS Pathog, 2015. **11**(5): p. e1004888.
90. Assefa, S., et al., *Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi.* Proc Natl Acad Sci U S A, 2015. **112**(42): p. 13027-32.
91. Chin, W., et al., *Experimental mosquito-transmission of Plasmodium knowlesi to man and monkey.* Am J Trop Med Hyg, 1968. **17**(3): p. 355-8.
92. Fooden, J., *Systematic Review of Southeast Asian Longtail Macaques Macaca fascicularis.* Fieldiana, 1995. **81**.
93. Young, H., et al., *Does habitat disturbance increase infectious disease risk for primates?* Ecol Lett, 2013. **16**(5): p. 656-63.
94. Nunn, C.L., et al., *Interactions between micro- and macroparasites predict microparasite species richness across primates.* Am Nat, 2014. **183**(4): p. 494-505.
95. Nunn, C.L., et al., *Comparative tests of parasite species richness in primates.* Am Nat, 2003. **162**(5): p. 597-614.
96. Kowalewski, M.M., et al., *Black and gold howler monkeys (Alouatta caraya) as sentinels of ecosystem health: patterns of zoonotic protozoa infection relative to degree of human-primate contact.* Am J Primatol, 2011. **73**(1): p. 75-83.
97. Gillespie, T.R. and C.A. Chapman, *Prediction of parasite infection dynamics in primate metapopulations based on attributes of forest fragmentation.* Conserv Biol, 2006. **20**(2): p. 441-8.
98. Mbora, D.N., J. Wieczkowski, and E. Munene, *Links between habitat degradation, and social group size, ranging, fecundity, and parasite prevalence in the Tana River mangabey (Cercocebus galeritus).* Am J Phys Anthropol, 2009. **140**(3): p. 562-71.
99. Hausermann, H., et al., *Contours of risk: spatialising human behaviours to understand disease dynamics in changing landscapes.* Ecohealth, 2012. **9**: p. 251-255.
100. Stoddard, S.T., et al., *The role of human movement in the transmission of vector-borne pathogens.* PLoS Negl Trop Dis, 2009. **3**(7): p. e481.
101. Pindolia, D.K., et al., *Human movement data for malaria control and elimination strategic planning.* Malar J, 2012. **11**(205).
102. Wesolowski, A., et al., *Quantifying the impact of human mobility on malaria.* Science, 2012. **338**(6104): p. 267-70.
103. Grigg, M.J., et al., *Individual-level factors associated with the risk of acquiring human Plasmodium knowlesi malaria in Malaysia: a case control study.* Lancet Planetary Health, 2017. **1**: p. e97-104.
104. Alias, H., et al., *Spatial distribution of malaria in Peninsular Malaysia from 2000 to 2009.* Parasit Vectors, 2014. **7**: p. 186.
105. Singh, R., et al., *High prevalence of asymptomatic malaria in apparently healthy schoolchildren in Aliero, Kebbi state, Nigeria.* J Vector Borne Dis, 2014. **51**(2): p. 128-32.
106. Stresman, G.H., et al., *High levels of asymptomatic and subpatent Plasmodium falciparum parasite carriage at health facilities in an area of heterogeneous malaria transmission intensity in the Kenyan highlands.* Am J Trop Med Hyg, 2014. **91**(6): p. 1101-8.
107. Starzengruber, P., et al., *High prevalence of asymptomatic malaria in south-eastern Bangladesh.* Malar J, 2014. **13**: p. 16.
108. Zhou, G., et al., *Active case surveillance, passive case surveillance and asymptomatic malaria parasite screening illustrate different age distribution, spatial clustering and seasonality in western Kenya.* Malar J, 2015. **14**(1): p. 41.
109. Golassa, L., et al., *Detection of a substantial number of sub-microscopic Plasmodium falciparum infections by polymerase chain reaction: a potential threat to malaria control and diagnosis in Ethiopia.* Malar J, 2013. **12**: p. 352.
110. Okell, L.C., et al., *Factors determining the occurrence of submicroscopic malaria infections and their relevance for control.* Nat Commun, 2012. **3**: p. 1237.

111. Okell, L.C., et al., *Submicroscopic infection in Plasmodium falciparum-endemic populations: a systematic review and meta-analysis.* J Infect Dis, 2009. **200**(10): p. 1509-17.
112. Marchand, R.P., et al., *Co-infections of Plasmodium knowlesi, P. falciparum, and P. vivax among Humans and Anopheles dirus Mosquitoes, Southern Vietnam.* Emerg Infect Dis, 2011. **17**(7): p. 1232-9.
113. Herdiana, H., et al., *Malaria risk factor assessment using active and passive surveillance data from Aceh Besar, Indonesia, a low endemic, malaria elimination setting with Plasmodium knowlesi, Plasmodium vivax, and Plasmodium falciparum.* Malar J, 2016. **15**: p. 468.
114. Bousema, T., et al., *Serologic markers for detecting malaria in areas of low endemicity, Somalia, 2008.* Emerg Infect Dis, 2010. **16**(3): p. 392-9.
115. Corran, P., et al., *Serology: a robust indicator of malaria transmission intensity?* Trends Parasitol, 2007. **23**(12): p. 575-82.
116. Wilson, S., et al., *Age-adjusted Plasmodium falciparum antibody levels in school-aged children are a stable marker of microgeographical variations in exposure to Plasmodium infection.* BMC Infect Dis, 2007. **7**: p. 67.
117. Drakeley, C.J., et al., *Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure.* Proc Natl Acad Sci U S A, 2005. **102**(14): p. 5108-13.
118. Cook, J., et al., *Using serological measures to monitor changes in malaria transmission in Vanuatu.* Malar J, 2010. **9**: p. 169.
119. Herman, L.S., et al., *Identification and validation of a novel panel of Plasmodium knowlesi biomarkers of serological exposure.* PLoS Negl Trop Dis, 2018. **12**(6): p. e0006457.
120. Stoddard, S.T., et al., *House-to-house human movement drives dengue virus transmission.* Proc Natl Acad Sci U S A, 2013. **110**(3): p. 994-9.
121. Gonzalez, M.C., C.A. Hidalgo, and A.L. Barabasi, *Understanding individual human mobility patterns.* Nature, 2008. **453**(7196): p. 779-82.
122. Song, C., et al., *Limits of predictability in human mobility.* Science, 2010. **327**(5968): p. 1018-21.
123. Bengtsson, L., et al., *Using mobile phone data to predict the spatial spread of cholera.* Sci Rep, 2015. **5**: p. 8923.
124. Schneider, C.M., et al., *Unravelling daily human mobility motifs.* J R Soc Interface, 2013. **10**(84): p. 20130246.
125. Deville, P., et al., *Dynamic population mapping using mobile phone data.* Proc Natl Acad Sci U S A, 2014. **111**(45): p. 15888-93.
126. Chaix, B., et al., *GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference?* Health Place, 2013. **21**: p. 46-51.
127. Tatem, A.J., et al., *The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents.* Malar J, 2009. **8**: p. 287.
128. Seto, E.Y., et al., *The use of a vest equipped with a global positioning system to assess water-contact patterns associated with schistosomiasis.* Geospat Health, 2007. **1**(2): p. 233-41.
129. Stothard, J.R., et al., *Investigating the spatial micro-epidemiology of diseases within a point-prevalence sample: a field applicable method for rapid mapping of households using low-cost GPS-dataloggers.* Trans R Soc Trop Med Hyg, 2011. **105**(9): p. 500-6.
130. Paz-Soldan, V.A., et al., *Strengths and weaknesses of Global Positioning System (GPS) data-loggers and semi-structured interviews for capturing fine-scale human mobility: findings from Iquitos, Peru.* PLoS Negl Trop Dis, 2014. **8**(6): p. e2888.
131. Vazquez-Prokopec, G.M., et al., *Using GPS technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment.* PLoS One, 2013. **8**(4): p. e58802.

132. Kaewwaen, W. and A. Bhumiratana, *Landscape ecology and epidemiology of malaria associated with rubber plantations in Thailand: integrated approaches to malaria ecotoping.* Interdiscip Perspect Infect Dis, 2015. **2015**: p. 909106.
133. Parsons, M.B., et al., *Global positioning system data-loggers: a tool to quantify fine-scale movement of domestic animals to evaluate potential for zoonotic transmission to an endangered wildlife population.* PLoS One, 2014. **9**(11): p. e110984.
134. Hay, S.I., et al., *Global mapping of infectious disease.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1614): p. 20120250.
135. Bhatt, S., et al., *The global distribution and burden of dengue.* Nature, 2013. **496**(7446): p. 504-7.
136. Kraemer, M.U., et al., *The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus.* Elife, 2015. **4**: p. e08347.
137. Messina, J.P., et al., *The global distribution of Crimean-Congo hemorrhagic fever.* Trans R Soc Trop Med Hyg, 2015. **109**(8): p. 503-13.
138. Nsoesie, E.O., et al., *Global distribution and environmental suitability for chikungunya virus, 1952 to 2015.* Euro Surveill, 2016. **21**(20).
139. Magalhaes, R.J., et al., *The applications of model-based geostatistics in helminth epidemiology and control.* Adv Parasitol, 2011. **74**: p. 267-96.
140. Department of Statistics, *Population and Housing Census, Malaysia*, D.o. Statistics, Editor. 2010: Malaysia.
141. Hay, S., *An overview of remote-sensing and geodesy for epidemiology and public health application.* Advances in Parasitology, 2000. **47**: p. 1-35.
142. Hamm, N.A., R.J. Soares Magalhaes, and A.C. Clements, *Earth Observation, Spatial Data Quality, and Neglected Tropical Diseases.* PLoS Negl Trop Dis, 2015. **9**(12): p. e0004164.
143. Mather, P.M. and M. Koch, *Computer processing of remotely-sensed images (4th edition)*. 2010, Chichester: Wiley.
144. Vanden Eng, J.L., et al., *Use of handheld computers with global positioning systems for probability sampling and data entry in household surveys.* Am J Trop Med Hyg, 2007. **77**(2): p. 393-9.
145. Kelly, G.C., et al., *Modern geographical reconnaissance of target populations in malaria elimination zones.* Malar J, 2010. **9**: p. 289.
146. Schiffman, R., *Wildlife conservation. Drones flying high as new tool for field biologists.* Science, 2014. **344**(6183): p. 459.
147. Dong, J., et al., *A 50-m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment.* PLoS One, 2014. **9**(1): p. e85801.
148. Turner, B.L., 2nd, E.F. Lambin, and A. Reenberg, *The emergence of land change science for global environmental change and sustainability.* Proc Natl Acad Sci U S A, 2007. **104**(52): p. 20666-71.
149. Hay, S.I., *An overview of remote sensing and geodesy for epidemiology and public health application.* Adv Parasitol, 2000. **47**: p. 1-35.
150. Pigott, D.M., et al., *Prioritising Infectious Disease Mapping.* PLoS Negl Trop Dis, 2015. **9**(6): p. e0003756.
151. Fornace, K.M., et al., *Mapping infectious disease landscapes: unmanned aerial vehicles and epidemiology.* Trends Parasitol, 2014. **30**(11): p. 514-519.
152. Wong, M.L., et al., *Seasonal and Spatial Dynamics of the Primary Vector of Plasmodium knowlesi within a Major Transmission Focus in Sabah, Malaysia.* PLoS Negl Trop Dis, 2015. **9**(10): p. e0004135.
153. Stark, D.J., et al., *Response of a group of long-tailed macaques (Macaca fascicularis) to habitat clearing in Sabah, Malaysian Borneo: implications for Plasmodium knowlesi risk.* (submitted).

154. Stark, D. and M. Salgado-Lynn, *Interim Report on Primatology Component, MONKEYBAR project* 2014.
155. Shimada, M., et al., *New Global Forest/Non-forest Maps from ALOS PALSAR Data (2007-2010).* Remote Sensing of Environment, 2014. **155**(13-31).
156. Di Gregorio, A. and L.J.M. Jansen, *Land cover classification system (LCCS): Classification concepts and user manul*. 2000, Food and Agriculture Organization of the United Nations: Rome.
157. Giri, S., et al., *Status and distribution of mangrove forests of the world using earth observation satellite data (version 1.3, updated by UNEP-WCMC).* Global Ecology and Biogeography: A Journal of Macroecology, 2011. **20**: p. 154-159.
158. Potapov, P.V., et al., *Mapping the world's intact forest landscapes by remote sensing.* Ecology and Society, 2008. **13**(2): p. 51.
159. Land Processes Distributed Active Archive Center (LP DAAC), *Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER GDEM) Version 2*. 2015, NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observatoin and Science (EROS) Center: Sioux Falls, South Dakota.
160. Land Processes Distributed Active Archive Center (LP DAAC), *Landsat 8 Operational Land Imager*. 2014, NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center: Sioux Falls, South Dakota.
161. ALOS-2/ALOS Science Project Earth Observation Research Center (EORC), *Global PALSAR-2/PALSAR/JERS-1 Mosaic and Forest/Non-Forest map*. 2017, Japan Aerospace Exploration Agency (JAXA).
162. Krischke, M., W. Niemeyer, and S. Scherer, *RapidEye satellite based geo-information system.* Acta Astronautica, 2000. **46**(2-6): p. 307-312.
163. Handwerk, B. *Five Surprising Drone Uses (besides Amazon delivery)*. National Geographic, 2013.
164. Cummings, M., *A drone in every driveway.* Sci Am, 2013. **308**(1): p. 28-9.
165. Dillow, C. *Drones are invading the Arctic!* What's Working 2014 18 May 2014]; Available from: www.cnbc.com/id/101417956.
166. Mulero-Pazmany, M., et al., *Remotely piloted aircraft systems as a rhinoceros anti-poaching tool in Africa.* PLoS One, 2014. **9**(1): p. e83873.
167. Koh, L. and S. Wich, *Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation.* Tropical Conservation Science, 2012. **5**(2): p. 121-132.
168. Getzin, S., K. Wiegand, and I. Schoning, *Assessing biodiversity in forests using very high resolution images and unmanned aerial vehicles.* Methods in Ecology and Evolution, 2012. **3**: p. 397-404.
169. Hodgson, A., N. Kelly, and D. Peel, *Unmanned aerial vehicles (UAVs) for surveying marine fauna: a dugong case study.* PLoS One, 2013. **8**(11): p. e79556.
170. Martin, J., et al., *Estimating distribution of hidden objects with drones: from tennis balls to manatees.* PLoS One, 2012. **7**(6): p. e38882.
171. Pena, J.M., et al., *Weed mapping in early-season maize fields using object-based analysis of unmanned aerial vehicle (UAV) images.* PLoS One, 2013. **8**(10): p. e77151.
172. Torres-Sanchez, J., et al., *Configuration and specifications of an Unmanned Aerial Vehicle (UAV) for early site specific weed management.* PLoS One, 2013. **8**(3): p. e58210.
173. Doherty, P. and P. Rudol, *A UAV search and rescue scenario with human body detection and geolocation*, in *AI 2007: Advances in Artificial Intelligence*, M. Orgun and J. Thornton, Editors. 2007, Springer: Berlin. p. 1-13.
174. Mendelow, B., et al., *Development of e-Juba, a preliminary proof of concept unmanned aerial vehicle designed to facilitate the transportation of microbiological test samples from remote rural clinics to National Health Laboratory Service laboratories.* S Afr Med J, 2007. **97**(11 Pt 3): p. 1215-8.

175. Klaptocz, A. *Mapping the Philippines after Typhoon Haiyan*. Drone Adventures, 2014.
176. Hickey, S., *Humanitarian drones to deliver medical supplies to roadless areas*, in *The Guardian*. 2014.
177. Pollanen, R., et al., *Radiation surveillance using an unmanned aerial vehicle.* Appl Radiat Isot, 2009. **67**(2): p. 340-4.
178. Corrigan, C., et al., *Capturing vertical profiles of aerosols and black carbon over the Indian Ocean using autonomous unmanned aerial vehicles.* Atmospheric Chemistry & Physics, 2008. **8**(3): p. 737-747.
179. Foley, J., et al., *Global consequences of land use.* Science, 2005. **309**(5734): p. 570-574.
180. Kilpatrick, A. and S. Randolph, *Drivers, dynamics and control of emerging vector-borne zoonotic diseases.* Lancet, 2012. **380**(9857): p. 1946-1955.
181. Wolfe, N., C. Dunavan, and J. Diamond, *Origins of major human infectious diseases.* Nature, 2007. **447**: p. 279-283.
182. Patz, J., et al., *Disease emergence from global climate and land use change.* Medical Clinics of North America, 2008. **92**(6): p. 1473-1491.
183. Jones, K., et al., *Global trends in emerging infectious diseases.* Nature, 2008. **451**: p. 990-993.
184. Patz, J., et al., *Unhealthy landscapes: policy recommendations on land use chance and infectious disease emergence.* Environmental Health Perspectives, 2004. **112**(10): p. 1092-1098.
185. Lambin, E., et al., *Pathogenic landscapes: interactions between land, people, disease vectors and their animal hosts.* International Journal of Health Geographics, 2010. **9**(1): p. 54.
186. Guerra, C., R. Snow, and S. Hay, *A global assessment of closed forests, deforestation and malaria risk.* Annals of Tropical Medicine and Parasitology, 2006. **100**(3): p. 189-204.
187. Yasuoka, J. and R. Levins, *Impact of deforestation and agricultural development on anopheline ecology and malaria epidemiology.* American Journal of Tropical Medicine and Hygiene, 2007. **76**(3): p. 450-460.
188. Barbieri, A., D. Sawyer, and B. Soares-Filho, *Population and land use effects on malaria prevalence in the Southern Brazilian Amazon.* Human Ecology, 2005. **33**(6): p. 847-874.
189. Hay, S., et al., *Global mapping of infectious disease.* Philosophical Transactions of the Royal Society B, 2012. **4**(368): p. 1814.
190. Machault, V., et al., *The use of remotely sensed environmental data in the study of malaria.* Geospatial Health, 2011. **5**(2): p. 151-168.
191. Dong, J., et al., *A 50-m forest cover map in Southeast Asia from ALOS/PALSAR and its application on forest fragmentation assessment.* PLoS One, 2014. **9**(1): p. e85801.
192. Chuang, T. and M. Wimberly, *Remote sensing of climatic anomalies and West Nile Virus incidence in the Northern Great Plains of the United States.* PLoS One, 2012. **7**(e46882).
193. Lourenco, P., et al., *Anopheles atroparvus density modeling using MODIS NDVI in a former malarious area in Portugal.* J Vector Ecology, 2011. **36**(279-291).
194. Weiss, D., et al., *Air temperature suitability for Plasmodium falciparum malaria transmission in Africa 2000- 2012: a high resolution spatiotemporal predication.* Malaria Journal, 2014. **13**(1): p. 171.
195. Beck, L., et al., *Assessment of a remote sensing-based model for predicting malaria transmission risk in villages of Chiapas, Mexico.* American Journal of Tropical Medicine and Hygiene, 1997. **56**(1): p. 99-106.
196. Dambach, P., et al., *Utilization of combined remote sensing techniques to develop environmental variables influencing malaria vector densities in West Africa.* International Journal of Health Geographics, 2012. **11**(8).
197. Glass, G., et al., *Satellite imagery characterises local animal reservoir populations of Sin Nombre virus in the southwestern United States.* Proc Natl Acad Sci, 2002. **99**(26): p. 16817-22.

198. Emch, M., et al., *Local population and regional environmental drivers of cholera in Bangladesh.* Environmental Health, 2010. **9**(2): p. doi: 10.1186.
199. Rogers, D., J. Suk, and J. Semenza, *Using global maps to predict the risk of dengue in Europe.* Acta Tropica, 2014. **129**: p. 1-14.
200. Randolph, S., et al., *Seasonal synchrony: the key to tick-borne encephalitis foci identified by satellite data.* Parasitology, 2000. **121**(1): p. 15-23.
201. Caughley, G., *Sampling in aerial survey.* J Wildlife Management, 1977. **41**(4): p. 605-615.
202. Zharikov, Y., et al., *Mapping and characterising subtropical estuarine landscapes using aerial photography and GIS for potential application in wildlife conservation and management.* Biological Conservation, 2005. **125**(1): p. 87-100.
203. Reiter, M. and D. LaPointe, *Larval habitat for avian malaria vector Culex quinquefasciatus (Diptera: Culicidae) in altered mid-elevation mesic-dry forests in Hawai'i.* Journal of Vector Ecology, 2009. **34**(2): p. 208-216.
204. Suarez, J., et al., *Use of airborne LiDAR and aerial photography in the estimation of individual tree heights in forestry.* Computers & Geoscience, 2005. **31**(2): p. 253-262.
205. Asner, G., et al., *A universal airborne LiDAR approach for tropical forest carbon mapping.* Oecologia, 2011. **168**(4): p. 1147-1160.
206. Erhart, A., et al., *Epidemiology of forest malaria in central Vietnam: a large scale cross-sectional study.* Malaria Journal, 2005. **4**(58): p. doi:10.1186.
207. Lee, K., et al., *Plasmodium knowlesi: reservoir hosts and tracking the emergence in humans and macaques.* PloS Pathogens, 2011. **4**(4): p. e1002015.
208. Singh, B. and C. Daneshvar, *Human infections and detection of knowlesi.* Clinical Microbiology Reviews, 2013. **26**(2): p. 165-184.
209. ASTER, *ASTER Global Digital Elevation Model (ASTER GDEM)*. 2014.
210. Grigg, M., et al., *Which factors are associated with the risk of acquiring Plasmodium knowlesi malaria in Sabah, Malaysia: a case control study protocol.* BMJ Open, In press.
211. Makynen, J., et al., *Multi- and hyperspectral UAV imaging system for forest and agriculture applications.* SPIE Proceedings, 2012. **Next-Generation Spectroscopic Technologies V**(doi:10.1117).
212. Duan, S.B., et al., *Land surface reflectance retrieval from hyperspectral data collected by an unmanned aerial vehicle over the Baotou test site.* PLoS One, 2013. **8**(6): p. e66972.
213. Lu, D. and Q. Weng, *A survey of image classification methods and techniques for improving classification performance.* International Journal of Remote Sensing, 2007. **28**(5): p. 823-870.
214. Millard, K. and M. Richardson, *On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping.* Remote Sensing, 2015. **7**: p. 8489-8515.
215. Foody, G.M. and A. Mathur, *Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification.* Remote Sensing of Environment, 2004. **93**: p. 107-117.
216. Foody, G.M., *Status of land cover classification accuracy assessment.* Remote Sensing of Environment, 1999. **80**: p. 185-201.
217. Foody, G.M., M.B. McCulloch, and W.B. Yates, *The effect of training set size and composition on artificial neural network classification.* International Journal of Remote Sensing, 1995. **16**(1707-1723).
218. Puissant, A., S. Rouguer, and A. Strumpf, *Object-oriented mapping of urban trees using remote sensing classifiers.* Int J Appl Earth Obs Geoinf, 2014. **26**: p. 235-245.
219. Li, C., et al., *Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery.* Remote Sensing, 2014. **6**: p. 964-983.
220. Bangko, G., *A review of assessing the accuracy of classifications of remotely sensed data and of methods including remote sensing data in forest inventory*. 1998, International Insitute for Applied Systems Analysis: Laxenburg, Austria.

221. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**: p. 5-32.
222. Belgiu, M. and L. Dragut, *Random forest in remote sensing: a review of applications and future directions.* ISPRS Journal of Photogrammetry and Remote Sensing, 2016. **114**: p. 24-31.
223. Colditz, R., *An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms.* Remote Sensing, 2015. **7**: p. 9655-9681.
224. Ahmed, O.S., et al., *Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR and the Random Forest algorithm.* ISPRS Journal of Photogrammetry and Remote Sensing, 2015. **101**: p. 89-101.
225. Gislason, P.O., J.A. Benediktsson, and J.R. Sveinsson, *Random forests for land cover classification.* Pattern Recognition Letters, 2006. **27**(4): p. 294-300.
226. Chan, J.C., et al., *An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery.* International Journal of Applied Earth Observation and Geoinformation, 2012. **18**: p. 13-22.
227. Waske, B. and M. Braun, *Classifier ensembles for land cover mapping using multitemporal SAR imagery.* ISPRS Journal of Photogrammetry and Remote Sensing, 2009. **64**: p. 450-457.
228. Corcoran, J.M., J.F. Knight, and A.L. Gallant, *Influence of multi-source and multi-temporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in Northern Minnesota.* Remote Sensing, 2013. **5**(7): p. 3212-3238.
229. Diaz-Uriarte, R., *GeneSrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest.* BMC Bioinformatics, 2007. **8**: p. 328.
230. Deng, H. and G. Runger, *Feature selection via regularized trees.* arXiv, 2012.
231. Rozenstein, O. and A. Karnieli, *Comparison of methods for land-use classification incorporating remote sensing and GIS inputs.* Applied Geography, 2011. **31**(2): p. 533-544.
232. Cohen, J.M., et al., *Spatial scale modulates the strength of ecological processes driving disease distributions.* Proc Natl Acad Sci U S A, 2016. **113**(24): p. E3359-64.
233. Stefani, A., et al., *Studying relationships between environment and malaria incidence in Camopi (French Guiana) through the objective selection of buffer-based landscape characterisations.* International Journal of Health Geographics, 2011. **10**(65).
234. Brock, P.M., et al., *Data mining across spatial scales links zoonotic malaria to deforestation.* (submitted).
235. Brock, P.M., et al., *Plasmodium knowlesi transmission: integrating quantitative approaches from epidemiology and ecology to understand malaria as a zoonosis.* Parasitology, 2016. **143**(4): p. 389-400.
236. Chapman, C.A., et al., *Life on the edge: gastrointestinal parasites from the forest edge and interior primate groups.* Am J Primatol, 2006. **68**(4): p. 397-409.
237. Ogrzewalska, M., et al., *Effect of forest fragmentation on tick infestations of birds and tick infection rates by rickettsia in the Atlantic forest of Brazil.* Ecohealth, 2011. **8**(3): p. 320-31.
238. Ewers, R.M. and C. Banks-Leite, *Fragmentation impairs the microclimate buffering effect of tropical forests.* PLoS One, 2013. **8**(3): p. e58093.
239. Overgaard, H.J., et al., *Effect of landscape structure on anopheline mosquito density and diversity in northern Thailand: implications for malaria transmission and control.* Landscape Ecology, 2003. **18**(6): p. 605-619.
240. Mcgarigal, K. and S. Cushman, *Quantifying Terrestrial Habitat Loss and Fragmentation : A Protocol.* 2005.
241. Mcgarigal, K., S. Cushman, and E. Ene. *FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps*. 2012; Available from: http://www.umass.edu/landeco/research/fragstats/fragstats.html.
242. Tropek, R., et al., *Comment on "High-resolution global maps of 21st-century forest cover change".* Science, 2014. **344**(6187): p. 981.

243. Hii, J. and L.M. Rueda, *Malaria vectors in the Greater Mekong Subregion: overview of malaria vectors and remaining challenges.* Southeast Asian J Trop Med Public Health, 2013. **44 Suppl 1**: p. 73-165; discussion 306-7.
244. Hutchings, R.S., M.A. Sallum, and R.W. Hutchings, *Mosquito (Diptera: Culicidae) diversity of a forest-fragment mosaic in the Amazon rain forest.* J Med Entomol, 2011. **48**(2): p. 173-87.
245. Vanwambeke, S.O., et al., *Impact of Land-use Change on Dengue and Malaria in Northern Thailand.* EcoHealth, 2007. **4**(1): p. 37-51.
246. UNDATA. *Malaysia*. 2015; Available from: https://data.un.org/CountryProfile.aspx?crName=MALAYSIA.
247. Grigg, M.J., et al., *Factors that are associated with the risk of acquiring Plasmodium knowlesi malaria in Sabah, Malaysia: a case-control study protocol.* BMJ Open, 2014. **4**(8): p. e006004.
248. Speybroeck, N., et al., *Misclassification errors in prevalence estimation: Bayesian handling with care.* Int J Public Health, 2013. **58**(5): p. 791-5.
249. NASA, *Moderate Resolution Imaging Spectroradiometer (MODIS)*. 2015.
250. Liu, J. and X.P. Chen, *Relationship of remote sensing normalized differential vegetation index to Anopheles density and malaria incidence rate.* Biomed Environ Sci, 2006. **19**(2): p. 130-2.
251. Wayant, N.M., et al., *Correlation between normalized difference vegetation index and malaria in a subtropical rain forest undergoing rapid anthropogenic alteration.* Geospat Health, 2010. **4**(2): p. 179-90.
252. Jung, M. *LecoS- A QGIS plugin to conduct landscape ecology statistics*. 2012; Available from: http://plugins.qgis.org/plugins/LecoS.
253. Hilbe, J.M., *Negative Binomial Regression*. 2007, Cambridge: Cambridge University Press.
254. Sallum, M.A.M., E.L. Peyton, and R.C. Wilkerson, *Six new species of the Anopheles leucosphyrus group, reinterpretation of An. elegans and vector implications.* Medical and Veterinary Entomology, 2005. **19**(158- 199).
255. Fornace, K.M., et al., *Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia.* Emerg Infect Dis, 2016. **22**(2): p. 201-9.
256. Fornace, K.M., et al., *Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines.* PLoS Negl Trop Dis, 2018. **12**(6): p. e0006432.
257. DAAC, N.L., *MODIS/ Terra Vegetation Indices 16-Day L3 Global 250m Grid SIN V006*. USGS Earth Resources Observation and Science (EROS) Center: Sioux Falls, South Dakota.
258. Tropical Rainfall Measurement Mission Project (TRMM), *Daily TRMM and other satellites precipitation product (3B42 V6 derived)*, G.S.F.C.D.A.A.C.G. DAAC), Editor. 2015.
259. Lloyd, C.T., A. Sorichetta, and A.J. Tatem, *High resolution global gridded data for use in population studies.* Sci Data, 2017. **4**: p. 170001.
260. Benhamou, S., *Dynamic approach to space and habitat use based on biased random bridges.* PLoS One, 2011. **6**(1): p. e14592.
261. Hooten, M.B., et al., *Reconciling resource utilization and resource selection functions.* J Anim Ecol, 2013. **82**(6): p. 1146-54.
262. Blangiardo, M. and M. Cameletti, *Spatial and Spatio-temporal Bayesian Models with R-INLA*. 2015, Chichester, United Kingdom: John Wiley & Sons, Ltd.
263. Sadykova, D., et al., *Bayesian joint models with INLA exploring marine mobile predator-prey and competitor species habitat overlap.* Ecology and Evolution, 2017. **7**: p. 5212-5226.
264. Barbet-Massin, M., et al., *Selecting pseudo-absences for species distribution models: how, where and how many?* Methods in Ecology and Evolution, 2012. **3**: p. 327-338.
265. Rue, H., S. Martino, and N. Chopin, *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.* Journal of the Royal Statistical Society, 2009. **71**(Part 2): p. 319-392.

266. Hawkes, F., et al., *Evaluation of electric nets as means to sample mosquito vectors host-seeking on humans and primates.* Parasit Vectors, 2017. **10**(1): p. 338.
267. Lindgren, F., H. Rue, and J. Lindstrom, *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach.* Statistical Methodology, 2011. **73**(4): p. 423-498.
268. Lindgren, F. and H. Rue, *Bayesian Spatial Modelling with R-INLA.* Journal of Statistical Software, 2015. **63**(19).
269. Adrion, C. and U. Mansmann, *Bayesian model selection techniques as decision support for shaping a statistical analysis plan of a clinical trial: an example from a vertigo phase III study with longitudinal count data as primary endpoint.* BMC Med Res Methodol, 2012. **12**: p. 137.
270. Simpson, D.P., et al., *Penalising model component complexity: a principled, practical approach to constructing priors.* Statistical Science, 2017.
271. Fornace, K.M., et al., *Asymptomatic and Submicroscopic Carriage of Plasmodium knowlesi Malaria in Household and Community Members of Clinical Cases in Sabah, Malaysia.* J Infect Dis, 2016. **213**(5): p. 784-7.
272. Weiss, D.J., et al., *Re-examining environmental correlates of Plasmodium falciparum malaria endemicity: a data-intensive variable selection approach.* Malar J, 2015. **14**: p. 68.
273. Reiner, R.C., Jr., et al., *A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010.* J R Soc Interface, 2013. **10**(81): p. 20120921.
274. Brady, O.J., et al., *Vectorial capacity and vector control: reconsidering sensitivity to parameters for malaria elimination.* Trans R Soc Trop Med Hyg, 2016. **110**(2): p. 107-17.
275. Yakob, L., M.B. Bonsall, and G. Yan, *Modelling knowlesi malaria transmission in humans: vector preference and host competence.* Malar J, 2010. **9**: p. 329.
276. Singh, B., et al., *A genus- and species-specific nested polymerase chain reaction malaria detection assay for epidemiologic studies.* Am J Trop Med Hyg, 1999. **60**(4): p. 687-92.
277. Imwong, M., et al., *Spurious amplification of a Plasmodium vivax small-subunit RNA gene by use of primers currently used to detect P. knowlesi.* J Clin Microbiol, 2009. **47**(12): p. 4173-5.
278. Reller, M.E., et al., *Multiplex 5' nuclease quantitative real-time PCR for clinical diagnosis of malaria and species-level identification and epidemiologic evaluation of malaria-causing parasites, including Plasmodium knowlesi.* J Clin Microbiol, 2013. **51**(9): p. 2931-8.
279. Kawai, S., et al., *Detection of Plasmodium knowlesi DNA in the urine and faeces of a Japanese macaque (Macaca fuscata) over the course of an experimentally induced infection.* Malar J, 2014. **13**: p. 373.
280. Divis, P.C., et al., *A TaqMan real-time PCR assay for the detection and quantitation of Plasmodium knowlesi.* Malar J, 2010. **9**: p. 344.
281. Putaporntip, C., P. Buppan, and S. Jongwutiwes, *Improved performance with saliva and urine as alternative DNA sources for malaria diagnosis by mitochondrial DNA-based PCR assays.* Clin Microbiol Infect, 2011. **17**(10): p. 1484-91.
282. Tanizaki, R., et al., *First case of Plasmodium knowlesi infection in a Japanese traveller returning from Malaysia.* Malar J, 2013. **12**: p. 128.
283. Schneider, P., et al., *Submicroscopic Plasmodium falciparum gametocyte densities frequently result in mosquito infection.* Am J Trop Med Hyg, 2007. **76**(3): p. 470-4.
284. van Smeden, M., et al., *Latent class models in diagnostic studies when there is no reference standard--a systematic review.* Am J Epidemiol, 2014. **179**(4): p. 423-31.
285. Lau, Y.L., et al., *Comparison of three molecular methods for the detection and speciation of five human Plasmodium species.* Am J Trop Med Hyg, 2015. **92**(1): p. 28-33.
286. Fornace, K.M., et al., *Asymptomatic and Submicroscopic Carriage of Plasmodium knowlesi Malaria in Household and Community Members of Clinical Cases in Sabah, Malaysia.* J Infect Dis, 2015.

287. Helb, D.A., et al., *Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities.* Proc Natl Acad Sci U S A, 2015. **112**(32): p. E4438-47.
288. Lubis, I.N., et al., *Contribution of Plasmodium knowlesi to multi-species human malaria infections in North Sumatera, Indonesia.* J Infect Dis, 2017.
289. Stewart, L., et al., *Rapid assessment of malaria transmission using age-specific sero-conversion rates.* PLoS One, 2009. **4**(6): p. e6083.
290. Corran, P.H., et al., *Dried blood spots as a source of anti-malarial antibodies for epidemiological studies.* Malar J, 2008. **7**: p. 195.
291. Collins, C.R., et al., *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): p. 7431-41.
292. Collins, C.R., et al., *An inhibitory antibody blocks interactions between components of the malarial invasion machinery.* PLoS Pathog, 2009. **5**(1): p. e1000273.
293. Burghaus, P.A. and A.A. Holder, *Expression of the 19-kilodalton carboxy-terminal fragment of the Plasmodium falciparum merozoite surface protein-1 in Escherichia coli as a correctly folded protein.* Mol Biochem Parasitol, 1994. **64**(1): p. 165-9.
294. Arisue, N., et al., *Clues to evolution of the SERA multigene family in 18 Plasmodium species.* PLoS One, 2011. **6**(3): p. e17775.
295. Yagi, M., et al., *Protective epitopes of the Plasmodium falciparum SERA5 malaria vaccine reside in intrinsically unstructured N-terminal repetitive sequences.* PLoS One, 2014. **9**(6): p. e98460.
296. Tougan, T., et al., *Immunogenicity and protection from malaria infection in BK-SE36 vaccinated volunteers in Uganda is not influenced by HLA-DRB1 alleles.* Parasitol Int, 2016. **65**(5 Pt A): p. 455-8.
297. Mantel, N., *The detection of disease clustering and a generalized regression approach.* Cancer Res, 1967. **27**(2): p. 209-20.
298. Goslee, S.C. and D.L. Urban, *The ecodist package for dissimilarity-based analysis of ecological data.* J Stat Software, 2007. **22**: p. 1-19.
299. DOH, P., WHO, and UCSF, *Eliminating malaria: case-study 6, Progress towards subnational elimination in the Philippines*. 2014, World Health Organisation: Geneva.
300. Hii, J., et al., *Transmission dynamics and estimates of malarial vectorial capacity for An. balabacensis and An. flavirostris on Banggi island, Sabah, Malaysia.* Ann Trop Med Parasitol, 1988. **82**: p. 91-101.
301. Metcalf, C.J., et al., *Use of serological surveys to generate key insights into the changing global landscape of infectious disease.* Lancet, 2016. **388**(10045): p. 728-30.
302. Department of Statistics, M., *Population and housing census*. 2015: Putrajaya (Malaysia).
303. Field, R.D., et al., *Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to El Nino-induced drought.* Proc Natl Acad Sci U S A, 2016. **113**(33): p. 9204-9.
304. van der Laan, M.J., E.C. Polley, and A.E. Hubbard, *Super learner.* Statistical Applications in Genetics and Molecular Biology, 2007. **6**(25).
305. Shete, P.B., et al., *Evaluation of antibody responses to panels of M. tuberculosis antigens as a screening tool for active tuberculosis in Uganda.* PLoS One, 2017. **12**(8): p. e0180122.
306. Polley, E.C. and M.J. van der Laan, *Super Learner in Prediction*, in *U. C. Berkeley Division of Biostatistics Working Paper Series*. 2010, University of California, Berkeley: Berkeley.
307. Elith, J., J.R. Leathwick, and T. Hastie, *A working guide to boosted regression trees.* J Anim Ecol, 2008. **77**(4): p. 802-13.
308. Suykens, J.A.K. and J. Vandewalle, *Least squares support vector machine classifiers.* Neural Processing Letters, 1999. **9**: p. 293 - 300.

309. Venables, W.N. and B.D. Ripley, *Modern applied statistics with S. Fourth Edition*. 2002, New York: Springer.
310. Friedman, J., T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent.* Journal of Statistical Software, 2010. **33**(1).
311. LeDell, E., M.J. van der Laan, and M. Petersen, *AUC-Maximizing Ensembles through Metalearning.* Int J Biostat, 2016. **12**(1): p. 203-18.
312. van Buuren, S. and K. Groothuis-Oudshoorn, *mice: Multivariate Imputation by Chained Equations in R.* Journal of Statistical Software, 2011. **45**(3): p. 1-67.
313. Vyas, S. and L. Kumaranayake, *Constructing socio-economic status indices: how to use principal components analysis.* Health Policy Plan, 2006. **21**(6): p. 459-68.
314. Delamater, P.L., et al., *Measuring geographic access to health care: raster and network-based methods.* Int J Health Geogr, 2012. **11**(1): p. 15.
315. Degenhardt, F., S. Seifert, and S. Szymczak, *Evaluation of variable selection methods for random forests and omics data sets.* Brief Bioinform, 2017.
316. Kursa, M. and W. Rudnicki, *Feature selection with the Boruta package.* J Stat Software, 2010. **36**: p. 1-13.
317. R Core Team, *R: A Language and Environment for Statistical Computing*. 2017, R Foundation for Statistical Computing: Vienna, Austria.
318. Siraj, A.S., et al., *Altitudinal changes in malaria incidence in highlands of Ethiopia and Colombia.* Science, 2014. **343**(6175): p. 1154-8.
319. Hasyim, H., et al., *Spatial modelling of malaria cases associated with environmental factors in South Sumatra, Indonesia.* Malar J, 2018. **17**(1): p. 87.
320. Premasiri, D.A., et al., *Malarial vectors in an irrigated rice cultivation area in southern Sri Lanka.* Trans R Soc Trop Med Hyg, 2005. **99**(2): p. 106-14.
321. Sanchez-Ribas, J., G. Parra-Henao, and A.E. Guimaraes, *Impact of dams and irrigation schemes in Anopheline (Diptera: Culicidae) bionomics and malaria epidemiology.* Rev Inst Med Trop Sao Paulo, 2012. **54**(4): p. 179-91.
322. Baeza, A., et al., *Long-lasting transition toward sustainable elimination of desert malaria under irrigation development.* Proc Natl Acad Sci U S A, 2013. **110**(37): p. 15157-62.
323. Ministry of Health Malaysia, World Health Organisation, and University of California San Francisco, *Eliminating malaria: Case study 8 Progress towards elimination in Malaysia*. 2015, World Health Organisation: Geneva.
324. Ministry of Health Malaysia, *Management guidelines of malaria in Malaysia*. 2014, Vector Borne Disease Sector, Disease Control Division, Ministry of Health, Malaysia: Putra Jaya.
325. World Health Organisation Regional Office for Western Pacific, *Expert consultation on Plasmodium knowlesi malaria to guide malaria elimination strategies*. 2017, World Health Organization: Manila, Philippines.
326. Siner, A., et al., *Absence of Plasmodium inui and Plasmodium cynomolgi, but detection of Plasmodium knowlesi and Plasmodium vivax infections in asymptomatic humans in the Betong division of Sarawak, Malaysian Borneo.* Malar J, 2017. **16**(1): p. 417.
327. Wu, W., J. Stamey, and D. Kahle, *A Bayesian Approach to Account for Misclassification and Overdispersion in Count Data.* Int J Environ Res Public Health, 2015. **12**(9): p. 10648-61.
328. Speybroeck, N., et al., *True versus apparent malaria infection prevalence: the contribution of a Bayesian approach.* PLoS One, 2011. **6**(2): p. e16705.
329. Siregar, J.E., et al., *Non-invasive surveillance for Plasmodium in reservoir macaque species.* Malar J, 2015. **14**: p. 404.
330. Chua, T.H., et al., *Phylogenetic analysis of simian Plasmodium spp. infecting Anopheles balabacensis Baisas in Sabah, Malaysia.* PLoS Negl Trop Dis, 2017. **11**(10): p. e0005991.
331. Wardrop, N.A., et al., *Interpreting predictive maps of disease: highlighting the pitfalls of distribution models in epidemiology.* Geospat Health, 2014. **9**(1): p. 237-46.

332. Cazelles, B. and S. Hales, *Infectious diseases, climate influences, and nonstationarity.* PLoS Med, 2006. **3**(8): p. e328.
333. Cazelles, B., K. Cazelles, and M. Chavez, *Wavelet analysis in ecology and epidemiology: impact of statistical tests.* J R Soc Interface, 2014. **11**(91): p. 20130585.
334. Cazelles, B., et al., *Time-dependent spectral analysis of epidemiological time-series with wavelets.* J R Soc Interface, 2007. **4**(15): p. 625-36.
335. Baeza, A., et al., *The rise and fall of malaria under land-use change in frontier regions.* Nat Ecol Evol, 2017. **1**(5): p. 108.
336. World Economic Forum, *The Global Risks Report 2017*. 2017, World Economic Forum: Geneva.
337. Fornace, K., et al., *Effects of land-use changes and agricultural practices on the emergence and re-emergence of human viral diseases*, in *Viral infections and global change*, S.K. Singh, Editor. 2013, John Wiley & Sons, Inc: Hoboken, New Jersey. p. 133-150.
338. Holmes, E.C., A. Rambaut, and K.G. Andersen, *Pandemics: spend on surveillance, not prediction.* Nature, 2018. **558**(7709): p. 180-182.
339. Han, B.A. and J.M. Drake, *Future directions in analytics for infectious disease intelligence: Toward an integrated warning system for emerging pathogens.* EMBO Rep, 2016. **17**(6): p. 785-9.
340. Bousema, T., et al., *Asymptomatic malaria infections: detectability, transmissibility and public health relevance.* Nat Rev Microbiol, 2014. **12**(12): p. 833-40.
341. Allen, T., et al., *Global hotspots and correlates of emerging zoonotic diseases.* Nat Commun, 2017. **8**(1): p. 1124.
342. Pulliam, J.R., et al., *Agricultural intensification, priming for persistence and the emergence of Nipah virus: a lethal bat-borne zoonosis.* J R Soc Interface, 2012. **9**(66): p. 89-101.
343. Lo, M.K. and P.A. Rota, *The emergence of Nipah virus, a highly pathogenic paramyxovirus.* J Clin Virol, 2008. **43**(4): p. 396-400.
344. Lee, M.H., et al., *Macacine Herpesvirus 1 in Long-Tailed Macaques, Malaysia, 2009-2011.* Emerg Infect Dis, 2015. **21**(7): p. 1107-13.
345. Brown, R., et al., *Evaluation of resting traps to examine the behaviour and ecology of mosquito vectors in an area of rapidly changing land use in Sabah, Malaysian Borneo.* Parasit Vectors, 2018. **11**(1): p. 346.
346. Cheong, Y.L., P.J. Leitao, and T. Lakes, *Assessment of land use factors associated with dengue cases in Malaysia using Boosted Regression Trees.* Spat Spatiotemporal Epidemiol, 2014. **10**: p. 75-84.
347. Muhammad Azami, N.A., et al., *Dengue epidemic in Malaysia: Not a predominantly urban disease anymore.* BMC Res Notes, 2011. **4**: p. 216.
348. Haddow, A.D., et al., *Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage.* PLoS Negl Trop Dis, 2012. **6**(2): p. e1477.
349. Dony, J.J.F., et al., *The first outbreak of autochthonous Zika virus in Sabah, Malaysian Borneo.* PLoS Currents Outbreaks, 2018.
350. Tappe, D., et al., *Acute Zika virus infection after travel to Malaysian Borneo, September 2014.* Emerg Infect Dis, 2015. **21**(5): p. 911-3.
351. Marchette, N.J., R. Garcia, and A. Rudnick, *Isolation of Zika virus from Aedes aegypti mosquitoes in Malaysia.* Am J Trop Med Hyg, 1969. **18**(3): p. 411-5.
352. Kilbourn, A.M., et al., *Health evaluation of free-ranging and semi-captive orangutans (Pongo pygmaeus pygmaeus) in Sabah, Malaysia.* J Wildl Dis, 2003. **39**(1): p. 73-83.
353. Wolfe, N.D., et al., *Sylvatic transmission of arboviruses among Bornean orangutans.* Am J Trop Med Hyg, 2001. **64**(5-6): p. 310-6.

12 Annex 1 Additional Publications

12.1 Identification and validation of a novel panel of *Plasmodium knowlesi* serological biomarkers of exposure.

Lou S Herman[1], Kimberly Fornace[1], Jody Phelan[1], Matthew J Grigg[2], Nicholas M Anstey[2], Timothy William[3], Robert W Moon[1], Michael J. Blackman[1,4], Chris J Drakeley[1] and Kevin K A Tetteh[1*].

1. Department Immunology and Infection, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom
2. Menzies School of Health Research and Charles Darwin University, Darwin, Northern Territory, Australia; Infectious Diseases Society Sabah-Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Sabah, Malaysia.
3. Clinical Research Centre, Queen Elizabeth Hospital, Kota Kinabalu, Sabah, Malaysia; Infectious Diseases Society Sabah-Menzies School of Health Research Clinical Research Unit, Kota Kinabalu, Sabah, Malaysia; Jesselton Medical Centre, Kota Kinabalu, Sabah, Malaysia
4. Malaria Biochemistry Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, United Kingdom

∗Corresponding author

E-mail: Kevin.tetteh@lshtm.ac.uk

Abstract

Background

Plasmodium knowlesi is the most common cause of malaria in Malaysian Borneo, with reporting limited to clinical cases presenting to health facilities and scarce data on the true extent of transmission. Serological estimations of transmission have been used with other malaria species to garner information about epidemiological patterns. However, there are a distinct lack of suitable serosurveillance tools for this neglected disease.

Methodology/Principal Findings

Using *in silico* tools, we designed and expressed four novel *P. knowlesi* protein products to address the distinct lack of suitable serosurveillance tools: *Pk*SERA3 antigens 1 and 2, *Pk*SSP2 and *Pk*TSERA2 antigen 1. Antibody prevalence to these antigens was determined by ELISA for three time-points post-treatment from a hospital-based clinical treatment trial in Sabah, East Malaysia (n=97 individuals; 241 total samples for all time points). Responses were observed for the *Pk*SERA3 antigen 2 (67%, 65/97 across all time-points (day 0: 36.9% 34/92; day 7: 63.8% 46/72; day 28: 58.4% 45/77)) between the clinical samples and the *P. knowlesi*-negative controls (n=55, mean plus 3 SD) (day 0 $p<0.0001$; day 7 $p<0.0001$; day 28 $p<0.0001$). Using boosted regression trees, we developed models to classify *P. knowlesi* exposure (cross-validated AUC 88.9%; IQR 86.1 – 91.3%) and identified the most predictive antibody responses.

Conclusions/Significance

The *Pk*SERA3 antigen had the highest relative variable importance in all models. Further validation of these antigens is underway to determine the specificity of these tools in the context of multi-species infections at the population level.

Author Summary

Malaria caused by *Plasmodium knowlesi* is the most common form of the disease in Malaysia. The parasite is transmitted from monkeys to humans via the bite of an infected mosquito, with the resulting *P. knowlesi* infection potentially leading to severe symptoms and in some cases, death. Although adult males working close to areas with infected monkeys are at the greatest risk of infection, the true extent of the geographical boundaries of *P. knowlesi* transmission is as yet unknown. The ability to measure antibodies to infection is a powerful technique that would help to address this deficit. However, currently available recombinant proteins lack the required specificity for this role. Here, we have developed a panel of recombinant proteins for eventual use as serological tools, strongly supported by robust statistical methods. We envisage that these tools will complement existing approaches to identifying the geographical limits of *P. knowlesi* transmission

Introduction

Plasmodium knowlesi is a simian parasite which can cause zoonotic malaria in humans [1]. Recent evidence suggests that human *P. knowlesi* infections are a growing public health threat in South East Asia, particularly in Malaysia [1-5]. *P. knowlesi* has the potential to cause severe disease in endemic regions [6], and is now the most common cause of clinical malaria in Malaysia [7, 8]. *P. knowlesi* is morphologically similar to *P. malariae* [9, 10], historically leading to the misdiagnosis of *P. knowlesi* infections as *P. malariae* [11]. Recent publications have also demonstrated misdiagnosis of *P. knowlesi* as *P. vivax* and *P. falciparum* [3, 12, 13] with potential delay of appropriate treatment associated with case fatalities [6, 14, 15]. Studies have shown that antibodies to *Plasmodium* proteins persist for long periods [16, 17], even in the context of limited exposure or absence of infection. Such antibodies can be utilised in serological assays, accurately estimating the incidence and exposure to *Plasmodium* parasites [18, 19].

One key requirement for serological studies is the identification of *Plasmodium* species-specific biomarkers, particularly in regions where multi-species infections are likely to occur. It is important to distinguish between human serological responses to different *Plasmodium* species to improve our understanding of immunity to these infections, as well as define the geographical spread of infection. Such information can also help to evaluate the impact of how control measures targeting a single species might affect the transmission and immunological profile of other co-endemic species. The few recombinant protein reagents that do exist for *P. knowlesi* have a high level of sequence homology with orthologues from other *Plasmodium* species and, as such, are not applicable to identifying species-specific antibody responses. For example, PK66 (*Pk*AMA1) [20] and *Pk*SPATR (secreted protein with altered thrombospondin repeat) [21] share 86% and 85% amino acid identity respectively with *P. vivax* (https://is.gd/MzISez), potentially making it difficult to distinguish between the two species where infections are co-endemic.

The 2011 WHO consultation panel on the public health importance of *P. knowlesi* recommended the urgent development of *P. knowlesi*-specific diagnostic tools [22]. Key to achieving this goal would be the development of sensitive and accurate tools to help monitor the transmission of infection.

In this study, we describe the development and evaluation of a panel of novel recombinant antigens based on *P. knowlesi*-specific amino acid sequences, using publicly available *in silico* tools. The development of such well-validated species-specific tools represent a potentially important serosurveillance tool to help monitor for historical *P. knowlesi* infections in endemic areas. To illustrate how these data can be used to identify seropositive individuals, we utilise data-adaptive statistical methods (boosted regression trees) to classify exposed individuals. By assessing relative variable importance within these models, we identify the antigen responses contributing most to model predictions and potential serological tools for use in epidemiological studies. These reagents will also serve as an important set of tools to help identify correlates of immunity to *P. knowlesi*.

Methods

Identification and screening of target sequences

Figure 1 outlines the experimental strategy used in the identification of the target sequences of interest. Known markers of seroincidence were selected based on available evidence from *P. falciparum:* AMA1 [23, 24], MSP1 [25-27], TRAP/SSP2 [28, 29] and SERA [30-32] (*Pk*AMA1 (PKNH_0931500), *Pk*MSP1 (PKNH_0728900), *Pk*TRAP/SSP2 (PKNH_1265400), SERA3 (PKNH_0413400) and TSERA2 (PKNH_0413500), respectively). Full-length protein sequences for each gene were initially screened using the BlastP search tool (Plasmodb: https://is.gd/XOs7vd [33] and NCBI: https://is.gd/MzISez). Amino acid sequences were used to generate maximum likelihood phylograms to summarise the relatedness of each gene target between species (S1 Figures a-e). Alignments were also generated for each target using amino acid sequences from other plasmodia matching the query sequence using the MUltiple Sequence Comparison by Log-Expectation (MUSCLE) software (http://www.ebi.ac.uk/Tools/msa/muscle/) [34] (S2 Figures a-e). Each alignment was then interrogated to identify regions of identity primarily with *P. vivax* and *P. falciparum* but also with *P. malariae* and *P. ovale*. Regions or entire sequences showing high levels of identity were excluded from further analysis and the *P. knowlesi*-specific truncated regions were again screened using BlastP to validate sequence specificity (Figure 1 and S1 Table). Each target sequence was analysed using domain prediction software (http://gene3d.biochem.ucl.ac.uk/ and http://smart.embl-heidelberg.de/) to help define putative domain boundaries, where possible. The aim was to limit the level of potential antibody cross-reactivity, which would limit the usefulness of the antigens as serological tools due to the high level of identical amino acids between species. A particular problem in co-endemic settings. Simultaneously, sequences were also screened using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) to help confirm the presence, or absence, of signal peptides and transmembrane regions. Previous experience from our group and others has shown that the presence of signal peptides and/or transmembrane domains can significantly impede protein expression and solubility [35]. Based on this, each confirmed target

construct was designed to exclude both the signal peptide and transmembrane domains, which in addition to the GST tag solubility tag was designed to aid expression of soluble proteins [36, 37].

Figure 1. Flowchart summarising the experimental strategy used in the identification and validation of the *P. knowlesi*-specific candidates.

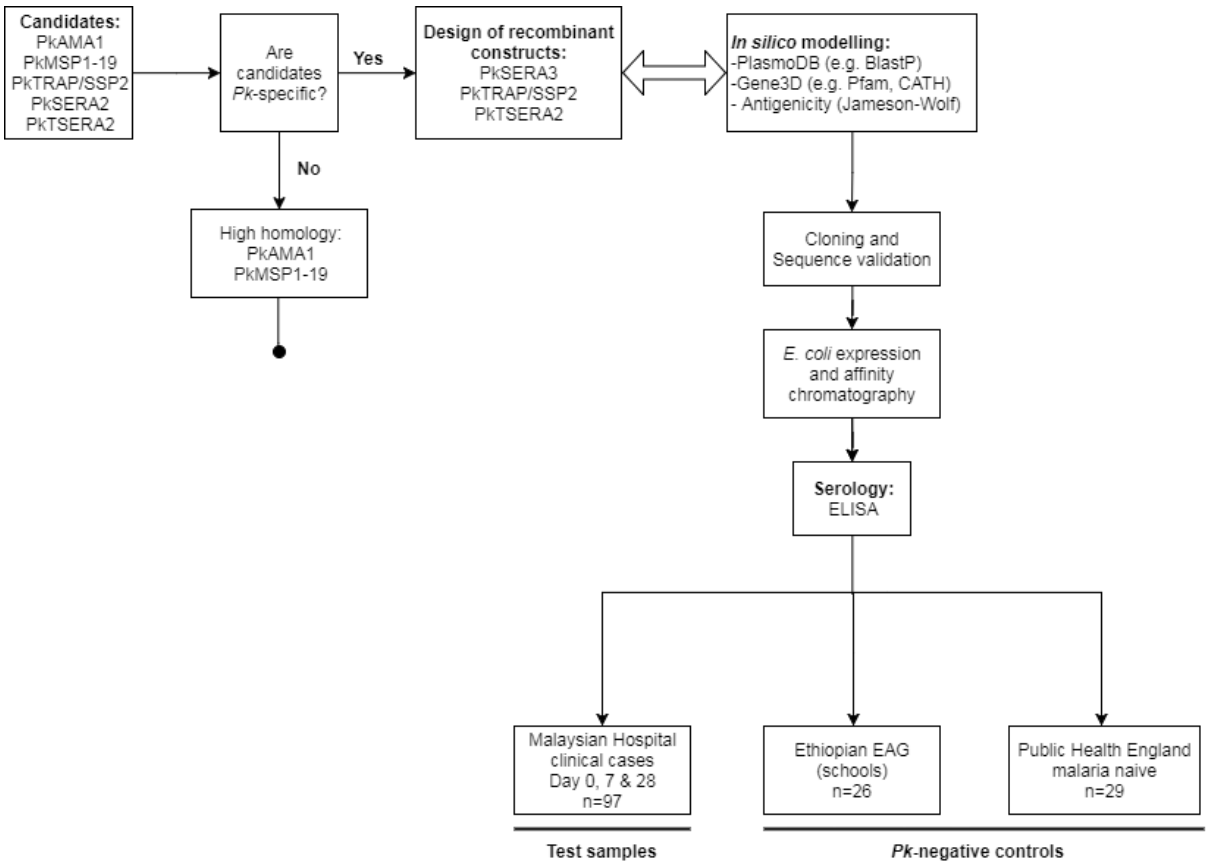


Table 1. Summary of recombinant antigen construct characteristics

Gene ID	Antigen	Description	Chromossome	AA position	Expression (mg/L)		Size (kDa)	
							Predicted	Empirical (aggregate)
PKNH_0413400	SERA3 ag1	cysteine protease (Serine repeat-like antigen)	4	25-140	20.5	w/ GST	37.7	49.6 (118.9)
						w/o GST	11.3	n/a
PKNH_0413400	SERA3 ag2	cysteine protease (Serine repeat-like antigen)	4	826-998	15	w/ GST	44.9	68.7 (162.7)
						w/o GST	18.4	n/a
PKNH_1265400	SSP2	sporozoite surface protein 2, putative, thrombospondin-related anonymous protein (TRAP)	12	381-500	17	w/ GST	39.7	53.1 (132.5)
						w/o GST	13.2	n/a
PKNH_0413500	TSERA2 ag1	Truncated cysteine protease	4	60-251	11.9	w/ GST	46.8	59.7 (117.9)
						w/o GST	20.4	n/a

An additional selection criteria step was to determine the transcriptional status of the candidate genes. Blood stage messenger RNA was collected and analysed using the human red blood cell culture adapted *P. knowlesi* A1-H1 line [38], grown in human blood obtained from the United Kingdom National Blood Transfusion Service. First strand synthesis was carried out using SuperScript® IV Reverse Transcriptase (RT) (Thermo Fisher Scientific) using oligo d(T)20 for priming (RT+) according to the manufacturer's instructions. As a negative control (RT-), a second identical reaction was set up in parallel without the addition of the SuperScript® IV RT. For PCR analysis of cDNA transcripts, RT+ and RT- samples were used as templates for transcript specific PCR primers for the candidate gene sequences alongside genomic DNA controls. In addition, both CTRP (circumsporozoite protein and thrombospondin-related adhesive protein [TRAP]-related protein) and CSP (circumsporozoite protein), both shown to be pre-erythrocytic stage targets, were included in the panel as negative controls. Where possible, primer pairs were designed to flank introns so that amplicons from cDNA and gDNA could be distinguished. Sequences of primer pairs used to amplify each transcript are listed in S2 Table alongside the expected cDNA and gDNA amplicon size. Amplicons were PCR amplified using GoTaq Green Master Mix (Promega) and analysed on a 1.2% agarose gel (S3 Figure).

Cloning and expression of *Plasmodium knowlesi*-specific recombinant antigens in *Escherichia coli*

Four new constructs were designed (Table 1 and Fig 2) based on three genes. Two sequences within *Pk*SERA3 (PKNH_0413400; nucleotide positions 73-419 (Antigen 1) and 2476 – 2994 (Antigen 2) based on the reference *P. knowlesi* H strain), SSP2 (PKNH_1265400; nucleotide positons 1141-1500) and TSERA2 (PKNH_0413500; nucleotide positons 178 – 751 (Antigen 1)) and were PCR amplified from *P. knowlesi* genomic DNA (H strain). Vector compatible primers were designed for each completed target sequence (S3 Table). Cloning of amplified sequences is as described previously [39]. Briefly, purified inserts were cloned into a TA vector (pGEM-T Easy,

Promega) and sequence verified. Correct sequences were restriction digested and sub-cloned into a GST expression vector (pGEX-2T, GE Healthcare) and sequence verified before transforming into BL21 (DE3) *Escherichia coli* expression cells (Novagen). Validated expression clones were expressed automatically using an autoinduction media based on established protocols [40, 41]. Following expression, protein purification was as described [42]. Briefly, GST-tagged proteins from clarified bacterial lysate were purified by affinity chromatography (Glutathione sepharose 4B; GE Healthcare) and fractions from each protein analysed (Bradford assay reagent, BioRad) to identify protein-containing fractions. Pooled protein positive fractions were dialysed against PBS and the protein content quantified (Bicinchoninic acid assay (BCA), Thermo Fisher). The dialysed purified proteins were analysed on a 4-20% gradient gel (NuPAGE® Bis-Tris acetate) under both native and denaturing conditions and visualised using the Coomassie blue staining method (BioRad BioSafe, USA) (Fig 3). The empirical sizes of each protein were calculated using the ImageLab (BioRad) software with the PageRuler prestained marker (Fermentas).

Figure 2. *Plasmodium knowlesi*-specific recombinant antigen constructs. Schematic representations for each protein are shown with key features labelled (a) PkSERA3 shows the location of the putative pro-enzyme and enzyme domains. The predicted subtilisin (SUB) 1 cleavage sites in relation to variable regions 1 and 2 and the cytoplasmic domain [43], (b) PkSSP2 contains a von Willebrandt A domain (vWA), thrombospondin type (TSP) 1 motif, a C-terminal transmembrane (TM) region and a cytoplasmic terminal domain (CTD). Also highlighted are putative T-cell (*) and B-cell (■) [44] and (c) PkTSERA2 shows the lack of central enzyme domain due to truncation of the sequence [43]. Predicted secondary structures are shown above each scheme. Red boxes represent helices, blue arrows sheets and the black line coils. Regions highlighting the position of recombinant proteins are highlighted below each scheme with the N- and C-terminal amino acid positions indicated. The

overall length of each protein is referenced by the amino acid ruler above each secondary structure prediction. For all protein SP refers to the signal peptide.

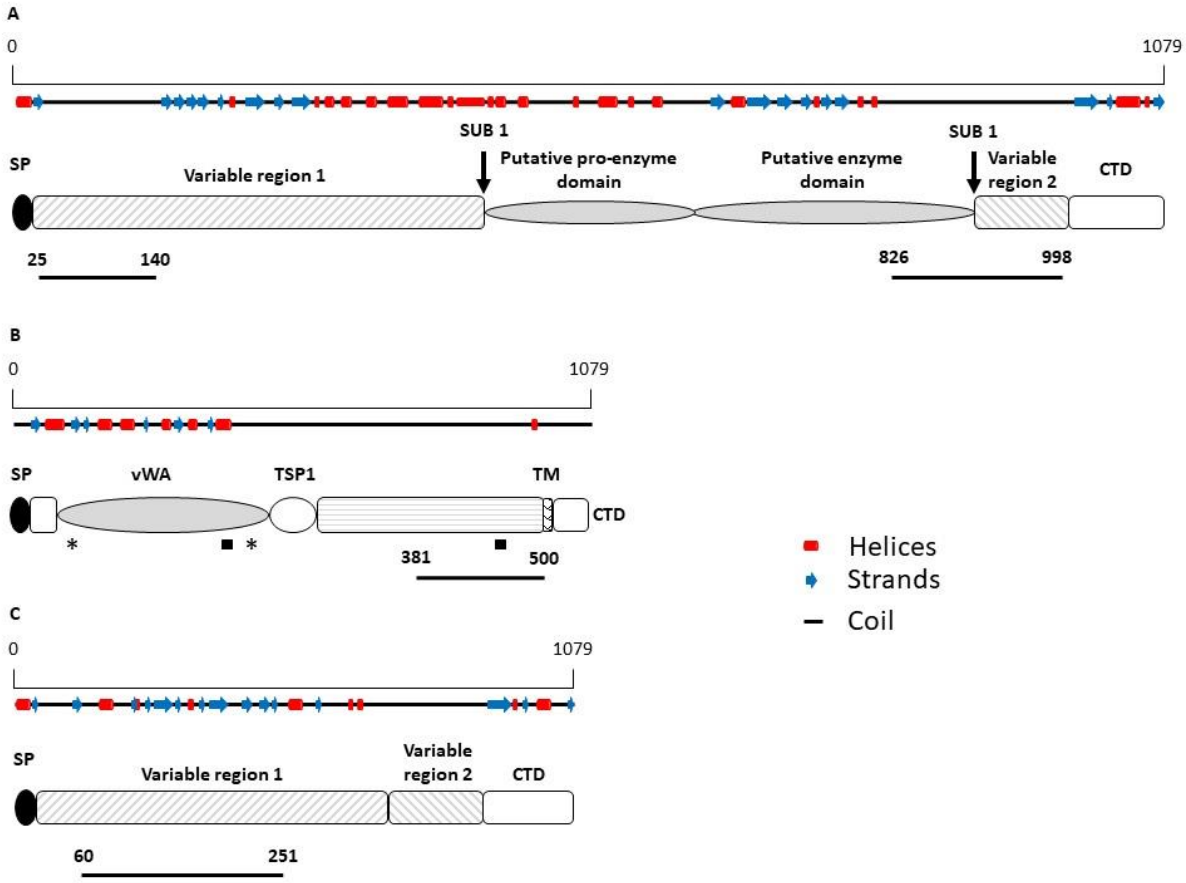
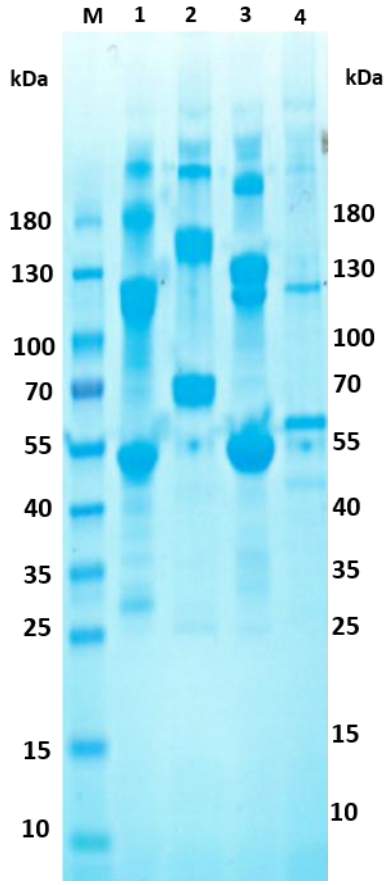


Figure 3. SDS-PAGE of purified recombinants. Lane 1: SERA3 ag1; Lane 2: SERA3 ag2; Lane 3: SSP2 and Lane 4: TSERA2ag1. Band sizes are indicated on one side in kDa (Fermentas PageRuler™ Prestained Protein Ladder). Samples were ran under reducing conditions at approximately 0.4 mg/ml per lane on a 4-20% gradient gel (NuPAGE, BioRad) and stained with Coomassie Blue (BioSafe, BioRad).



SNP and phylogenetic analysis

Full-length sequence data from Plasmodb and construct-specific truncated sequences generated in-house using Sanger sequencing were mapped to an in-house reference sequence strain (Pk-H strain) using the Burrows-Wheeler Aligner (BWA) software (v0.7.5a-r405) [45]. Single nucleotide polymorphisms (SNPs) (S4 Table) were called using the SAMtools (v1.3) (Sequence Alignment/Map) software using default settings [46] and were filtered to increase stringency and

target only high quality variants (missingness<10%, mixed calls<10%). Custom Perl scripts identified overlap between these SNPs and each gene candidate. Variants were annotated using snpEFF (v4.3i) (http://snpeff.sourceforge.net/) [47] *to retrieve the amino acid position and type effect of the variant.* Maximum likelihood phylogenetic trees were constructed from protein sequences using RAxML [48] with a fixed empirical substitution matrix and 200 bootstraps and was visualised using iTOL (http://itol.embl.de) [49].

Enzyme-linked immunosorbent assay (ELISA) and sera collection

The Indirect enzyme-linked immunosorbent assay was performed to screen for antibodies to *P. falciparum*, *P. vivax* and *P. knowlesi* antigens using previously described methods [50]. Briefly, antigens were coated at 50 ng/well and serum samples assayed at 1/1000 dilution for both the *P. knowlesi* recombinants and the *Pv*MSP1-19 (donated as a kind gift from Tony Holder) positive control antigen. Polyclonal rabbit anti-human IgG-HRP (Dako, Denmark) was used at 1/15,000 dilution and plates were developed using TMB (One component HRP microwell substrate, Tebu-bio). Negative and positive controls, including blank (buffer only) wells were used to help standardise across assay runs. Values in excess of 1.5 CV between duplicates were considered fails and re-ran.

Written informed consent was obtained from all study participants [25, 51]. Samples were collected as part of a hospital-based clinical trial in Malaysia, Sabah (www.clinicaltrials.gov: #NCT01708876) (Fig 1) [51]. Serum samples were collected at Day 0 (n=92), 7 (n=72) and 28 (n=77) following hospital admission, with drug treatment also taking place at Day 0. The human research ethics committees of Malaysia (MREC) (#NMRR-12-537-12568), the Menzies School of Health and Research (Australia) (#HREC-2012-1814) and the London School of Hygiene and Tropical Medicine (UK) (#6244) approved the study. Filter paper bloodspots and plasma spots samples were collected and processed as previously described [50]. Twenty-six *P. vivax*-positive Ethiopian samples [25] based on positive responses to *Pv*AMA1 and *Pv*MSP1-19 were used as the Pv-positive, Pk-negative control group. In addition, twenty-nine malaria naïve (Public Health England; LSHTM ethics approval

#11684) serum samples were used as the *P. knowlesi*-negative control group. All samples used in the study were anonymised.

Statistical and sequence analysis

Descriptive analysis of serological data was performed using STATA/IC 14.2 (StataCorp LP, USA) and PRISM (GraphPad PRISM 7). P values were generated using the Wilcoxon signed rank and Wilcoxon-Mann Whitney tests (STATA/IC 14.2). Scatter plots showing reactivity between *P. knowlesi* recombinant antigens and *P. vivax* MSP1-19 were created using STATA (Fig 4) and dot plots showing reactivity to *P. knowlesi* recombinant antigens were created using GraphPad PRISM (Fig 5 and S4 Figure). Final optical density (OD) values were obtained by subtracting blank OD values, reducing background reactivity. Cut off values for each *P. knowlesi*-specific antigen were calculated based on the average ODs of Public Health England negative control sera samples ± (3xSD).

Ensemble boosted regression trees were fit to determine predictive power of antibody responses for classification of *P. knowlesi* exposure. To quantify uncertainty around estimates, 100 datasets were assembled including all seronegative individuals from the malaria unexposed population and an equal number of randomly selected *P. knowlesi* seropositive individuals (from all time points). All models were fit using stratified 10-fold cross validation with model predictive ability assessed by the area under the receiver operating curve (AUC). The learning rate was set at 0.001 and tree complexity set at 4, to allow for interactions within the dataset. Contribution of responses to each antigen to models was assessed using relative variable importance as described by Elith *et. al*.[52]. In this method, the relative importance of individual predictor variables is calculated as the number of times a variable is selected for splitting, weighted by the squared improvement to the model and averaged over all trees and scaled to 100%. Boosted regression tree analysis was completed in R statistical software (v 3.4.2) using the gbm package.

Amino acid sequence alignments were generated using MUltiple Sequence Comparison by Log-Expectation (MUSCLE) (http://www.ebi.ac.uk/Tools/msa/muscle/) [34].

Figure 4. Endemic and *P. knowlesi*-negative sera reactivity to *Plasmodium knowlesi*-specific antigens. Scatter plots showing sera reactivity to: *P. vivax* MSP1-19 with *P. knowlesi* SERA3 ag1 (column 1), SERA3 ag2 (column 2), SSP2 (column 3) and TSERA2 ag1 (column 4) antigens. Sera samples from *P. knowlesi*-negative controls n=55 (row 1; PHE UK malaria naïve (blue), Ethiopian children (red)) and Malaysian hospital case sera samples from days 0 (n=92), 7 (n=72) and 28 (n=77) of diagnosis (rows 2-4, respectively). The red line in each graph represent the cut off values for the respective *P. knowlesi* antigen and was calculated based on Public Health England negative control sera samples (average ODs ± (3xSD)): The vertical cut off line is based on *Pv*MSP1-19=0.501. The horizontal cut off line for the *P. knowlesi* antigens were based on the following values: SERA3 ag1=0.292; SERA3 ag2=0.366; SSP2=0.322 and TSERA2 ag1=0.208.

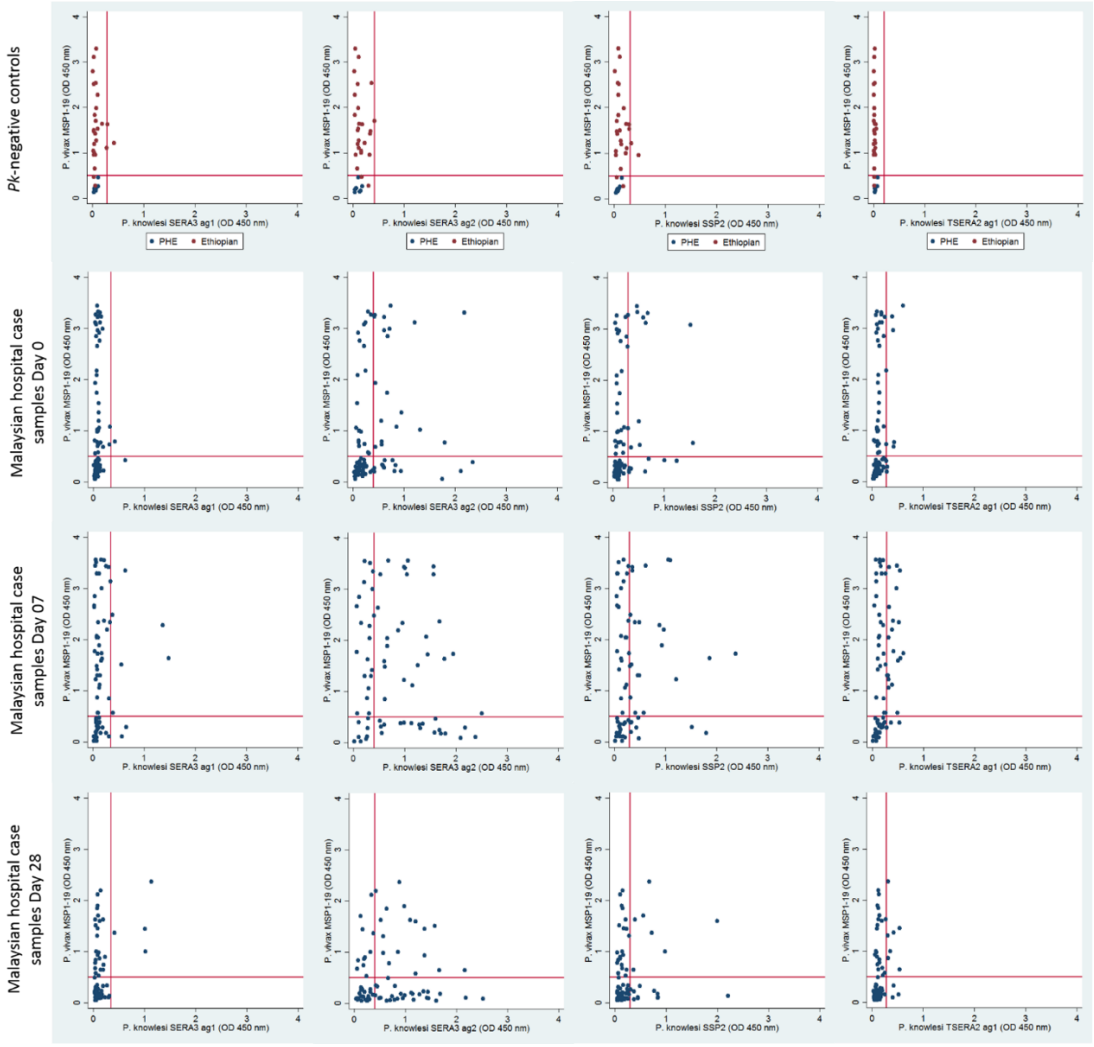
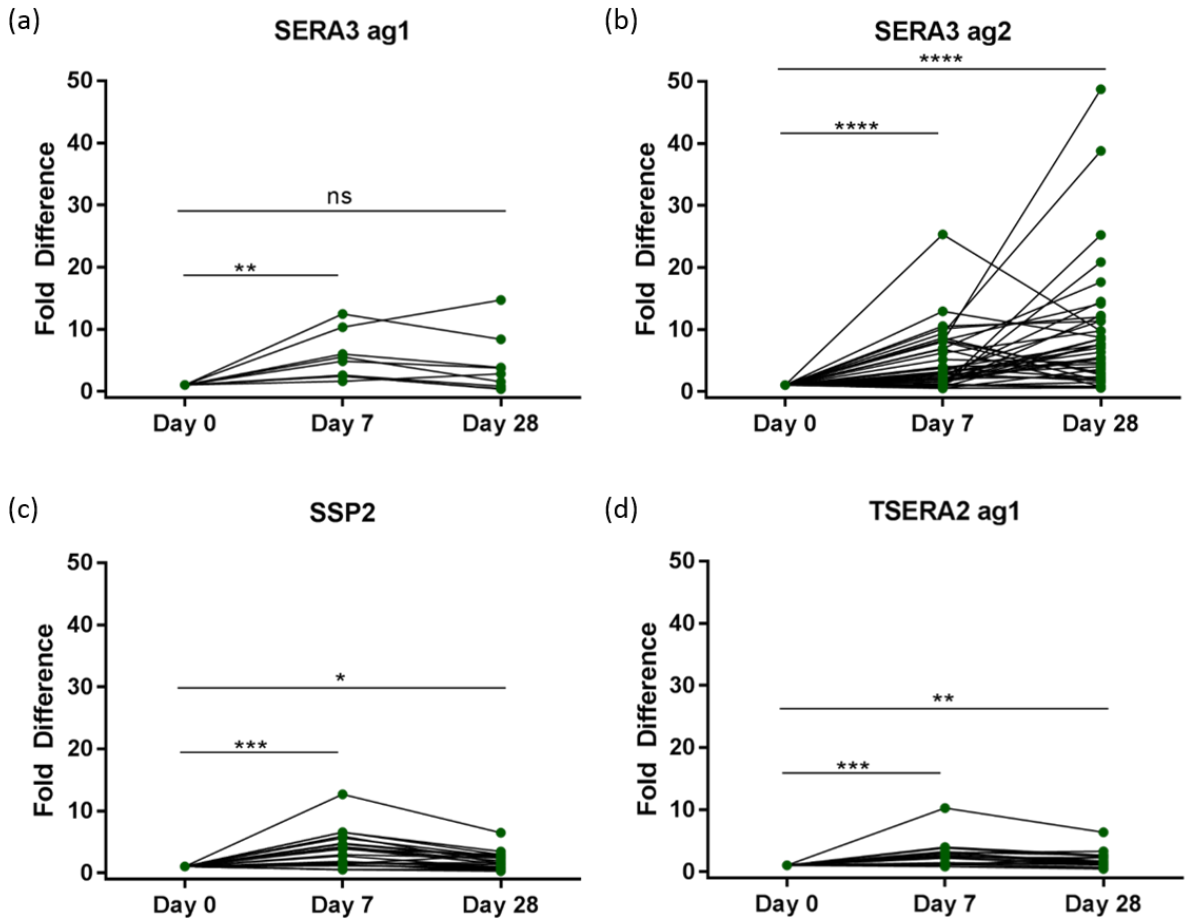


Figure 5. Serial fold increase in antibody reactivity for each antigen following treatment of knowlesi malaria. (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2 and (d) TSERA2 ag1**.** Asterisks indicate level of significance, ns denotes non-significant values (p≤0.0001 ****; p≤0.001 ***; p≤0.01 **; p≤0.05 * and p>0.05 ns).



Results:

***In silico* identification, design and expression of target sequences**

Sequences associated with known immunological markers in *P. falciparum* were selected based on existing evidence (AMA1 [24, 26], MSP1 [26, 27], TRAP/SSP2 [29] and SERA antigens [30, 31]), by interrogating existing *P. knowlesi* databases[33, 53] and supporting literature [54] (Fig 1). AMA1 is expressed in the micronemes of both the merozoite (invasive asexual blood stage) and sporozoite (invasive pre-erythrocytive stage) forms [24]. MSP1 is a major protein located on the surface of the merozoite[27]. TRAP/SSP2 is also expressed on the surface of the sporozoite forms [29], and the SERA antigens are soluble parasitophorous vacuole proteins [30, 31]. Each sequence was processed using available *in silico* analytical tools (Fig 1). Gene3D [55] and SMART (http://smart.embl-heidelberg.de/) were used to obtain domain prediction information for each gene which helped with the design of truncated fragments (Fig 2). This approach ensured that the design of truncated sequences properly accounted for the presence of any potential domains within each sequence, avoiding unintended truncation of domains which could impact on the solubility of the recombinant proteins. To ensure that expressed products would be specific for *P. knowlesi*, target sequences were interrogated multiple times using the BlastP algorithm [56] against both the *Plasmodium* specific (Plasmodb: https://is.gd/XOs7vd [33]) and non-redundant databases (NCBI: https://is.gd/MzISez).

Maximum likelihood phylogenetic trees were constructed using the *P. knowlesi* H reference strain, highlighting the relationship of each gene between *Plasmodium* species (S1 Figures a-e). Specifically, for both *Pv*AMA1 (bootstrap value: 100%) and *Pv*MSP1-19 (bootstrap value: 87%), there is a strong relationship between different *Plasmodium* species, particularly between *P. knowlesi* and *P. vivax* (S1 Figure a), highlighted further by corresponding near identical amino acid alignments (S2 Figure a). Amino acid alignments were generated using available sequences for human-pathogenic *Plasmodium* spp., which clearly highlight the level of sequence identity for both genes between *P. knowlesi* and *P. vivax* (S2 Figures a-e). Although the bootstrap value strongly supports the grouping

of *P. knowlesi* with *P. vivax* and *P. simiovale (P. simiovale* was used when data for *P. ovale* was lacking) (S1 Figures c-e; bootstrap value: 100%), the alignments for SSP2 and the SERA antigens (PKNH_0413400 and PKNH_0413500), help identify regions specific for *P. knowlesi* (S2 Figures c-e). Based on these screens, any sequences showing high amino acid sequence identity to other *Plasmodium* spp., specifically *P. ovale*, *P. malariae*, *P. falciparum* and *P. vivax*, were re-edited to focus on *P. knowlesi*-specific regions only, where possible. All the antigens were expressed in *Escherichia coli* as soluble products with final yields ranging from 11.9 – 20.5 mg/L (Fig 3, Table 1).

Based on their predicted molecular masses (including the GST tag), SDS PAGE analysis of the purified proteins clearly suggested multimerisation of the purified products (both monomer and dimer) (Fig 3). The Coomassie stained profiles also illustrated that there is very little non-specific degradation of the recombinant proteins (Fig 3), suggesting that the proteins are stable under the conditions used.

The results of the Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR) confirmed that both the SERA3 and TSERA2 candidate genes were actively transcribed in the blood stage (S3 Figure). By contrast, SSP2, a sporozoite stage along with the CTRP and CSP pre-erythrocytic stage controls, were negative by RT-PCR (S3 Figure).

SNP analysis: capturing polymorphic epitopes in target genes

The existence of three major subpopulations of *P. knowlesi* have been recently described, two associated with clinical human infections from separate macaque species reservoir hosts and the third from long-term laboratory isolates [57]. The presence of amino acid polymorphisms biased towards a single cluster would likely limit the utility of any reagents generated to function as *P. knowlesi*-specific, for all *P. knowlesi*-strains. Therefore, we characterised the presence of SNPs associated with the clusters, focussing on non-synonymous positions within the *P. knowlesi*-specific truncated constructs. Supplementary Table 4 summarises both the synonymous and non-synonymous SNPs associated with the three clusters. For all antigens, the vast majority of the non-

synonymous SNPs lie in regions not covered by the antigen design. By omitting the majority of these cluster-specific SNPs we hoped to avoid segregation of detectable antibodies according to the defined clusters. The relevance of these genetic clusters in the context of immunity, and the potential relevance to host preferences is yet to be defined.

Serum reactivity to recombinant antigen panel

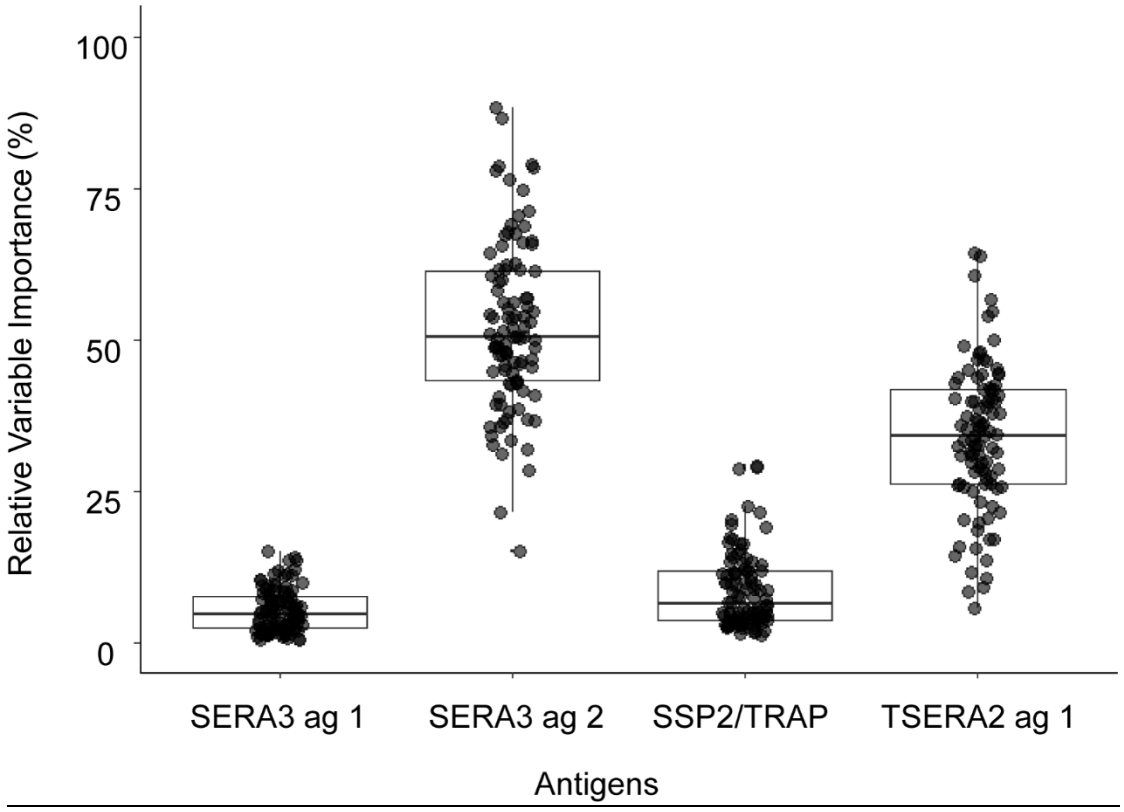
Samples were collected from 97 Malaysian adults and children hospitalised with *P. knowlesi* malaria on day of diagnosis (day 0), 7 and 28 days post-treatment. Hospital case samples were assayed by enzyme-linked immunosorbent assay (ELISA) using the *P. knowlesi*-specific protein panel. Ethiopian non- *P. knowlesi* malaria endemic children's sera (n=26) and adult UK malaria naïve sera (n=29) were used as a *P. knowlesi*-negative control panel. The *P. knowlesi*-negative malaria endemic controls were all reactive with the *Pv*MSP1-19 antigen due to previous *P. vivax* exposure. The malaria naïve controls showed no reactivity to any of the antigens tested (Fig 4, top row and S4 Figure). With the exception of one weakly positive sample to SERA3 ag 1 and SSP2, there was no other detectable antibody reactivity in the control group to the *P. knowlesi*-specific antigens (Fig 4). Antibody reactivity to all four antigens appeared to peak at day 7 (Fig 4, 5 and S4 Figure), although prevalence of antibody responses to SERA3 antigen 1, *Pk*SSP2 and TSERA2 antigen 1 remained relatively low (18.1% (13/72); 33.3% (45/72) and 43.1% (31/72) respectively) (Fig 4, columns 1, 3 and 4), compared to SERA 3 antigen 2 (63.8% (46/72)). The *Pk*SERA3 antigen 2 had a higher prevalence compared to controls at all time-points ($p<0.001$) (Fig 4 and S4 Figure). Antibody responses measured at day 7 and 28 to SERA3 antigen 2 demonstrated a significant increase when compared to day 0 ($p<0.001$ for both comparisons), with fold changes as high as 50 observed for some samples (Fig 5). In comparison, the fold changes observed in serum responses to the TSERA2 antigen 1 (day 7 and 28; p=<0.001 and p=0.005 respectively), SERA3 antigen 1 (day 7; p=0.008), and *Pk*SSP2 (day 7

and 28; p=0.001 and p=0.013), although statistically significant had comparatively lower fold changes with a maximum of 15 (Fig 5).

Identification of *P. knowlesi* exposed individuals

To assess the predictive ability of responses to these antigens to identify *P. knowlesi* exposed individuals, we used boosted regression tree analysis, an ensemble modelling method combining aspects of machine learning and statistical analysis shown to have strong predictive performance and reliable identification of variable importance [52]. Similar data-adaptive statistical models are increasingly being used for classification and identification of patterns in large datasets and have previously been applied to identify predictive antigens [58]. Out of 100 models fit for randomly sampled equal numbers of exposed and unexposed individuals, the median classification accuracy was 88.9% (IQR: 86.1-91.3%), calculated as the cross-validated area under the receiver operator curve (AUC). Relative variable importance was calculated for all models. SERA3 antigen 2 contributed most to the models (median relative variable importance: 50.4% (IQR 43.3- 61.4%)), followed by TSERA2 antigen 1, PkSSP2 and SERA3 antigen 1 (Fig 6).

Figure 6. Relative variable importance of responses to each antigen from 100 boosted regression tree models predicting *P. knowlesi* seropositivity. Median values for the relative variable importance and interquartile ranges are shown for all antigens tested: SERA3 ag 1 (4.8%; IQR 2.5 – 7.8%); SERA3 ag 2 (50.4%; IQR 43.3- 61.4%); PkSSP2 (6.5%; IQR 3.7 – 11.8%) and TSERA ag 1 (34.2%; IQR 26.2- 41.8%).



Discussion:

P. knowlesi is a naturally occurring infection of long-tailed and pig-tailed macaques, historically associated with forested areas of Southeast Asia [10]. Increased deforestation of their natural habitat is thought to have led to increased interaction between macaques and the human population in endemic areas [59]. Changes in village level forest cover and historical forest loss has been associated with an increase in *P. knowlesi* clinical cases in Sabah [60], with malaria caused by *P. knowlesi* increasingly reported in Southeast Asia [3]. Conversely, there has also been a steady decline in the prevalence of *P. falciparum* and *P. vivax* infections in the same region [61].

Currently, there are no specifically designed biomarkers for the serosurveillance of *P. knowlesi* infections. Recombinant proteins are available [PkCSP [62], PkAMA1 [63], PkDBP [64], PkSPATR [21], PkLDH [65], Pk1-Cys peroxiredoxin [66], Pk knowpains [67], PkMSP1-42 [68], PkMSP1-33 [69], PkMSP1-19 [70], Pk tryptophan-rich antigens (PkTrags) [71], PkMSP3 [72] and PkSBP1 [73], but are limited in number and are generally not species-specific. As a result, their utility as serological diagnostic tools is generally secondary to their original design. The reported level of amino acid sequence conservation to other *Plasmodium* spp. in some currently available *P. knowlesi* proteins is > 60%. Such reagents could not be specific to *P. knowlesi* [68, 74-76] and would be unable to reliably discriminate between antibody responses to different parasite species in co-endemic settings.

The recent efforts of the malaria community towards achieving malaria elimination means that tools to help monitor the impact and effectiveness of intervention strategies are an urgent requirement [77]. The development of species-specific tools for *P. knowlesi* would allow accurate assessment of the levels and geographical limits of infection with this zoonotic species [78]. There is an urgent need to develop a comprehensive discovery strategy to help identify *P. knowlesi* unique antigenic markers of exposure in order to further characterise this organism and develop stronger and better identification methods.

High levels of amino acid identity (83%) between PvMSP1-19 and PkMSP1-19, meant we were unable to use these reagents to dissect the species-specific immune responses due to the inevitable cross-reactive antibody responses. This is consistent with a proportion (48.9% (45/92)) of the confirmed *P. knowlesi*-exposed clinical samples in this study reacting with *Pv*MSP1-19 at day 0, although it is unknown whether these participants had previously been exposed to *P. vivax*. However, this limitation simply reflects the paucity of available serological reagents for use in assessing exposure to infection, a deficit this study aims to address. Although the small number of clinical case samples do not give sufficient statistical power to assess either the duration of antibody responses to the panel of antigens or population-level exposure, the *P. knowlesi* clinical case samples represent a unique dataset with which to validate the immunogenicity of our antigen panel. The SERA3 antigen 2 recombinant was recently assayed as part of a population-based survey of case study communities in Sabah, Malaysia and Palawan, Philippines [79].

The panel of reagents developed for this study focussed on immunologically relevant orthologous targets previously described in *P. falciparum*. The serine repeat antigen (SERA) family had previously attracted attention as a source of both drug and vaccine candidates [80]. In *P. falciparum*, SERA 5 is the most abundant parasitophorous vacuole protein and is essential to blood stage growth of the parasite [81], with antibodies against this antigen thought to inhibit parasite growth [82]. Although possessing a papain-like enzymatic domain, recent evidence suggests that the protein plays a non-enzymatic role [81]. SERA 3 has also been shown to be a highly immunogenic antigen with an important, although not essential role in the erythrocytic cycle [83] and has also been implicated as having a role in liver stage merozoite release in *P. berghei* [84]. Similarly, evidence for the sporozoite surface protein 2 (SSP2) suggested an immunogenic antigen involved in protection from disease in mice [85]. Although we were unable to confirm active transcription of SSP2 due to the lack of available material, we were able to validate active transcription of both the SERA3 and TSERA2 candidate genes. Collectively, the evidence provided by studies on *Plasmodium* supports the design of seroepidemiology tools based on these targets. Despite the targeted

approach used in designing the recombinant constructs, the SERA3 antigen 2 construct was by far the most promising candidate. The differences in the performances of the antigens could be due to a number of factors: (1) variation in the inherent immunogenicity of the regions selected, despite the predicted antigenicity score of the Jameson-Wolf antigenicity index or (2) variations in the expression status of the *P. knowlesi* antigens compared to *P. falciparum*.

The potential limitations of the study, are in the small number of clinical samples used in the study and the lack of a longer longitudinal follow-up. This would have allowed us to investigate the longevity of antibody responses to each target, across individuals and age groups. In addition, the lack of supporting information on *P. knowlesi*, such as functional data, transcriptional or RNA seq data hampered the rational selection of candidates.

This is the first study to describe the development a panel of *P. knowlesi*-specific serological tools using freely available *in silico* software. We have demonstrated the importance of targeting species-specific reagents at the amino acid level and highlighted the potential of such proteins as serosurveillance tools. Using these tools we have been able to measure specific immune responses to these reagents and described the change in antibody profile following treatment. As such, we have already demonstrated the utility of the SERA3 antigen 2 reagents as a potential seroepidemiological tool. Studies are also currently in development to expand the existing panel of *P. knowlesi* species-specific reagents to identify additional serological tools. Beyond this we envisage employing high throughput antigen discovery approaches such as the protein microarray to help identify additional important targets of immunity (Helb et al. PMID: 26216993; Uplekar et al. PMID: 28118367;. Further validation of the SERA3 antigen 2 at the population level has recently been performed [79]. Further studies are also planned to characterise the wider immunoglobulin responses, such as IgG subclasses and IgM, to these and future antigens.

Acknowledgments

We are grateful to all study participants for their cooperation.

References:

1. White NJ. Plasmodium knowlesi: the fifth human malaria parasite. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2008;46(2):172-3. Epub 2008/01/04. doi: 10.1086/524889. PubMed PMID: 18171246.
2. Moyes CL, Henry AJ, Golding N, Huang Z, Singh B, Baird JK, et al. Defining the geographical range of the Plasmodium knowlesi reservoir. PLoS neglected tropical diseases. 2014;8(3):e2780. Epub 2014/03/29. doi: 10.1371/journal.pntd.0002780. PubMed PMID: 24676231; PubMed Central PMCID: PMCPMC3967999.
3. Singh B, Daneshvar C. Human infections and detection of Plasmodium knowlesi. Clin Microbiol Rev. 2013;26(2):165-84. Epub 2013/04/05. doi: 10.1128/cmr.00079-12. PubMed PMID: 23554413; PubMed Central PMCID: PMCPmc3623376.
4. William T, Jelip J, Menon J, Anderios F, Mohammad R, Awang Mohammad TA, et al. Changing epidemiology of malaria in Sabah, Malaysia: increasing incidence of Plasmodium knowlesi. Malaria journal. 2014;13:390. Epub 2014/10/03. doi: 10.1186/1475-2875-13-390. PubMed PMID: 25272973; PubMed Central PMCID: PMCPMC4195888.
5. Lubis IN, Wijaya H, Lubis M, Lubis CP, Divis PC, Beshir KB, et al. Contribution of Plasmodium knowlesi to multi-species human malaria infections in North Sumatera, Indonesia. The Journal of infectious diseases. 2017. Epub 2017/02/16. doi: 10.1093/infdis/jix091. PubMed PMID: 28201638.
6. Cox-Singh J, Davis TM, Lee KS, Shamsul SS, Matusop A, Ratnam S, et al. Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2008;46(2):165-71. Epub 2008/01/04. doi: 10.1086/524888. PubMed PMID: 18171245; PubMed Central PMCID: PMCPmc2533694.
7. Barber BE, William T, Jikal M, Jilip J, Dhararaj P, Menon J, et al. Plasmodium knowlesi malaria in children. Emerging infectious diseases. 2011;17(5):814-20. Epub 2011/05/03. doi: 10.3201/eid1705.101489. PubMed PMID: 21529389; PubMed Central PMCID: PMCPmc3321776.
8. Daneshvar C, Davis TM, Cox-Singh J, Rafa'ee MZ, Zakaria SK, Divis PC, et al. Clinical and laboratory features of human Plasmodium knowlesi infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2009;49(6):852-60. Epub 2009/07/29. doi: 10.1086/605439. PubMed PMID: 19635025; PubMed Central PMCID: PMCPmc2843824.
9. Garnham PCC. Malaria parasites and other haemosporidia: Blackwell Scientific; 1966.
10. Coatney GR, Collins WE, Contacos PG. The Primate malarias. Bethesda, Md.: U.S. National Institute of Allergy and Infectious Diseases; 1971.
11. Singh B, Kim Sung L, Matusop A, Radhakrishnan A, Shamsul SS, Cox-Singh J, et al. A large focus of naturally acquired Plasmodium knowlesi infections in human beings. Lancet. 2004;363(9414):1017-24. Epub 2004/03/31. doi: 10.1016/s0140-6736(04)15836-4. PubMed PMID: 15051281.
12. Barber BE, William T, Grigg MJ, Menon J, Auburn S, Marfurt J, et al. A prospective comparative study of knowlesi, falciparum, and vivax malaria in Sabah, Malaysia: high proportion with severe disease from Plasmodium knowlesi and Plasmodium vivax but no mortality with early referral and artesunate therapy. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2013;56(3):383-97. Epub 2012/10/23. doi: 10.1093/cid/cis902. PubMed PMID: 23087389.
13. Barber BE, William T, Grigg MJ, Yeo TW, Anstey NM. Limitations of microscopy to differentiate Plasmodium species in a region co-endemic for Plasmodium falciparum, Plasmodium vivax and Plasmodium knowlesi. Malaria journal. 2013;12:8. Epub 2013/01/09. doi: 10.1186/1475-2875-12-8. PubMed PMID: 23294844; PubMed Central PMCID: PMCPMC3544591.
14. William T, Menon J, Rajahram G, Chan L, Ma G, Donaldson S, et al. Severe Plasmodium knowlesi malaria in a tertiary care hospital, Sabah, Malaysia. Emerging infectious diseases.

2011;17(7):1248-55. Epub 2011/07/19. doi: 10.3201/eid1707.101017. PubMed PMID: 21762579; PubMed Central PMCID: PMCPmc3381373.
15. Rajahram GS, Barber BE, William T, Menon J, Anstey NM, Yeo TW. Deaths due to Plasmodium knowlesi malaria in Sabah, Malaysia: association with reporting as Plasmodium malariae and delayed parenteral artesunate. Malaria journal. 2012;11:284. Epub 2012/08/22. doi: 10.1186/1475-2875-11-284. PubMed PMID: 22905799; PubMed Central PMCID: PMCPMC3472242.
16. Wipasa J, Suphavilai C, Okell LC, Cook J, Corran PH, Thaikla K, et al. Long-lived antibody and B Cell memory responses to the human malaria parasites, Plasmodium falciparum and Plasmodium vivax. PLoS Pathog. 2010;6(2):e1000770. Epub 2010/02/23. doi: 10.1371/journal.ppat.1000770. PubMed PMID: 20174609; PubMed Central PMCID: PMCPMC2824751.
17. Akpogheneta OJ, Duah NO, Tetteh KK, Dunyo S, Lanar DE, Pinder M, et al. Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. Infection and immunity. 2008;76(4):1748-55. Epub 2008/01/24. doi: 10.1128/iai.01333-07. PubMed PMID: 18212081; PubMed Central PMCID: PMCPMC2292892.
18. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? Trends in Parasitology. 2007;23(12):575-82. doi: http://dx.doi.org/10.1016/j.pt.2007.08.023.
19. Drakeley C, Cook J. Chapter 5 Potential Contribution of Sero-Epidemiological Analysis for Monitoring Malaria Control and Elimination: Historical and Current Perspectives. Advances in parasitology. 69: Academic Press; 2009. p. 299-352.
20. Waters AP, Thomas AW, Deans JA, Mitchell GH, Hudson DE, Miller LH, et al. A merozoite receptor protein from Plasmodium knowlesi is highly conserved and distributed throughout Plasmodium. J Biol Chem. 1990;265(29):17974-9. Epub 1990/10/15. PubMed PMID: 2211675.
21. Mahajan B, Jani D, Chattopadhyay R, Nagarkatti R, Zheng H, Majam V, et al. Identification, cloning, expression, and characterization of the gene for Plasmodium knowlesi surface protein containing an altered thrombospondin repeat domain. Infection and immunity. 2005;73(9):5402-9. Epub 2005/08/23. doi: 10.1128/iai.73.9.5402-5409.2005. PubMed PMID: 16113256; PubMed Central PMCID: PMCPmc1231135.
22. World Health Organization. Informal Consultation on the Public Health Importance of Plasmodium knowlesi: Kota Samaharan, Sarawak, Malaysia; 2011 [3 May 2017]. Available from: http://www.wpro.who.int/mvp/documents/docs/Pknowlesi_final_report.pdf.
23. Laurens MB, Kouriba B, Bergmann-Leitner E, Angov E, Coulibaly D, Diarra I, et al. Strain-specific Plasmodium falciparum growth inhibition among Malian children immunized with a blood-stage malaria vaccine. PLoS One. 2017;12(3):e0173294. Epub 2017/03/11. doi: 10.1371/journal.pone.0173294. PubMed PMID: 28282396; PubMed Central PMCID: PMCPMC5345808.
24. Yang AS, Lopaticki S, O'Neill MT, Erickson SM, Douglas DN, Kneteman NM, et al. AMA1 and MAEBL are important for Plasmodium falciparum sporozoite infection of the liver. Cell Microbiol. 2017. Epub 2017/04/04. doi: 10.1111/cmi.12745. PubMed PMID: 28371168.
25. Tadesse FG, van den Hoogen L, Lanke K, Schildkraut J, Tetteh K, Aseffa A, et al. The shape of the iceberg: quantification of submicroscopic Plasmodium falciparum and Plasmodium vivax parasitaemia and gametocytaemia in five low endemic settings in Ethiopia. Malaria journal. 2017;16(1):99. Epub 2017/03/04. doi: 10.1186/s12936-017-1749-4. PubMed PMID: 28253867; PubMed Central PMCID: PMCPMC5335517.
26. Ssewanyana I, Arinaitwe E, Nankabirwa JI, Yeka A, Sullivan R, Kamya MR, et al. Avidity of anti-malarial antibodies inversely related to transmission intensity at three sites in Uganda. Malaria journal. 2017;16(1):67. Epub 2017/02/12. doi: 10.1186/s12936-017-1721-3. PubMed PMID: 28183299; PubMed Central PMCID: PMCPMC5301436.
27. Moss DK, Remarque EJ, Faber BW, Cavanagh DR, Arnot DE, Thomas AW, et al. Plasmodium falciparum 19-kilodalton merozoite surface protein 1 (MSP1)-specific antibodies that interfere with parasite growth in vitro can inhibit MSP1 processing, merozoite invasion, and intracellular parasite

development. Infection and immunity. 2012;80(3):1280-7. Epub 2011/12/29. doi: 10.1128/iai.05887-11. PubMed PMID: 22202121; PubMed Central PMCID: PMCPMC3294643.
28. Bargieri DY, Thiberge S, Tay CL, Carey AF, Rantz A, Hischen F, et al. Plasmodium Merozoite TRAP Family Protein Is Essential for Vacuole Membrane Disruption and Gamete Egress from Erythrocytes. Cell Host Microbe. 2016;20(5):618-30. Epub 2016/11/11. doi: 10.1016/j.chom.2016.10.015. PubMed PMID: 27832590; PubMed Central PMCID: PMCPMC5104695.
29. Rogers WO, Malik A, Mellouk S, Nakamura K, Rogers MD, Szarfman A, et al. Characterization of Plasmodium falciparum sporozoite surface protein 2. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(19):9176-80. Epub 1992/10/01. PubMed PMID: 1409621; PubMed Central PMCID: PMCPmc50088.
30. Owalla TJ, Palacpac NM, Shirai H, Horii T, Egwang TG. Association of naturally acquired IgG antibodies against Plasmodium falciparum serine repeat antigen-5 with reduced placental parasitemia and normal birth weight in pregnant Ugandan women: a pilot study. Parasitology international. 2013;62(3):237-9. Epub 2013/02/12. doi: 10.1016/j.parint.2013.01.006. PubMed PMID: 23395684.
31. Yagi M, Palacpac NM, Ito K, Oishi Y, Itagaki S, Balikagala B, et al. Antibody titres and boosting after natural malaria infection in BK-SE36 vaccine responders during a follow-up study in Uganda. Sci Rep. 2016;6:34363. Epub 2016/10/06. doi: 10.1038/srep34363. PubMed PMID: 27703240; PubMed Central PMCID: PMCPMC5050508 following received renumeration from BIKEN: N.M.Q.P., B.B., E.H.N., A.Y., B.N.K., O.K., H.S. and T.G.E. M.Y., K.I., Y.O., S.I., W.F. and Y.H. declare no conflict of interest.
32. Arisue N, Kawai S, Hirai M, Palacpac NMQ, Jia M, Kaneko A, et al. Clues to Evolution of the SERA Multigene Family in 18 Plasmodium Species. PLoS ONE. 2011;6(3):e17775. doi: 10.1371/journal.pone.0017775. PubMed PMID: PMC3058004.
33. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: a functional genomic database for malaria parasites. Nucleic acids research. 2009;37(Database issue):D539-43. Epub 2008/10/30. doi: 10.1093/nar/gkn814. PubMed PMID: 18957442; PubMed Central PMCID: PMCPMC2686598.
34. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research. 2004;32(5):1792-7. Epub 2004/03/23. doi: 10.1093/nar/gkh340. PubMed PMID: 15034147; PubMed Central PMCID: PMCPMC390337.
35. Vedadi M, Lew J, Artz J, Amani M, Zhao Y, Dong A, et al. Genome-scale protein expression and structural biology of Plasmodium falciparum and related Apicomplexan organisms. Molecular and biochemical parasitology. 2007;151(1):100-10. Epub 2006/11/28. doi: 10.1016/j.molbiopara.2006.10.011. PubMed PMID: 17125854.
36. Gopal GJ, Kumar A. Strategies for the production of recombinant protein in Escherichia coli. The protein journal. 2013;32(6):419-25. Epub 2013/07/31. doi: 10.1007/s10930-013-9502-5. PubMed PMID: 23897421.
37. Carpenter EP, Beis K, Cameron AD, Iwata S. Overcoming the challenges of membrane protein crystallography. Current opinion in structural biology. 2008;18(5):581-6. Epub 2008/08/05. doi: 10.1016/j.sbi.2008.07.001. PubMed PMID: 18674618; PubMed Central PMCID: PMCPMC2580798.
38. Moon RW, Hall J, Rangkuti F, Ho YS, Almond N, Mitchell GH, et al. Adaptation of the genetically tractable malaria pathogen Plasmodium knowlesi to continuous culture in human erythrocytes. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(2):531-6. Epub 2012/12/26. doi: 10.1073/pnas.1216457110. PubMed PMID: 23267069; PubMed Central PMCID: PMCPmc3545754.
39. Tetteh KK, Osier FH, Salanti A, Kamuyu G, Drought L, Failly M, et al. Analysis of antibodies to newly described Plasmodium falciparum merozoite antigens supports MSPDBL2 as a predicted target of naturally acquired immunity. Infection and immunity. 2013;81(10):3835-42. Epub 2013/07/31. doi: 10.1128/iai.00301-13. PubMed PMID: 23897617; PubMed Central PMCID: PMCPmc3811751.

40. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein expression and purification. 2005;41(1):207-34. Epub 2005/05/26. PubMed PMID: 15915565.
41. Studier FW. Stable expression clones and auto-induction for protein production in E. coli. Methods in molecular biology (Clifton, NJ). 2014;1091:17-32. Epub 2013/11/10. doi: 10.1007/978-1-62703-691-7_2. PubMed PMID: 24203322.
42. Polley SD, Tetteh KK, Cavanagh DR, Pearce RJ, Lloyd JM, Bojang KA, et al. Repeat sequences in block 2 of Plasmodium falciparum merozoite surface protein 1 are targets of antibodies associated with protection from malaria. Infection and immunity. 2003;71(4):1833-42. Epub 2003/03/26. PubMed PMID: 12654798; PubMed Central PMCID: PMCPMC152097.
43. Arisue N, Hirai M, Arai M, Matsuoka H, Horii T. Phylogeny and evolution of the SERA multigene family in the genus Plasmodium. Journal of molecular evolution. 2007;65(1):82-91. Epub 2007/07/05. doi: 10.1007/s00239-006-0253-1. PubMed PMID: 17609844.
44. Offeddu V, Thathy V, Marsh K, Matuschewski K. Naturally acquired immune responses against Plasmodium falciparum sporozoites and liver infection. Int J Parasitol. 2012;42(6):535-48. Epub 2012/05/09. doi: 10.1016/j.ijpara.2012.03.011. PubMed PMID: 22561398.
45. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. PubMed PMID: PMC2705234.
46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMCPmc2723002.
47. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80-92. Epub 2012/06/26. doi: 10.4161/fly.19695. PubMed PMID: 22728672; PubMed Central PMCID: PMCPmc3679285.
48. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3. Epub 2014/01/24. doi: 10.1093/bioinformatics/btu033. PubMed PMID: 24451623; PubMed Central PMCID: PMCPMC3998144.
49. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic acids research. 2016;44(W1):W242-5. Epub 2016/04/21. doi: 10.1093/nar/gkw290. PubMed PMID: 27095192; PubMed Central PMCID: PMCPMC4987883.
50. Corran PH, Cook J, Lynch C, Leendertse H, Manjurano A, Griffin J, et al. Dried blood spots as a source of anti-malarial antibodies for epidemiological studies. Malaria journal. 2008;7:195. Epub 2008/10/02. doi: 10.1186/1475-2875-7-195. PubMed PMID: 18826573; PubMed Central PMCID: PMCPmc2567984.
51. Grigg MJ, William T, Menon J, Dhanaraj P, Barber BE, Wilkes CS, et al. Artesunate-mefloquine versus chloroquine for treatment of uncomplicated Plasmodium knowlesi malaria in Malaysia (ACT KNOW): an open-label, randomised controlled trial. The Lancet infectious diseases. 2016. Epub 18 November 2015. doi: 10.1016/s1473-3099(15)00415-6. PubMed PMID: 26603174.
52. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. J Anim Ecol. 2008;77(4):802-13. doi: 10.1111/j.1365-2656.2008.01390.x. PubMed PMID: 18397250.
53. Logan-Klumpler FJ, De Silva N, Boehme U, Rogers MB, Velarde G, McQuillan JA, et al. GeneDB--an annotation database for pathogens. Nucleic acids research. 2012;40(Database issue):D98-108. Epub 2011/11/26. doi: 10.1093/nar/gkr1032. PubMed PMID: 22116062; PubMed Central PMCID: PMCPMC3245030.
54. Pain A, Bohme U, Berry AE, Mungall K, Finn RD, Jackson AP, et al. The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature. 2008;455(7214):799-803. Epub

2008/10/10. doi: 10.1038/nature07306. PubMed PMID: 18843368; PubMed Central PMCID: PMCPmc2656934.
55. Lewis TE, Sillitoe I, Andreeva A, Blundell TL, Buchan DW, Chothia C, et al. Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains. Nucleic acids research. 2013;41(Database issue):D499-507. Epub 2012/12/04. doi: 10.1093/nar/gks1266. PubMed PMID: 23203986; PubMed Central PMCID: PMCPMC3531217.
56. Mount DW. Using the Basic Local Alignment Search Tool (BLAST). CSH protocols. 2007;2007:pdb.top17. Epub 2007/01/01. doi: 10.1101/pdb.top17. PubMed PMID: 21357135.
57. Assefa S, Lim C, Preston MD, Duffy CW, Nair MB, Adroub SA, et al. Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi. Proceedings of the National Academy of Sciences. 2015. doi: 10.1073/pnas.1509534112.
58. Helb DA, Tetteh KK, Felgner PL, Skinner J, Hubbard A, Arinaitwe E, et al. Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(32):E4438-47. Epub 2015/07/29. doi: 10.1073/pnas.1501705112. PubMed PMID: 26216993; PubMed Central PMCID: PMCPMC4538641.
59. Imai N, White MT, Ghani AC, Drakeley CJ. Transmission and control of Plasmodium knowlesi: a mathematical modelling study. PLoS neglected tropical diseases. 2014;8(7):e2978. Epub 2014/07/25. doi: 10.1371/journal.pntd.0002978. PubMed PMID: 25058400; PubMed Central PMCID: PMCPmc4109903.
60. Fornace KM, Abidin TR, Alexander N, Brock P, Grigg MJ, Murphy A, et al. Association between Landscape Factors and Spatial Patterns of Plasmodium knowlesi Infections in Sabah, Malaysia. Emerging infectious diseases. 2016;22(2):201-9. doi: 10.3201/eid2202.150656. PubMed PMID: PMC4734530.
61. Mohd Abd Razak MR, Sastu UR, Norahmad NA, Abdul-Karim A, Muhammad A, Muniandy PK, et al. Genetic Diversity of Plasmodium falciparum Populations in Malaria Declining Areas of Sabah, East Malaysia. PLoS One. 2016;11(3):e0152415. Epub 2016/03/31. doi: 10.1371/journal.pone.0152415. PubMed PMID: 27023787; PubMed Central PMCID: PMCPMC4811561.
62. Sharma S, Godson GN. Expression of the major surface antigen of Plasmodium knowlesi sporozoites in yeast. Science. 1985;228(4701):879-82. Epub 1985/05/17. PubMed PMID: 3890178.
63. Mahdi Abdel Hamid M, Remarque EJ, van Duivenvoorde LM, van der Werff N, Walraven V, Faber BW, et al. 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. Epub 2011/06/10. doi: 10.1371/journal.pone.0020547. PubMed PMID: 21655233; PubMed Central PMCID: PMCPMC3105089.
64. Singh AP, Puri SK, Chitnis CE. Antibodies raised against receptor-binding domain of Plasmodium knowlesi Duffy binding protein inhibit erythrocyte invasion. Molecular and biochemical parasitology. 2002;121(1):21-31. Epub 2002/05/03. PubMed PMID: 11985860.
65. Singh V, Kaushal DC, Rathaur S, Kumar N, Kaushal NA. Cloning, overexpression, purification and characterization of Plasmodium knowlesi lactate dehydrogenase. Protein expression and purification. 2012;84(2):195-203. Epub 2012/06/12. doi: 10.1016/j.pep.2012.05.008. PubMed PMID: 22683723.
66. Hakimi H, Asada M, Angeles JM, Kawai S, Inoue N, Kawazu S. Plasmodium vivax and Plasmodium knowlesi: cloning, expression and functional analysis of 1-Cys peroxiredoxin. Experimental parasitology. 2013;133(1):101-5. Epub 2012/11/28. doi: 10.1016/j.exppara.2012.10.018. PubMed PMID: 23178658.
67. Prasad R, Atul, Soni A, Puri SK, Sijwali PS. Expression, characterization, and cellular localization of knowpains, papain-like cysteine proteases of the Plasmodium knowlesi malaria parasite. PLoS One. 2012;7(12):e51619. Epub 2012/12/20. doi: 10.1371/journal.pone.0051619. PubMed PMID: 23251596; PubMed Central PMCID: PMCPMC3520923.

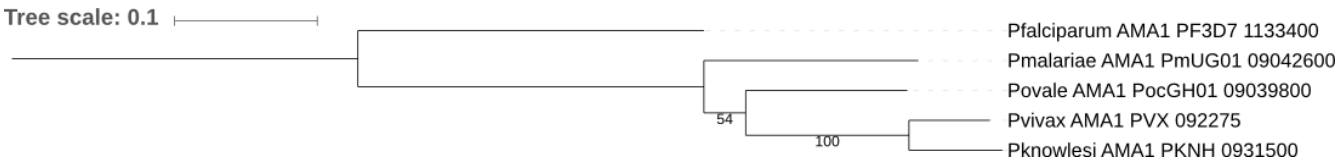
68. Cheong FW, Fong MY, Lau YL, Mahmud R. Immunogenicity of bacterial-expressed recombinant Plasmodium knowlesi merozoite surface protein-142 (MSP-142). Malaria journal. 2013;12:454. Epub 2013/12/21. doi: 10.1186/1475-2875-12-454. PubMed PMID: 24354660; PubMed Central PMCID: PMCPmc3878241.
69. Cheong FW, Lau YL, Fong MY, Mahmud R. Evaluation of recombinant Plasmodium knowlesi merozoite surface protein-1(33) for detection of human malaria. Am J Trop Med Hyg. 2013;88(5):835-40. Epub 2013/03/20. doi: 10.4269/ajtmh.12-0250. PubMed PMID: 23509118; PubMed Central PMCID: PMCPMC3752745.
70. Lau YL, Cheong FW, Chin LC, Mahmud R, Chen Y, Fong MY. Evaluation of codon optimized recombinant Plasmodium knowlesi Merozoite Surface Protein-119 (pkMSP-119) expressed in Pichia pastoris. Tropical biomedicine. 2014;31(4):749-59. Epub 2015/03/18. PubMed PMID: 25776601.
71. Tyagi K, Gupta D, Saini E, Choudhary S, Jamwal A, Alam MS, et al. Recognition of Human Erythrocyte Receptors by the Tryptophan-Rich Antigens of Monkey Malaria Parasite Plasmodium knowlesi. PLoS One. 2015;10(9):e0138691. Epub 2015/09/24. doi: 10.1371/journal.pone.0138691. PubMed PMID: 26393350; PubMed Central PMCID: PMCPmc4579084.
72. De Silva JR, Lau YL, Fong MY. Expression and Evaluation of Recombinant Plasmodium knowlesi Merozoite Surface Protein-3 (MSP-3) for Detection of Human Malaria. PLoS One. 2016;11(7):e0158998. Epub 2016/07/09. doi: 10.1371/journal.pone.0158998. PubMed PMID: 27391270; PubMed Central PMCID: PMCPMC4938616.
73. Lucky AB, Sakaguchi M, Katakai Y, Kawai S, Yahata K, Templeton TJ, et al. Plasmodium knowlesi Skeleton-Binding Protein 1 Localizes to the 'Sinton and Mulligan' Stipplings in the Cytoplasm of Monkey and Human Erythrocytes. PLoS One. 2016;11(10):e0164272. Epub 2016/10/13. doi: 10.1371/journal.pone.0164272. PubMed PMID: 27732628; PubMed Central PMCID: PMCPMC5061513.
74. Sonaimuthu P, Cheong FW, Chin LC, Mahmud R, Fong MY, Lau YL. Detection of human malaria using recombinant Plasmodium knowlesi merozoire surface protein-1 (MSP-1(1)(9)) expressed in Escherichia coli. Experimental parasitology. 2015;153:118-22. Epub 2015/03/31. doi: 10.1016/j.exppara.2015.03.010. PubMed PMID: 25812552.
75. Ahmed AM, Pinheiro MM, Divis PC, Siner A, Zainudin R, Wong IT, et al. Disease progression in Plasmodium knowlesi malaria is linked to variation in invasion gene family members. PLoS neglected tropical diseases. 2014;8(8):e3086. Epub 2014/08/15. doi: 10.1371/journal.pntd.0003086. PubMed PMID: 25121807; PubMed Central PMCID: PMCPmc4133233.
76. Rawa MS, Fong MY, Lau YL. Genetic diversity and natural selection in the rhoptry-associated protein 1 (RAP-1) of recent Plasmodium knowlesi clinical isolates from Malaysia. Malaria journal. 2016;15:62. Epub 2016/02/06. doi: 10.1186/s12936-016-1127-7. PubMed PMID: 26847346; PubMed Central PMCID: PMCPMC4743133.
77. World Health Organization. World Malaria report: 2016. 2016.
78. World Health Organization. WHO Malaria Policy Advisory Committee (MPAC) meeting 2017 [9 May 2017]. Available from: http://www.who.int/malaria/publications/atoz/WHO-HTM-GMP-2017.8-eng.pdf?ua=1.
79. Fornace KM, Herman LS, Abidin TR, Chua TH, Daim S, Lorenzo PJ, et al. Exposure and infection to Plasmodium knowlesi in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. (Submitted).
80. McCoubrie JE, Miller SK, Sargeant T, Good RT, Hodder AN, Speed TP, et al. Evidence for a common role for the serine-type Plasmodium falciparum serine repeat antigen proteases: implications for vaccine and drug design. Infection and immunity. 2007;75(12):5565-74. Epub 2007/09/26. doi: 10.1128/iai.00405-07. PubMed PMID: 17893128; PubMed Central PMCID: PMCPmc2168336.
81. Stallmach R, Kavishwar M, Withers-Martinez C, Hackett F, Collins CR, Howell SA, et al. Plasmodium falciparum SERA5 plays a non-enzymatic role in the malarial asexual blood-stage

lifecycle. Mol Microbiol. 2015;96(2):368-87. Epub 2015/01/21. doi: 10.1111/mmi.12941. PubMed PMID: 25599609; PubMed Central PMCID: PMCPMC4671257.
82. Aoki S, Li J, Itagaki S, Okech BA, Egwang TG, Matsuoka H, et al. Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of Plasmodium falciparum, and the acquired antibody titers correlate with serum inhibition of the parasite growth. J Biol Chem. 2002;277(49):47533-40. Epub 2002/09/24. doi: 10.1074/jbc.M207145200. PubMed PMID: 12244052.
83. Miller SK, Good RT, Drew DR, Delorenzi M, Sanders PR, Hodder AN, et al. A subset of Plasmodium falciparum SERA genes are expressed and appear to play an important role in the erythrocytic cycle. J Biol Chem. 2002;277(49):47524-32. Epub 2002/09/14. doi: 10.1074/jbc.M206974200. PubMed PMID: 12228245.
84. Schmidt-Christensen A, Sturm A, Horstmann S, Heussler VT. Expression and processing of Plasmodium berghei SERA3 during liver stages. Cell Microbiol. 2008;10(8):1723-34. Epub 2008/04/19. doi: 10.1111/j.1462-5822.2008.01162.x. PubMed PMID: 18419771; PubMed Central PMCID: PMCPMC2613260.
85. Khusmith S, Charoenvit Y, Kumar S, Sedegah M, Beaudoin RL, Hoffman SL. Protection against malaria by vaccination with sporozoite surface protein 2 plus CS protein. Science. 1991;252(5006):715-8. Epub 1991/05/03. PubMed PMID: 1827210.

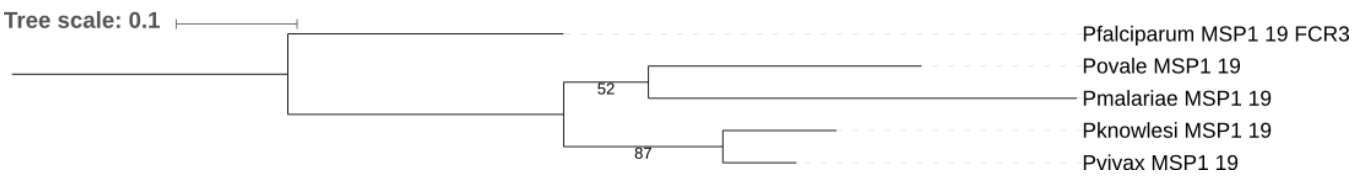
Supplementary Figures

Supplementary Figure 1: Maximum likelihood phylogenetic analysis of the amino acid sequences of AMA1 (a), MSP1-19 (b), SERA3 (c), TSERA2 (d) and TRAP/SSP2 (e) gene sequences between *P. knowlesi, P. falciparum, P. vivax, P. malariae* and *P. ovale/P. simiovale*. Bootstrap values are given in percentages.

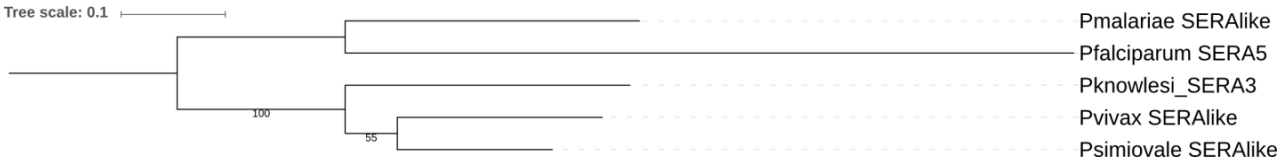
Supplementary Figure 1a



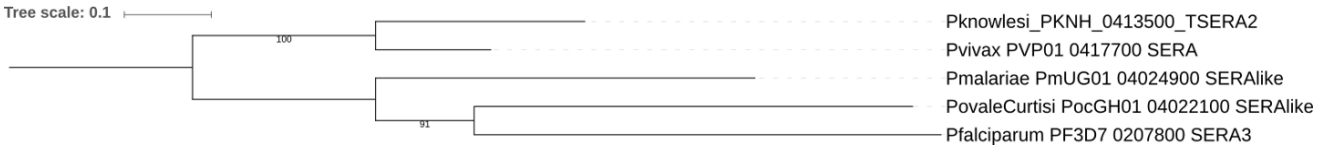
Supplementary Figure 1b



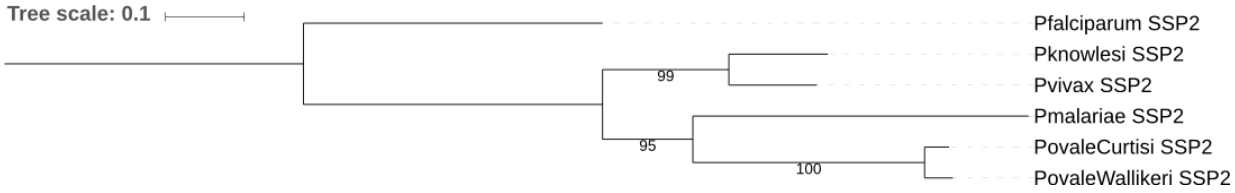
Supplementary Figure 1c



Supplementary Figure 1d



Supplementary Figure 1e



Supplementary Figure 2: Amino acid sequences alignments for AMA1 (a), MSP1-19 (b), SERA3 (c), TRAP/SSP2 (d) and TSERA2 (e) gene sequences between *P. knowlesi, P. falciparum, P. vivax, P. malariae* and *P. ovale/P. simiovale*. *P. knowlesi*-specific sequences are highlighted in yellow. Asterisks indicate fully conserved residues, colons indicates strong residue conservation (>0.5, Gonnet PAM 250 matrix), period indicates weak residue conservation (=<0.5, Gonnet PAM 250 matrix). Blank spaces indicate no residue conservation.

Supplementary Figure 2

Supp. Fig 2a: Apical Membrane Antigen 1 (AMA1)

```
Pfalciparum_AMA1_PF3D7_1133400      MRKLYCVLLLSAFEFTYMINFGRGQNYWEHPYQNSDVYRPINEHREHPKEYEYPLHQEHT
Pmalariae_AMA1_PmUG01_09042600      MKKLYYILLLST---QYLIHVYA-------------------------------------
Povale_AMA1_PocGH01_09039800        MKKIYYIFLLSA---HYLINVGK-------------------------------------
Pknowlesi_AMA1_PKNH_0931500         MNKIYYILFLSA---QCLVHMGK-------------------------------------
Pvivax_AMA1_PVX_092275              MNKIYYIIFLSA---QCLVHIGK-------------------------------------
                                    *.*:* :::**:     :::.

Pfalciparum_AMA1_PF3D7_1133400      YQQEDSGEDENTLQHAYPIDHEGAEPAPQEQNLFSSIEIVERSNYMGNPWTEYMAKYDIE
Pmalariae_AMA1_PmUG01_09042600      -----SPRNAKHGRLNGSG-GTLIEKGQ----------VVERSTRMSNPWKKYMEKYDVE
Povale_AMA1_PocGH01_09039800        -----CTRNQKQGRLTRSG-SAMVEKNP----------TIERSTRMINPWKKYMEKFDVE
Pknowlesi_AMA1_PKNH_0931500         -----CERNQKTTRLTRSANNASLEKGP----------IIERSIRMSNPWKAFMEKYDLE
Pvivax_AMA1_PVX_092275              -----CGRNQKPSRLTRSANNVLLEKGP----------TVERSTRMSNPWKAFMEKYDIE
                                         .  : :  .   .      *             :***  * ***. :* *:*:*

Pfalciparum_AMA1_PF3D7_1133400      EVHGSGIRVDLGEDAEVAGTQYRLPSGKCPVFGKGIIIENSNTTFLTPVATGNQYLKDGG
Pmalariae_AMA1_PmUG01_09042600      KTHGAGVRVDLGEDAEVKNSKYRIPGGKCPVFGKGIIIENSNVDFLTPVATGNRNLKSGG
Povale_AMA1_PocGH01_09039800        KTHGSGIRVDLGEDAEVKNSKYRIPSGRCPVFGKGITIENSEVSFLKPVATGNEKLKSGG
Pknowlesi_AMA1_PKNH_0931500         RAHNSGIRIDLGEDAEVGNSKYRIPAGKCPVFGKGIVIENSNVSFLTPVATGAQRLKEGG
Pvivax_AMA1_PVX_092275              RTHSSGVRVDLGEDAEVENAKYRIPAGRCPVFGKGIVIENSDVSFLRPVATGDQKLKDGG
                                     .*.:*:*:******** .::**:*.*.******** ****:. ** *****   **.**

Pfalciparum_AMA1_PF3D7_1133400      FAFPPTEPLMSPMTLDEMRHFYKDNKYVKNLDELTLCSRHAGNMIPDNDKNSNYKYPAVY
Pmalariae_AMA1_PmUG01_09042600      FAFPATDDHISPVTIEVLRKRYEEHADLMNLNDLSLCSKHASSFVISDDLNTSYRHPAVY
Povale_AMA1_PocGH01_09039800        FAFPLTDYHISPISLQNLKRRYNENVELMKLNDMSLCAKHASSFVISEDQNTTYRHPAVY
Pknowlesi_AMA1_PKNH_0931500         FAFPNADDHISPITIANLKERYKENADLMKLNDIALCKTHAASFVIAEDQNTSYRHPAVY
Pvivax_AMA1_PVX_092275              FAFPNANDHISPMTLANLKERYKDNVEMMKLNDIALCRTHAASFVMAGDQNSSYRHPAVY
                                    **** ::  :**:::  :.  *:::  : :*::::**  **..::   * *:.*.:****

Pfalciparum_AMA1_PF3D7_1133400      DDKDKKCHILYIAAQENNGPRYCNKDESKRNSMFCFRPAKDISFQNYTYLSKNVVDNWEK
Pmalariae_AMA1_PmUG01_09042600      DEKTKTCYILYLSAQENIGPRYCSKDAADKDTMFCFKPAKTDNFKHYAYLSKNVVSDWDV
```

```
Povale_AMA1_PocGH01_09039800        DEKEQTCYILYLSAQENLGPRYCSNDAADKDSIFCFKPEKNESFQNYVYLSKNLRDDWSS
Pknowlesi_AMA1_PKNH_0931500         DEKNKTCYMLYLSAQENMGPRYCSPDSQNKDAMFCFKPDKNEKFDNLVYLSKNVSNDWEN
Pvivax_AMA1_PVX_092275              DEKEKTCHMLYLSAQENMGPRYCSPDAQNRDAVFCFKPDKNESFENLVYLSKNVRNDWDK
                                    *:* :.*::**::**** *****. *  ..:::***.* *  .*.: .*****: .:*.

Pfalciparum_AMA1_PF3D7_1133400      VCPRKNLQNAKFGLWVDGNCEDIPHVNEFPAIDLFECNKLVFELSASDQPKQYEQHLTDY
Pmalariae_AMA1_PmUG01_09042600      KCPRKSLGVAKFGLWVDGNCEEIPSVKAFYADNLTECNRIVFEASASDQPTQYEENMTDY
Povale_AMA1_PocGH01_09039800        KCPRNNLTNSKFGLWVDGNCEDIPYVKEFQANTLRECNRIVFEASASDQPRQYEEELTDY
Pknowlesi_AMA1_PKNH_0931500         KCPRKNLGNAKFGLWVDGNCEEIPYVNEVEARSLRECNRIVFEASASDQPRQYEEELTDY
Pvivax_AMA1_PVX_092275              KCPRKNLGNAKFGLWVDGNCEEIPYVKEVEAEDLRECNRIVFGASASDQPTQYEEEMTDY
                                     ***:.*  :***********:** *: . *  * ***.:**  ****** ***: :***

Pfalciparum_AMA1_PF3D7_1133400      EKIKEGFKNKNASMIKSAFLPTGAFKADRYKSHGKGYNWGNYNTETQKCEIFNVKPTCLI
Pmalariae_AMA1_PmUG01_09042600      KKLEQGFRDNNPDMIKGAFLPVGAFNANFNKSKGKGFNWGNYDKINKKCFIFNVKPTCLI
Povale_AMA1_PocGH01_09039800        EKIQEGFRQNNPDMIKGAFFPVGAYKSDNFKSRGKGFNWGNFDIVNKKCYIFSAKPTCLI
Pknowlesi_AMA1_PKNH_0931500         EKIQEGFRQNNRDMIKSAFLPVGAFNSDNFKSKGRGYNWANFDSVNNKCYIFNTKPTCLI
Pvivax_AMA1_PVX_092275              QKIQQGFRQNNREMIKSAFLPVGAFNSDNFKSKGRGFNWANFDSVKKKCYIFNTKPTCLI
                                    :*:::**.::* .***.**:*.**::::  **.*.*:**.*::  .:** **..******

Pfalciparum_AMA1_PF3D7_1133400      NNSSYIATTALSHPIEVENNFPCSLYKDEIMKEIERESKRIKLNDNDDEGNKKIIAPRIF
Pmalariae_AMA1_PmUG01_09042600      NNKDFIATTALSHPEEVQEDFPCDIYKNEIEKELKRNSGNVKLYSLDGE---KIVLPRIF
Povale_AMA1_PocGH01_09039800        NDKNYIATTALSHPEDVERNFPCEIYKNEIEKEIEKQNRKAKLYSTDGD---RVVLPRIF
Pknowlesi_AMA1_PKNH_0931500         NDKNFFATTALSHPQEVDNEFPCSIYKDEIEREIKKQSRNMNLYSVDKE---RIVLPRIF
Pvivax_AMA1_PVX_092275              NDKNFIATTALSHPQEVDLEFPCSIYKDEIEREIKKQSRNMNLYSVDGE---RIVLPRIF
                                    *:..::******** :*: :***.:**:** .*::.:. . :* . * :   .:: ****

Pfalciparum_AMA1_PF3D7_1133400      ISDDKDSLKCPCDPEMVSNSTCRFFVCKCVERRAEVTSNNEVVVKEEYKDEYADIPEHKP
Pmalariae_AMA1_PmUG01_09042600      ISNNKDSLNCPCEPEKITNSSCDFYLCNCVEKRAEIKENNEVVIKDEFKEEYEYNEGNS-
Povale_AMA1_PocGH01_09039800        ISDDKDSLKCPCEPERITNSTCNYYVCNCVEKRAEIKENNEVIIKDEFKEDYENEEGENT
Pknowlesi_AMA1_PKNH_0931500         ISTDKESIKCPCEPEHISNSTCNFYVCNCVEKRAEIKENNEVIIKEEFKEDYENPDGKH-
Pvivax_AMA1_PVX_092275              ISNDKESIKCPCEPERISNSTCNFYVCNCVEKRAEIKENNQVVIKEEFRDYYENGEEKS-
                                    ** :*:*::***:** ::**:* :::*:***.***:..**:*::*:*:.: *

Pfalciparum_AMA1_PF3D7_1133400      TYDKMKIIIASSAAVAVLATILMVYLYKRKGNAEKYDKMDEPQDYGKS-NSRNDEMLDPE
Pmalariae_AMA1_PmUG01_09042600      NNKKTLIIIGLAGGVGILALASSFFFFKKKTENEKYDKMDQADVYGKS-TTRKDEMLDPE
Povale_AMA1_PocGH01_09039800        NRQRTIIIIGLAGGVAVLGCASFFFFFKKKAQGKEYDKMDQTDGYGKP-KSRKDEMLDPE
Pknowlesi_AMA1_PKNH_0931500         KKKMLLIIIGVTGAVCVVAVASLFY-FRKKAQDDKYDKMDQAEAYGKTANTRKDEMLDPE
Pvivax_AMA1_PVX_092275              NKQMLLIIIGITGGVCVVALASMAY-FRKKANNDKYDKMDQAEGYGKP-TTRKDEMLDPE
                                    . .   ***. :..* ::.     : :..* : .:*****:.: ***. .:*:*******

Pfalciparum_AMA1_PF3D7_1133400      ASFWGEEKRASHTTPVLMEKPYY
Pmalariae_AMA1_PmUG01_09042600      ASFWGEEKRASHTTPVLMEKPYY
Povale_AMA1_PocGH01_09039800        ASFWGEEKRASHTTPVLMEKPYY
Pknowlesi_AMA1_PKNH_0931500         ASFWGEDKRASHTTPVLMEKPYY
```

```
Pvivax_AMA1_PVX_092275              ASFWGEDKRASHTTPVLMEKPYY
                                    ******:****************
```

Supp. Fig 2b: Merozoite Surface Protein (MSP) 1-19

```
Pfalciparum_MSP1_19_FCR3      -NISQHQCVKKQCPQNSGCFRHLDEREECKCLLNYKQEGDKCVENPNPTCNENNGGCDAD
Pmalariae_MSP1_19             NISAKHACTETKYPENAGCYRYEDGKEVWRCLLNYKLVDGGCVEDEEPSCQVNNGGCAPE
Povale_MSP1_19                -MGSKHKCIDITYPDNAGCYRFSDGREEWRCLLNFKKVGETCVPNNNPTCAENNGGCDPT
Pknowlesi_MSP1_19             NMSSAHKCIDTNVPENAACYRYLDGTEEWRCLLGFKEVGGKCVPAS-ITCEENNGGCAPE
Pvivax_MSP1_19                -MSSEHTCIDTNVPDNAACYRYLDGTEEWRCLLTFKEEGGKCVPASNVTCKDNNGGCAPE
                                 : * * .   *:*:.*:*. *  *  .*** :*  .  **     :*  ***** .

Pfalciparum_MSP1_19_FCR3      AKCTEEDSGSNGKKITCECTKPDSYPLFDGIFCSSSN
Pmalariae_MSP1_19             ANCTKGDD----NKIVCACNAPYSEPIFEGVFCGSSS
Povale_MSP1_19                ADCAESEN----NKITCTCTGQ-NESFFEGVFCGSSS
Pknowlesi_MSP1_19             AECTMDDK----KEVECKCTKEGSEPLFEGVFCSSSS
Pvivax_MSP1_19                AECKMTDS----NKIVCKCTKEGSEPLFEGVFCSSSS
                              *.*   :.    ::: * *.   . .:*:*:**.**.
```

Supp. Fig 2c: Serine Repeat Antigen (SERA) 3 protein

Antigen 1

```
Pfalciparum_SERA5       MKSYISLFFILCVIFNKNVIKCT--GESQTGNTG---------GGQAGNTGGDQAGSTGG
Pmalariae_SERAlike      MKYGILYIFMICISFGSNTIKCTTVSVSDNRGNEASEQPLQPAQPGPQTHEPSNSQVQNS
Pknowlesi_SERA3         MKSSFLLLLALCATYGNNLAICTTEGTAQSGVSSDSQHSLSSSETETGSHGAPGAEAQSV
Pvivax_SERAlike         MKSSVLLLLALGATYGNNVAMCT--ATPPSGGPHASLPNPGGPGTGAENQGQSQAGQQLP
Psimiovale_SERAlike     MKSSFLLLLALGTAYGNNVVICTTGQTPSSGESGVASSSSSGPGTGSDNQQESGQGPQPE
                        **  .  :: :   :..*   **    . .               . .

Pfalciparum_SERA5       SPQGSTG---------------ASPQGSTGASPQGSTGASQPGSSEPS---NPVSSGHSV
Pmalariae_SERAlike      SNPNISDLTVTTPPVAQNLSHETPKNGSSQSPPQNGPLLSSPSAVNNGQ--PNVSSGGAV
Pknowlesi_SERA3         SPEGGQD---AVHSTN-ESAESDAESPTEPNPPQEDGTSNEDGNGGHSESSAPSVPGGSA
Pvivax_SERAlike         SPSGEPG---AASPTHTPSPGPLPPNPAQPNSLPAVETLSQQGGGAPAASSNALTAGGNV
Psimiovale_SERAlike     SAGNGQN---VEESTNHQQPVPGAPNSAGSNLSQGGATIKSNGDGNPSAISPAPNAEGSV
                        *  .  .               . . :          .. .    .       .   .

Pfalciparum_SERA5       STVSVSQTSTSSEK------------------QDTIQVKSALLKDYMGLKVTGPCNENF
Pmalariae_SERAlike      SP-NLSSAGNSNGATQLSAESQNGAVSPKVPNYHNMAKIESALLKNHTGVRITGPCNEEV
Pknowlesi_SERA3         SPSSVENSNEQAAGTQLQVA-------------PQKAQVKSALLKNFTGVKVTGPCDTEV
Pvivax_SERAlike         PPGSQVNSGEQGGATQLQAT-------------PKKAELQSSLLKNFTGVKVTGPCDTEV
Psimiovale_SERAlike     SPSSSGSTRGPGGATQLQAS-------------HKKAELQSALLKNFTGVKVTGPCDTEV
                        .. .  .:                         .  :::*:***:. *:.:****: :.
```

Antigen 2

```
Pfalciparum_SERA5       IYDYYLKASPEFYHNLYFKNFNVGKKNLFSEKEDNENN---KKLGNNYIIFGQDT-----
Pmalariae_SERAlike      LYNYYLKTSPDFYSNLYFNSLSAEKANDLST----------NKVLDQMTVHGQAVEESSE
Pknowlesi_SERA3         LHSYYLKNSPDFYKNLYYNALDGECGNAPCNTVEGQDA-PGEKATDQVGASGAGVATVTT
Pvivax_SERAlike         LYSYHLKTSPDFYKNLYYNAVGGEKGSVLSNAVQGQDTPPGEEALPGAKVDGGGT-----
Psimiovale_SERAlike     LYSYHLNSSPDFYKNLYYNALGEKSGSALSHAVHGQDA-PQEEGESLGTVVGEGISESTV
                        ::.*:*: **:** ***:: ..    .  .          ::        *

Pfalciparum_SERA5       ------------------------------------------------------AGSGQSGK
Pmalariae_SERAlike      ------------------------------------------------------GTSGQHGQ
Pknowlesi_SERA3         ----------------------------------------------TGTGAAPGTGAEAK
Pvivax_SERAlike         ------------------------------------------------------PGPTGP
Psimiovale_SERAlike     QEVQQTQLQVPSVMSTNPRVEGQQEQAVVVDQASSHQRAEQVEASTLGAANTQESGPKVT
                                                                              .*

Pfalciparum_SERA5       ESNT----------------------------------------------------------
Pmalariae_SERAlike      HGQP----------------------------------------------------------
Pknowlesi_SERA3         AGAEAGAGAETE--ATKATEVLEPKEQQAQSQVTVVENSVSKDQPQPQPQPQPQPQPQLQ
Pvivax_SERAlike         EGQPQPHASSVG--------GQIP----QESRQEVEEKQVVDGQ----------------
Psimiovale_SERAlike     EGAQNGDAQSTGIAGVQTAPGTAPAIGAQTAQNTVGGETAAGGQGEGSVVSQLAADGTDS
                         .
```

```
Pfalciparum_SERA5       ------------------------------------------------------------ALE
Pmalariae_SERAlike      -------------------GQPESTS-----------------SSETVAESSA---QGLD
Pknowlesi_SERA3         QEQQQEQQQEQQQEQQQEQQQQQEQQQQQPQPPQPALQDLTNEHSPHIGETEVKVEPEGE
Pvivax_SERAlike         ----------PQPVTVE---GHPPSQPLQPTSPLQQSTQPGVGGAASTPLSQGT---PGQA
Psimiovale_SERAlike     RGITTDISSPQAQTLQPPGGQGVSTPAQPPSP-QQPTEP---VHASTSLVTGT---PGAA


Pfalciparum_SERA5       SAGTSNEVSERVHVYHILKHIKDGKIRMGMRKYIDTQDVNKKHSCTRSYAFNPENYEKCV
Pmalariae_SERAlike      SAASDVPDVQKFEVVHILKHIKNSKSKTTLVKYDYYYDFG-DHACSRTQASNPEKLGDCI
Pknowlesi_SERA3         NSNAELQKAKMVQIIHVLKHIKQTKMVTRVVTYQGNYELG-EHSCSRTEASSVEKLDECI
Pvivax_SERAlike         NSNGEAENANISQIIHVLKHIKKTKMVTRIVTYEGEYDLG-DHSCSRTQASSLEKLDDCI
Psimiovale_SERAlike     NPNAGVQSAKISQIIHVLKHIKQTKMVTRVVTYEGEYELG-DQSCSRTQALSLEKLDECI
                        ..       :   : *:*****. *    : .*    :.. .::*:*: * . *:  .*:
```

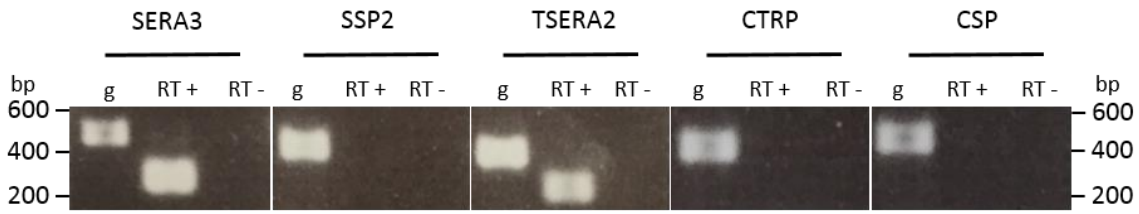
Supp. Fig 2d: Sporozoite Surface Protein (SSP) 2

```
Pfalciparum_SSP2          ENPPNPD--IPEQEPNIPEDSEKEVPSD------VPKNPEDD---RE----------
Pknowlesi_SSP2            TVPDESN--VIPVPPTVPGGSNSEFSSDVENAAQYPENPENPEN-PENSENPENPENQ--
Pvivax_SSP2               SVPDESN--VLPLPPAVPGGSSEEFPAD------VQNNPD--------------------
Pmalariae_SSP2            PNSESKDKLLPSDKEDEENKNNSNLPEG------LEERPQEGESLPVEPEGNENVEDNFP
PovaleWallikeri_SSP2      FVPDEEV--LPEAPANVPEERGNDVPEE------FPQNPENEQNVPE----------
PovaleCurtisi_SSP2        FILDEEN--LPEAPANVPEG-GNDVPEE------FPQNPENEQNVPE----------
                              .    :            .:..          :.*:

Pfalciparum_SSP2          -ENFDIPKKPENKH-------DN--QNNLPNDKSDR---SIPYSPLPPKVLDNERKQSDP
Pknowlesi_SSP2            NNPEDFPMEPDMSA-------DN--KINEPTNPSDSGQ-GIPENVIPTPINNEKDIINKN
Pvivax_SSP2               -SPEELPMEQEVPQ-------DN--NVNEPERSDSNGY-GVNEKVIPNPLDNERDMANKN
Pmalariae_SSP2            QAPNDLPGKQGQPDILNPDGGPNFGNNEHPGSPSNNDYSGKAYTHIPSPIGNEKNRSNYN
PovaleWallikeri_SSP2      -KPNDLPIEQEKPQDDGNNKVDY--KKNDMYKPEKGGYVIENDHRAPKPSNSYSDSKGKA
PovaleCurtisi_SSP2        -KPNDLPIEQEKPQDDGNNKGDH--KKSDIHVPETAGYVIENGHRVPKPLVNYSDDKGKA
                              ::* :               : .     .           *    .     .

Pfalciparum_SSP2          QSQDNNGNRHVPNSEDRETRPHGRNNENRSYNRKYNDTP---KHPEREEHEKPDNNKKKG
Pknowlesi_SSP2            KAVYPNGS---NQSHDRYPKPHRNAGGYDNNPNANSDIP---EGPFSSEEEQPEDKGKK-
Pvivax_SSP2               KTVHPDRK---DSARDRYARPHGSTHVNNNRANENSDIP---NNPVPSDYEQPEDKAKK-
Pmalariae_SSP2            H--NYSKSPNNNGPEDRVARPHKVDTNTESPRDSYNANPEYDETRESPNYEQREDNGKR-
PovaleWallikeri_SSP2      QSMNYSSNQYNNIPEERYPKPHKSIGRKDNSRNNYPSAP---YTPEEPTDDEYANKGKK-
PovaleCurtisi_SSP2        QSMNNEGNKNKNTSEERYPKPHKSIGRNDNSRKNYPSAP---YTPEEPTDDEYANKGKK-
                          :    . .     . :* ..**       .        *           ::  :: *.
```

Supp. Fig 2e: Truncated Serine Repeat Antigen (TSERA) 2 protein

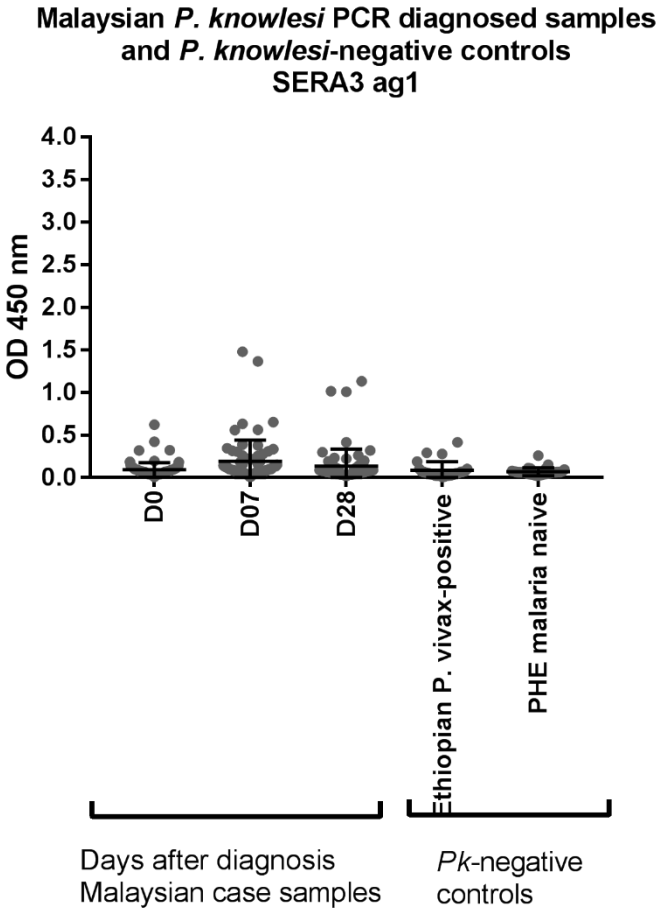
```
Pknowlesi_PKNH_0413500_TSERA2        MKARLSLILILCAVCRECTVRCTDTAIN-------QGQDAVQQPEESLSQDASDNSLPGQ
Pvivax_PVP01_0417700_SERA            MKARLSLILILCVVCRDCAVRCTGTTEA------QGAVEGAKGPKPEAEEAGAGKEEGAN
PovaleCurtisi_PocGH01_04022100_S     MKSHVSIIVTLYGIFSIHVTQCTGGGAHWPPHNMHTSNPGNGNKTGDSAIPPSSGNLST-
Pfalciparum_PF3D7_0207800_SERA3      MKFSISLFLILCVLFCKNDIKCTTVDES--------TKEGSQNPKNSSSTTPASGSQKG-
Pmalariae_PmUG01_04024900_SERAli     MRSCISLLLTIA---------CASDSNN--------GAPGSTG--NSSTTISSSNSERS-
                                     *.  :*::: :          *:               .      .     :. .

Pknowlesi_PKNH_0413500_ TSERA2       PVASPNGADQA----EVSQLQEGAVESSSNSADASNPNANADVTDVEGEKAATPSEGS--
Pvivax_PVP01_0417700_SERA            VGEAGTGGPGADGGTEAGARAEEGEGAGTEAEPARGPEPEAGGEGINRDAAGNQREGQLE
PovaleCurtisi_PocGH01_04022100_S     --SA-SGSTGS----ILPECTGESKGEDVKGGACSNSAPIEDLTLDRIEVVSIDKDGE--
Pfalciparum_PF3D7_0207800_SERA3      --SS-SESPGS----SVEKQSQESNKESTNGGNVVSQGTPANTFGQNSNNPSDSPQGT--
Pmalariae_PmUG01_04024900_SERAli     --ASPSESPSS------------GGQATSNGG--------SNLVGSSTAVPSNPQPSQ--
                                        : . .  :            .     :.          .         .    .

Pknowlesi_PKNH_0413500_ TSERA2       ----------------------------------------------------KEGMQ
Pvivax_PVP01_0417700_SERA            APSDSARPGAIPQVAPRDTVETSSDAADSSSPDQNPLPGADNTKVGNAATPPEGAKEETQ
PovaleCurtisi_PocGH01_04022100_S     -----------------DVEEGQNNEEKTENYIQKIIHSRTHSTVSHIRERTTDEPNIIE
Pfalciparum_PF3D7_0207800_SERA3      -----------------ST-------------------------------LPSPPKSID
Pmalariae_PmUG01_04024900_SERAli     -----------------DV-------------------------------STSIANKAK
                                                                                          :  .

Pknowlesi_PKNH_0413500_ TSERA2       VKSSLLKGYKGVKVTGPCSASFLVFFAPYLFIDVDADSSNIYLGTDLNDLE---------
Pvivax_PVP01_0417700_SERA            VKSSLLKGHKGVKVTGPCGASFLVFFAPYLFIDVDTDSSNVYLGTDLSDLE---------
PovaleCurtisi_PocGH01_04022100_S     IKSALLRDYNGVKVTGPCKAVFQMFLVPHITVNVETNKNSITLGPKLVQAHKKERVTTDG
Pfalciparum_PF3D7_0207800_SERA3      VKSAFLKHYKGVKVTGSCNANFQLFLVPHIFINVETKENNIQLDVKFLK-----------
Pmalariae_PmUG01_04024900_SERAli     IQSALLTDSNGVMVTGPCNEIFQVFFVPNIFINVQTDKNTVEMGNKFKS-----------
                                     ::*::*   :** ***.*   * :*:.* : ::*::....: :. .: .

Pknowlesi_PKNH_0413500_ TSERA2       -------ITEKMGKGKDE---KNKCQEGKT--FKFVAFVVNDHLTIKWKVYDSEDQTPTP
Pvivax_PVP01_0417700_SERA            -------VTEKMGIQDNG---KNKCEDKKT--FKFVALIGEDHLTIKWKVYDPSVKTPTP
PovaleCurtisi_PocGH01_04022100_S     VDAYTEHIEKGLMFEKEEKKLLNKCADGKS--FKFVLFIEGNKLTVKWKVYD-ATEAENS
Pfalciparum_PF3D7_0207800_SERA3      -------LTKRIDFAKDKSMLKNKCESGKNQTFKFVLYFKDDILTIKWKVYEEKSATPQK
Pmalariae_PmUG01_04024900_SERAli     -------LSNSITLKTFEYS-KNECAGGKT--FKFVALIQENKLTLKWKVYDAPNQNKTT
                                            : : :          *:*   *.  ****  .  : **:*****:
```

Supplementary Figure 3: *Plasmodium knowlesi* candidate gene transcriptional status in parasite mixed blood stage. Panel 1: SERA3; panel 2: SSP2; panel 3: TSERA2; panel 4: CTRP; panel 5: CSP. g refers to genomic DNA, RT+ refers to presence of RT enzyme and RT- refers to absence of RT enzyme. Samples were run on a 1.2% agarose gel. The DNA ladder is indicated in bp (Hyperladder 1Kb, Bioline).

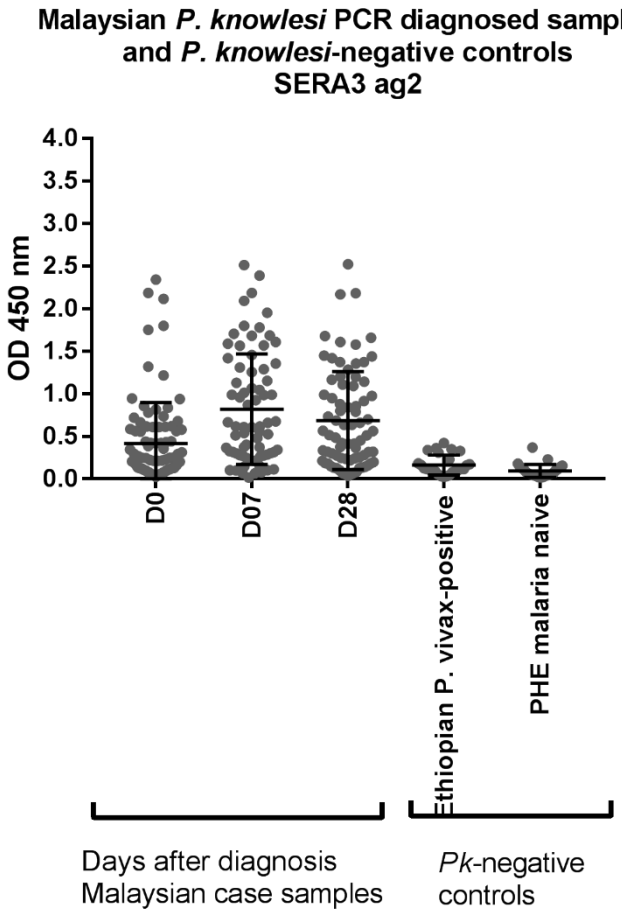


Supplementary Figure 4: *Plasmodium knowlesi* antigen reactivity to Malaysian hospital case serum samples and negative control serum samples. Dot plot of Malaysian hospital case serum samples from days 0 (n=92), 7 (n=72) and 28 (n=77) of PCR diagnosis and *Pk*-negative control serum samples (Ethiopian *Pv*-positive n=26; PHE malaria naïve n=29). Antibody reactivity to the *P. knowlesi*-specific antigens (a) SERA3 ag1, (b) SERA3 ag2, (c) SSP2 and (d) TSERA2 ag1 are shown.

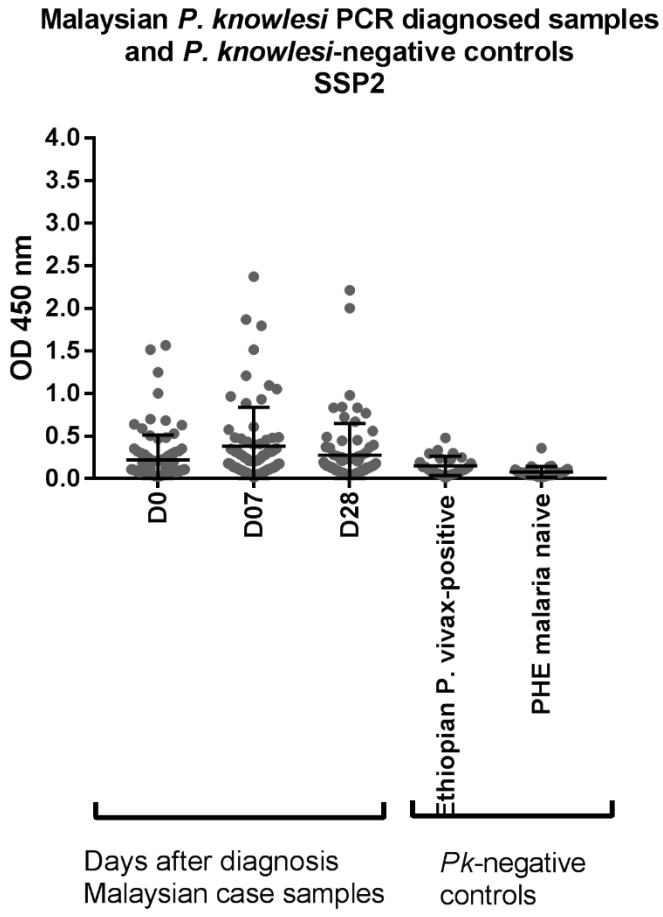
Supplementary Figure 4a



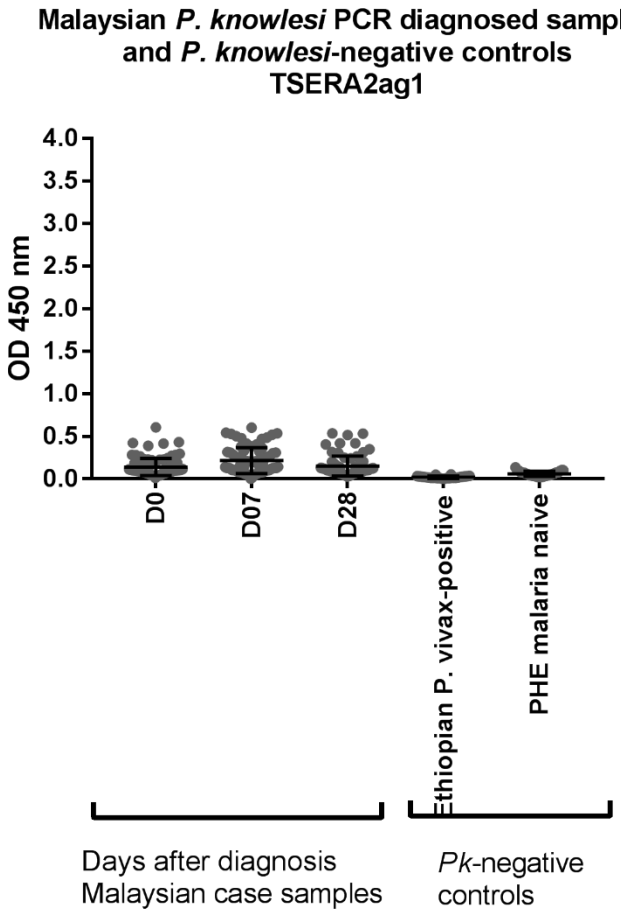
Supplementary Figure 4b



Supplementary Figure 4c



Supplementary Figure 4d



Supplementary Tables

Supplementary Table 1: Summary of the percentage amino acid identity between P. knowlesi and the other Plasmodium for all five candidates

P. knowlesi gene candidate	Homology (%)			
	P. vivax	***P. falciparum***	***P. malariae***	***P. ovale***
	PlasmoDB Identitities (%)	PlasmoDB Identitities (%)	PlasmoDB Identitities (%)	PlasmoDB Identitities (%)
AMA-1	85	58	69	73
MSP1	63	40	45	46
MSP1-19	83	50	56	56
SERA3	58	40	49	46
SERA3 ag1	40	25	28	29
SERA3 ag2	25	29	50	-
TRAP/SSP2	69	41	45	48
SSP2 frag	44	-	-	31
TSERA2	67	33	40	43
TSERA2 ag1	61	48	33	42

Supplementary Table 2: *P. knowlesi* gene name and ID, primer sequences, primer length, fragment size with and without intron.

Gene	Gene ID	Oligo name	Sequence: (5' to 3')	length (bp)	Size with intron (bp)	Size without intron (bp)
SERA3	PKNH_0413400	PKH041230F1	GCTCCAAGGGGAAAACATTA	20	401	250
SERA3	PKNH_0413400	PKH041230R1	GCAAGTTTGTATTCGCCAGA	20		
SSP2	PKNH_1265400	PKH121770F1	GGAAACCCATTCGAAGAAAA	20	419	-
SSP2	PKNH_1265400	PKH121770R1	TTTCTGTGTGGTTTGGGGTA	20		

TSERA2	PKNH_0413500	PKH041240F1	TGTGCAAGTACACAGCACGA	20	468	264
TSERA2	PKNH_0413500	PKH041240R1	GTCGTATATTGGGGCTTCCT	20		
CTRP	PKNH_0826900	PKCTRPF1	TCCTCACGAGAACAATGCAG	20	436	-
CTRP	PKNH_0826900	PKCTRPR1	TCACTGTCGCTTCCACTGTC	20		
CSP	PKNH_0838500	PKCSPF1	GAAACAACCGAATGAAGGACA	21	467	-
CSP	PKNH_0838500	PKCSPR1	CGTTTCTTCCTTGTCGTGGT	20		

Supplementary Table 3: *P. knowlesi* candidate name, primer sequences and primer length. The vector portion of each primer sequence (pGEX-2T) are highlighted in bold and the candidate portion of the sequence in italics. Stop codons are underlined

Antigen	Oligo name	Sequence : (5' to 3')	Length (bp)
SERA3 ag1	PKH041230_1_F1	**GTTCCGCGTGGATCC***GAGGGAACGGCGCAATCTGGA*	36
SERA3 ag1	PKH041230_1_R1	**GTCACGATGAATTCC**TTA*AGCTTTCTGGGGGGCGACTTG*	39
SERA3 ag2	PKH041230_2_F1	**GTTCCGCGTGGATCC***AACACTGTTGAGGGTCAAGAC*	36
SERA3 ag2	PKH041230_2_R1	**GTCACGATGAATTCC**TTA*GACCATTTTTGCCTTTTGCAA*	39
SSP2	PKH121770F1	**CCAAAATCGGATCTGGTTCCGCGTGGATCC***GCGGCACAGTACCCAGAAAAT*	51
SSP2	PKH121770R1	**GCAGATCGTCAGTCAGTCACGATGAATTCC**TTA*CTCCTCAGATGAGAAGGGACC*	54
TSERA2 ag1	PKH041240F1	**GTTCCGCGTGGATCC***CGATGCACGGATACTGCTATA*	36
TSERA2 ag1	PKH041240R1	**ATGAATTCCCGGGGA**TCA*TGGGGTTGGTGTCTGATCTTC*	39

Supplementary Table 4: Single-nucleotide polymorphism frequencies of Malaysian clinical isolates sequences within *P. knowlesi* candidate genes.

				cluster 1	cluster 2	cluster 3	Type		
Gene ID	**Candidate name**	**Chromosome**	**SNPs**	**Lab adapted**	**Pig-tailed**	**Long-tailed**	**Non-synonymous**	**Synonymous**	**Stop-Gained**
PKNH_0413400	Full gene		**146**	**55**	**75**	**126**	**58**	**85**	**0**
	SERA3 ag1	4	5	2	1	4	4	1	0
	SERA3 ag2	4	3	1	1	3	1	2	0
PKNH_1265400	Full gene		**71**	**23**	**12**	**57**	**46**	**24**	**1**
	SSP2	12	17	6	4	14	12	5	0
PKNH_0413500	Full gene		**122**	**56**	**48**	**107**	**69**	**52**	**0**
	TSERA2 ag1	4	39	21	16	36	17	22	0

13 Annex 2 Questionnaires and survey tools

13.1 GPS Tracking

13.1.1 GPS Participant Information

Examples of participant information given on use and function of GPS in addition to written consent

Mengenai GPS

Apa itu GPS?

- GPS adalah singkatan kepada **Global Positioning System** atau dalam Bahasa Melayu, Sistem Penentududuk Sejagat.
- Ia adalah sejenis peranti kecil yang membantu mengenalpasti posisi anda melalui penggunaan isyarat daripada satelit
- Apabila maklumat dalam peranti ini digunakan dengan satu program khas, anda akan melihat tanda-tanda kecil pada peta yang menunjukkan lokasi di mana alat ini telah berada.

Kenapa kita perlu mengunakan GPS?

- Oleh kerana terdapatnya penyakit malaria di Kudat, Pulau Banggi dan Matunggong, segala maklumat yang dikumpul daripada unit GPS akan membantu kita mencari kawasan di mana penyebaran penyakit malaria kepada manusia boleh berlaku.
- Kami ingin tahu ke manakah anda pergi supaya kami boleh melawat tempat/tapak yang sama nanti.

Adakah GPS memberi kesan terhadap kesihatan saya?

- Tidak, unit GPS tidak akan memberi sebarang masalah kesihatan.
- Unit GPS juga tidak akan menganggu atau menyusahkan anda
- Penggunaan unit GPS ini juga ibarat menggunakan sebuah telefon bimbit.

Bagaimana unit GPS boleh membantu saya?

- Unit GPS boleh membantu kita mencegah malaria di Kudat, Pulau Banggi dan Matunggong. Maklumat yang diperoleh membantu kita mempromosikan kesejahteraan masyarakat
- Selepas selesainya penyertaan anda, anda akan mendapat sagu-hati khas daripada kami sebagai tanda terima kasih.

Siapakah yang tahu akan kedudukan saya?

- Hanya pihak penyelidik yang terlibat dalam kajian ini.
- Maklumat ini tidak akan dikongsikan kepada sesiapa pun.

Jika seseorang itu mempunyai komputer atau program, bolehkah mereka melihat kedudukan saya?

- Tiada sesiapa pun yang akan mengetahui kedudukan anda **KECUALI** pihak penyelidik kajian ini yang mempunyai akses kepada maklumat yang diperolehi dengan unit GPS ini.

Adakah GPS ini merakam apa-apa sahaja?

- GPS tidak boleh merakam perbualan.
- GPS tidak boleh mengambil gambar seperti kamera.
- GPS tidak boleh menentukan perlakuan seseorang itu di tempat yang dikenalpasti kedudukannya.
- GPS tidak boleh memberitahu di manakah anda berada atau membawa sesiapa pun ke lokasi anda.
- GPS **HANYA** boleh membuat tanda-tanda kecil pada peta yang menunjukkan tempat yang anda lalui.

Apa yang akan berlaku sekiranya alat ini rosak?

- Tidak perlu risau jika unit ini berfungsi atau sebaliknya.
- Sekiranya peranti ini rosak, ia adalah dibawah tanggungan kami.
- Adalah menjadi tanggungjawab anda untuk menggunakannya.

Apa yang akan berlaku sekiranya alat ini dicuri atau hilang?

- Anda tidak perlu membayar unit GPS tersebut, tetapi anda tidak akan menerima sebarang sagu hati.

Apa yang perlu saya buat?

- Anda perlu memakai sepanjang hari, setiap hari **KECUALI** apabila anda berada di rumah. Anda boleh menanggalkannya di rumah.
- Setiap 3-4 hari, seseorang akan membawa sebuah unit yang telah dicas-semula dan mengambil unit yang telah anda pakai.
- Anda boleh memakai unit ini dengan meletakkannya di leher (seperti dalam gambar di bawah) atau membawanya dalam beg.
- Anda tidak perlu menghidupkan atau mematikan unit GPS ini.

13.1.2 Questionnaire

*** Data from this questionnaire will be electronically collected using the Pendragon Forms VI software. Bahasa Melayu translation in red*

1. Study ID

2. Interviewer initials Penemuduga

3. Are the activities in your household any different than normal? Y/N
   Adakah aktiviti anda di rumah mempunyai kelainan dari yang biasa?

    a. How?: Bagaimana?

4. Are your activities affected by any current or recent illness? Y/N
   Adakah aktiviti anda dipengaruhi dengan mana-mana penyakit yang terkini?

5. GPS tracker:

6. Start date: Tarikh mula:

For each visit (every 2-3 days):

1. Date: Tarikh:

2. Was the participant observed using the GPS correctly (worn outside the house/ kept by the front door in the house)? Y/N
   Adakah orang itu menggunakan GPS dengan betul (membawa atau menyimpan di dalam rumah)?

3. Is the GPS working (light flashing)? Y/N Adakah GPS berfungsi (cahaya berkelip)?

   If no:
   - GPS replaced? Adakah GPS digantikan?
     - Replacement S/N:
   - GPS collected and not replaced? Y/N (include final section)
   - Comments on GPS

Describe your activities since the last visit:

Day #: Date (subform)

1. Location visited (subform)

    a. Location:

    b. Date: Tarikh

    c. Start time: Masa mula

d. End time: Masa tamat

e. Purpose: Tujuan

f. Did you wear the GPS? Y/N Adakah anda membawa GPS?

g. For mothers only, was your child/ infant with you during this activity?

2. Did you see any macaques during this day? Y/N (if yes: subform) Adakah anda nampak monyet?

    a. Were you wearing the GPS? Adakah anda membawa GPS?

    b. Did you press the red button when you saw the monkeys? Adakah anda menekan butang merah apabila anda nampak monyet?

    c. Were they wild or pet macaques? Adakah liar atau monyet peliharaan?

        i. Wild macaques Liar
        ii. Pet macaques (this house) Monyet peliharaan (rumah ini)
        iii. Pet macaques (other house) Monyet peliharaan (rumah lain)

    d. Where did you see the macaques? Di mana anda nampak monyet?

        i. Around the house Di kawasan rumah
        ii. Around the kampung Di kawasan kampung
        iii. Around the garden/ farm Di kawasan kebun atau ladang
            1. Type of crop: Jenis tanaman?
        iv. In the forest Dalam hutan
            1. Inside the forest Dalam hutan
            2. At the edge of the forest Sempadan hutan
        v. At the rice paddy (dry) Di sawah padi
        vi. At the rice paddy (wet) Di padi bukit
        vii. On the beach Dekat pantai
        viii. In the bamboo Dekat buluh
        ix. In the mangroves Dekat paya bakau
        x. Other, specify: Lain-lain
        xi. Don't know/ can't remember Tidak tahu

    e. How many macaques did you see? Berapa banyak monyet anda nampak?

        i. Less than 5 1-5
        ii. 6-20
        iii. More than 20 Lebih 20

    f. What were the macaques doing when you saw them? Apa yang monyet lakukan apabila anda nampak?

        i. Sitting/ resting Duduk/ berehat
        ii. Running Lari

iii. Eating Makan
    1. What are they eating? Apa yang mereka makan?
iv. Other Lain-lain

g. What did you do when you saw the macaques? Apa yang anda lakukan apabila anda nampak monyet?

i. Ignore them/ nothing Biarkan sahaja/ tidak ada apa-apa
ii. Chase them away Halau mereka
iii. Set dogs on them Guna anjing
iv. Kill them/ try to kill them Bunuh mereka
    1. How many killed Berapa?
v. Hunt them Buru mereka
    1. How many killed Berapa?
vi. Other (specify) Lain-lain

Malaria prevention practices (to be asked when GPS is collected):

3. Have you used any methods to prevent mosquito bites in the last 2 weeks? Y/N Adakah anda mempunyai kaedah untuk mencegah daripada gigitan nyamuk dalam dua minggu ini?

a. What did you use?: Apa yang anda gunakam?
    i. Mosquito coils Racun nyamuk
    ii. Insect repellent (skin lotion) Pencegah serangga
    iii. Smoke from egg tray (cardboard) Kulit telur
    iv. Smoke from wood Asap dari kayu
    v. Kerosene lamp or candles Lampu pom atau lilin
    vi. Insect spray before sleeping Penyembur serangga sebelum tidur
    vii. Full length clothing Pakaian lengkap
    viii. Mosquito bednet Jaring nyamuk
    ix. Other Lain-lain

b. Where did you use this method? Di mana anda menggunakan kaedah ini?
    i. Inside the house Di dalam rumah
    ii. Around the house or in the village Di kawasan rumah atau di dalam kampung
    iii. At the farm/ garden Di ladang/ kebun
    iv. In the forest Dalam hutan
    v. Other (specify) Lain-lain

c. What type of bed net? Apa jenis kelambu?
    i. Long lasting treated bed net Kelambu dirawat yang tahan lama
    ii. Net requiring treatment every 6 months Kelambu yang perlu dirawat setiap 6 bulan
    iii. Net untreated (bought from local shop) Kelambu yang tidak dirawat contohnya yang dibeli dari kedai runcit
    iv. Don't know Tidak tahu

d. Where do you use this bed net? Di mana anda menggunakan kelambu?
    i. In the house Dalam rumah
    ii. Around the house (outside) Di kawasan rumah (luar)

iii. At the farm Di ladang/ kebun
iv. In the forest Dalam hutan
v. Other (specify) Lain-lain

e. Does the bed net have any tears or holes? Adakah kelambu anda koyak atau mempunyai lubang?
   i. No holes Tiada lubang
   ii. A few small holes Mempunyai lubang kecil
   iii. Lots of tears and holes Banyak koyak dan lubang

f. How long have you used this bednet? Berapa lama anda telah menggunakan kelambu ini?
   i. Less than 6 months Kurang dari 6 bulan
   ii. 6-12 months 6-12 bulan
   iii. 1-3 years 1-3 tahun
   iv. Over 3 years Lebih 3 tahun

13.2 Intensive site survey questionnaire

*** Data from this questionnaire will be electronically collected using the Pendragon Forms VI software. Bahasa Melayu translation in red*

Participant present:

1. Is the individual present in the house? Y/N

2. If no:
    a. House abandoned (no people)
    b. Participant moved (permanent)
    c. Participant travelling outside kampung for study period
    d. Death
    e. Record not correct

Participant information:

1. Have you had a fever in the past 24 hours? Y/N Kau ada demam dalam 24 jam yang lepas?

2. Temperature: Suhu:

3. Have you/ the participant ever been diagnosed by clinic or hospital staff as having malaria? Y/N Kau pernah dibagi tahu klinik atau pekerja hospital yang kau ada malaria?
    a. If yes, how many times? Kalau ya, berapa kali?

4. Have you/ the participant taken anti-malarial medicine in the past 4 weeks? Y/N Kau ada makan ubat anti-malaria dalam 4 minggu yang lepas?
    a. If yes, why? Kalau ya, kenapa?
        i. For malaria treatment Untuk rawatan malaria
        ii. For malaria prevention Supaya tidak kena malaria
        iii. Other reason (specify) Sebab lain

Daily activities: Aktiviti harian:

4. What is your current occupation? Pekerjaan
    a. Farmer - Petani
    b. Rubber tapper – Penoreh getah
    c. Palm oil plantation worker – Pekerja ladang kelapa sawit
    d. Fisherman - nelayan
    e. Housewife – suri rumah
    f. Construction/ contractor - Kontraktor
    g. Shopkeeper - pekedai
    h. Office worker (business/ government) – pekerja pejabat (swasta/kerajaan)
    i. None - Tiada
    j. Retired - Pesara
    k. Student - pelajar
    l. Other

5. Where did you/ the participant mainly go to work or school in the past 4 weeks? Di mana kau biasanya pergi untuk kerja atau sekolah dalam 4 minggu yang lepas?
    a. Within or around the house Dalam atau sekeliling rumah
    b. Within or near home kampung Dalam atau dekat kampung
        i. Specify names of places Nama tempat:
    c. Within or near different kampung in same district, specify Dalam atau dekat kampung lain dalam daerah sama, nama:
    d. Outside of district, within Sabah, specify Di luar daerah, dalam Sabah, nama:
    e. Outside of Sabah, specify Di luar Sabah, nama:
    f. Overseas, specify Luar negara, nama:
    g. No work/ school Tidak kerja atau sekolah

6. When do you usually go to work or school? Bila kau biansanya pergi ke kerja atau sekolah?
    a. Night/ early morning (1am – 6am) Malam/ awal pagi
    b. Morning (7am – 12am) Pagi
    c. Afternoon (1pm – 6pm) Tengah hari/ petang
    d. Evening/ night (7pm – 12am) Malam
    e. Don't know Tidak tahu

7. When do you usually come back from work or school? Bila kau biasanya balik dan kerja atau sekolah?
    a. Night/ early morning (1am – 6am) Malam/ awal pagi
    b. Morning (7am – 12am) Pagi
    c. Afternoon (1pm – 6pm) Tengah hari/ petang
    d. Evening/ night (7pm – 12am) Malam
    e. Don't know Tidak tahu

8. What type of vegetation do you pass on the way to/ from work or school? Tumbuhan jenis apa kau limpas/ lalu bila kau pergi atau pulang dari kerja atau sekolah?
    a. Palm oil trees Pokok kelapa sawit
    b. Banana trees Pokok pisang
    c. Coconut trees Pokok kelapa
    d. Rice paddies- wet Sawah padi- basah
    e. Rice paddies- dry (hill) Sawah padi- kering (bukit)
    f. Rubber trees Pokok getah
    g. Mango trees Pokok mangga
    h. Mangroves Paya bakau
    i. Corn field Padang jagung
    j. Bush forest Hutan semak
    k. Fallow forest
    l. Thick forest
    m. None Tiada
    n. Don't know Tidak tahu

9. In the past four weeks, how have you usually travelled from work or school? Dalam 4 minggu yang lepas, macam mana kau biasanya pergi dan balik ke kerja atau sekolah?
    a. Walk Jalan
        i. Main road Jalan besar
        ii. Path Denai/ jalan hutan

b. Bicycle Basikal
c. Own car Kereta sendiri
d. Friend/ family car Kereta kawan atau keluarga
e. School bus Bas sekolah
f. Public transport (kereta sewa) Pengangkutan awam (kereta sewa)
g. Public transport (kereta sapu) Pengangkutan awam (kereta sewa)
h. Own boat Bot sendiri
i. Rented boat Bot sewa
j. Other Lain-lain

10. How long did it take you to get from home to work or school? HH:MM Berapa lama masa kau ambil untuk pergi dan balik ke kerja atau sekolah?

11. Have you/ the participant done any work in a plantation in the last 4 weeks? Y/N Kau ada buat kerja di ladang/ estet dalam 4 minggu yang lepas?

    a. How many days in the past 4 weeks did you work there? Berapa hari dalam 4 minggu yang lepas kau kerja di sana?

    b. What type of plantation? Apa jenis ladang atau estet?
        i. Rubber Getah
        ii. Oil palm Kelapa sawit
        iii. Fruit Buah-buahan
        iv. Coconut Kelapa
        v. Vegetable Sayur-sayuran
        vi. Other; specify: Lain-lain, jenis:

    c. Name and location of plantation Nama dan tempat ladang/ estet

12. Have you/ the participant gone to the forest in the last 4 weeks? Y/N Kau ada pergi ke hutan dalam 4 minggu yang lepas?

    a. How many days in the past 4 weeks did you go there? Berapa hari kau pergi ke hutan dalam 4 minggu yang lepas?

    b. What time do you usually go there? Bila kau biansanya pergi ke hutan?
        i. Night/ early morning (1am – 6am) Malam/ awal pagi
        ii. Morning (7am – 12am) Pagi
        iii. Afternoon (1pm – 6pm) Tengah hari/ petang
        iv. Evening/ night (7pm – 12am) Malam
        v. Don't know Tidak tahu

    c. What time do you usually come back? Bila kau biasanya balik hutan?
        i. Night/ early morning (1am – 6am) Malam/ awal pagi
        ii. Morning (7am – 12am) Pagi
        iii. Afternoon (1pm – 6pm) Tengah hari/ petang
        iv. Evening/ night (7pm – 12am) Malam
        v. Don't know Tidak tahu

    d. What activities did you do there? Apa aktiviti kau buat di sana?
        i. Hunting Memburu

ii. Fishing Memancing ikan
iii. Collecting wood Mengumpal kayu
iv. Collecting other plants or food Mengumpal tumbuhan lain atau buah-buahan
v. Other Lain-lain

13. What other activities do you usually do? Apa aktiviti lain kau biasanya buat?
    a. Fishing Memancing ikan
    b. Hunting Memburu
    c. Sport, specify: Sukan, jenis:
    d. Other: Lain-lain
    e. None of the above Tiada

14. If yes: Kalau ya
    a. How many days in the past 4 weeks did you do this activity? Berapa hari kau buat aktiviti ini dalam 4 minggu yang lepas?

    b. When did you usually do this activity? Bila kau biasanya buat aktiviti ini?
        i. Night/ early morning (1am – 6am) Malam/ awal pagi
        ii. Morning (7am – 12am) Pagi
        iii. Afternoon (1pm – 6pm) Tengah hari/ petang
        iv. Evening/ night (7pm – 12am) Malam
        v. Don't know Tidak tahu

    c. Where do you usually do this activity? Di mana kau biasanya buat aktiviti ini?
        i. Near house Dekat rumah
        ii. Near ocean Dekat laut
        iii. Near river Dekat sungai
        iv. In forest; specify type Dalam hutan, jenis:
        v. In plantation; specify type Di ladang, jenis:
        vi. Other; specify Lain-lain
        vii. Don't know Tidak tahu

    d. Name and description of place Nama tempat:
15. Where do you usually bathe? Di mana kamu biasanya mandi?
    a. Bathroom Bilik mandi (di dalam atau di luar)
    b. Well Perigi (di luar)
    c. Water pipe Air paip (di luar)
    d. River Sungai
    e. Other, specify Lain-lain, tempat:

16. When do you usually bathe? Bila masa biasanya kamu mandi?
    a. Malam/ awal pagi (12am – 6am)
    b. Pagi (7am – 12pm)
    c. Tengahari/ petang (1pm – 6pm)
    d. Malam (6pm – 12am)

Travel history:

1. Is this the place you have slept most often in the past month? Y/N Adakah tempat ini tempat yang kau tidur paling kerap (sering) dalam bulan yang lepas?

2. Have you/ the participant stayed overnight outside this kampung in the last 4 weeks? Y/N Adakah kau bermalam di luar kampung ini dalam 4 minggu yang lepas?

    a. Where outside the kampung? Di mana di luar kampung ini?
        i. Forest near this kampung, specify Hutan dekat kampung ini, nama
        ii. Plantation near this kampung, specify Ladang/ estet dekat kampung ini, nama:
        iii. Within or near different kampung in same district, specify Dalam atau dekat kampung lain dalam daerah sama, nama:
        iv. Outside of district, within Sabah, specify Di luar daerah, dalam Sabah, nama:
        v. Outside of Sabah, specify Di luar Sabah, nama:

    b. How many nights in the past 4 weeks? (number) Berapa malam dalam 4 minggu yang lepas? (nombor)

    c. Did you sleep outside the walls during this trip? Y/N Ada kau tidur di luar dinding semasa perjalanan ini?
        i. If yes, was it: Kalau ya, ini
            1. In the forest Dalam hutan
            2. In a plantation Dalam ladang
            3. Did you use any mosquito prevention? Kau ada guna apa-apa supaya tidak digigit nyamuk?
                a. Mosquito net Kelambu
                b. Mosquito coils Ubat halau nyamuk
                c. Other Lain-lain

Malaria prevention practices:

17. Have you used any methods to prevent mosquito bites in the last 2 weeks? Y/N Kau ada guna cara supaya tidak digigit nyamuk dalam 2 minggu yang lepas?

    a. What did you use?: Apa kau guna?
        i. Mosquito coils Ubat halau nyamuk
        ii. Insect repellent (skin lotion) Ubat halau serangga (losyen kulit)
        iii. Smoke from egg tray (cardboard) Asap dari kadbod (simpan telur)
        iv. Smoke from wood Asap dari kayu
        v. Kerosene lamp or candles Lampu kerosin atau lilin
        vi. Insect spray before sleeping Ubat sembur serangga sebelum tidur
        vii. Full length clothing Pakaian panjang
        viii. Mosquito bednet Kelambu
        ix. Other Lain-lain

    b. Where did you use this method? Di mana kau guna cara ini?
        i. Inside the house Dalam rumah
        ii. Around the house or in the village Sekeliling rumah atau dalam kampung
        iii. At the farm/ garden Di ladang atau kebun
        iv. In the forest Dalam hutan
        v. Other (specify) Lain-lain, nama:

13. What do you do when you get a fever?Apa anda buat apabila anda demam?

I. Nothing. Tiada apa-apa
II. Use traditional medicine. Menggunakan ubat-ubatan tradisional
III. Use western medicine. Menggunakan ubat-ubatan dari klinik.
IV. Consult traditional healer. Mendapatkan bantuan pawang
V. Go to western doctor. Berjumpa dengan doctor swasta
VI. Go to government / hospital doctor. Berjumpa dengan doctor hospital / kerajaan
VII. Other. Lain-lain

Monkey and mosquito interactions:

1. When do you usually get bitten by mosquitoes? Bila biasa digigit dari nyamuk?
    a. Awal pagi/ malam (12am – 6am)
    b. Pagi (7am – 12pm)
    c. Tengahari/ petang (1pm – 6pm)
    d. Malam (6pm – 12am)

2. How often are you bitten by mosquitoes at night (6pm – 6am)? Berapa hari kamu digigit dari nyamuk?
    a. Daily Tiap-tiap hari
    b. Weekly Setiap minggu
    c. Several times a month Berapa kali sebulan
    d. Several times a year Berapa kali setahun
    e. Other Lain-lain

3. How many times do you usually get bitten by mosquitoes at night? Berapa kali kamu digigit dari nyamuk?
    a. Never
    b. 1
    c. 2 – 5
    d. 6 – 20
    e. > 20

4. Do you ever see monkeys? Kau pernah nampak monyet?

5. Where do you usually see monkeys? Di mana kau biasanya nampak monyet?
    a. Around the house Di sekitar rumah (sekeliling)
    b. Around the garden Di sekitar kebun
    c. Around the village Di sekitar kampung
    d. In the forest Dalam hutan
    e. In the plantation Dalam ladang
    f. In the paddy fields Di sawah padi
    g. In the hill rice Di padi bukit
    h. Near the beach Dekat pantai
    i. Near the mangroves Dekat paya bakau
    j. Other Lain-lain

k. Tidak sedar kehadiran monyet Tidak tahu kalau ada monyet

6. How frequently do you see monkeys? **–** Berapa kerap kau nampak monyet?
    a. Every day – Tiap-tiap hari
    b. Weekly – Minggu- minggu
    c. Several times a month – Beberapa kali dalam satu bulan
    d. Several times a year – Beberapa kali dalam satu tahun
    e. Other (example during harvest, etc) – lain-lain (macam semasa menuai)

7. What time of day do you usually see monkeys? – Bila masa kau biasanya nampak monyet?

    a. Night/ early morning (1am – 6am) Malam/ awal pagi
    b. Morning (7am – 12am) Pagi
    c. Afternoon (1pm – 6pm) Tengah hari/ petang
    d. Evening/ night (7pm – 12am) Malam
    e. Don't know Tidak tahu

Household questionnaire (to be answered by head of household):

A. Migration/ Migrasi

1. How long has this family lived here? Berapa lama sudah keluarga ini tinggal di sini?
    a. Less than 6 months Kurang 6 bulan
    b. 6 months to 1 year 6 bulan hingga 1 tahun
    c. 1-5 years 1 hingga 5 tahun
    d. 5-10 years 5 hingga 10 tahun
    e. Over 10 years Lebih 10 tahun

2. How old is this house (when was this house built)? Berapa lama sudah rumah ini (bila rumah ini dibina)?
    a. Less than 6 months Kurang 6 bulan
    b. 6 months to 1 year 6 bulan hingga 1 tahun
    c. 1-5 years 1-5 tahun
    d. 5-10 years 5-10 tahun
    e. Over 10 years Lebih 10 tahun

3. If the family has lived other places in the last 10 years, where else have they lived? Kalau keluarga tinggal di tempat lain dalam 10 tahun yang lepas, di mana mereka tinggal?
    a. Location Tempat
        i. Other location within this kampung Tempat lain dalam kampung ini
        ii. Other kampung within this district Kampung lain dalam daerah ini
        iii. Outside district, within Sabah Di luar daerah, dalam Sabah
        iv. Outside Sabah Di luar Sabah
    b. Start date Tarikh mula
    c. End date Tarikh habis

B. Farming Berkebun

1. Does this household do any farming activities? Y/N Keluarga ini buat aktiviti berkebun?

2. If yes, what do you farm? Kalau ya, apa kau tanam?
    a. Oil palm Kelapa sawit
    b. Coconut Kelapa
    c. Rubber Getah
    d. Rice Padi
        i. Wet rice padi basah
        ii. Hill rice padi bukit
    e. Corn Jagung
    f. Vegetables/ garden Sayur-sayuran/ kebun
    g. Fruit Buah-buahan
    h. Livestock, specify: Haiwan/ binatang ternakan, jenis:
    i. Other, specify: Lain-lain, jenis:

3. Is the crop sold or consumed by the household? Tanaman atau haiwan ternakan dijual atau dimakan oleh keluarga?
    a. Entirely consumed by the household Semuanya dimakan oleh keluarga
    b. Some sold at market, some consumed by household Ada yang dijual di pasar, ada yang dimakan oleh keluarga

c. All sold at market Semuanya dijual di pasar

4. What amount of land is currently cultivated for each crop? Berapa banyak tanah yang diguna untuk setiap tanaman atau haiwan ternakan?
    a. Crop/ animal: Tanaman/ haiwan ternakan
    b. Amount of land Berapa banyak tanah (acres?)

5. What amount of land was cultivated for each crop/ animal 5 years ago? Berapa banyak tanah yang diguna untuk setiap tanaman atau haiwan ternakan 5 tahun yang lepas?

6. What amount of land was cultivated for each crop/ animal 10 years ago? Berapa banyak tanah yang diguna untuk setiap tanaman atau haiwan ternakan 10 tahun yang lepas?

7. Who owns the land that is currently farmed? Siapa miliki tanah yang deguna untuk berkebun sekarang?
    a. Communal land Tanah awam
    b. Owned by self or family Dimiliki sendiri atau keluarga
    c. Owned by other individual, company or government Dimiliki orang, syarikat atau kerajaan lain
    d. Other: Lain-lain

8. How far is the farm from the house? (majority of land) Berapa jauh kebun dari rumah? (kebanyakan tanah)
    a. Around the house (within 200m of the house) Sekeliling rumah (dalam 200m dari rumah)
    b. Within the kampung Dalam kampung
    c. Outside kampung Di luar kampung

9. Are any fertilisers used? Y/N Kau ada guna baja?

10. Are any pesticides used? Y/N Kau ada guna racun serangga?

11. Are any tractors or other mechanical equipment used on the farm? Y/N specify: Kau ada guna traktor atau alat mekanikal lain di kebun?

12. Are any additional workers hired to help on the farm? Y/N Ada pekerja lebih yang diambil kerja di kebun?

13. Is swidden farming used (burning the land to clear)? Y/N Kau ada bakar tanah untuk berkebun?
    a. Length of fallow period: # years Masa tanah kosong: berapa tahun?

C. Forest use: Kegunaan hutan

14. Does the household regularly collect wood from the forest? Y/N Keluarga kerap kumpul kayu dari hutan?
    a. If yes, purpose of collection: Kalau ya, kenapa?
        i. Fuel/ cooking Bahan api/ masak
        ii. Timber/ construction Kayu balak/ binaan
        iii. Other, specify: Lain-lain, apa?

15. Does the household consume any food from the forest? Y/N Keluarga ada makan makanan dari hutan?
    a. Hunting Haiwan atau ikan liar
    b. Forest fruits Buah-buahan hutan
    c. Honey Madu
    d. Other, specify: Lain-lain, apa?

16. Does the household collect any medicinal plants from the forest? Y/N Keluarga ada kumpul tumbuhan ubatan dari hutan?

17. Types of forest used: Jenis hutan yang diguna

18. Who owns the land of the forests used? Siapa miliki tanah hutan yang diguna?

    a. Government Kerajaan
    b. Community Awam
    c. Individual Individu
    d. No one Tiada orang

13.3 Cross-sectional survey questionnaire

Participant present:

1. Is the individual present in the house during the study period? Y/N

2. Return house: Y/N

3. Consent form:

4. Consent blood:

5. If no:
    a. Participant moved (permanent)
    b. Participant travelling outside kampung for study period
    c. Travelling for work or school (RETURN)
    d. Death

Participant information:

5. Name: (M)

6. Gender: M/F (M)

7. Date of birth (M)
    a. If unknown, age:

8. IC Nombor:
    a. N/A

9. Ethnicity (M)
    a. Rungus
    b. Dusun
    c. Kadazan
    d. Bajau
    e. Kegayan
    f. Ubian
    g. Sungoi
    h. Chinese **- Cina**
    i. Malay - **Melayu**
    j. Filipino
    k. Indonesian
    l. Lain-lain

Medical history:

10. For women, are you pregnant? Y/N (Hide if Lelaki)
    a. Number of weeks

11. Have you had a fever in the past 24 hours? Y/N Kau ada demam dalam 24 jam yang lepas?

12. Have you/ the participant ever been diagnosed by clinic or hospital staff as having malaria? Y/N Kau pernah dibagi tahu klinik atau pekerja hospital yang kau ada malaria?

13. Have you/ the participant taken anti-malarial medicine in the past 4 weeks? Y/N Kau ada makan ubat anti-malaria dalam 1 bulan yang lepas?
    a. If yes, why? Kalau ya, kenapa?
        i. For malaria treatment Untuk rawatan malaria
        ii. For malaria prevention Supaya tidak kena malaria
        iii. Other reason (specify) Sebab lain

14. The last time you had a fever, what did you do? Kali terakhir kamu demam, apa kamu buat?
    a. Nothing
    b. Use traditional medicine. Menggunakan ubat-ubatan tradisional
    c. Pharmacy/ use western medicine. Menggunakan ubat-ubatan dari klinik.
    d. Consult traditional healer. Mendapatkan bantuan pawing
    e. Private doctor.
    f. Go to government / hospital doctor. Berjumpa dengan doctor hospital / kerajaan

    g. What was the total cost for this consultation: RM
    h. Total transport time: hours

Daily activities: Aktiviti harian:

18. What is your current occupation? Pekerjaan
    a. Farmer – Petani (multi-selection)

        i. Oil palm Kelapa sawit
        ii. Coconut Kelapa
        iii. Rubber Getah
        iv. Rice Padi
            1. Wet rice padi basah
            2. Hill rice padi bukit
        v. Corn Jagung
        vi. Vegetables/ garden Sayur-sayuran/ kebun
        vii. Fruit Buah-buahan
        viii. Livestock, specify: Haiwan/ binatang ternakan, jenis:
        ix. Other, specify: Lain-lain, jenis:

    b. Rubber tapper – Pekerja ladang getah
    c. Palm oil plantation worker – Pekerja ladang kelapa sawit
    d. Fisherman - Nelayan
    e. Housewife – Suri rumah
    f. Construction/ contractor - Kontraktor
    g. Shopkeeper - Pekedai
    h. Office worker (business/ government/school) – Pekerja pejabat (swasta/kerajaan/sekolah)

i. None/ Retired – Tiada/ pesara
j. Student - Pelajar
k. Driver/ transport - Pemandu
l. Other

19. Where did you/ the participant mainly go to work or school in the past 4 weeks? Di mana kamu biasanya pergi untuk kerja atau sekolah dalam 1 bulan yang lepas?
    a. Within or around the house Dalam atau sekeliling rumah
    b. Within or near home kampung Dalam atau dekat kampung
        i. Specify names of places Nama tempat:
    c. Within or near different kampung in same district, specify Dalam atau dekat kampung lain dalam daerah sama, nama:
    d. Outside of district, within Sabah, specify Di luar daerah, dalam Sabah, nama:
    e. Outside of Sabah, specify Di luar Sabah, nama:
    f. Overseas, specify Luar negara, nama:
    g. No work/ school Tidak kerja atau sekolah

20. When do you usually go to work or school? Bila kau biansanya pergi ke kerja atau sekolah?
    a. Night/ early morning (1am – 6am) Malam/ awal pagi
    b. Morning (7am – 12am) Pagi
    c. Afternoon (1pm – 6pm) Tengah hari/ petang
    d. Evening/ night (7pm – 12am) Malam
    e. Don't know Tidak tahu

21. When do you usually come back from work or school? Bila kau biasanya balik dan kerja atau sekolah?
    a. Night/ early morning (1am – 6am) Malam/ awal pagi
    b. Morning (7am – 12am) Pagi
    c. Afternoon (1pm – 6pm) Tengah hari/ petang
    d. Evening/ night (7pm – 12am) Malam
    e. Don't know Tidak tahu

22. In the past four weeks, how have you usually travelled from work or school? Dalam 4 minggu yang lepas, macam mana kau biasanya pergi dan balik ke kerja atau sekolah?
    a. Walk Jalan kaki
        i. Main road Jalan besar
        ii. Path Denai/ jalan hutan
    b. Bicycle/ motorbike Basikal
    c. Car/ bus
    d. Boat

23. How long did it take you to get from home to work or school? HH:MM Berapa lama masa kau ambil untuk pergi dan balik ke kerja atau sekolah?

24. Have you/ the participant done any work in a plantation in the last 4 weeks? Y/N Kau ada buat kerja di ladang/ estet dalam 4 minggu yang lepas?

    a. How many days in the past 4 weeks did you work there? Berapa hari dalam 4 minggu yang lepas kau kerja di sana?

b. What type of plantation? Apa jenis ladang atau estet?
    i. Rubber Getah
    ii. Oil palm Kelapa sawit
    iii. Fruit Buah-buahan
    iv. Coconut Kelapa
    v. Vegetable Sayur-sayuran
    vi. Other; specify: Lain-lain, jenis:

25. Have you/ the participant gone to the forest in the last 4 weeks? Y/N Kau ada pergi ke hutan dalam 4 minggu yang lepas?

    a. How many days in the past 4 weeks did you go there? Berapa hari kau pergi ke hutan dalam 4 minggu yang lepas?

    b. What time do you usually go there? Bila kau biansanya pergi ke hutan?
        i. Night/ early morning (1am – 6am) Malam/ awal pagi
        ii. Morning (7am – 12am) Pagi
        iii. Afternoon (1pm – 6pm) Tengah hari/ petang
        iv. Evening/ night (7pm – 12am) Malam
        v. Don't know Tidak tahu

    c. What time do you usually come back? Bila kau biasanya balik hutan?
        i. Night/ early morning (1am – 6am) Malam/ awal pagi
        ii. Morning (7am – 12am) Pagi
        iii. Afternoon (1pm – 6pm) Tengah hari/ petang
        iv. Evening/ night (7pm – 12am) Malam
        v. Don't know Tidak tahu

    d. What activities did you do there? Apa aktiviti kau buat di sana?
        i. Hunting Memburu
        ii. Fishing Memancing ikan
        iii. Collecting wood Mengumpal kayu
        iv. Collecting other plants or food Mengumpal tumbuhan lain atau buah-buahan
        v. Other Lain-lain

26. In the past year, have you worked clearing land for farming or building? Dalam tahun lepas, kamu bersih tanah atau bina?

27. Have you been involved in any construction activities? Y/N
    a. Building houses Bina rumah
    b. Roadworks Bina jalan
    c. Waterworks, including dams or water pipes Projek air
    d. Other Lain-lain

28. What other activities do you usually do? Apa aktiviti lain kau biasanya buat?
    a. Fishing Memancing ikan
    b. Hunting Memburu
    c. Sport, specify: Sukan, jenis:
    d. Visiting neighbour: Melawat jiran

i. Outside- Di luar
ii. Inside house – Di dalam rumah

e. Other: Lain-lain
f. None of the above Tiada

29. If yes: Kalau ya
    a. How many days in the past 4 weeks did you do this activity? Berapa hari kau buat aktiviti ini dalam 4 minggu yang lepas?

    b. When did you usually do this activity? Bila kau biasanya buat aktiviti ini?
        i. Night/ early morning (1am – 6am) Malam/ awal pagi
        ii. Morning (7am – 12am) Pagi
        iii. Afternoon (1pm – 6pm) Tengah hari/ petang
        iv. Evening/ night (7pm – 12am) Malam
        v. Don't know Tidak tahu

30. Where do you usually bathe? Di mana kamu biasanya mandi?
    a. Bathroom Bilik mandi (di dalam atau di luar)
    b. Well Perigi (di luar)
    c. Water pipe Air paip (di luar)
    d. River Sungai
    e. Other, specify Lain-lain, tempat:

31. When do you usually bathe? Bila masa biasanya kamu mandi?
    a. Malam/ awal pagi (12am – 6am)
    b. Pagi (7am – 12pm)
    c. Tengahari/ petang (1pm – 6pm)
    d. Malam (6pm – 12am)

32. When do you usually go to bed? Bila masa biasanya kamu tidur?
    a. Around sunset (6-7pm)
    b. Within 2 hours of sunset (8-9pm)
    c. Between 3-4 hours after sunset (10-11pm)
    d. More than 4 hours after sunset (after 11pm)
    e. Other, specify
    f. Don't know

33. How much time do you usually spend outside around the house at night time (6pm – 6am) Berapa lama kamu biasanya di luar rumah di petang dan malam (jam)?
    a. Less than 1 hour
    b. 1-3 hours
    c. Over 3 hours
    d. More than 3 hours
    e. Don't know

Travel history:

3. Is this the place you have slept most often in the past month? Y/N Adakah tempat ini tempat yang kau tidur paling kerap (sering) dalam bulan yang lepas?

4. Have you/ the participant stayed overnight outside this kampung in the last 4 weeks? Y/N Adakah kau bermalam di luar kampung ini dalam 4 minggu yang lepas?

    a. Where outside the kampung? Di mana di luar kampung ini?
        i. Forest near this kampung, specify Hutan dekat kampung ini, nama
        ii. Plantation near this kampung, specify Ladang/ estet dekat kampung ini, nama:
        iii. Within or near different kampung in same district, specify Dalam atau dekat kampung lain dalam daerah sama, nama:
        iv. Outside of district, within Sabah, specify Di luar daerah, dalam Sabah, nama:
        v. Outside of Sabah, specify Di luar Sabah, nama:

    b. How many nights in the past 4 weeks? (number) Berapa malam dalam 4 minggu yang lepas? (nombor)

    c. Did you sleep outside the walls during this trip? Y/N Ada kau tidur di luar dinding semasa perjalanan ini?
        i. If yes, was it: Kalau ya, ini
            1. In the forest Dalam hutan
            2. In a plantation Dalam ladang

Malaria prevention practices:

34. Have you used any methods to prevent mosquito bites in the last 4 weeks? Y/N Kau ada guna cara supaya tidak digigit nyamuk dalam 1 bulan yang lepas?

    a. What did you use?: Apa kau guna?
        i. Mosquito coils Ubat halau nyamuk
        ii. Insect repellent (skin lotion) Ubat halau serangga (losyen kulit)
        iii. Smoke from egg tray (cardboard) Asap dari kadbod (simpan telur)
        iv. Smoke from wood Asap dari kayu
        v. Kerosene lamp or candles Lampu kerosin atau lilin
        vi. Insect spray before sleeping Ubat sembur serangga sebelum tidur
        vii. Full length clothing Pakaian panjang
        viii. Mosquito bednet Kelambu
            1. Long-lasting insecticide treated net
            2. Net treated every 6 months
            3. Untreated net
            4. Don't know
        ix. Window net Jaring tingkap
        x. Fan- Kipas angin
        xi. Other Lain-lain

35. When do you usually get bitten by mosquitoes?
    a. Awal pagi/ malam (12am – 6am)
    b. Pagi (7am – 12pm)
    c. Tengahari/ petang (1pm – 6pm)
    d. Malam (6pm – 12am)

e. Tidak tahu

Interactions with monkeys, mosquitoes and other animals:

36. Do you ever see monkeys? Kau pernah nampak monyet?

37. What type of monkeys do you see? Apa jenis monyet nampak?

38. Where do you usually see monkeys? Di mana kau biasanya nampak monyet?
    a. Around the house Di sekitar rumah (sekeliling)
    b. Around the garden Di sekitar kebun
    c. Around the village Di sekitar kampung
    d. In the forest Dalam hutan
    e. At the farm/ plantation Dalam ladang
    f. In the rice paddy Di sawah padi
    g. In hill rice Di padi bukit
    h. At the beach Dekat pantai
    i. At the mangroves Dekat paya bakau
    j. Lain-lain
    k. Don't know Tidak tahu kalau ada monyet

39. How frequently do you see monkeys? – Berapa kerap kau nampak monyet?
    a. Every day – Tiap-tiap hari
    b. Weekly – Minggu- minggu
    c. Several times a month – Beberapa kali dalam satu bulan
    d. Several times a year – Beberapa kali dalam satu tahun
    e. Other (example during harvest, etc) – lain-lain (macam semasa menuai)

40. Do you ever come into contact with: Kamu pernah jumpa:
    a. Rats or mice Tikus
    b. Bats Kelawar
    c. Wild dogs Anjing liar
    d. Ticks Kutu

Household questionnaire (to be answered by head of household):

A. Migration/ Migrasi

4. How long has this family lived here? Berapa lama sudah keluarga ini tinggal di sini?
    a. Less than 6 months Kurang 6 bulan
    b. 6 months to 1 year 6 bulan hingga 1 tahun
    c. 1-5 years 1 hingga 5 tahun
    d. 5-10 years 5 hingga 10 tahun
    e. Over 10 years Lebih 10 tahun

5. How old is this house (when was this house built)? Berapa lama sudah rumah ini (bila rumah ini dibina)?
    a. Less than 6 months Kurang 6 bulan
    b. 6 months to 1 year 6 bulan hingga 1 tahun
    c. 1-5 years 1-5 tahun
    d. 5-10 years 5-10 tahun
    e. Over 10 years Lebih 10 tahun

6. If the family has lived other places in the last 10 years, where else have they lived? Kalau keluarga tinggal di tempat lain dalam 10 tahun yang lepas, di mana mereka tinggal?
    a. Location Tempat
        i. Other location within this kampung Tempat lain dalam kampung ini
        ii. Other kampung within this district Kampung lain dalam daerah ini
        iii. Outside district, within Sabah Di luar daerah, dalam Sabah
        iv. Outside Sabah Di luar Sabah
            1. Tempat:

Assets:

1. Education level of the head of household
    a. None Tiada
    b. Completed primary school Sekolah rendah
    c. Completed secondary school Sekolah tinggi
    d. Completed tertiary Universiti

2. Who owns the place where you stay? Permilik tempat kediaman ini
    a. Self or family Sendiri/ keluarga
    b. Other individual Milik individu lain
    c. Government Kerajaan/ Badan Berkanun
    d. Company Swasta
    e. Other Lain-lain

For the following questions, were these observed or reported?

3. House has electricity Rumah disambung dengan letrik (YES/NO)

4. Refrigerator Peti ais (YES/NO)

5. Car Kereta (YES/NO)

6. Motorcycle Motorsikal (YES/NO)

7. TV Televisyen (YES/NO)

8. Generator Alat generator letrik (YES/NO

9. Roof type Jenis atap?

    a. Zing (corrugated iron)
    b. Aluminum
    c. Simen (cement/ concrete)
    d. Kayu (wooden)
    e. Spandeks
    f. Lain-lain

10. Are there open eaves between the roof and walls?

11. What type of walls does the house have? Apakah jenis dinding terdapat di rumah anda?
    a. Wood/ bamboo Kayu/ bambu
    b. Simen (cement/ concrete)
    c. Bata (brick)
    d. Lain-lain

12. What type of floor does the house have? Apakah jenis lantai terdapat di rumah anda?
    a. Papan (wood)
    b. Tiles
    c. Simen (cement/ concrete)
    d. Tanah liat (earth)
    e. Bambu (bamboo)
    f. Lain-lain

13. Are there slats or openings in the floor? Y/N

14. How high is the house from the ground?
    a. On ground level
    b. Raised concrete (less than 1 m)
    c. Stilts (less than 1 m)
    d. Stilts (more than 1m)
    e. Over water

15. Are there openings or gaps in the bedrooms (over 10cm x 10cm)?

16. Number of windows that close in the house
    a. None
    b. 1 or more
    c. All of them
    d. No windows

17. Are there insect screens fitted to the windows?
    a. In the bedroom
    b. In living room
    c. In kitchen/ cooking area

d. Other

18. Is there a bed net in the house? Ada kelambu dalam rumah?
    a. If yes, what condition is the bed net in?
        i. Good (no holes)
        ii. Moderate
        iii. Poor
    b. Have you ever treated the net with an insecticide to kill or repel mosquitoes? Pernah guna racun serangga untuk kelambu?
        i. If yes, when was the net treated? Bila?
            1. Less than 1 year ago
            2. More than 1 year ago
            3. Don't know
    c. Has the net ever been used for other purpose (ex. Fishing)?

19. What do you use to cook? Jenis bahan api untuk memasak?
    a. Gas
    b. Kerosene
    c. Wood Kayu api
    d. Charcoal Arang
    e. Eletrik
    f. Lain-lain

20. Where is most of the cooking done? Di mana masak?
    a. Inside the house Dalam rumah
    b. Outside Di luar rumah
    c. Other Lain-lain

21. What type of toilets does the house have? Jenis tandas?
    a. Tandas tarik (flush)
    b. Tandas curah/ siram
    c. Pit Lubang
    d. Place outside Ruang tertutup di permukaan air
    e. None Tiada

22. Where are the toilets located? Di mana tandas?
    a. Inside house Dalam rumah
    b. Around the house Dekat rumah
    c. Far from house Juah dari rumah

23. Where do you get drinking water? Kemudahan bekalan air minum
    a. Pipe inside the house Paip dalam rumah
    b. Pipe outside house Paip luar rumah
    c. Well Perigi
    d. River Sungai
    e. Rain - Hujan
    f. Other Lain-lain

24. How is household rubbish disposed of? Apa kamu buat dengan sampah dari rumah?
    a. Burnt or buried daily Bakar atau bumi setiap hari

b. Kept by house for few days before disposal (bury or burn) Simpang dekat rumah untuk beberapa hari sebelum bakar atau bumi
c. In central location in kampung Hantar ke tempat dalam kampung
d. Throw outside house or in water Buang dekat rumah atau dalam laut/ sungai

25. Is the household rubbish kept in a covered container before disposal? Sampah dari rumah disimpang dalam tempat ditutup?
    a. Yes
    b. No
    c. Don't know

26. Number of animals Bilangan haiwan
    a. Cattle - Lembu
    b. Buffalo - Kerbau
    c. Goats/ sheep – Kambing/biri-biri
    d. Pigs - Babi
    e. Chickens/ ducks/ poultry Ayam/itik
    f. Cats Kucing
    g. Dogs Anjing

27. Does anyone in this house have a pet monkey now? Ada orang dalam rumah ini ada monyet pelihara?

28. Do any people in this kampung currently have a pet monkey? Y/N/ Don't know Ada orang dalam kampung ini ada monyet pelihara?

29. Do the monkeys eat the crops or food around the house or farm? Y/N/Don't know Monyet biasanya makan tanaman atau makanan dekat rumah atau kebun?

B. Farming Berkebun

19. What water sources are around the house (within 5 minutes)? (direct observation)
    a. Stream or river
    b. Stagnant water (pond, puddle)
    c. Well or gravity tank
    d. Plastic container or basin
    e. Lake
    f. Ocean
    g. Other

20. Are uncovered containers containing water observed around the house? (direct observation)
    a. Yes
    b. No
    c. Don't know

21. Does this household do any farming activities? Y/N Keluarga ini buat aktiviti berkebun?

22. If yes, what do you farm? Kalau ya, apa kau tanam?
    j. Oil palm Kelapa sawit
    k. Coconut Kelapa

l. Rubber Getah
m. Rice Padi
    i. Wet rice padi basah
    ii. Hill rice padi bukit
n. Corn Jagung
o. Vegetables/ garden Sayur-sayuran/ kebun
p. Fruit Buah-buahan
q. Livestock, specify: Haiwan/ binatang ternakan, jenis:
r. Other, specify: Lain-lain, jenis:

23. How much land does the family use now? Berapa banyak tanah yang diguna? (ekar) – nombor
    a. Less than 1 ekar
    b. 1-5 ekar
    c. 5-10 ekar
    d. More than 10 ekar

24. What amount of land was cultivated 5 years ago? Berapa banyak tanah yang diguna 5 tahun yang lepas? (ekar)- nombor
    a. Less than 1 ekar
    b. 1-5 ekar
    c. 5-10 ekar
    d. More than 10 ekar

25. What amount of land was cultivated 10 years ago? Berapa banyak tanah yang diguna 10 tahun yang lepas? (ekar)- nombor
    a. Less than 1 ekar
    b. 1-5 ekar
    c. 5-10 ekar
    d. More than 10 ekar

26. How far is the farm from the house? (majority of land) Berapa jauh kebun dari rumah? (kebanyakan tanah)
    a. Around the house (within 200m of the house) Sekeliling rumah (dalam 200m dari rumah)
    b. Within the kampung Dalam kampung
    c. Outside kampung Di luar kampung

27. Are any pesticides used? Y/N Kau ada guna racun serangga?

28. Is swidden farming used (burning the land to clear)? Y/N Kau ada bakar tanah untuk berkebun?
    b. Length of fallow period: # years Masa tanah kosong: berapa tahun?

Household Forest Use:

29. Does the household regularly collect wood from the forest? Y/N Keluarga kerap kumpul kayu dari hutan?
    b. If yes, purpose of collection: Kalau ya, kenapa?
        i. Fuel/ cooking Bahan api/ masak

ii. Timber/ construction Kayu balak/ binaan
iii. Other, specify: Lain-lain, apa?

30. Does the household consume any food from the forest? Y/N Keluarga ada makan makanan dari hutan?
    e. Hunting Haiwan atau ikan liar
    f. Forest fruits Buah-buahan hutan
    g. Honey Madu
    h. Other, specify: Lain-lain, apa?

31. Does the household collect any medicinal plants from the forest? Y/N Keluarga ada kumpul tumbuhan ubatan dari hutan?