The burden of *Plasmodium vivax* malaria

A thesis submitted for the degree of *Doctor of Philosophy*

Katherine Elizabeth Battle

Green Templeton College, University of Oxford

Trinity 2015

The burden of *Plasmodium vivax* malaria

A thesis submitted for the degree of *Doctor of Philosophy*

Katherine Elizabeth Battle

Green Templeton College, University of Oxford

Trinity 2015

Abstract

Plasmodium vivax is the most geographically widespread of the human malarias and is capable of causing severe debilitating disease. The parasite's unique biology poses challenges to control of the disease and the understanding of its epidemiology. It is less researched and well understood than the more deadly *P. falciparum*. In this thesis, spatial relapse patterns and models of endemicity and clinical disease were applied to generate robust estimates of the *P. vivax* burden to address a key knowledge gap in malaria epidemiology.

First, a review of the distribution of the parasite, its vectors and populations at risk found nearly one third of the global population living at risk, and more potential vectors than *P. falciparum*. In spite of low observed endemicity, the public health impact of *P. vivax* is likely to have been seriously underestimated in the past.

To accurately define the burden of *P. vivax* it was necessary to improve understanding of one of the parasite's most unique and challenging aspects, its ability to relapse. A meta-analysis of individual records of relapse showed that relapse periodicity varied systematically by geographic region and could be categorized by nine global regions. The nine regions were applied to a model to quantify the relationship between prevalence of infection and incidence of clinical disease. As relapse would have an influence on both measures, separate relationships were drawn for each relapse zone.

The prevalence-incidence model was used to translate maps of predicted endemicity into measures of clinical burden. The evidence-base of *P. vivax* prevalence was poor in some regions and therefore a burden estimate based on surveillance reports was also derived. Reported cases must be adjusted for parameters such as under-reporting and treatment-seeking behaviours. A model used to fill gaps in treatment-seeking data available from national household surveys was developed to allow burden to be estimated using both cartographic modelling and surveillance reporting methods. To improve fidelity, the results of the two approaches were combined to enumerate *P. vivax* burden globally.

The results and conclusions of these studies are discussed with recommendations for how these findings influence our understanding of *P. vivax* epidemiology and implications for future control and elimination efforts.

Statement of Contribution and Associated Publications

The chapters presented in this thesis (excluding Chapters 1 & 7) have been published, submitted or are in preparation for submission to peer-review journals. These chapters represent collaborative efforts including members of the Malaria Atlas Project (MAP) as well as colleagues from other institutions. I have summarised my personal contributions to each chapter here.

Chapter 2: The global public health significance of *Plasmodium vivax*.

This chapter has been published and has been included here in its final form. I was responsible for all aspects of this work. I conceived the scope of the review (with SIH and PWG), synthesized the relevant data, produced all figures and wrote the first draft of the manuscript. The final manuscript was edited and approved by all authors.

Battle, K.E., Gething, P.W., Elyazar, I.R., Moyes, C.L., Sinka, M.E., Howes, R.E., Guerra, C.A., Price, R.N., Baird, K.J. and Hay, S.I. (2012). The global public health significance of *Plasmodium vivax*. *Advances in Parasitology* **80**, 1-111.

Chapter 3: Geographical variation in *Plasmodium vivax* relapse.

This chapter is included in its final published form. I oversaw all aspects of this work: conceived the study (with SIH and JKB), assembled the data, helped implement the modelling (with MSK and SB), performed all geographic analysis (with PWG and TVB), and wrote the first draft of the manuscript. The final manuscript was edited and approved by all authors.

Battle, K.E., Karhunen, M.S., Bhatt, S., Gething, P.W., Howes, R.E., Golding, N., Van Boeckel, T., Messina, J.P., Shanks, G.D., Smith, D.L., Baird, J.K. and Hay, S.I. (2014). Geographical variation in *Plasmodium vivax* relapse. *Malaria Journal* **13**, 144.

Chapter 4: Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease.

This chapter has been published and is included in its final form. The associated data descriptor publication is included in the Appendix. I was responsible for overseeing all aspects of this work. I performed the literature search and data abstraction (with CAG and KAD), assisted with the design and implementation of the modelling framework and analysis (with EC and NG), and wrote the first draft of both manuscripts. All authors edited and approved the final version of the research article and data descriptor.

Battle, K.E., Cameron, E., Guerra, C.A., Golding, N., Duda, K.A., Howes, R.E., Elyazar, I.R.F., Baird, J.K., Reiner Jr, R.C., Smith, D.L., Gething, P.W. and Hay, S.I. (2015). Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease. *Malaria Journal* **14**, 191.

Battle, K.E., Guerra, C.A., Golding, N., Duda, K.A., Cameron, E., Howes, R.E., Elyazar, I.R.F., Baird, J.K., Reiner Jr, R.C., Smith, D.L., Gething, P.W. and Hay, S.I. (2015). Global database of *Plasmodium falciparum* and *P. vivax* incidence records, 1985-2013. *Scientific Data* **2**, 150012.

Chapter 5: Modelling national treatment-seeking rates to improve interpretation of malaria case reporting.

This chapter has been submitted for publication. I was responsible for all aspects of this work. I conceived the study (with PWG), assembled all data (with HSG), performed all modelling and analyses (with DB) and wrote the first draft of the paper. All authors edited and approved the final manuscript.

Battle, K.E., Bisanzio, D., Gibson, H.S., Bhatt, S., Cameron, E., Weiss, D.J., Mappin, B., Dalrymple, U., Howes, R.E., Hay, S.I. and Gething, P.W. (2015). Modelling national treatment-seeking rates to improve interpretation of malaria case reporting. *Malaria Journal* **submitted**; MS ID: MALJ-S-15-00880.

Chapter 6: The global burden of *Plasmodium vivax* malaria in 2013

This chapter has been prepared for submission to a peer-review journal. I was responsible for all aspects of this work: conceived the study (with PWG), assembled all new input data (with contributions from REC), helped with the modelling analyses (with SB and EC and input from DLS), generated all maps, and wrote the first draft of the manuscript. The manuscript has been edited and approved by BS, EC, SIH and PWG.

Battle, K.E., Bhatt, S., Cameron, E., Smith, D.L., Cibulskis, R.E., Hay, S.I. and Gething, P.W. (2015). The global burden of *Plasmodium vivax* malaria in 2013. **In preparation**.

* * *

The data, methods and outputs of some of these chapters have gone on to be used by other projects which I have contributed to, but was not chiefly responsible for. These are listed and summarised below.

Related manuscript for Chapter 3:

White, M.T., Karl, S., Battle, K.E., Hay, S.I., Mueller, I. and Ghani, A.C. (2014). Modelling the contribution of the hypnozoite reservoir to *Plasmodium vivax* transmission. *Elife* **3**.

Related manuscripts for Chapter 4:

Cameron, E., Battle, K.E., Bhatt, S., Weiss, D.J., Bisanzio, D., Mappin, B., Dalrymple, U., Hay, S.I., Smith, D.L., Griffin, J.T., Wenger, E.A., Eckhoff, P.A., Smith, T.A., Penny, M.A. and Gething, P.W. (2015). An emulator ensemble for modelling the relationship between *Plasmodium falciparum* parasite prevalence and clinical incidence. *Nature Communications* **6**, 8170.

Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K.E., Moyes, C.L., Henry, A., Eckhoff, P.A., Wenger, E.A., Briet, O., Penny, M.A., Smith, T.A., Bennett, A., Yukich, J., Eisele, T.P., Griffin, J.T., Fergus, C.A., Lynch, M.,

Lindgren, F., Murray, C.J., Smith, D.L., Hay, S.I., Cibulskis, R.E. and Gething, P.W. (2015). The impact of malaria control on *Plasmodium falciparum* in Africa, 2000–2015. *Nature* **Accepted**; MS ID: 2015-06-08556B.

Related manuscripts for Chapters 2-4:

Howes, R.E., Battle, K.E., Golding, N. and Hay, S.I. (2014). Epidemiology. *Thematic review for the Writing Committee for the development of the Global Strategic Plan on* Plasmodium vivax *control and elimination.*

Howes, R.E.*, Battle, K.E*., Mendis, K.N., Smith, D.L., Cibulskis, R.E., Baird, J.K. and Hay, S.I. (2015). The epidemiology of *Plasmodium vivax*. *The American Journal of Tropical Medicine and Hygiene* **In preparation**. *These authors share first authorship for this work.

Related manuscript for Chapter 6:

Howes, R.E., Reiner Jr, R.C., Battle, K.E., Longbottom, J., Mappin, B., Ordanovich, D., Tatem, A.J., Drakeley, C., Gething, P.W., Zimmerman, P.A., Smith, D.L. and Hay, S.I. (2015). *Plasmodium vivax* transmission in Africa. *PLoS Neglected Tropical Diseases* **Under review**.

* * *

Finally, I have been involved in various projects, the publications from which are listed below, that were not directly related to my thesis, but were conducted in parallel to the aforementioned work:

Hay, S.I., Battle, K.E., Pigott, D.M., Smith, D.L., Moyes, C.L., Bhatt, S., Brownstein, J.S., Collier, N., Myers, M.F., George, D.B. and Gething, P.W. (2013). Global mapping of infectious disease. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **368**, 20120250.

Howes, R.E., Battle, K.E., Satyagraha, A.W., Baird, J.K. and Hay, S.I. (2013). G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Advances in Parasitology* **81**, 133-201.

Howes, R.E., Dewi, M., Piel, F.B., Monteiro, W.M., Battle, K.E., Messina, J.P., Sakuntabhai, A., Satyagraha, A.W., Williams, T.N., Baird, J.K. and Hay, S.I. (2013). Spatial distribution of G6PD deficiency variants across malaria-endemic regions. *Malaria Journal* **12**, 418.

Durham, J., Battle, K., Rickart, K. and Shanks, G.D. (2014). Geographical origin of post-landmine injury malaria infections. *Disaster Medicine and Public Health Preparedness* **8**, 417-421.

Messina, J.P., Brady, O.J., Scott, T.W., Zou, C., Pigott, D.M., Duda, K.A., Bhatt, S., Katzelnick, L., Howes, R.E., Battle, K.E., Simmons, C.P. and Hay, S.I. (2014). Global spread of dengue virus types: mapping the 70 year history. *Trends in Microbiology* **22**, 138-146.

Pigott, D.M., Bhatt, S., Golding, N., Duda, K.A., Battle, K.E., Brady, O.J., Messina, J.P., Balard, Y., Bastien, P., Pratlong, F., Brownstein, J.S., Freifeld, C.C., Mekaru, S.R., Gething, P.W., George, D.B., Myers, M.F., Reithinger, R. and Hay, S.I. (2014). Global distribution maps of the leishmaniases. *Elife* **3**.

Pigott, D.M., Golding, N., Messina, J.P., Battle, K.E., Duda, K.A., Balard, Y., Bastien, P., Pratlong, F., Brownstein, J.S., Freifeld, C.C., Mekaru, S.R., Madoff, L.C., George, D.B., Myers, M.F. and Hay, S.I. (2014). Global database of leishmaniasis occurrence locations, 1960-2012. *Scientific Data* **1**, 140036.

Pigott, D.M.*, Howes, R.E.*, Wiebe, A., Battle, K.E., Golding, N., Gething, P.W., Dowell, S.F., Farag, T.H., Garcia, A.J., Kimball, A.M., Krause, L.K., Smith, C.H., Brooker, S.J., Kyu, H.H., Vos, T., Murray, C.J., Moyes, C.L. and Hay, S.I. (2015). Prioritising infectious disease mapping. *PLoS Neglected Tropical Diseases* **9**, e0003756. *These authors share first authorship for this work.

* * *

As her supervisors, I certify that the statements of contribution listed here are a fair representation of Katherine Battle's work:

Prof. Simon I. Hay	Prof. Peter W. Gething	Dr. Andrew W. Farlow
Date: 31/08/2015	30/08/2015	31/08/2015

Table of Contents

Figures

Tables

Acknowledgments

The path to the production of this thesis has been an inspiring, challenging and exciting one. I'm extremely grateful to Simon Hay for responding to an email I sent him out of the blue from the highlands of Western Kenya and for all the opportunities that came after, as part of the Wellcome Trust Senior Research Fellowship and the many other exciting projects you lead. Thank you also to Pete Gething for taking me and all my *P. vivax* problems on in addition to your other responsibilities. Andrew Farlow, I regret that we didn't get to work together more and explore our proposed questions, but I'm grateful for your support.

I must also acknowledge the mentorship that I have received from abroad. Kevin Baird, Ric Price and Iqbal Elyazar have offered invaluable insight throughout this project. Kevin, from our first foray into the relapse work you inspired a genuine passion for *P. vivax* in me and I thank you for your mentorship. Thank you also to Dave Smith for your guidance in nearly all the analyses conducted here.

Being a part of the Malaria Atlas Project has benefited me with a dynamic interdisciplinary group of lovely people. Sam and Ewan, you have been integral to this work. The statistical wizardry and patience of yourselves, Donal and Nick is greatly appreciated. Dan, Harry, Zhi and Andy, I am also thankful for the GIS platform that you provide the group, which I greatly benefitted from. My wonderful office mates (old and new) – David, Oli, Bonnie, Ursula and Janey (by extension) – and the rest of the DPhil gang – Mortiz and Freya – you are brilliant friends and colleagues. I especially have to thank Ros, my *P. vivax* partner in crime. You are an inspiring colleague and great friend. Thank you for all of your input.

Of course none of this would have been possible without the support from my friends and family, who put up with my insatiable appetite for education and living far away. I'm grateful to have had support of my parents throughout this endeavour – mom even got stuck with some proofreading on more than one occasion. Most of all I must thank my lovely husband Tim. You take great care of me and kept me laughing (and fed!) even though the more stressful times of this DPhil. Thank you for all that you do for us.

Abbreviations

ACT	Artemisinin-based combination therapy
95% CI	95% confidence interval
95% CrI	95% credible interval
ACD	Active case detection
ADMIN (Admin)	Administrative unit
Africa+	Africa, Yemen and Saudi Arabia
AFRO	African Regional Office (WHO)
AFRO-C	African Regional Office (WHO) - Central Africa
AFRO-E	African Regional Office (WHO) - East Africa and high-transmission areas in Southern Africa
AFRO-S	African Regional Office (WHO) - low-transmission Southern African countries
AFRO-W	African Regional Office (WHO) - West Africa
AIC	Akaike information criterion
AICc	Corrected Akaike information criterion
API	Annual parasite incidence (in cases per 1000 people per year)
ARI	Acute respiratory infection
Asia	Mainland Asia excluding the Malaysian Peninsula
Asia Pacific	The Malaysian Peninsula and the islands of Asia and the Pacific

BIC	Bayesian information criterion
CAR	Central Africa Republic
CSE Asia	Central and South East Asia
DHS	Demographic and Health Survey
DPRK (Korea DPR)	Democratic People's Republic of Korea
DVS	Dominant vector species or species complexes
ELISA	Enzyme-linked immunosorbent assay
EMRO	Eastern Mediterranean Regional Office (WHO)
EURO	European Regional Office (WHO)
EVI	Enhanced vegetation index
G6PD	Glucose-6-phosphate dehydrogenase
GAMM	Generalized additive mixed models
GDP	Gross domestic product
GMRF	Gaussian Markov random field
GNI	Gross national income
GPW	Gridded population of the world
GRUMP	Global Rural-Urban Mapping Project
IGBP	International Geosphere-Biosphere Programme
INLA	Integrated nested Laplace approximation

IQR	Inter-quartile range
JAGS	Just Another Gibbs Sampler
Lao PDR	Lao People's Democratic Republic
LGM	Latent Gaussian model
LST	Land surface temperature
MAP	Malaria Atlas Project
MBG	Model-based geostatistics
MCMC	Markov chain Monte Carlo
MIS	Malaria Indicator Survey
ML	Maximum likelihood
MOD1	Model 1 (Chapter 5)
MOD2	Model 2 (Chapter 5)
MoH	Ministry of Health
NMCP	National malaria control programme
PAHO	Pan-American Health Organization (WHO Region of the Americas)
PAR	Populations at risk
PCD	Passive case detection
PCR	Polymerase chain reaction
PNG	Papua New Guinea

*Pv*API	*P. vivax* annual parasite incidence
*Pv*MEC	*P. vivax* malaria endemic country
*Pv*PR	*P. vivax* parasite rate
$PvPR_{1-99}$	*P. vivax* parasite rate in one-99 year olds
RC	Reporting completeness
RDT	Rapid diagnostic test
REML	Restricted maximum likelihood
RMSE	Root mean square error
s.l.	*sensu lato*
s.s.	*sensu stricto*
SEARO	South-East Asia Regional Office (WHO)
SPR	Slide positivity rate
SVR	*P. vivax* slide positivity rate
TAG	Technical advisory group
WHO	World Health Organization
WMR	World Malaria Report
WPRO	Western Pacific Regional Office (WHO)
XSS	Cross-sectional survey

Chapter 1 – Introduction

1.1. The human malarias

Malaria is the most important parasitic disease of mankind with a long history of human suffering, genetic and societal change, and innovation both on the part of the parasite and the human interventions devised to control its impact. It is a parasitic protozoan disease of the genus *Plasmodium* that is transmitted by the *Anopheles* mosquito. There are six species of *Plasmodium* that infect humans: *P. falciparum* and *P. vivax*, which are attributable for nearly all mortality and morbidity [1], as well as *P. malariae*, *P. ovale* (sympatric species *P. ovale curtisi* and *P. ovale wallikeri* [2]) and *P. knowlesi*. *Plasmodium knowlesi* is a primate species restricted to Southeast Asia that was only relatively recently shown to be endemic in human populations [3-5], and natural transmission of two other simian species to humans has been observed in the past few years: *P. brasilianum* [6] and the close relative of *P. vivax*, *P. cynomolgi* [7].

Each of the human malarias bears unique biological traits making the species more or less vulnerable to public health control interventions. Frequently considered the most challenging to control [8], *P. vivax* presents lifecycle characteristics able to overcome the main components of the malaria control toolkit. Despite this, the species has been relatively understudied. The epidemiological idiosyncrasies of *P. vivax*, and the pathogen's resulting global public health burden, represent the focus of this thesis.

1.2. *Plasmodium vivax* origins

The origin of *P. vivax* as a human malaria begins with transmission in primates. *Plasmodium vivax* is the predominant malaria species outside of Africa and it was long

thought that its phylogenetic roots were in Asia. The lineage that *P. vivax* and *P. cynomolgi* share is with other *Plasmodium* species that infect primates only found in Asia [9-12]. Recent sequence analyses of *Plasmodium* samples gathered from great apes show that *P. vivax* instead originated in Africa [13]. Approximately 30,000 years ago, ancestral *P. vivax* likely infected humans, chimpanzees and gorillas across Africa, until the Duffy negative mutation began to spread [14]. *Plasmodium vivax* appears to be dependent on the Duffy antigen being expressed on the surface of red blood cells for invasion (step 5, Figure 1.1) [15]. As the mutation that causes the antigen not to be expressed spread, the African population essentially became immune to *P. vivax*. This is one of the many ways that malaria has shaped the human genome [16]. Other heritable traits such as sickle cell disease [17, 18], haemoglobins C [19] and E, Southeast Asian ovalocytosis [20], alpha and beta thalassemia [21] and glucose-6-phosphate dehydrogenase (G6PD) deficiency [22] share geographic distributions that are roughly similar to that of malaria prior to the era of control campaigns which started around the turn of the twentieth century [23, 24].

Population genetic analyses of human samples show that *P. vivax*, after being driven from Africa, now thrives in Asia and the Americas with genetically diverse parasite populations [25]. *Plasmodium vivax* has proven to be highly adaptable. Human migration brought the parasite to the Americas [26] and back to Africa [13]. Its ability to form a dormant liver stage allowed the geographic range of the parasite to extend well beyond the tropics into temperate ranges as far north as Russia and Finland [27]: the most geographically extensive distribution of the human malarias.

1.3. Biology and epidemiology

The dormant liver stage hypnozoites can generate relapses of blood-stage infection weeks, months, or even years after the primary infection [28]. While the biological mechanism that triggers the relapses to occur remains unknown [29], it has long been understood that there are distinct geographic patterns in relapse phenotypes [28], with tropical strains relapsing quickly following the primary infection from the mosquito bite, and temperate strains relapsing after several months or more. All *Plasmodium* have a liver-stage, as seen in phase A in Figure 1.1, but only *P. vivax* and *P. ovale* form the long-term dormant stages that are capable of evading routine drug treatment and which therefore pose important challenges for control.

The liver stages cannot be detected by common diagnostic tools, such as rapid diagnostic tests (RDTs) and microscopy, or by high sensitivity laboratory-based molecular assays [30]. Detection of *P. vivax* is further complicated by low parasite densities in blood-stage infection. Vivax parasites invade young red blood cells (reticulocytes), which are less common than mature erythrocytes, leading to low parasitaemia infections and false negative diagnoses [31]. *Plasmodium vivax* gametocytes are more efficiently transmitted to mosquitoes at low densities than falciparum gametocytes [32], and infectious gametocytes also appear earlier in the course of infection than in *P. falciparum* infections, making transmission-blocking treatment less effective for *P. vivax* as transmission may already have occurred at the time of the onset of clinical symptoms and therefore treatment-seeking [33].

Finally, parasite development within the mosquito (the sporogonic cycle; Stage C, Figure 1.1) occurs more rapidly for *P. vivax* at equivalent temperatures, which also contributes to its wider geographic range [34-36].

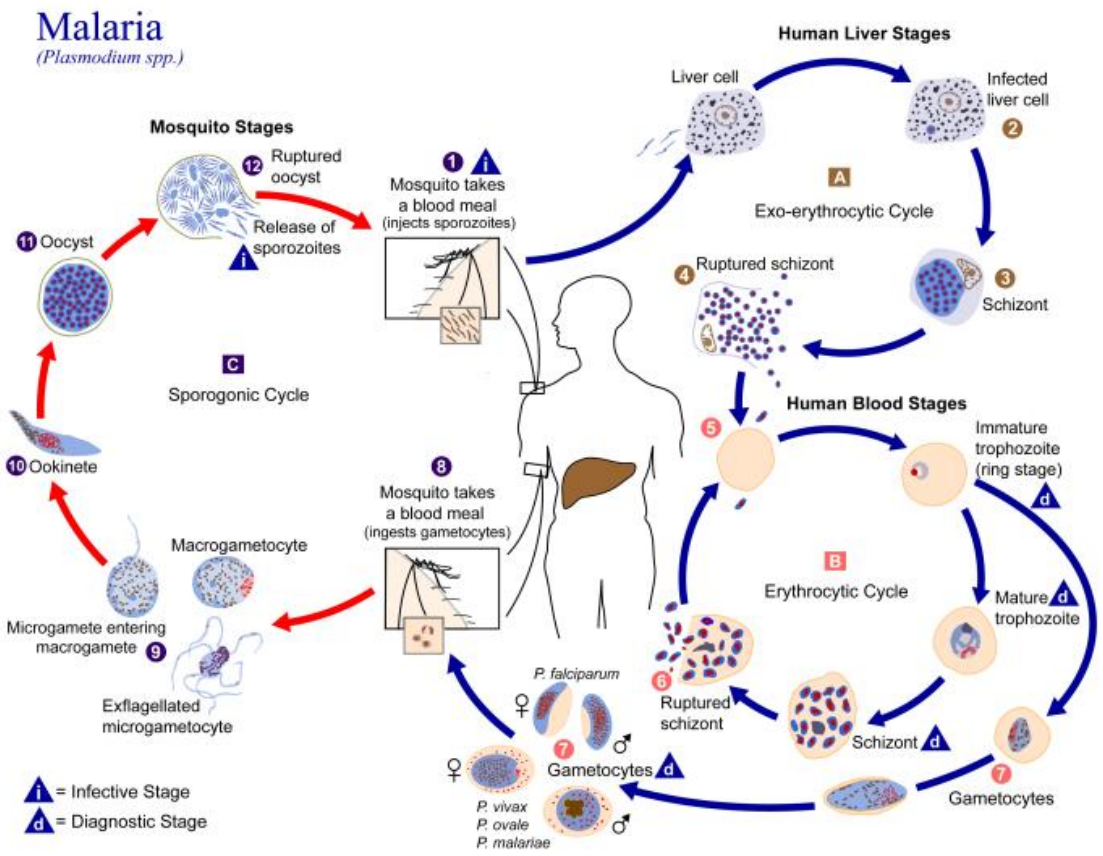


Figure 1.1. Life cycle of *Plasmodium spp.* Source: Centers for Disease Control and Prevention, CC license.

1.4. The public health importance of *Plasmodium vivax*

While elimination goals ostensibly address all human malarias, the allocation of funding and resources between *P. falciparum* and *P. vivax* has been highly disparate [37], in part because of the dogma that *P. vivax* is less clinically significant than its cousin, *P. falciparum*. However, *P. vivax* has been shown to be the most widespread of the malaria parasites and to cause severe disease and death [38].

The ability of *P. vivax* to cause relapses of infection makes vivax malaria more difficult to control than *P. falciparum* [32, 39]. The only class of drug that can treat the liver stage, the 8-aminoquinolines, is contraindicated for individuals with a G6PD deficiency and for pregnant women and infants because it may cause haemolytic anaemia [40-42].

Primaquine, is the only 8-aminoquinoline currently licensed and has a 14-day treatment schedule which challenges adherence [43, 44]; though, a single-dose alternative drug tafenoquine is currently in trials [45]. Further issues arise where resistance to the gold standard blood-stage therapy, chloroquine, is increasing [46]. These biological features often make *P. vivax* the final challenge in elimination efforts [47-50]. The map below (Figure 1.2) indicates that all countries mono-endemic with exclusively *P. vivax* (outlined in red on the map) are in elimination phases (hatched overlay).

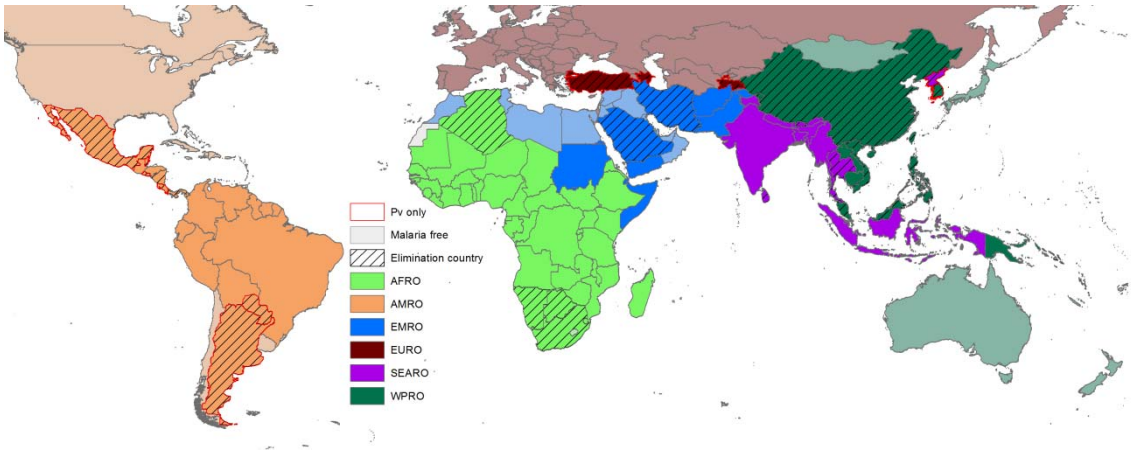


Figure 1.2. *Plasmodium vivax* endemic and eliminating countries. Countries are shaded by their WHO region and those countries outlined in red are endemic only with *P. vivax* and those hatched with black lines are listed as eliminating countries [49].

Malaria is one of the leading causes of death in low- and middle-income countries and a primary cause of death in children aged 1-4 years [51]. Such impact estimates are not stratified by parasite species, and recent initiatives calling for elimination and eradication [52-54] similarly refer to malaria as a single disease [55] and are often steered towards *P. falciparum*. The prioritization of *P. falciparum* and the relative neglect of *P. vivax* malaria [31, 50, 56, 57] stems from *P. falciparum* being historically known as ‘malignant tertian malaria’, and *P. vivax* as ‘benign tertian malaria’ because of the tendency for *P. falciparum* infections to be more severe and deadly than *P. vivax*

[58]. An increasing body of evidence has shown, however, that *P. vivax* should no longer be thought of as a benign and rarely fatal disease [57, 59-63], but instead that *P. vivax* is capable of causing severe disease and fatality, particularly in children [63-66]. The extent to which the public health burden of *P. vivax* has been underestimated is unknown.

1.5. Mapping *Plasmodium vivax* and public health planning

Maps are the principal output in spatial epidemiology because they allow for the visualization of the extent and foci of public health problems [67-70]. Evidence-based maps of malaria have proven to be useful tools for decision makers assessing prospective courses of action to control the disease in different parts of the world [35, 71, 72]. Knowledge of the geographic limits and burden of infection are essential for control and elimination planning [73, 74]. Maps of clinical and economic burden serve as benchmarks from which organizations can set control and elimination targets and measure success [52, 53]. Modern methods used to provide mapped estimates are based on data of known origin and fidelity, and display associated uncertainty [75, 76]. This is considered an improvement on surveillance-based approaches which are affected by diagnostic and reporting biases that are often difficult to quantify. Cartographic evidence of malaria endemicity and burden have already been accepted by the malaria research community and successfully applied to national level and regional control priority settings [35, 71, 72, 77-79].

The limits of *P. vivax* transmission, and the parasite's endemicity within those limits, have been mapped [35]. A review of the predicted levels of endemicity and the associated uncertainty highlights regions where more intense control measures and improved monitoring need to be applied. Dominant anopheline vector species of

'malaria' have been described [80-83], but it is not clear if all named dominant vector species are capable of transmitting vivax malaria as well as falciparum, and how the distributions of those species correspond to the map of predicted *P. vivax* prevalence.

Estimates of parasite endemicity provide predictions of prevalence in all endemic areas, whether prevalence surveys have been conducted in those locations or not. Endemicity is a measure of disease intensity based on prevalence. Prevalence, also commonly referred to as parasite rate (PR) in the context of malaria, is the most commonly available malaria metric [84]. Prevalence refers to the proportion of the population found to be infected with *P. vivax* (without necessarily having clinical symptoms) at a given time through a cross-sectional survey. Incidence, on the other hand, measures the number of people who acquire a new clinical infection in a specified time period. These data are collected through longitudinal studies. Clinical incidence is obtained by only testing those patients with suspected symptoms of disease, such as fever [72]. Work on *P. falciparum* serves as a proof of concept that the relationship between the common metric of prevalence and variably-reported clinical incidence measures can be described through modelling [72, 85].

1.6. Thesis aims

Plasmodium vivax has until very recently been considered a neglected disease. The tide has begun to turn with mounting evidence of its severity and wide distribution – over one third of the world is predicted to live at risk of infection. In order to address global elimination goals, the current knowledge gaps in the understanding of the transmission and burden of *P. vivax* must be addressed [31, 56, 86]. The objective of this thesis is to present up-to-date geographic descriptions of *P. vivax* transmission, endemicity as well as a global estimate of the burden of *P. vivax* disease with quantifiable uncertainty.

Chapter 2 will provide a review of baseline regional descriptions of the populations at risk of *P. vivax* infection, its endemicity and an audit of *Anopheles* species incriminated as dominant *P. vivax* vectors, along with a description of their distributions and bionomics by region.

Chapter 3 will assess the hypothesis that periodicity of *P. vivax* relapse differs geographically to synchronise blood-stage parasitaemia re-emergence with local seasonal mosquito abundance, and to ensure clearance of previous blood-stage infection. Meta-analysis of observed relapse patterns in thousands of individuals within various geographic classifications characterise the global patterns of variation and timing of relapse within different transmission settings.

The relapse zones identified in Chapter 3 will be applied to the analysis in Chapter 4, which will describe the relationship between prevalence of infection and incidence of clinical disease. An assembled database of spatially and temporally matched prevalence and incidence measures fit a model that is stratified by relapse zone that will allow measures and maps of prevalence to be translated into incidence.

There are presently two primary methods to estimate malaria burden, the surveillance-based approach, which adjusts cases reported by country control programmes and ministries of health [87], and the cartographic approach, which models cases from estimates of prevalence. The aim in this thesis is to apply both methods to *P. vivax* malaria and determine which method best characterizes the burden in each endemic country. In order to generate surveillance-based estimates, a number of parameters are needed. Data on treatment-seeking behaviours cannot be derived from country reports and must instead be derived from large-scale household surveys. Chapter 5 assembles available survey data on treatment-seeking and also develops and applies a model to fill

gaps for countries lacking data by using social and economic indicator variables. In Chapter 6, burden is estimated using both the surveillance and cartographic methods. To achieve this, reports of *P. vivax* incidence and relative adjustment parameters are assembled, the limits of transmission and predicted endemicity of *P. vivax* is updated, and incidence of clinical disease is modelled. A framework is applied to assign each country a burden estimate from one of the approaches in order to generate a hybridized global burden estimate with associated uncertainty.

This thesis is presented as a collection of publications. Chapter 2 through 4 are included in their published form and Chapter 5 and 6 are manuscripts that have been submitted and are in preparation respectively. Additional information from both the published and unpublished papers is provided in the Appendix. These are co-authored publications and author contributions are provided at the end of each chapter.

1.7. References

1. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW: **The global distribution and population at risk of malaria: past, present, and future**. *Lancet Infect Dis* 2004, **4**(6):327-336.

2. Sutherland CJ, Tanomsing N, Nolder D, Oguike M, Jennison C, Pukrittayakamee S *et al,*: **Two nonrecombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally**. *J Infect Dis* 2010, **201**(10):1544-1550.

3. 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**. *Clin Infect Dis* 2008, **46**(2):165-171.

4. 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):1017-1024.

5. White NJ: ***Plasmodium knowlesi*: the fifth human malaria parasite**. *Clin Infect Dis* 2008, **46**(2):172-173.

6. Lalremruata A, Magris M, Vivas-Martínez S, Koehler M, Esen M, Kempaiah P *et al,*: **Natural infection of *Plasmodium brasilianum* in humans: Man and monkey share quartan malaria parasites in the Venezuelan Amazon**. *EBioMedicine* 2015, **In press**.

7. Ta TH, Hisam S, Lanza M, Jiram AI, Ismail N, Rubio JM: **First case of a naturally acquired human infection with *Plasmodium cynomolgi***. *Malar J* 2014, **13**:68.

8. Tanner M, Greenwood B, Whitty CJ, Ansah EK, Price RN, Dondorp AM *et al,*: **Malaria eradication and elimination: views on how to translate a vision into reality**. *BMC Med* 2015, **13**:167.

9. Carlton JM, Das A, Escalante AA: **Genomics, population genetics and evolutionary history of *Plasmodium vivax***. *Adv Parasitol* 2013, **81**:203-222.

10. Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE *et al,*: **A monkey's tale: The origin of *Plasmodium vivax* as a human malaria parasite**. *Proc Natl Acad Sci U S A* 2005, **102**(6):1980-1985.

11. Mu J, Joy DA, Duan J, Huang Y, Carlton J, Walker J *et al,*: **Host switch leads to emergence of *Plasmodium vivax* malaria in humans**. *Mol Biol Evol* 2005, **22**(8):1686-1693.

12. Tachibana SI, Sullivan SA, Kawai S, Nakamura S, Kim HR, Goto N *et al,*: ***Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade**. *Nat Genet* 2012, **44**(9):1051-+.

13. Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, Malenke JA *et al,*: **African origin of the malaria parasite *Plasmodium vivax***. *Nat Commun* 2014, **5**:3346.

14. Hamblin MT, Di Rienzo A: **Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus**. *Am J Hum Genet* 2000, **66**(5):1669-1679.

15. Miller LH, Mason SJ, Clyde DF, McGinniss MH: **The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy**. *N Engl J Med* 1976, **295**(6):302-304.

16. Kwiatkowski DP: **How malaria has affected the human genome and what human genetics can teach us about malaria**. *Am J Hum Genet* 2005, **77**(2):171-192.

17. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Dewi M *et al,*: **Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates**. *Lancet* 2012:doi:10.1016/S0140-6736(1012)61229-X.

18. Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Williams TN *et al,*: **Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis**. *Nat Commun* 2010, **1**:104.

19. Piel FB, Howes RE, Patil AP, Nyangiri OA, Gething PW, Bhatt S *et al,*: **The distribution of haemoglobin C and its prevalence in newborns in Africa**. *Sci Rep* 2013, **3**:1671.

20. Paquette AM, Harahap A, Laosombat V, Patnode JM, Satyagraha A, Sudoyo H *et al,*: **The evolutionary origins of Southeast Asian Ovalocytosis**. *Infect Genet Evol* 2015, **34**:153-159.

21. Piel FB, Weatherall DJ: **The alpha-Thalassemias**. *N Engl J Med* 2014, **371**(20):1908-1916.

22. Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, Dewi M *et al,*: **G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map**. *PLoS Med* 2012, **9**(11):e1001339.

23. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM: **Malaria**. *Lancet* 2013.

24. Carter R, Mendis KN: **Evolutionary and historical aspects of the burden of malaria**. *Clin Microbiol Rev* 2002, **15**(4):564-594.

25. Koepfli C, Rodrigues PT, Antao T, Orjuela-Sanchez P, Van den Eede P, Gamboa D *et al,*: ***Plasmodium vivax* diversity and population structure across four continents**. *PLoS Negl Trop Dis* 2015, **9**(6):e0003872.

26. Taylor JE, Pacheco MA, Bacon DJ, Beg MA, Machado RL, Fairhurst RM *et al,*: **The evolutionary history of *Plasmodium vivax* as inferred from**

mitochondrial genomes: parasite genetic diversity in the Americas. *Mol Biol Evol* 2013, **30**(9):2050-2064.

27. Hulden L, Heliovaara K: **Endemic malaria: an 'indoor' disease in northern Europe. Historical data analysed**. *Malar J* 2005, **4**(1):19.

28. White NJ: **Determinants of relapse periodicity in *Plasmodium vivax* malaria**. *Malar J* 2011, **10**:297.

29. Shanks GD, White NJ: **The activation of vivax malaria hypnozoites by infectious diseases**. *Lancet Infect Dis* 2013, **13**(10):900-906.

30. The malERA Consultative Group on Diagnoses and Diagnostics: **A Research Agenda for Malaria Eradication: Diagnoses and Diagnostics**. *PLoS Med* 2011, **8**(1):e1000396.

31. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL *et al,*: **Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite**. *Lancet Infect Dis* 2009, **9**(9):555-566.

32. Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J, Cui L: ***Plasmodium vivax* transmission: chances for control?** *Trends Parasitol* 2004, **20**(4):192-198.

33. Douglas NM, John GK, von Seidlein L, Anstey NM, Price RN: **Chemotherapeutic strategies for reducing transmission of *Plasmodium vivax* malaria**. *Adv Parasitol* 2012, **80**:271-300.

34. Detinova TS: **Age-grouping methods in Diptera of medical importance, with special reference to some vectors of malaria**. In. Geneva: World Health Organization; 1962.

35. Gething PW, Elyazar IRF, Moyes CM, Smith DL, Battle KE, Guerra CA *et al,*: **A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010** *PLoS Negl Trop Dis* 2012, **6**(9):e1814.

36. Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW *et al,*: **Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax***. *Parasit Vectors* 2011, **4**(1):92.

37. PATH: **Staying the Course? Malaria Research and Development in a Time of Economic Uncertainty**. In. Seattle, U.S.A.: PATH; 2011.

38. Baird JK: **Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria**. *Clin Microbiol Rev* 2013, **26**(1):36-57.

39. Wells TN, Burrows JN, Baird JK: **Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination**. *Trends Parasitol* 2010, **26**(3):145-151.

40. Beutler E: **G6PD deficiency**. *Blood* 1994, **84**(11):3613-3636.

41. Cappellini MD, Fiorelli G: **Glucose-6-phosphate dehydrogenase deficiency**. *Lancet* 2008, **371**(9606):64-74.

42. WHO: **Guidelines for the Treatment of Malaria, Third Edition**. Geneva, Switzerland: World Health Organization; 2015.

43. Baird JK, Hoffman SL: **Primaquine therapy for malaria**. *Clin Infect Dis* 2004, **39**(9):1336-1345.

44. Fernando D, Rodrigo C, Rajapakse S: **Primaquine in vivax malaria: an update and review on management issues**. *Malar J* 2011, **10**:351.

45. Llanos-Cuentas A, Lacerda MV, Rueangweerayut R, Krudsood S, Gupta SK, Kochar SK *et al,*: **Tafenoquine plus chloroquine for the treatment and relapse prevention of *Plasmodium vivax* malaria (DETECTIVE): a multicentre, double-blind, randomised, phase 2b dose-selection study**. *Lancet* 2014, **383**(9922):1049-1058.

46. Price RN, von Seidlein L, Valecha N, Nosten F, Baird JK, White NJ: **Global extent of chloroquine-resistant Plasmodium vivax: a systematic review and meta-analysis**. *Lancet Infect Dis* 2014, **14**(10):982-991.

47. Garnham PC: **Some effects on the community of malaria eradication with special reference to the relapse phenomenon**. *East Afr Med J* 1951, **28**(1):6-10.

48. Pampana E: **A textbook of malaria eradication**. London: Oxford University Press; 1969.

49. Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM *et al,*: **Shrinking the malaria map: progress and prospects**. *Lancet* 2010, **376**(9752):1566-1578.

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

51. GBD 2013 Mortality and Causes of Death Collaborators: **Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013**. *Lancet* 2015, **385**(9963):117-171.

52. Feachem R, Sabot O: **A new global malaria eradication strategy**. *Lancet* 2008, **371**(9624):1633-1635.

53. RBMP: **The global malaria action plan for a malaria free world**. In. Geneva: Roll Back Malaria Partnership (RBMP), World Health Organization; 2008.

54. WHO: **Malaria elimination: a field manual for low and moderate endemic countries**. In. Geneva: World Health Organization; 2007: 85.

55. Baird JK: **Eliminating malaria - all of them**. *Lancet* 2010, **376**(9756):1883-1885.

56. Baird JK: **Neglect of *Plasmodium vivax* malaria**. *Trends Parasitol* 2007, **23**(11):533-539.

57. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM: **Vivax malaria: neglected and not benign**. *Am J Trop Med Hyg* 2007, **77**(6 Suppl):79-87.

58. Rogerson SJ, Carter R: **Severe vivax malaria: newly recognised or rediscovered**. *PLoS Med* 2008, **5**(6):e136.

59. Anstey NM, Russell B, Yeo TW, Price RN: **The pathophysiology of vivax malaria**. *Trends Parasitol* 2009, **25**(5):220-227.

60. Barcus MJ, Basri H, Picarima H, Manyakori C, Sekartuti, Elyazar I *et al,*: **Demographic risk factors for severe and fatal vivax and falciparum malaria among hospital admissions in northeastern Indonesian Papua**. *Am J Trop Med Hyg* 2007, **77**(5):984-991.

61. Karyana M, Burdarm L, Yeung S, Kenangalem E, Wariker N, Maristela R *et al,*: **Malaria morbidity in Papua Indonesia, an area with multidrug resistant *Plasmodium vivax* and *Plasmodium falciparum***. *Malar J* 2008, **7**:148.

62. Kochar DK, Saxena V, Singh N, Kochar SK, Kumar SV, Das A: ***Plasmodium vivax* malaria**. *Emerg Infect Dis* 2005, **11**(1):132-134.

63. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M *et al,*: **Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia**. *PLoS Med* 2008, **5**(6):e128.

64. Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, Alpers MP *et al,*: ***Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea**. *PLoS Med* 2008, **5**(6):e127.

65. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Hasanuddin A, Warikar N *et al,*: **Vivax malaria: a major cause of morbidity in early infancy**. *Clin Infect Dis* 2009, **48**(12):1704-1712.

66. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Warikar N, Seal A *et al,*: **Adverse pregnancy outcomes in an area where multidrug-resistant *Plasmodium vivax* and *Plasmodium falciparum* infections are endemic**. *Clin Infect Dis* 2008, **46**(9):1374-1381.

67. Hay SI, Snow RW: **The Malaria Atlas Project: developing global maps of malaria risk**. *PLoS Med* 2006, **3**(12):e473.

68. Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ: **Global environmental data for mapping infectious disease distribution**. *Adv Parasitol* 2006, **62**:37-77.

69. Rogers DJ, Randolph SE: **Studying the global distribution of infectious diseases using GIS and RS**. *Nat Rev Microbiol* 2003, **1**(3):231-237.

70. Hay SI, Battle KE, Pigott DM, Smith DL, Moyes CL, Bhatt S *et al,*: **Global mapping of infectious disease**. *Philos Trans R Soc Lond B Biol Sci* 2013, **368**(1614):20120250.

71. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL *et al,*: **A new world malaria map: *Plasmodium falciparum* endemicity in 2010**. *Malar J* 2011, **10**:378.

72. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA *et al,*: **Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007**. *PLoS Med* 2010, **7**(6):e1000290.

73. Tatem AJ, Smith DL, Gething PW, Kabaria CW, Snow RW, Hay SI: **Ranking of elimination feasibility between malaria-endemic countries**. *Lancet* 2010, **376**(9752):1579-1591.

74. The Global Health Group and the Malaria Atlas Project: **Atlas of Malaria-Eliminating Countries, 2011**. San Francisco, U.S.A.: The Global Health Group, Global Health Sciences, University of California; 2011.

75. Patil AP, Gething PW, Piel FB, Hay SI: **Bayesian geostatistics in health cartography: the perspective of malaria**. *Trends Parasitol* 2011, **27**(6):246-253.

76. Diggle PJ, Thomson MC, Christensen OF, Rowlingson B, Obsomer V, Gardon J *et al,*: **Spatial modelling and the prediction of *Loa loa* risk: decision making under uncertainty**. *Ann Trop Med Parasitol* 2007, **101**(6):499-509.

77. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperely WH *et al,*: **The international limits and population at risk of *Plasmodium vivax* transmission in 2009**. *PLoS Negl Trop Dis* 2010, **4**(8):e774.

78. Guerra CA, Snow RW, Hay SI: **Defining the global spatial limits of malaria transmission in 2005**. *Adv Parasitol* 2006, **62**:157-179.

79. Guerra CA, Snow RW, Hay SI: **Mapping the global extent of malaria in 2005**. *Trends Parasitol* 2006, **22**(8):353-358.

80. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH *et al,*: **The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis**. *Parasit Vectors* 2011, **4**:89.

81. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J *et al,*: **The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis**. *Parasit Vectors* 2010, **3**:117.

82. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M *et al,*: **A global map of dominant malaria vectors**. *Parasit Vectors* 2012, **5**(1):69.

83. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW *et al,*: **The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic precis**. *Parasit Vectors* 2010, **3**:72.

84. Hay SI, Smith DL, Snow RW: **Measuring malaria endemicity from intense to interrupted transmission**. *Lancet Infect Dis* 2008, **8**(6):369-378.

85. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U *et al,*: **The impact of malaria control on *Plasmodium falciparum* in Africa, 2000–2015**. *Nature* 2015, **Accepted**.

86. Cotter C, Sturrock HJ, Hsiang MS, Liu J, Phillips AA, Hwang J *et al,*: **The changing epidemiology of malaria elimination: new strategies for new challenges**. *Lancet* 2013, **382**(9895):900-911.

87. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C: **Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods**. *PLoS Med* 2011, **8**(12):e1001142.

Chapter 2 – The global public health significance of *Plasmodium vivax*

This first research chapter is a review of the methods used to estimate the geographic limits of *Plasmodium vivax* transmission and the population at risk and prevalence of infection within those limits. The chapter also contains a literature review to identify the *Anopheles* species that have been specifically incriminated as *P. vivax* vectors. This chapter lays the groundwork of the thesis to show the extent of the *P. vivax* public health problem and give an indication of its complexity. The limits, populations at risk, and prevalence predictions described here are updated as part of Chapter 6. This description therefore serves a baseline for how the methods and the resulting estimates have changed during the course of this doctoral work.

CHAPTER ONE

The Global Public Health Significance of *Plasmodium vivax*

Katherine E. Battle*, Peter W. Gething*, Iqbal R.F. Elyazar†, Catherine L. Moyes*, Marianne E. Sinka*, Rosalind E. Howes*, Carlos A. Guerra‡, Ric N. Price§,¶, J. Kevin Baird†,§, Simon I. Hay*,‡

*Department of Zoology, University of Oxford, South Parks Road, Oxford, UK
†Eijkman-Oxford Clinical Research Unit, Jalan Diponegoro No. 69, Jakarta, Indonesia
‡Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
§Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK
¶Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia

Contents

Advances in Parasitology, Volume 80
ISSN 0065-308X, http://dx.doi.org/10.1016/B978-0-12-397900-1.00001-3

Abstract

Plasmodium vivax occurs globally and thrives in both temperate and tropical climates. Here, we review the evidence of the biological limits of its contemporary distribution and the global population at risk (PAR) of the disease within endemic countries. We also review the most recent evidence for the endemic level of transmission within its range and discuss the implications for burden of disease assessments. Finally, the evidence-base for defining the contemporary distribution and PAR of *P. vivax* are discussed alongside a description of the vectors of human malaria within the limits of risk. This information along with recent data documenting the severe morbid and fatal consequences of *P. vivax* infection indicates that the public health significance of *P. vivax* is likely to have been seriously underestimated.

1. INTRODUCTION

Malaria is a highly significant global public health problem. Its greatest burden is imposed on the world's poorest countries (Sachs and Malaney, 2002). It is the third leading cause of death from infectious diseases for children under age of five worldwide (Black et al., 2010) and the fourth leading cause for all ages (WHO, 2008). After decades of neglect, malaria control research and financing has experienced a resurgence in recent years (RBMP, 2008) and targets have been raised by international initiatives aiming for the goal of elimination (Feachem et al., 2009; Chitnis et al., 2010a; Moonen et al., 2010; Tanner and Hommel, 2010; Tatem et al., 2010; Alonso et al., 2011). While elimination goals ostensibly address all human malarias, the allocation of funding and resources between the two parasites of greatest significance, *Plasmodium falciparum* and *Plasmodium vivax*, has been highly disparate. From 2007 to 2009, only 3.1% of all malaria research and development funding was targeted at *P. vivax* (PATH, 2011). *Plasmodium vivax* is epidemiologically and biologically different to *P. falciparum* and it is not, therefore, possible to assume that control methods developed for falciparum malaria are transferable to vivax malaria (Luxemburger et al., 1994; Bockarie and Dagoro, 2006; Baird, 2010; Bousema and Drakeley, 2011). Evidence from *P. vivax* infections in carefully monitored populations show that vivax malaria should no longer be thought of as a benign and rarely fatal disease but one that can lead to severe disease and death (Baird, 2007; Price et al., 2007b; Anstey et al., 2009). The reader is referred to an accompanying review in this thematic volume of *Advances in Parasitology* (Chapter 3, Volume 80), which addresses the clinical severity of *P. vivax*. The clinical importance of *P. vivax*, along with its wide geographic distribution extending well beyond the limits of falciparum malaria (Guerra et al.,

2006a, 2006b, 2010; Gething et al., 2011b) and the challenges it presents for control (Sattabongkot et al., 2004; Wells et al., 2010), has led to a call for greater attention to be paid to understanding the global distribution and burden of this neglected parasite (Baird, 2007; Price et al., 2007b; Mueller et al., 2009a).

A key information gap impeding *P. vivax* control and progress towards elimination has been the lack of geographical estimates of risk (Mendis et al., 2001; Price et al., 2007b; Malaria Eradication Research Agenda, 2011b). Assessment of geographic variations in levels of endemicity of the parasite is essential to estimate the burden of the disease and measure the impact of control and the feasibility of elimination (Pampana, 1969; Hay et al., 2008; Tatem et al., 2010). The Malaria Atlas Project (MAP) was founded to address this evidence gap (Hay and Snow, 2006; Hay et al., 2009; Gething et al., 2011a), with an initial focus on *P. falciparum*, and the resulting global malaria distributions have been used to assess the adequacy and equity in control funding (Pigott et al., 2012), to inform international policy and resource allocation (Anonymous, 2009; Feachem et al., 2009; McLaughlin et al., 2009; World Bank, 2009; Zanzibar Malaria Control Program, 2009; DFID, 2010; Global Partnership to Roll Back Malaria et al., 2010) and to estimate the global burden of disease (Patil et al., 2009; Gething et al., 2010a; Hay et al., 2010b). A suite of modelled spatial data on *P. vivax* transmission and endemicity (Guerra et al. 2010; Gething et al. 2012) and *P. vivax* vectors (Sinka et al., 2010a, 2010b, 2011, 2012) add to this evidence base for strategic disease-control planning, implementation and monitoring. Here, for the first time, we bring all these mapped data on *P. vivax* together in one place. We explain how these products were generated, their limitations and their value, and provide an overview for each malaria-endemic region of the world. We also consider those areas where our lack of geographical knowledge is particularly acute and would benefit most from concerted future research efforts. A full review of the methodologies used to generate each mapped product is also provided at the end of this chapter.

2. THE GLOBAL DISTRIBUTION OF *P. VIVAX* INFECTIONS

Plasmodium vivax has the widest geographical distribution of the human malarias with an estimated 2.49 billion individuals living at risk of infection in 2010 (Gething et al., 2012). Biological features of *P. vivax* that distinguish it from *P. falciparum* present unique challenges to the control of

the parasite (Mendis et al., 2001; Sattabongkot et al., 2004; Wells et al., 2010); in elimination settings, *P. vivax* is often the 'last parasite standing' (Garnham, 1951; Yekutiel, 1960; Pampana, 1969; Wernsdorfer et al., 2009; Tatem et al., 2010). *Plasmodium vivax* gametocytes are present earlier in the progression of a primary or recrudescent infection than *P. falciparum* (Mendis et al., 2001; McKenzie et al., 2002), such that the majority of patients have sufficient gametocytaemia to allow transmission before the infection is diagnosed or treated (Ratcliff et al., 2007; Awab et al., 2010; Douglas et al., 2010). *Plasmodium vivax* gametocytes are transmitted more efficiently to *Anopheles* mosquito vectors (Boyd and Kitchen, 1937; Collins et al., 2002) than those of *P. falciparum* and are transmissible at lower parasite densities (Sattabongkot et al., 2004). Within the mosquito, vivax sporozoites develop faster than *P. falciparum* and with slightly wider viable temperature ranges, allowing for a greater geographical distribution (Gething et al., 2011b). In addition, due to vector bionomics in regions where *P. vivax* is most prevalent, methods of control that are broadly effective in reducing *P. falciparum* transmission, such as insecticide-treated bed nets (ITNs), show far less success in the control of *P. vivax* (Luxemburger et al., 1994; Bockarie and Dagoro, 2006). *Plasmodium vivax* malaria is typically carried with lower levels of parasitaemia, making it relatively difficult to diagnose (Mendis et al., 2001). However, there is evidence that, despite lower blood parasite loads, *P. vivax* immunity is acquired more rapidly than *P. falciparum* and may result in an earlier age-prevalence peak in areas of high transmission (Mueller et al., 2009a, 2009b). Detailed reviews regarding the biology (Chapter 2, Volume 80), control (Chapter 6, Volume 80), and acquired immunity (Chapter 3, Volume 81) of *P. vivax* are available elsewhere in this thematic issue of *Advances in Parasitology*.

Perhaps the most important feature of *P. vivax* biology is its ability to relapse in the weeks and months following a primary parasitaemia, via a dormant liver stage known as the hypnozoite (James, 1931; Coatney, 1976; Garnham, 1989; Prudencio et al., 2006; Chen et al., 2007; Imwong et al., 2007; White, 2011). It has long been known that there is significant geographical variation in the rate at which a 'strain' of *P. vivax* may relapse (Coatney and Cooper, 1948; Winckel, 1955; Garnham et al., 1975). The exact mechanism through which hypnozoite relapses are triggered is unknown (Cogswell, 1992; Prudencio et al., 2006; Baird, 2009; Mueller et al., 2009a). One theory is that the mechanism is an adaptive trait of the parasite to sequester or 'hibernate' during times when climatic conditions would be inhospitable to the *Anopheles* vector of the disease (Shute et al., 1976; Baird and Rieckmann, 2003; White, 2011). This theory is supported by observations that temperate strains of the parasite tend to exhibit

longer relapse intervals than tropical strains (Garnham et al., 1975; Shute et al., 1976; Cogswell, 1992; Collins and Jeffery, 1996; Adak et al., 1998; Baird et al., 2007; Imwong et al., 2007). Primaquine is currently the only widely available drug with activity on the hypnozoite stage capable of preventing relapse (Baird and Hoffman, 2004; Galappaththy et al., 2007), but is associated with haemolysis in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency (Beutler, 1994; Baird and Hoffman, 2004; Cappellini and Fiorelli, 2008). Primaquine is contraindicated in pregnant women because of the risk of haemolytic anaemia in the foetus of unknown G6PD status (WHO, 2010a). The reader is referred to reviews that address relapse (Chapter 2, Volume 80), treatment (Chapters 4 and 5, Volume 80) and G6PD deficiency (Chapter 4, Volume 81), which are provided elsewhere in this special issue *Advances in Parasitology*. Relapse also has implications for understanding the burden of *P. vivax* malaria based on prevalence rates derived from malariometric surveys and cartographic studies (Patil et al., 2009; Hay et al., 2010b).

To accurately illustrate *P. vivax* endemicity, it is necessary to incorporate the distribution of the Duffy-negative phenotype. The varying prevalence of Duffy negativity in populations throughout the world is a significant determinant of the distribution of *P. vivax* (Livingstone, 1984). Duffy-negative individuals are, for the most part, refractory to *P. vivax* infection and the phenotype is found at highest frequencies in Africa, whereas it is relatively rare elsewhere (Howes et al., 2011). A detailed review of the effect of Duffy negativity on the epidemiology of *P. vivax* (Chapter 2, Volume 81) is provided elsewhere in this series. The influence of Duffy negativity on *P. vivax* transmission reinforces the need to differentiate strategies employed to generate and interpret maps of *P. vivax* endemicity from those used for *P. falciparum* (Hay et al., 2009; Gething et al., 2011a).

Until recently, little work had been done to define the geographic limits and risk of *P. vivax* infection. The only map including vivax malaria endemicity was that of Lysenko from 1968 (Lysenko and Semashko, 1968). As shown in Fig. 1.1, Lysenko defined endemicity as the parasite rate (PR) in children aged between 2 and 10 years old (hypoendemic <10%, mesoendemic 11–50%, hyperendemic 51–75%), with the exception of the holoendemic class (>75%) where the PR was defined in the 1-year age cohort. The map was derived from a synthesis of historical records and malariometric indices of all four human malaria parasites: disease and vector presence and absence records; spleen, parasite, sporozoite and biting rates; sickle cell incidence; and others (Hay et al., 2004; Gething et al., 2010b). Lysenko interpolated the data globally to determine the distribution of

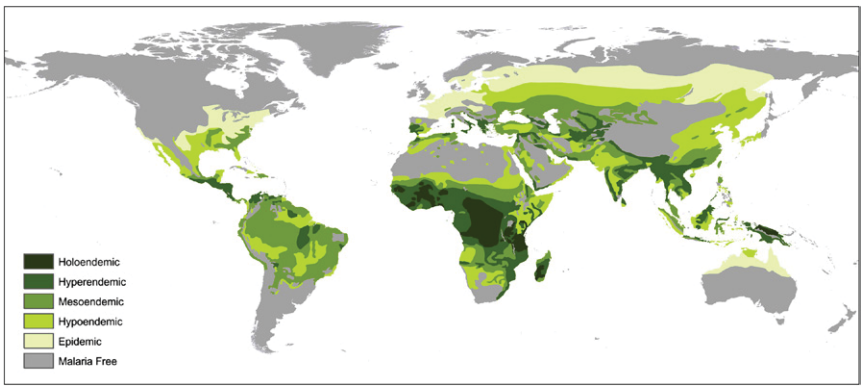


Figure 1.1 The Lysenko map of global malaria endemicity. The map was digitised from its original source (Lysenko and Semashko, 1968) where endemicity classes were defined by parasite rate (PR) in children between 2 and 10 years old (hypoendemic <10%, mesoendemic 11–50%, hyperendemic 51–75%), with the exception of the holoendemic class (>75%) where the parasite rate was defined in the 1-year age cohort. The 'epidemic' class is restricted to the temperate regions in these maps and it should be noted that this term is used differently today. For a colour version of this figure, the reader is referred to the online version of this book.

malaria at the peak of its historic distribution (*circa* 1900) using expert opinion, temperature ranges and rainfall isohyets (Lysenko and Semashko, 1968; Lysenko and Beljaev, 1969; Kaneko et al., 1998).

Lysenko's endemicity map was not specific to one human malaria parasite. The map was created before the advent of geographic information systems (GIS) and lacks geographic precision – as exemplified by a significant misplacement of the Nile River. Relative to the remote sensing data available today, the environmental data Lysenko and Semashko used to refine the endemicity boundaries were both limited and crude. The evidence base used was not described in detail and, since the map was not generated via a formal statistical method, there was no measure of uncertainty provided. It is therefore difficult to determine how robust the distribution estimates of Lysenko were.

3. SPATIAL DISTRIBUTION OF *P. VIVAX* MALARIA, POPULATIONS AT RISK AND ITS VECTORS

Here, we bring together maps at the regional level (Fig. 1.2) of the limits of transmission, endemicity and estimates of the populations at risk, along with the distribution of *P. vivax* vectors.

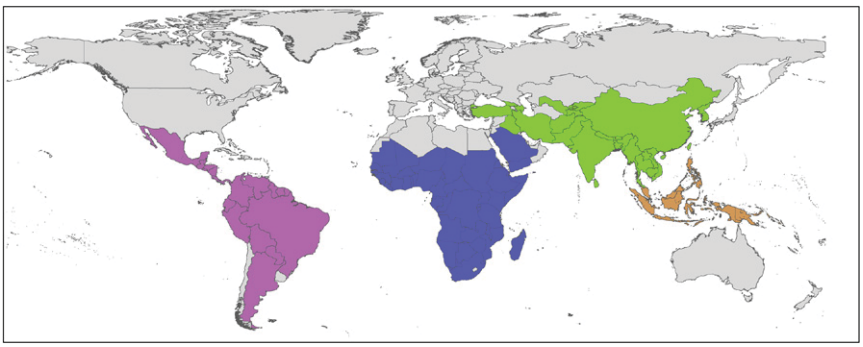


Figure 1.2 Regional tiles of *Plasmodium vivax* endemic countries. The regional tiles shown were used to stratify the modelling of *P. vivax* endemicity. Pink = the Americas; blue = Africa+; green = Asia; orange = Asia-Pacific; grey = non endemic for *P. vivax*. For interpretation of the references to colour in this figure legend, the reader is referred to the online version of this book.

3.1. *P. vivax* Malaria Limits and Endemicity

Figure 1.3 shows estimated limits of unstable (light grey) and stable (dark grey) transmission of *P. vivax* in Panel 1 (A1–D1) and, within Panel 2 (A2–D2), levels of endemicity (blue to yellow to red, red being >7% prevalence). The map is accompanied by two measures of uncertainty in our endemicity estimates: one absolute measure and the second weighted by population density in each location (Fig. 1.4).

A detailed review of the data and methodology used to create this map is given at the end of this chapter in Section 6. In brief, we have used 9970 spatiotemporally unique records of parasite prevalence (Fig. 1.3, Panel 1), environmental covariates and a Bayesian geostatistical model with seasonal and age-standardisation components to generate a smooth map of point estimates of parasite prevalence in all areas with stable transmission (Gething et al., 2012). Throughout the world, endemicity was predicted within a relatively narrow range, with the point estimate of the *P. vivax* PR age-standardised to the 1–99 year age range ($Pv\mathrm{PR}_{1-99}$) rarely exceeding 7%. Here, we highlight important aspects of the methodological approach taken that improve on preceding global estimates of *P. vivax* malaria risk, along with key assumptions and constraints:

Temporal influence. The maps presented (Figure 1.3, Panel 2) are of endemicity in 2010 but the data used to inform the model ranged from 1985 to 2010. The model was designed to down-weight older data, so that whilst they can help inform the contemporary estimates, particularly in areas lacking contemporary data, these had less influence on our estimates of prevalence. This was reflected by higher uncertainty those areas reliant on older data.

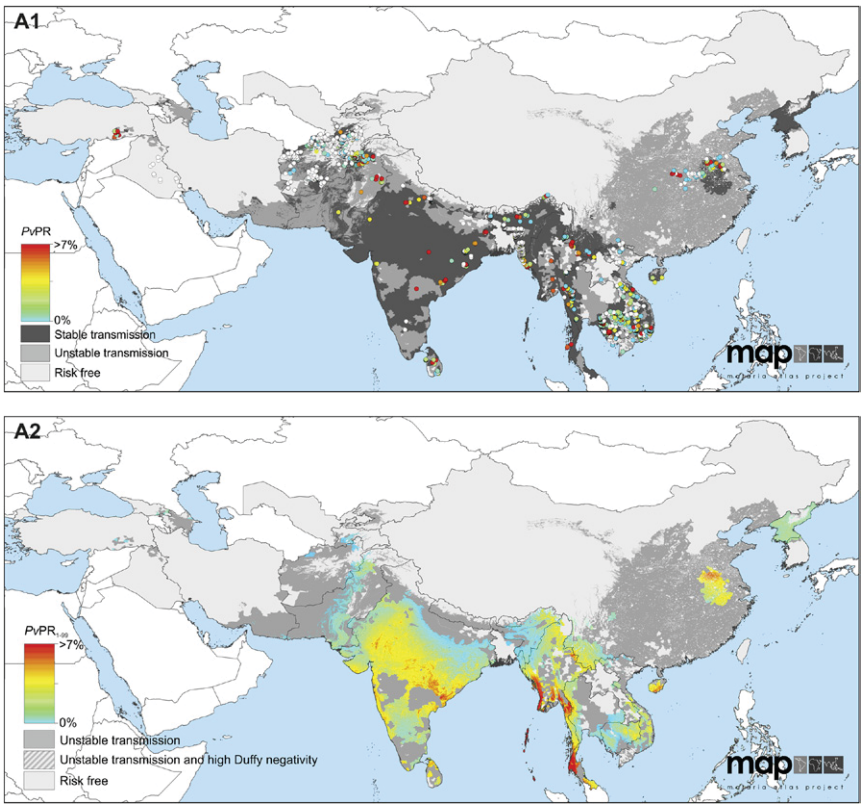


Figure 1.3 The spatial distribution of *Plasmodium vivax* malaria endemicity in 2010 in Asia, Asia-Pacific, the Americas and Africa+. The spatial distribution of *P. vivax* malaria endemicity is shown at the regional levels: Asia (A), Asia-Pacific (B), the Americas (C) and Africa+ (D). Panel 1 in A–D shows the 2010 spatial limits of *P. vivax* malaria risk defined by *Plasmodium vivax* annual parasite incidence (*Pv*API) with further medical intelligence, temperature and aridity masks. Areas were defined as stable (dark grey areas, where *Pv*API ≥ 0.1 per 1000 per annum), unstable (medium grey areas, where *Pv*API < 0.1 per 1000 p.a.) or no risk (light grey, where *Pv*API = 0 per 1000 p.a.). Only the *P. vivax* malaria endemic countries (*Pv*MECs) in each region are shaded in. The community surveys of *P. vivax* prevalence conducted between January 1985 and June 2010 are plotted. The survey data are presented as a continuum of light blue to red (see map legend), with zero-valued surveys shown in white. Panel 2 in each region shows the model-based geostatistics (MBG) point estimates of the annual mean $PvPR_{1-99}$ for 2010 within the spatial limits of stable *P. vivax* malaria transmission, displayed on the same colour scale. Areas within the stable limits (Panel 1) that were predicted with high certainty (>0.9) to have a $PvPR_{1-99}$ less than 1% were classed as unstable. Areas in which Duffy negativity gene frequency is predicted to exceed 90% (Howes et al., 2011) are shown in hatching for additional context. For interpretation of the references to colour in this figure legend, the reader is referred to the online version of this book.

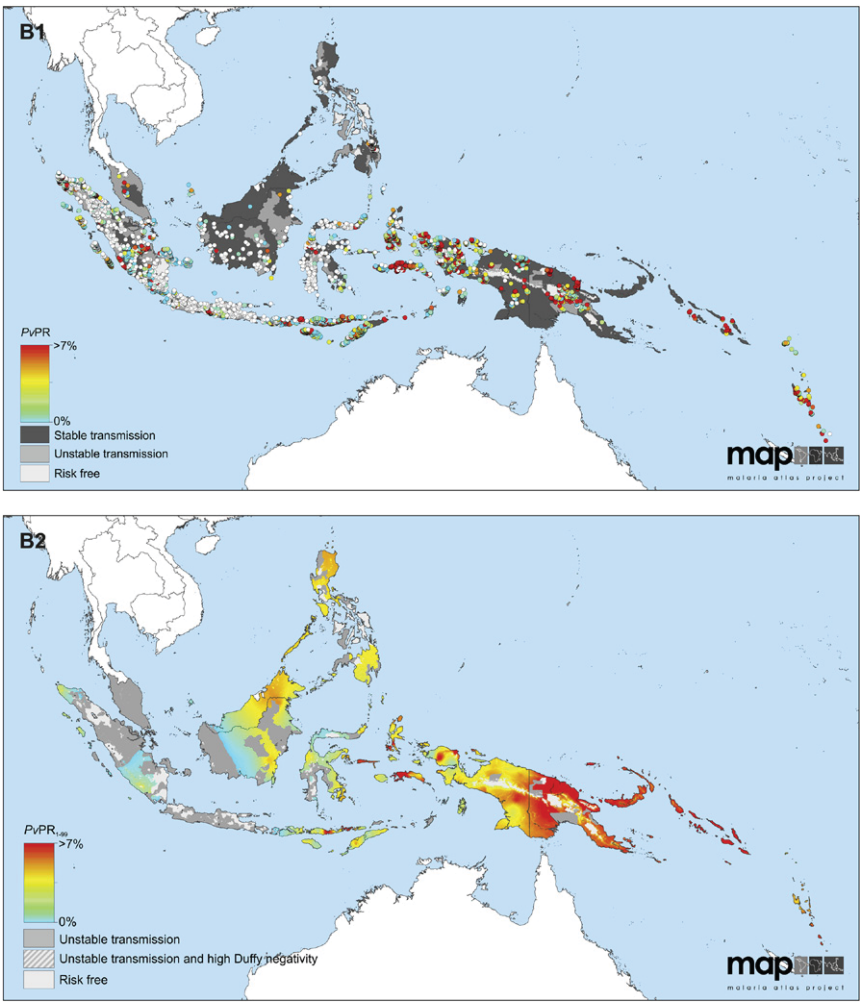


Figure 1.3, cont'd

Seasonal fluctuations. The estimates provide the average *P. vivax* endemicity over a year. Seasonal fluctuations of parasite prevalence were included in the model structure but aggregated in the outputs. Seasonal estimates are not currently available and would be confounded by relapses that can occur outside the transmission season (see the section 'Relapse of *Plasmodium vivax* blood infections' below).

Smoothness of the global map. Malaria endemicity can be highly heterogeneous over small distances and, therefore, our model produced candidate maps that are also highly heterogeneous. An artefact of presenting one map that summarises the full model output using either a mean or a median

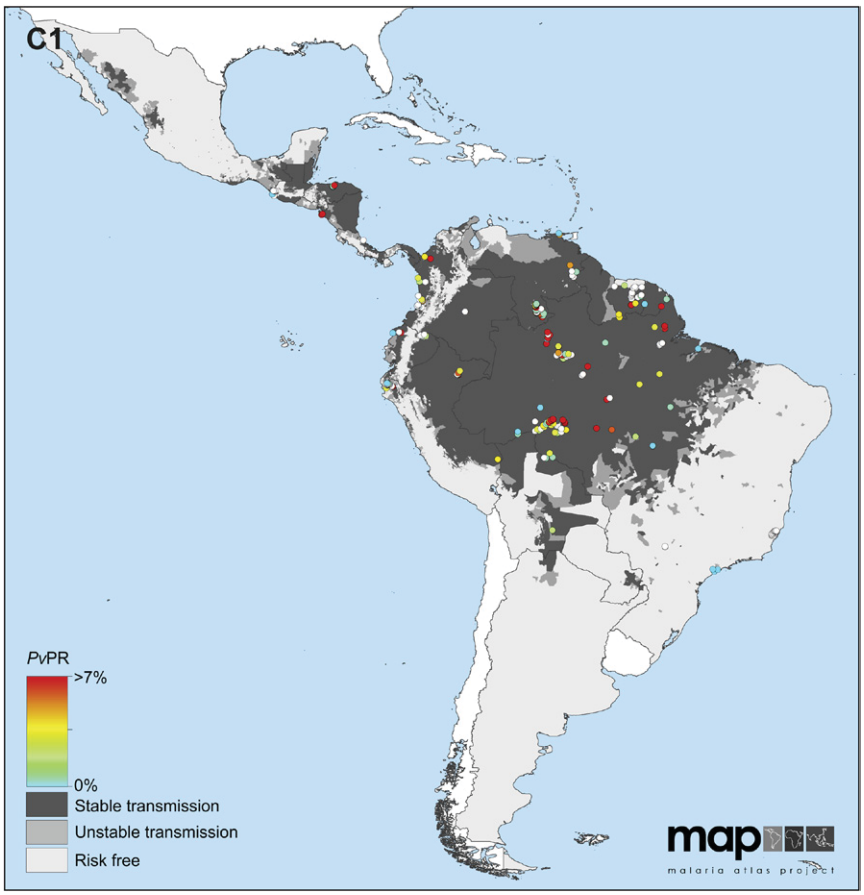


Figure 1.3, cont'd

of the estimates for each location was that this process smoothed out the patchy nature of malaria endemicity.

Use of parasite prevalence data. It is important that our estimates were based on real data, with a mathematical model that compensated for the fact that the real data were neither comprehensive nor universally contemporary. The most commonly available metric from almost all countries is prevalence (often referred to as parasite rate (PR)), estimated from surveys that are conducted using a common methodology (Hay et al., 2008). Prevalence data are relatively robust but have the disadvantage that required sample sizes become prohibitively large when prevalence rates are low.

Use of clinical incidence data. Annual clinical incidence per 1000 people (commonly referred to as annual parasite incidence (API) data) from public

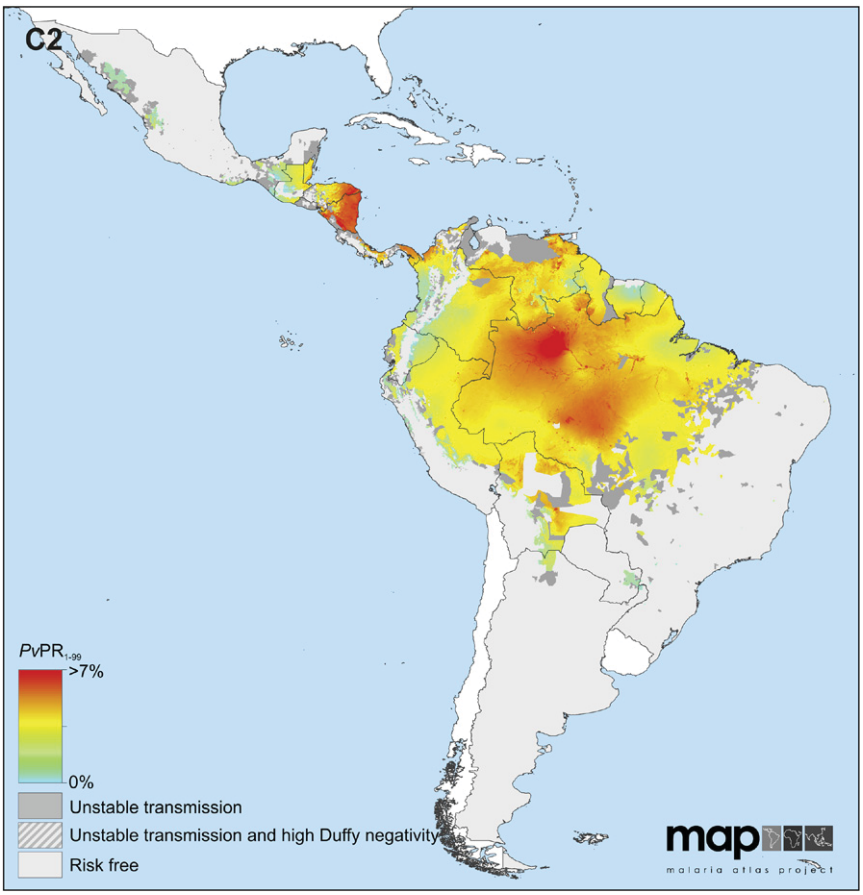


Figure 1.3, cont'd

health records is more sensitive at low parasite prevalence values, but is biased. For example, we do not know the proportion of people within every district who do not seek care for malaria at a public health facility and are therefore missed by the statistics. We used clinical incidence data to define the areas where we know very little, other than that malaria exists but is so low as to be essentially unquantifiable using available data. We referred to this level of risk as unstable malaria transmission and defined it as less than one case per 10,000 people per year (or <0.1 API). At very low levels of incidence, a small change is proportionally large so API is measured across an administrative (ADMIN) division and we averaged this figure over 4 years (providing data are available).

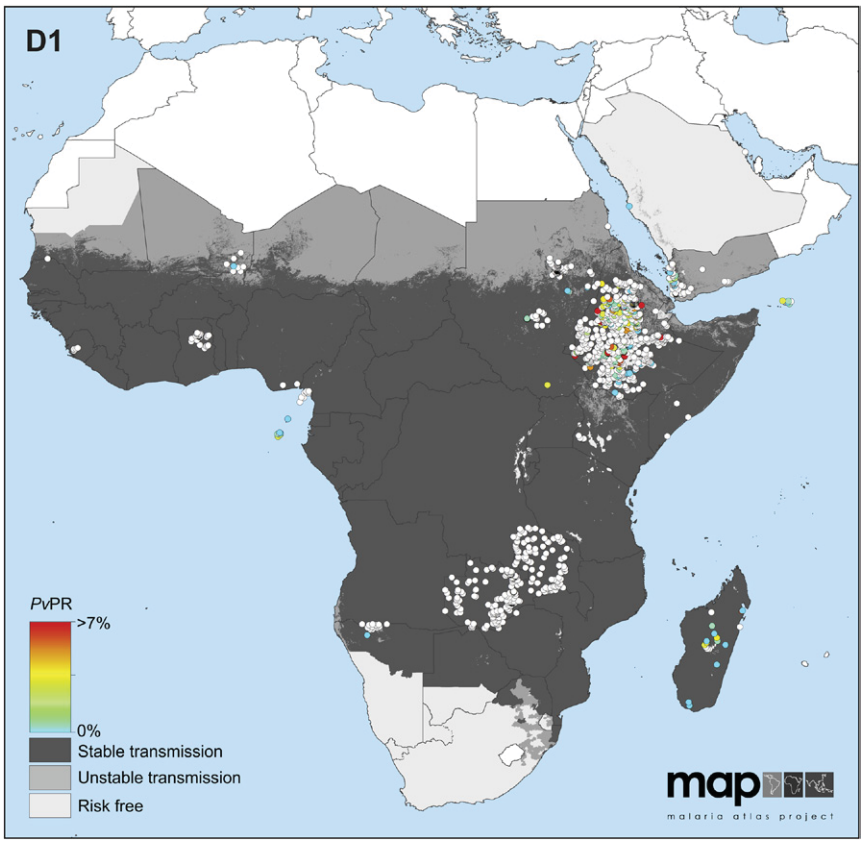


Figure 1.3, cont'd

Use of biological data to refine the limits defined by clinical incidence data. Remote sensing surfaces, showing temperature (Fig. 1.5) and aridity (Fig. 1.6), and biological models were used to identify areas where extreme temperature or aridity regimes would reduce or preclude *P. vivax* transmission. The limits of transmission are adjusted accordingly.

The effect of interventions. The PR data used came from many areas where interventions (ITNs, indoor residual spraying (IRS), etc.) have been implemented so that the effect of these interventions at these locations was included within the estimates produced by the model. This effect was not modelled separately and cannot be extracted from our results.

The impact of Duffy negativity. Since Duffy negative individuals are largely refractory to *P. vivax* infection (Miller et al., 1976), high population frequencies of this phenotype suppress endemicity, even where conditions are

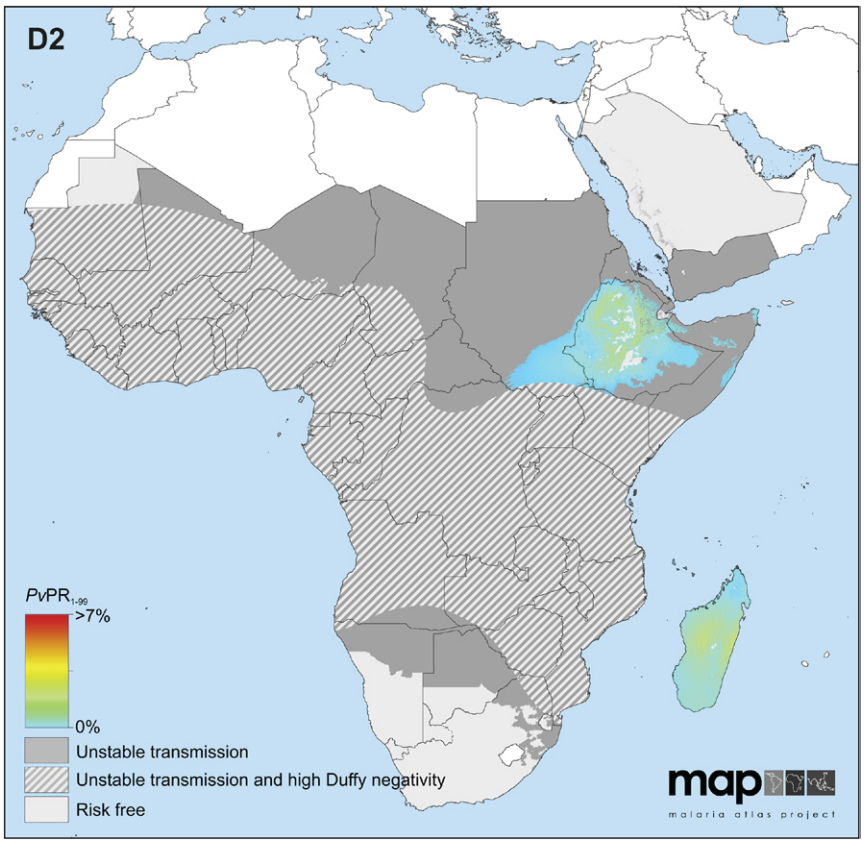


Figure 1.3, cont'd

otherwise well suited for transmission (Guerra et al., 2010). We incorporated recent global estimates of the frequencies of Duffy negativity (Howes et al., 2011) in our model. Regional maps of the predicted prevalence of Duffy negativity are shown in Fig. 1.7. The fraction of the population at each population that was Duffy-negative was excluded from the denominator in the prevalence data, such that any *P. vivax*-positive individuals are assumed to have arisen from the Duffy-positive population subset. Thus, in a location with 90% Duffy negativity, five positive individuals in a survey of 100 would give an assumed prevalence of 50% amongst Duffy positives. Correspondingly, prediction of prevalence was then restricted to the Duffy-positive proportion, with the final prevalence estimate re-converted to refer to the total population. This approach meant prevalence could never exceed the Duffy-positive proportion of a population and, where *P. vivax*

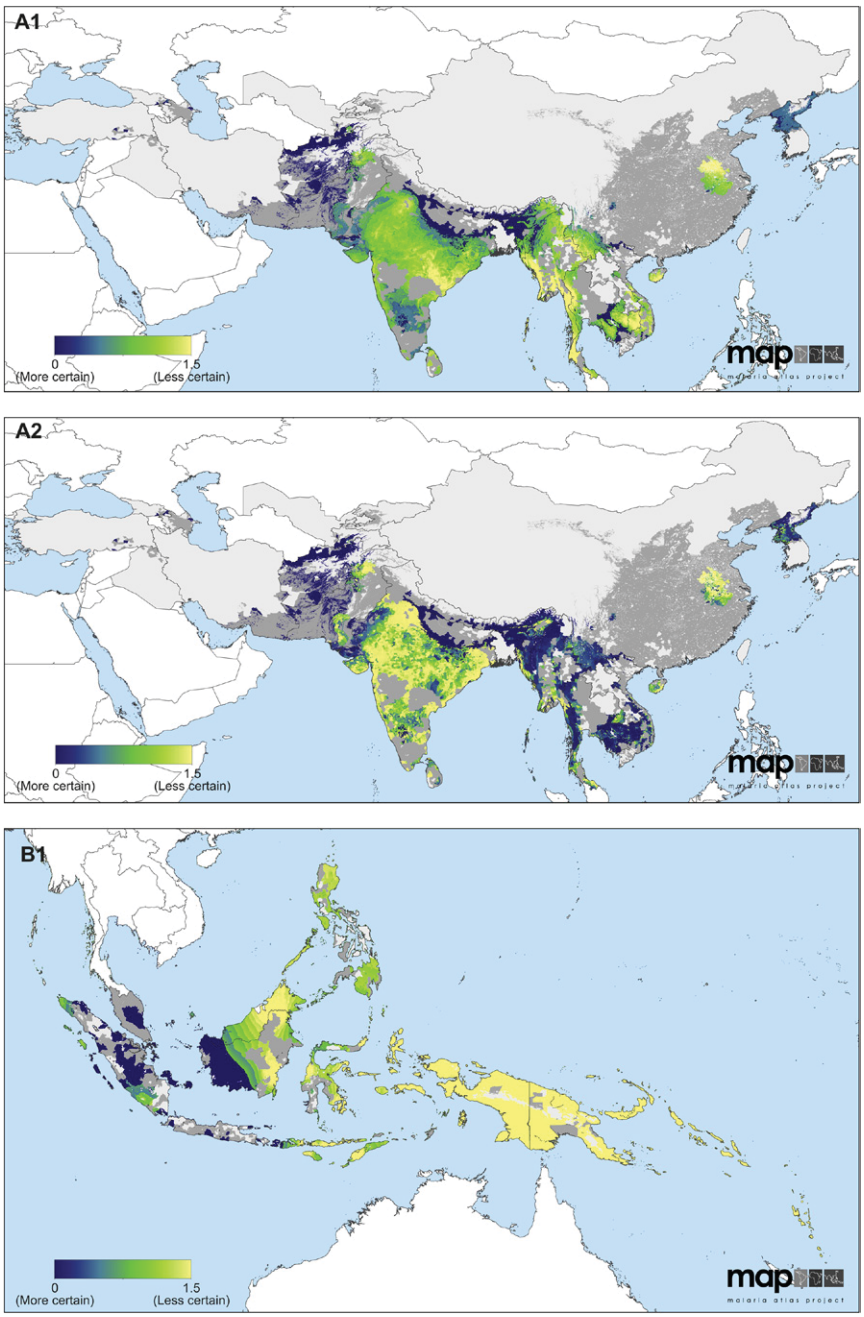
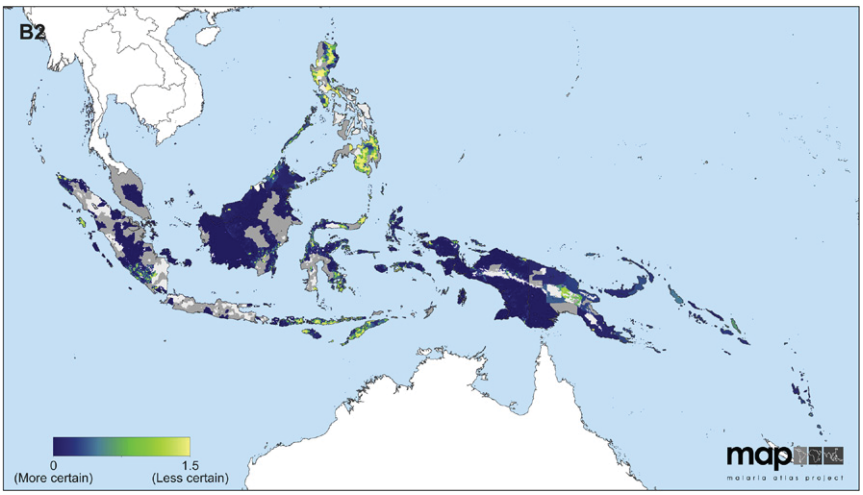


Figure 1.4 Uncertainty associated with predictions of *Plasmodium vivax* endemicity in Asia, Asia-Pacific, the Americas and Africa+.



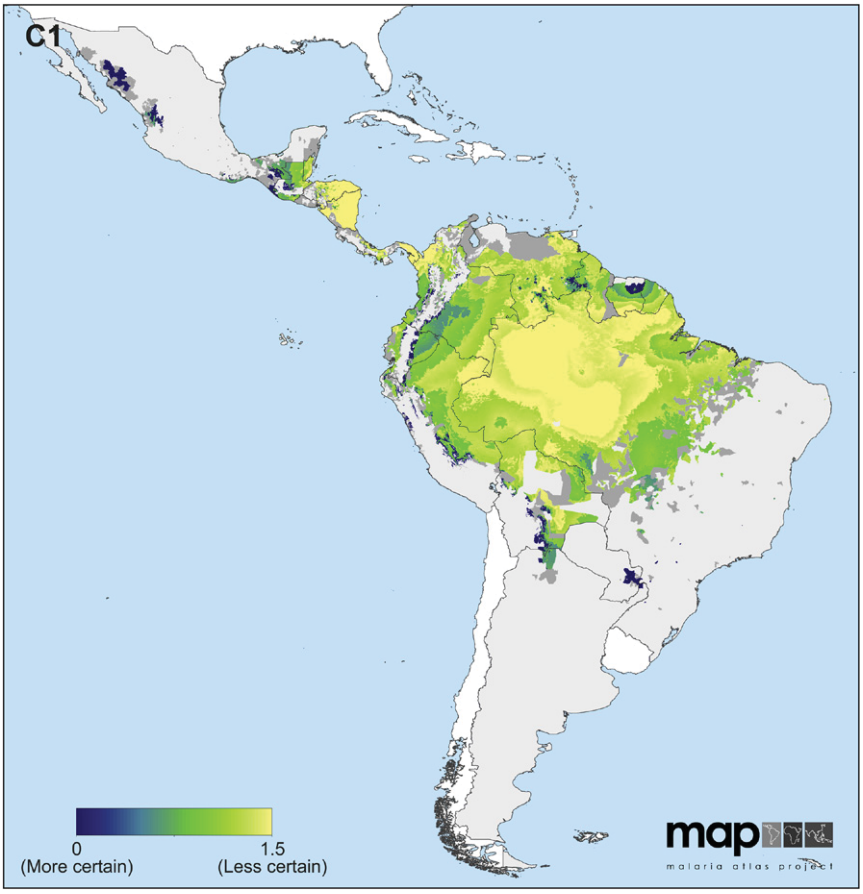


Figure 1.4, cont'd

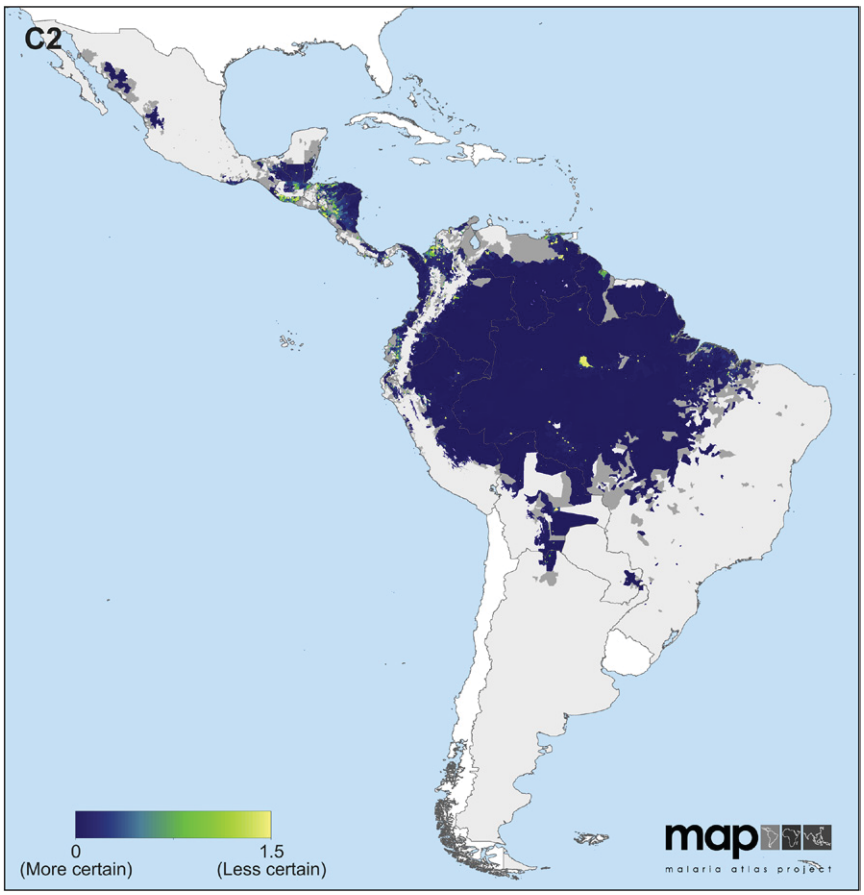


Figure 1.4, cont'd The uncertainty associated with *P. vivax* malaria endemicity predictions is shown at the regional levels: Asia (A), Asia-Pacific (B), the Americas (C) and Africa+ (D). Panel 1 in A–D shows the ratio of the posterior inter-quartile range to the posterior mean prediction at each pixel. Large values indicate greater uncertainty: the model predicts a relatively wide range of *Pv*PR_{1-99} as being equally plausible given the surrounding data. Conversely, smaller values indicate a tighter range of values have been predicted and, thus, a higher degree of certainty in the prediction. Panel 2 in each region shows the same index multiplied by the underlying population density and rescaled to 0–1 to correspond to Panel 1. Higher values indicate areas with high uncertainty and large populations. Areas of no risk within *Pv*MECs are shown in grey and countries not endemic for *P. vivax* or outside the named region are in white. For a colour version of this figure, the reader is referred to the online version of this book.

survey data were sparse across much of Africa, the prevalence predictions could borrow strength from the Duffy negativity map because predictions of prevalence were restricted to a narrower range of possible values.

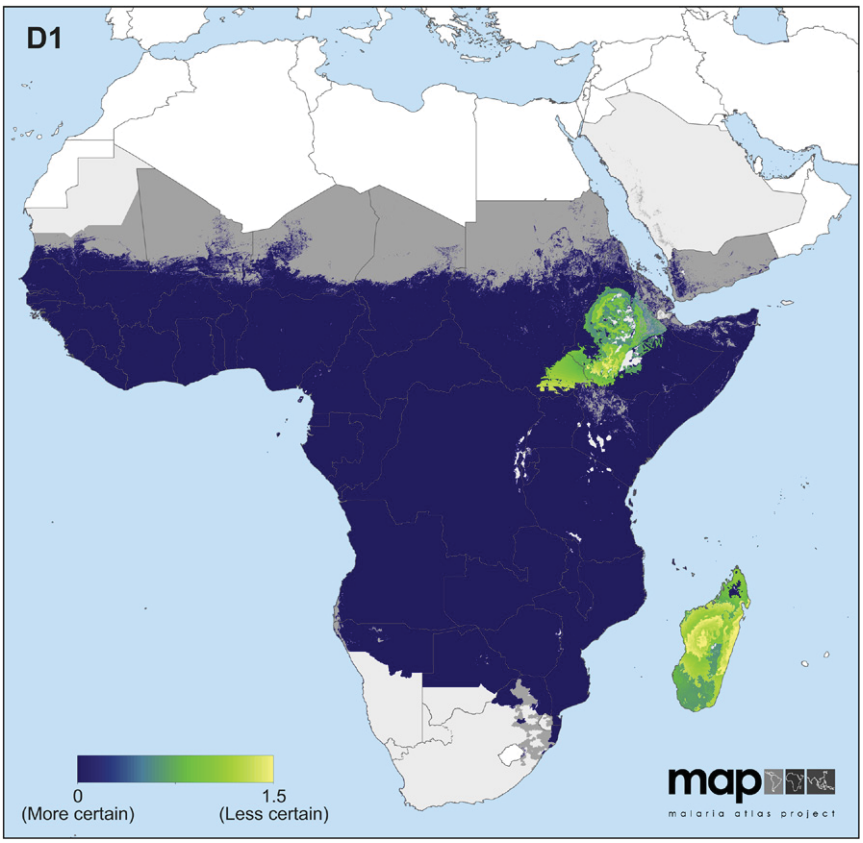


Figure 1.4, cont'd

Transmission in Africa. The predominance of Duffy negativity in Africa has led to a historical perception that *P. vivax* is absent from much of the continent (Rosenberg, 2007). Evidence exists, however, of autochthonous *P. vivax* transmission in nearly every African country (Guerra et al., 2010) and that *P. vivax* is capable of causing severe disease in the continent (Mahgoub et al., 2012). Therefore, we did not preclude any areas at risk before modelling endemicity. We initially assumed stable *P. vivax* transmission, unless the biological mask layers or clinical incidence (API) data confirmed otherwise. In some regions provisionally labelled as stable in this way, the endemicity model subsequently predicted extremely low prevalence, either because of survey data reporting zero infections or because of very high Duffy-negativity prevalence. Locations predicted with high certainty (probability >0.9) of being less than 1% prevalence were therefore re-assigned

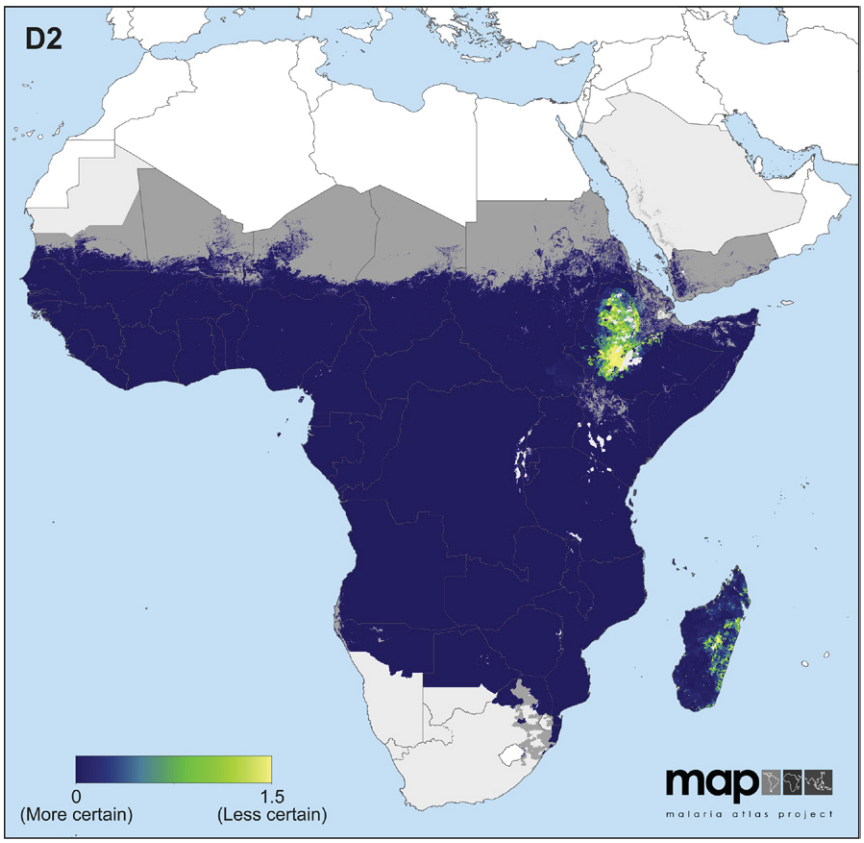


Figure 1.4, cont'd

to the unstable transmission class. This re-assignment is shown in the final limits presented in Fig. 1.3A1–D2.

Relapse of Plasmodium vivax *blood infections.* The prevalence surveys used detect parasites in the blood and not latent infections in the liver. Primary blood infections detected by these surveys are indistinguishable from relapses caused by previously dormant parasites from the liver. The estimates produced by the model were, therefore, the combined prevalence of primary infections and relapses.

The impact of low parasitaemia levels. *Plasmodium vivax* parasites are typically found in the blood at much lower densities than are found for *P. falciparum* and so are less likely to be detected by some diagnostic methods. We made no direct adjustments for these 'sub-patent' infections. It is important to note that *P. vivax* can cause fever and anaemia at lower

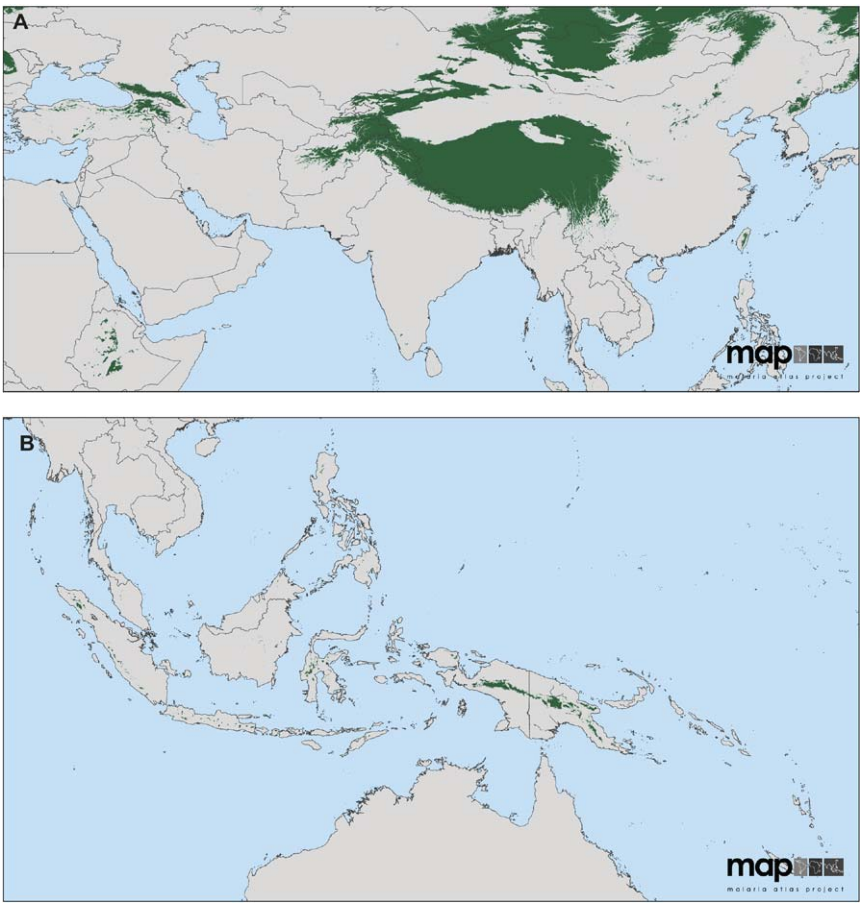


Figure 1.5 Environmental suitability for transmission of *P. vivax* as defined by temperature in Asia, Asia-Pacific, the Americas and Africa+. Areas shaded green in Asia (A), Asia-Pacific (B), the Americas (C) and Africa+ (D) are those in which no windows exist across an average year in which the annual temperature regime is likely to support the presence of infectious vectors. The temperature suitability model is described in full elsewhere (Gething et al., 2011b). For interpretation of the references to colour in this figure legend, the reader is referred to the online version of this book.

parasitaemia levels than *P. falciparum* (Mendis et al., 2001; Mueller et al., 2009a).

Uncertainty in the estimates of endemicity. The modelling framework allowed the uncertainty in predicted endemicity values to vary between locations, depending on the observed variation, density and sample size of surveys and the predictive utility of the suite of environmental covariates (described in detail in Section 1.6). The uncertainty associated with

Figure 1.5, cont'd

predictions is summarised by maps showing the ratio of the inter-quartile range (IQR) to its mean (Fig. 1.4A1–D1). The IQR is a simple measure of the precision with which each value is predicted, and standardisation by the mean produces an uncertainty index less affected by underlying prevalence levels and more illustrative of relative model performance. This index was then also weighted by the underlying population density to produce a second map indicative of those areas, where uncertainty is likely to be most operationally important (Fig. 1.4A2–D2).

3.2. Population at Risk of *P. vivax* Malaria

Table 1.1 shows our estimates of the population at risk (PAR) of infection and a detailed description of the methodology used to generate these

Figure 1.5, cont'd

estimates is given at the end of this chapter in Section 6. In brief, these estimates were generated by combining maps of the limits of transmission described above with 1 × 1 km resolution gridded population surfaces for 2010 projected from the Global Rural-Urban Mapping Project (GRUMP) year 2000 *beta* version population counts (Balk et al., 2006; CIESIN/IFPRI/WB/CIAT, 2007). Regional maps of 1 × 1 km gridded population estimates for 2010 are shown in Fig. 1.8. Fine-scale population data were used to ensure the detailed variations in risk level described in our limits maps are appropriately assigned to the underlying population.

3.3. Vectors of the *P. vivax* Parasite

Figure 1.9 shows the distribution of *Anopheles* species considered to be the dominant vector species/species complexes (DVS) of human malaria

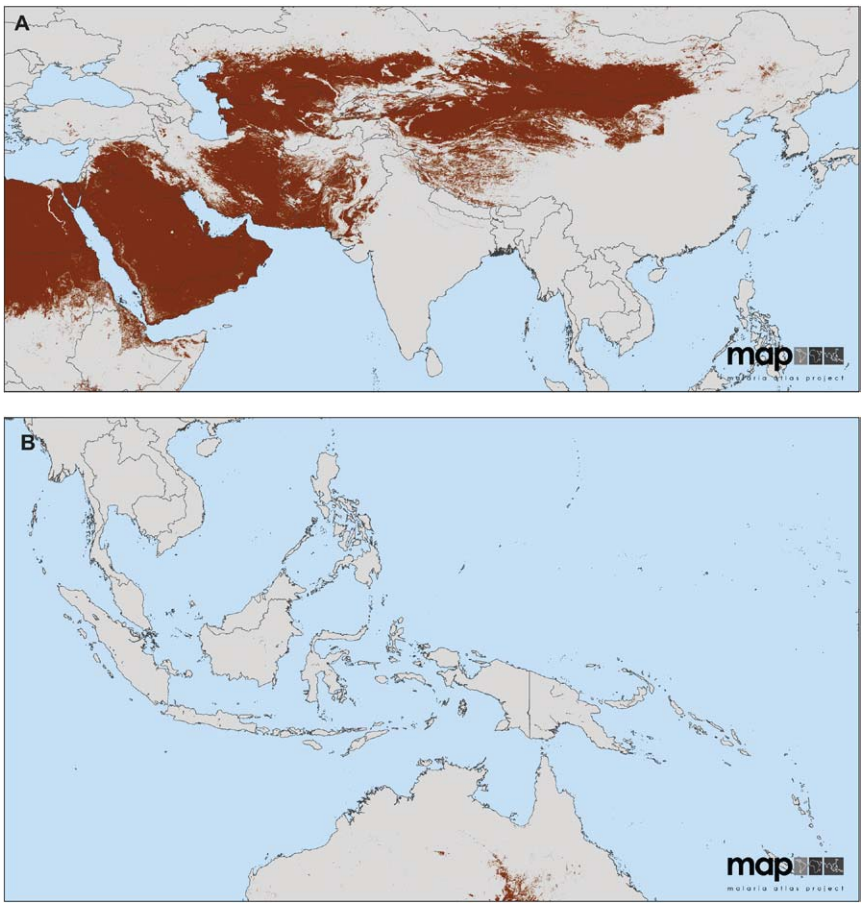


Figure 1.6 Environmental suitability for transmission of *P. vivax* as defined by extreme aridity in Asia, Asia-Pacific, the Americas and Africa+. Areas shaded in brown are those classified as bare areas by the GlobCover land cover product in Asia (A), Asia-Pacific (B), the Americas (C) and Africa+ (D), interpreted as lacking sufficient moisture to support populations of *Anopheles* necessary for transmission. For interpretation of the references to colour in this figure legend, the reader is referred to the online version of this book.

(*P. falciparum* and *P. vivax)* within each region of the world (Sinka et al., 2012), and a detailed description of the methods to generate these maps is given in Section 6.4. In brief, the distribution of each of the dominant species or species complex was individually predicted using occurrence data (mainly abstracted from published surveys), the Boosted Regression Trees (BRT) ecological niche modelling technique (Elith et al., 2008), environmental covariates, and species range expert opinion maps derived from

Figure 1.6, cont'd

consultation with vector experts (Hay et al., 2010c; Sinka et al., 2010a, 2010b, 2011). The resulting maps illustrate the predicted distribution of the DVS using a probability of occurrence metric. While the ecological covariates of the model are based on annualised means of temperature and precipitation, the maps cannot represent the season fluctuations that may occur in the distribution of the DVS, which may be significant in those that extend into temperate climes. Expert opinion was also used to classify the most important species / combinations of species per region (i.e. those with the highest impact) and their individual distributions were overlaid, with the most competent vectors species uppermost, to generate the maps shown. The combined vector map aims to aid vector control planning by showing which species need to be controlled in each area.

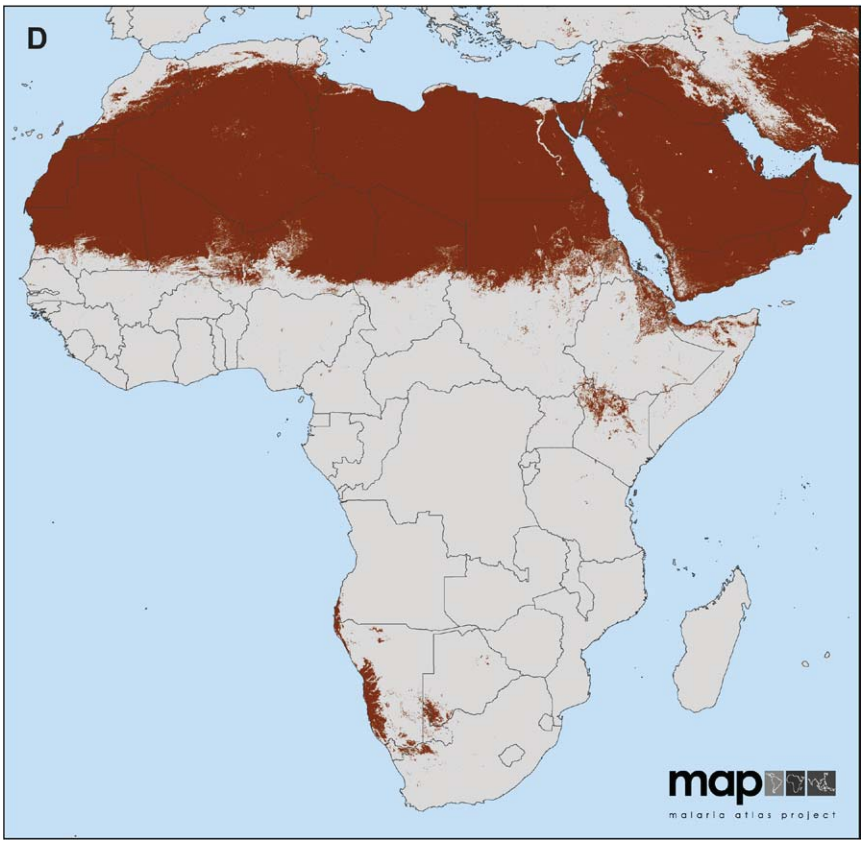


Figure 1.6, cont'd

A species complex is a group of closely related species that are often indistinguishable based on morphology alone. Reference to a complex may be indicated by referring to the species as '*sensu lato*' or '*s.l.*' (e.g. *Anopheles gambiae s.l.*) meaning 'in the broad sense', whereas '*sensu stricto*' or '*s.s.*' ('in the strict sense') indicates the species alone (often a sibling within a species complex has the same name as the complex). The presence of species complexes adds a level of complexity to vector control efforts. Sibling species that are morphologically indistinguishable and often sympatric within an area can have such varied bionomics that one sibling is rendered a dominant vector and the other a non-vector (Meek, 1995; Manguin et al., 2008). The proper identification of species and knowledge of their ranges, often rapidly altered by expanding agriculture and land use changes (Amerasinghe et al., 1991a; Amerasinghe and Indrajith, 1994; Lee, 1998; Singh and Mishra, 2000;

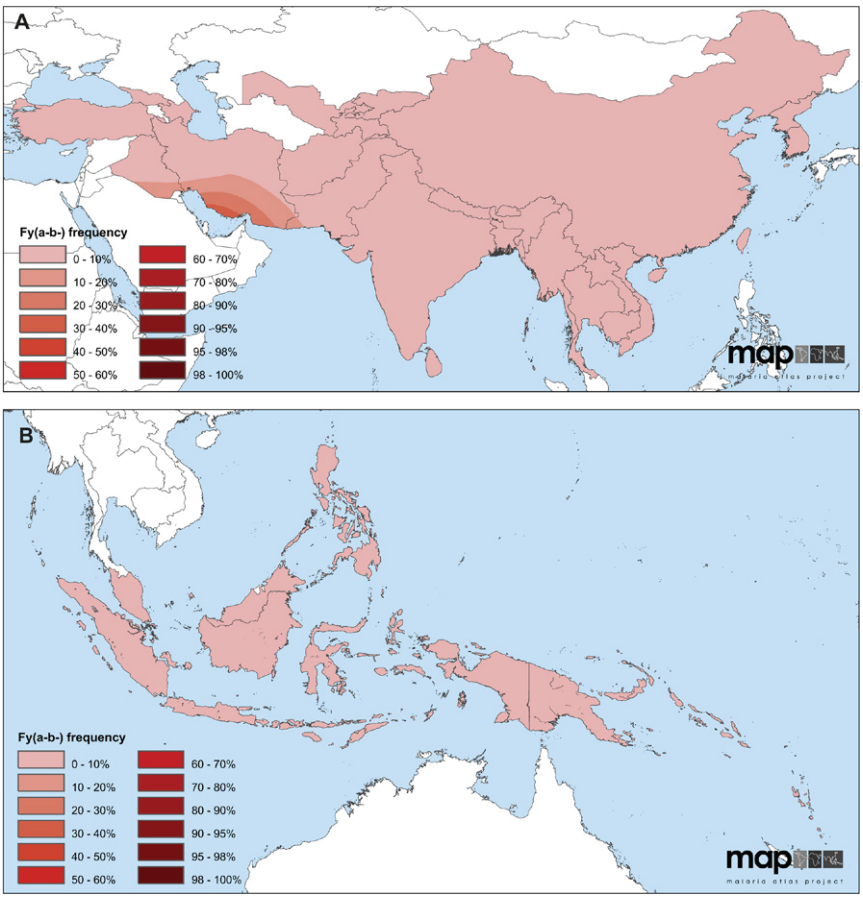


Figure 1.7 Distribution of the Duffy negative phenotype in Asia, Asia-Pacific, the Americas and Africa+. The prevalence of the Duffy-negative phenotype, Fy(a–b–), is shown in Asia (A), Asia-Pacific (B), the Americas (C) and Africa+ (D). The Duffy prevalence model is full described elsewhere (Howes et al., 2011). Only the *Pv*MECs in each region are shown in colour. For a colour version of this figure, the reader is referred to the online version of this book.

Vythilingam et al., 2005), is vital for appropriate allocation of vector control resources.

A search of the current literature has identified 71 species/species complexes with the potential ability to transmit *P. vivax* (Table 1.2), which includes all 41 species classified by Sinka et al. (2010a, 2010b, 2011) as DVS, 34 of which are shown in Fig. 1.9 (a number of species were not included in the global multi-species maps as they were of lesser importance when classified by region and were completely overlaid by higher impacting vectors).

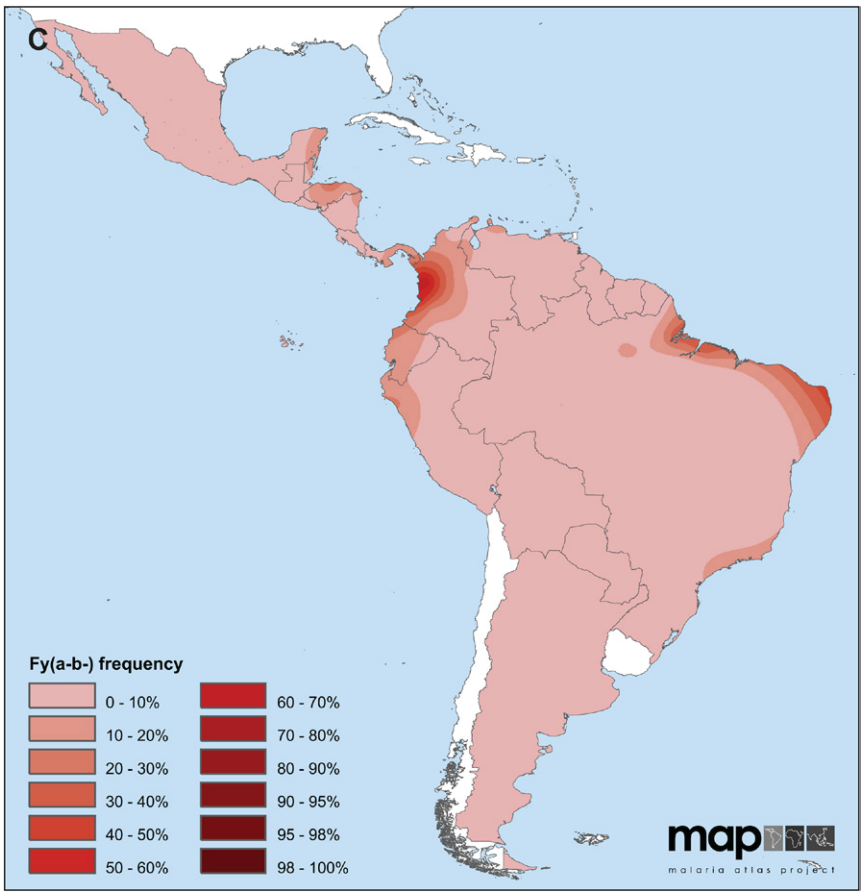


Figure 1.7, cont'd

This suggests, and the literature appears to corroborate, that all vectors capable of transmitting *P. falciparum* are also capable of transmitting *P. vivax*. Interestingly, the reverse (all *P. vivax* vectors being capable of transmitting *P. falciparum*) appears not to hold true.

4. REGIONAL SUMMARIES OF THE PUBLIC HEALTH SIGNIFICANCE OF *P. VIVAX* MALARIA

4.1. Asia

The Asia region (defined for these purposes as mainland Asia but excluding the Malaysian Peninsula; Fig. 1.2) has, amongst the highest endemicity estimates of *P. vivax* malaria globally (Fig. 1.3A2), a diverse range of

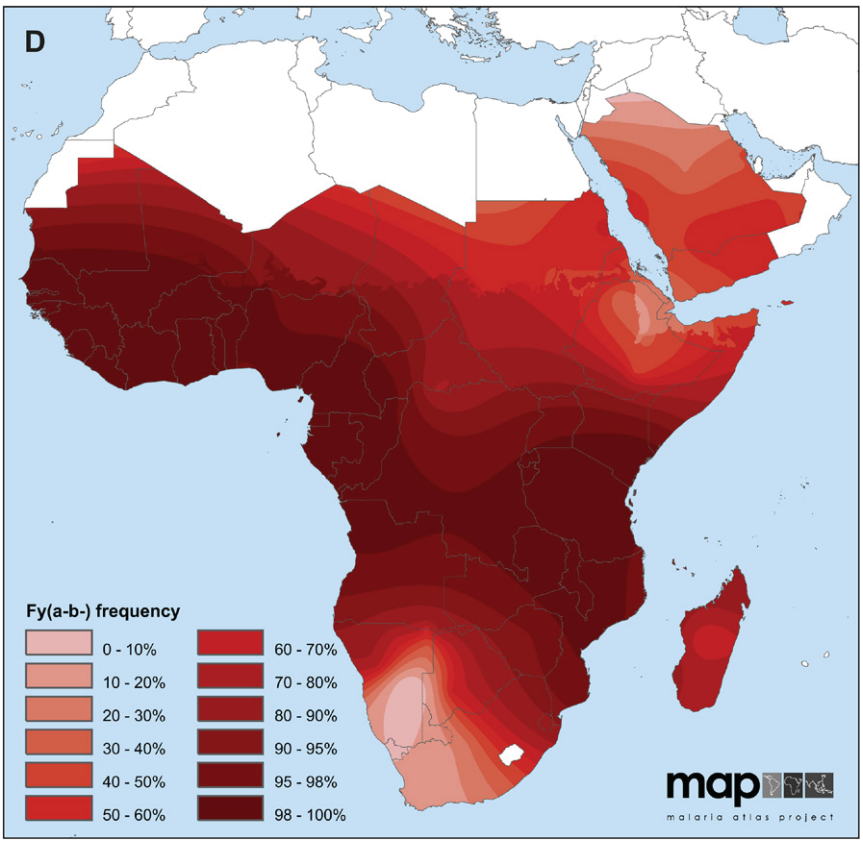


Figure 1.7, cont'd

Table 1.1 Area and Populations at Risk of *Plasmodium vivax* Malaria in 2010

	Area (million km²)			Population (million)		
Region	**Unstable**	**Stable**	**Any risk**	**Unstable**	**Stable**	**Any risk**
America	1.38	8.08	9.46	87.66	49.79	137.45
Africa+	20.60	1.86	22.46	48.72	37.66	86.38
Asia	5.60	3.63	9.24	1236.92	812.55	2049.47
Asia-Pacific	0.96	1.78	2.74	150.17	64.90	215.07
World	28.55	15.34	43.90	1523.47	964.90	2488.37

Risk is stratified into unstable risk (*Pv*API < 0.1 per 1000 people p.a.) and stable risk (*Pv*API ≥ 0.1 per 1000 people p.a.). America = Central and South America; Africa+ = Africa, Yemen and Saudi Arabia; Asia = mainland Asia excluding the Malaysian Peninsula; Asia-Pacific = southern islands of Asia-Pacific and the Malaysian Peninsula.

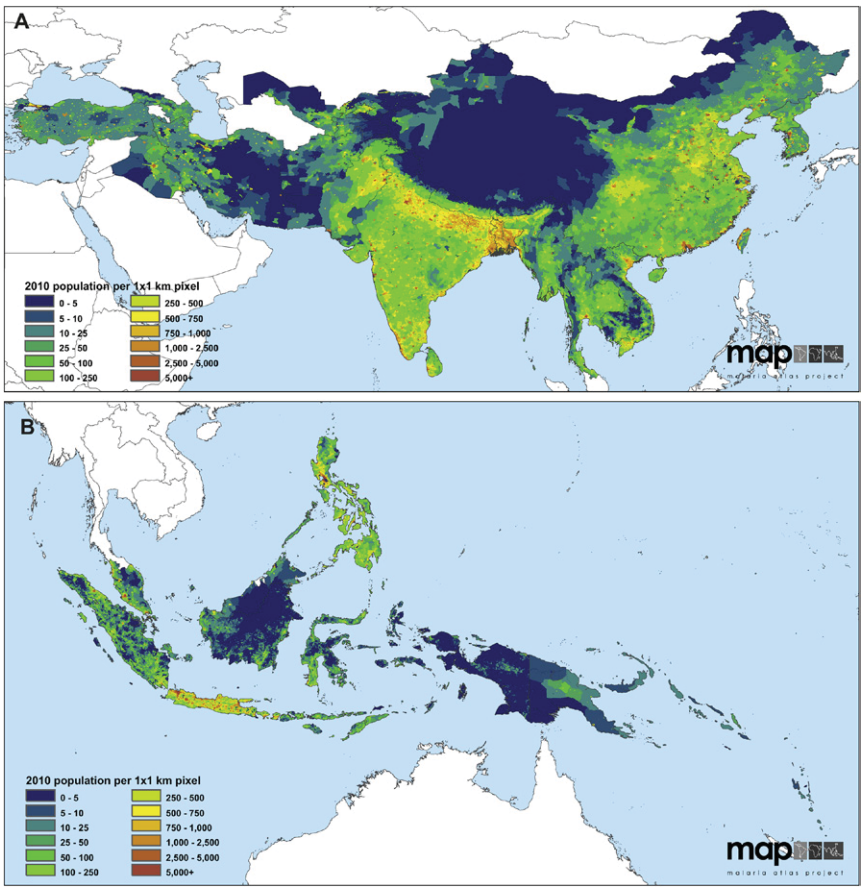


Figure 1.8 Population surfaces for 2010 in Asia, Asia-Pacific, the Americas and Africa+. Population estimates in Asia (A), Asia-Pacific (B), the Americas (C) and Africa+ (D) were estimated from the GRUMP beta 2000 data (Balk et al., 2006; CIESIN/IFPRI/WB/CIAT, 2007) and are shown here at 0.0083 × 0.0083 decimal degree resolution (0.0083 decimal degrees is approximately equal to 1 km at the equator). Only the *Pv*MECs in each region are shown in colour. For a colour version of this figure, the reader is referred to the online version of this book.

epidemiologically important vectors, and by far the largest populations at risk. A low prevalence of Duffy negativity and high population densities in many of the endemic countries in Asia make it an important global *P. vivax* transmission setting to understand and the vast majority (84%) of the global PAR of stable *P. vivax* transmission live in this region.

Defining the limits of transmission. API data were available for all countries except Uzbekistan and Democratic People's Republic of Korea (Korea

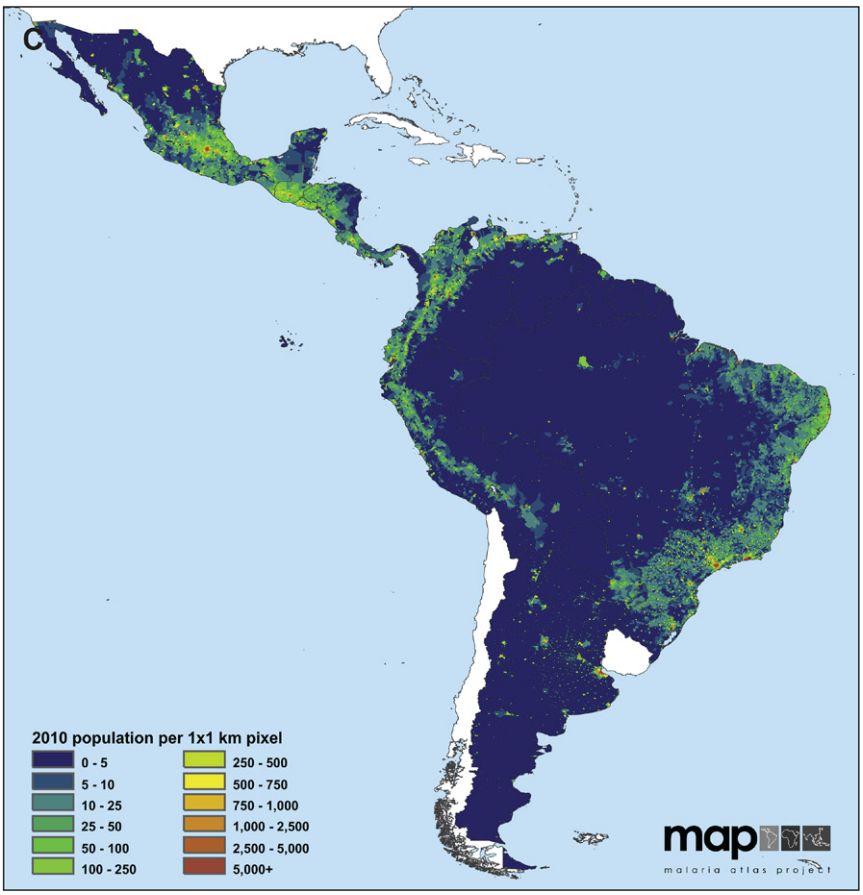


Figure 1.8, cont'd

DPR). Five countries reported data up to the year 2010 (Bhutan, Georgia, Nepal, Sri Lanka and Thailand) with the majority (14/23) providing data up to 2008. India and China's last year of reporting was 2007.

Transmission was estimated to span 9 million square kilometres of land in Asia (around 45% of the total land area in Asia). The majority of this area (61%; 5.60 million km^2) was at unstable risk. The areas at risk in India and China covered 2.97 and 2.82 million km^2 each and, therefore, comprised 63% of the total area at risk in Asia (Panel A1 of Fig. 1.3).

Estimating endemicity. There were 2665 records of prevalence data collected from the region. The three most data-rich countries were Viet Nam (n = 657), Afghanistan (n = 493) and Bangladesh (n = 365), as illustrated in

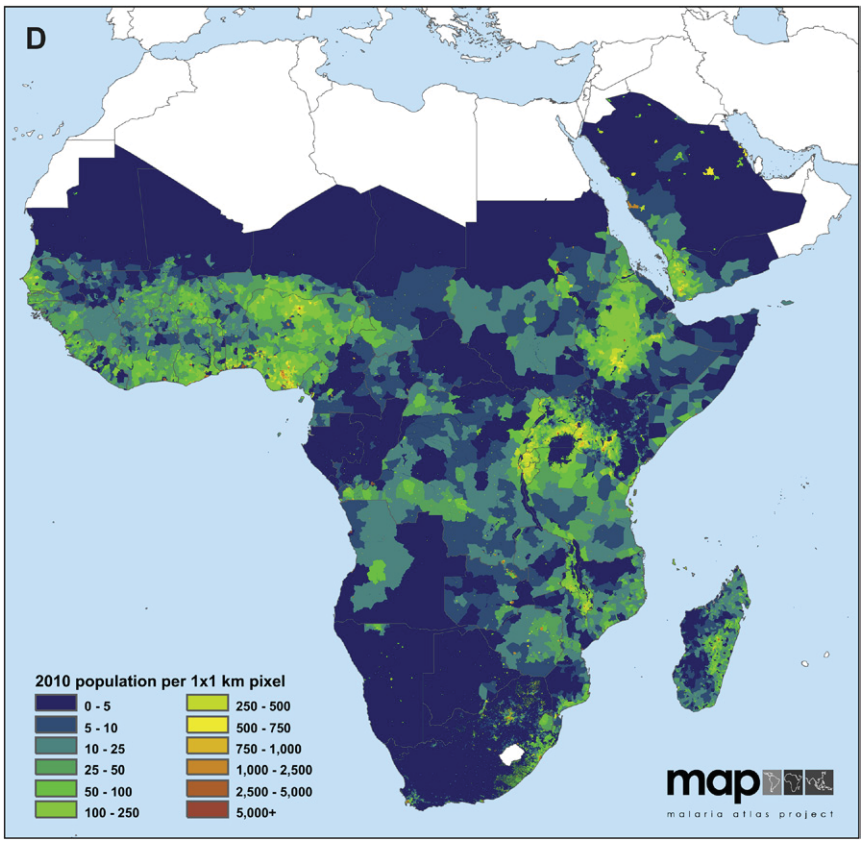


Figure 1.8, cont'd

Panel A1 of Fig. 1.3.The predicted prevalence estimates in stable transmission areas in Asia were highly heterogeneous. Areas with point estimates above 7% were found in small parts of India, Myanmar and Thailand. Regions with *Pv*$PR_{1–99}$ estimates between 3.5 and 7% were found in large areas of India and pockets of China, Myanmar, Thailand, Cambodia and Lao People's Democratic Republic (Lao PDR). Uncertainty maps reveal that areas of high uncertainty correspond with high transmission areas. These areas also had sparse prevalence survey data. Relative to its size, India had very little prevalence data available. Myanmar and Thailand also had large regions with stable transmission but with few surveys to support the prevalence estimates. While predictions on the Korean peninsula were made with relatively high certainty (Fig. 1.4A1), there were no survey data available from Korea DPR. Known prevalence values from this region as well as from India, Myanmar,

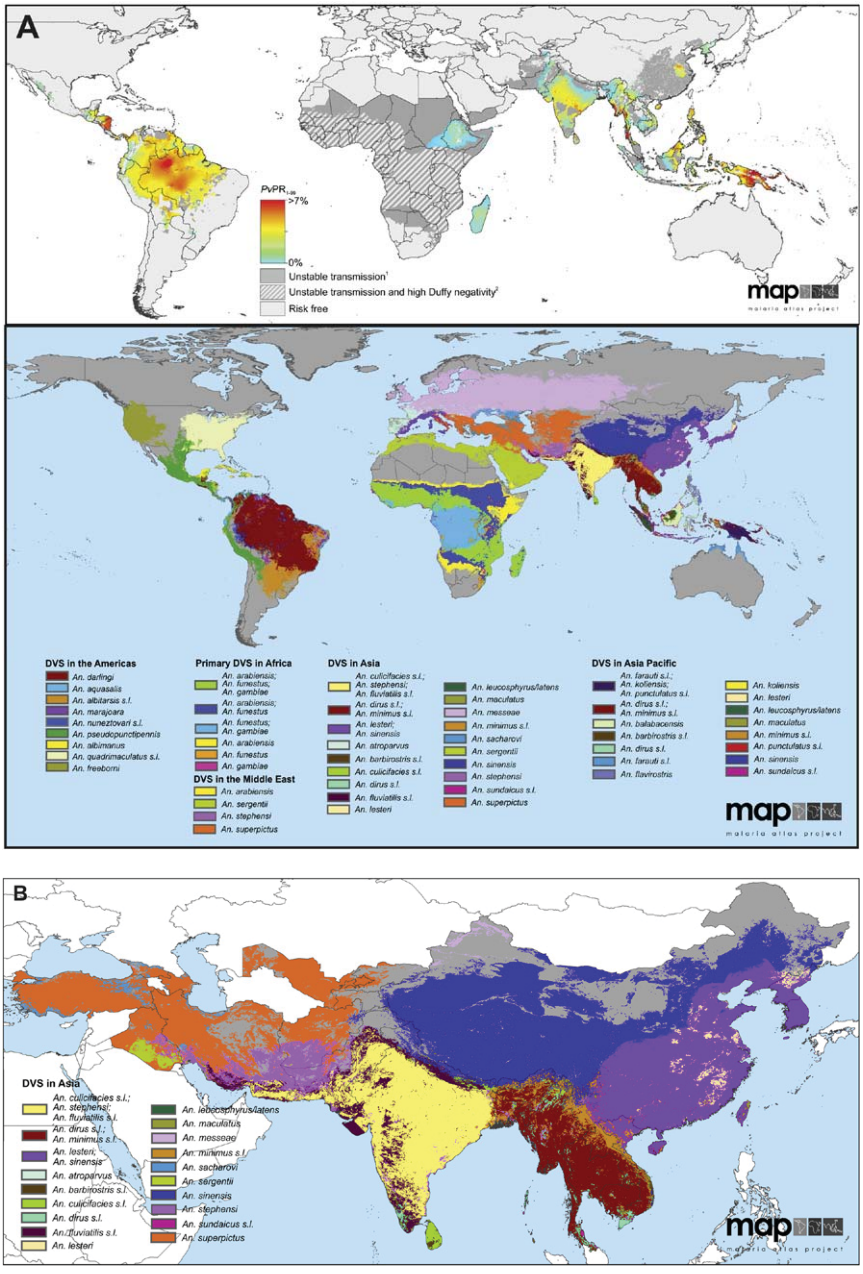


Figure 1.9 Distribution of dominant vector species globally and in Asia, Asia-Pacific, the Americas and Africa+.

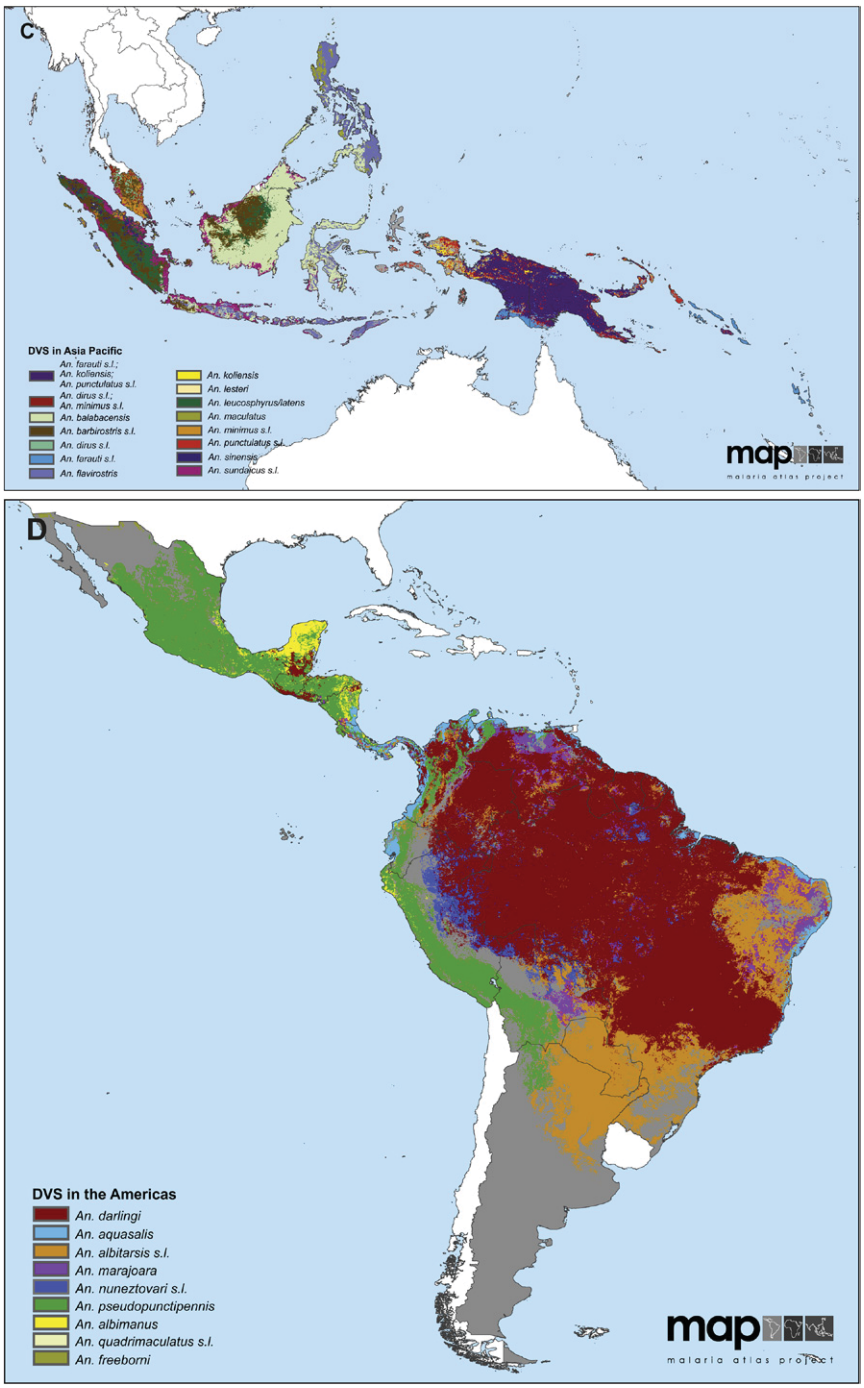


Figure 1.9, cont'd

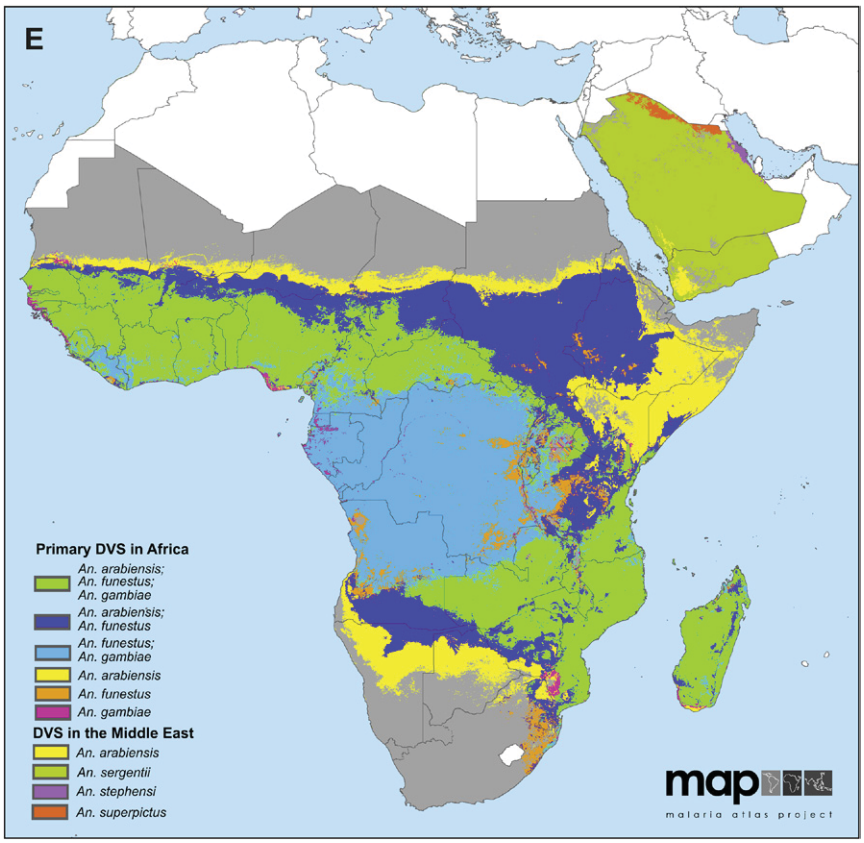


Figure 1.9, cont'd The global distribution of the main dominant vector species is shown in Panel A alongside the predicted endemicity of *P. vivax* in 2010. The distribution of the primary vector species are also shown at the regional level: Asia (B), Asia-Pacific (C), the Americas (D) and Africa+ (non-endemic countries are shaded white). The legend for Africa+ is disaggregated in to Africa and the Middle east because only the primary vector species in Africa are shown in the map. For colour version of this figure, the reader is referred to the online version of this book.

Thailand and parts of China with stable vivax transmission would improve the certainty of $Pv\text{PR}_{1\text{–}99}$ predictions for the area as a whole. Regions with low endemicity and high density of surveys, such as Afghanistan, Cambodia and the small region with stable transmission in Turkey, were predicted with high certainty. The population-weighted uncertainty map differs substantially for parts of this region (Fig. 1.4A2). Incorporating population estimates allows weighting to highlight areas where high uncertainty and large populations coincide. Thus, sparsely populated areas of Myanmar, Thailand,

Cambodia, Lao PDR, Viet Nam and Korea DPR displayed small values of this index despite uncertain predictions of prevalence. Conversely, values for highly populated areas across India, parts of Pakistan and the stable transmission region in China, were inflated substantially. Predictions in this region would be best improved, therefore, with more detailed prevalence data from these population centres.

Population at risk. The PAR in Asia was, by far, the largest of all of the regions due to large areas of risk and high population densities. There were over two billion people living at risk in this region in 2010, which constituted 82% of the total global PAR. The greatest PAR was found in India, China and Pakistan with 1.13 billion, 462 and 169 million at risk, respectively. India's PAR comprised more than half (55%) of the Asia-region PAR. Together, China and Pakistan made up another 30% (23 and 8%, respectively), indicating that 85% of the region's PAR was attributable to three nations. These nations also had the largest populations in this region. Compared to India and Pakistan, a smaller proportion of China's overall population was living at risk of *P. vivax*. Of China's estimated 1.4 billion people, 33% were at risk of *P. vivax*, whereas 95% of India's 1.2 billion and 90% of Pakistan's 188 million people were living at some level of risk. When we consider both levels of risk (unstable and stable transmission), we find that 60%, or 1.2 billion, of the PAR in Asia experienced unstable risk, indicating that over 800 million individuals in this region lived in stable transmission areas. Countries devoid of stable transmission areas were Iraq, Kyrgyzstan and Uzbekistan. Countries with the greatest population at stable risk in this region were again India (642.07 million), Pakistan (43.82 million) and China (31.92 million). Other countries that had large (>10 million) PAR living in stable transmission areas were Korea DPR (21.07 million), Myanmar (20.48 million), Thailand (17.11 million) and Bangladesh (16.04 million). Ninety eight percent of the population living in stable transmission areas was, therefore, found in 30% (7 out of 24) of the countries in Asia.

Vectors. Of the 35 potential *P. vivax* vector species found within Asia as a whole (Table 1.2), there are 19 DVS (Sinka et al., 2011) of which, at least seven are considered to be species complexes (Harbach, 2004). Nine DVS are found solely in the Asia region, whilst the remaining 10 have distributions extending into the Asia-Pacific region.

A total of 10,667 unique occurrence points were georeferenced for the 19 DVS across Asia. Records of species occurrence obtained from observations recorded to point (≤10 km^2) and wide area (10–25 km^2) locations

(sites) were used for analyses. In some cases, more than one occurrence point was obtained for a single site. The greatest number of points were found for the *Anopheles culicifacies* complex (*n* = 1568) followed by the *Anopheles subpictus* (*n* = 1143) and *Anopheles barbirostris* (*n* = 1064) complexes. From the 15 Asian *P. vivax* malaria endemic countries (*Pv*MECs), where anopheline occurrence data were found, data were obtained from 5388 sites. The greatest number of sites were in Myanmar (*n* = 1830), India (*n* = 1529) and China (*n* = 355). There was only one reported occurrence located in Turkey.

The predicted distributions of the DVS in Asia resulted in a complex multi-species map, where the main vector species overlap over large areas of land and exist independently in small areas (Fig. 1.9B). It is beyond the scope of this review to describe the individual distributions and bionomics of the 19 potential DVS in this region, however, a complete discussion of the ranges and bionomics of the global DVS may be found elsewhere (Sinka et al., 2010a, 2010b, 2011, 2012). The DVS described in the following paragraphs are those that were deemed to be primary malaria vectors by the project's technical advisory group (TAG), which are therefore shown prominently in Fig. 1.9A, and showed conclusive evidence for the ability to transmit *P. vivax* (Table 1.2). The potential primary DVS that met these criteria were the *An. culicifacies* complex (Culicifacies Complex; *An. culicifacies s.l.*), the *Anopheles dirus* complex (Dirus Complex; *An. dirus s.l.*), the *Anopheles minimus* complex (Minimus Complex; *An. minimus s.l.*), *Anopheles sinensis*, *Anopheles stephensi* and *Anopheles superpictus*.

Anopheles culicifacies s.l. is a prominent vector across the Indian subcontinent, found in sympatry with *An. stephensi* and *Anopheles fluviatilis s.l.* Species B of the Culicifacies Complex is considered a non-vector of *P. falciparum*, currently attributed to highly zoophilic behaviour. However, there is evidence shown in Table 1.2 that the species is also partially refractory to *P. vivax* and, hence, perhaps to malaria parasites overall (Adak et al., 2005; Vijay et al., 2011). Evidence of wild-infected populations and experimental infections suggest that at least species A and C of the complex are primary vectors of malaria on the Indian subcontinent. Bangladesh, Myanmar, Thailand, Cambodia, Lao PDR and Viet Nam are dominated by the Dirus and Minimus Complexes (Fig. 1.9B). The range of the *An. sinensis* complex covers the majority of China, where it is shown to co-exist with *Anopheles lesteri* along the eastern areas of the country and the Korean peninsula. However, the overlapping distributions of these two vector species may be an artefact of mis-identification in some areas rather than true sympatry

Table 1.2 Known and potential vector species of *Plasmodium vivax*

Species, species complex* or group	Distribution and bionomics reviewed by MAP†	*P. vivax* vector (Yes/No)‡	Notes and reference(s)
Asia			
An. atroparvus van Thiel	Yes	Yes(?)	*An. atroparvus* has been shown to be experimentally infected with *P. vivax* by humans (Daskova and Rasnicyn, 1982) and (Collins et al., 1980; Collins et al., 2009). This species has been shown to be refractory to *P. falciparum* infection (de Zulueta et al., 1975; Daskova and Rasnicyn, 1982; Romi et al., 2001).
An. baimaii Sallum & Peyton	Yes (part of the *An. dirus* complex)	Yes(?)	*P. vivax* and *P. falciparum* are commonly found in *An. baimaii* (Sallum et al., 2005; Obsomer et al., 2007), but original sources do not differentiate the members of the Dirus Complex (Prakash et al., 2001).
An. culicifacies complex	Yes	Yes	Naturally *P. vivax*-infected *An. culicifacies* were found in Madhya Pradesh (Species C and D) (Subbarao et al., 1992) and Uttar Pradesh (Species A) (Subbarao et al., 1988) in India, and in Sri Lanka (Amerasinghe et al., 1991b). *An. culicifacies* have been experimentally infected with *P. vivax* in from infected monkeys (Collins et al., 2009) and laboratory colonies established from wild caught *An. culicifacies* (Species A, B and C) were infected with vivax from blood drawn from infected humans (Adak et al., 1999). Species A and C showed relatively high oocyst infections but Species B did not. Sporozoite infections were relatively high in Species A compared to Species C and negligible in Species B. There is evidence that Species B may be refractory to *P. vivax* infection (Adak et al., 2006; Vijay et al., 2011).
An. fluviatilis complex	Yes	Yes(?)	Sporozoites were detected in specimens collected in Iran and the species is said to transmit both *P. vivax* and *P. falciparum* (Eshghi et al., 1976). Both oocytes and sporozoites were found in an *An. fluviatilis* (Species T) laboratory colony infected by human volunteers under controlled conditions (Adak et al., 2005).

Hyracanus Group	No	Yes	Wild *P. vivax*-infected Hyracanus Group mosquitoes were found in Assam, India (Rattanarithikul et al., 1996; Prakash et al., 2004).
An. messeae Falleroni	Yes	Yes(?)	*An. messeae* was implicated as a vector of *P. vivax* in eighteenth and nineteenth century Finland (Hulden et al., 2008; Hulden, 2009) and the authors refer to evidence of *An. messeae* as a vector in Russia (Sokolova and Snow, 2002), but only indirect evidence linking malaria incidence with vector density was provided.
An. nimpe Nguyen, Tran & Harbach	No (part of the Hyracanus Group)	Yes(?)	*An. nimpe* was implicated as a vector (Nguyen et al., 2000) of *P. vivax* and *P. falciparum* after being sampled in coastal Viet Nam, but direct evidence of infection was not available.
An. philippinensis-nivipes complex	No	Yes	Wild-infected *An. philippinensis-nivipes s.l.* specimens were found to be positive for *P. vivax* in Assam, India (Prakash et al., 2004).
An. pulcherrimus Theobald	No	Yes	*An. pulcherrimus* is thought to be the main vector in Afghanistan (Brooker et al., 2006). *Plasmodium vivax* (VK210 and VK247 subtypes) was detected in specimens collected in Afghanistan (Rowland et al., 2002) and *P. vivax* circumsporozoite proteins were detected though enzyme-linked immunosorbent assay (ELISA) in wild-infected wild-infected mosquitoes (the vast majority of which were *An. pulcherrimus*), but the authors do not explicitly state that *P. vivax* was found in *An. pulcherrimus* (Faulde et al., 2007).
An. sacharovi Favre	Yes (part of the Maculipennis Subgroup)	Yes	*P. vivax* circumsporozoite antigens were detected in wild populations sampled in Turkey (Simsek et al., 2010). *P. vivax* oocysts and sporozoites were detected in the salivary glands of a laboratory colony of *An. sacharovi* fed on vivax infected humans (Kasap, 1990).
An. sawadwongporni Rattanarithikul & Green	No (part of the Maculatus Group)	Yes	*P. vivax* (VK247) was detected in specimens collected in Thailand (Coleman *et al.*, 2002).
An. sergentii (Theobald)	Yes	Yes	*P. vivax* circumsporozoite proteins were detected in *An. sergentii* specimens sampled from desert oases in Egypt (Kenawy et al., 1990).

Continued

Table 1.2 Known and potential vector species of *Plasmodium vivax—cont'd*

Species, species complex* or group	Distribution and bionomics reviewed by MAP†	*P. vivax* vector (Yes/No)‡	Notes and reference(s)
An. stephensi Liston	Yes	Yes	*P. vivax* (VK210 and VK247) was detected in specimens collected in Afghanistan (Rowland et al., 2002). *An. stephensi* has been experimentally infected with *P. vivax* from infected monkeys (Basseri et al., 2008; Collins et al., 2009) and humans (Adak et al., 2005); both oocysts and sporozoites found were found in the mosquitoes.
An. superpictus Grassi	Yes	Yes	*P. vivax* (VK247) was detected in specimens collected in Afghanistan (Rowland et al., 2002). Laboratory colonies *An. superpictus* have been infected with *P. vivax*-infected humans; oocysts and sporozoites were in salivary glands (Kasap, 1990). This species seems to be generally implicated as a vivax vector in other articles.
An. varuna Iyengar	No (part of the Aconitus Subgroup)	Yes	Wild-infected mosquitoes were found to be positive for *P. vivax* in Assam, India (Prakash et al., 2004).
Asia and Asia-Pacific			
An. aconitus Dönitz	Yes	Yes	*P. vivax* sporozoites have been detected (by ELISA) in wild captured *An. aconitus* in Sri Lanka (Amerasinghe et al., 1991b). Laboratory studies revealed that *An. aconitus* forms B and C were susceptible to *P. vivax* and *P. falciparum*, and that *An. aconitus* form C was susceptible only to *P. vivax* (Junkum et al., 2005).
An. annularis van der Wulp	Yes	Yes	*P. vivax* sporozoites have been detected (by ELISA) in wild captured *An. annularis* in Sri Lanka (Amerasinghe et al., 1991b) and India (Prakash et al., 2004).
An. barbirostris complex	Yes	Yes	*P. vivax* circumsporozoite proteins were found in specimens collected in Thailand with infectivity rates of 0.24% (Rattanarithikul et al., 1996) and 4.8% (Frances et al., 1996).

An. campestris Reid	No	Yes	*P. vivax* (VK210) was detected in specimens collected in Thailand (Coleman et al., 2002).
An. dirus complex	Yes	Yes	*P. vivax* circumsporozoite proteins have been detected in an *An. dirus* specimen collected in Thailand (Baker et al., 1987). The vectorial capacity of *An. dirus* for *P. vivax* has been estimated (Prakash et al., 2001). Experimental infection from humans (Wirtz et al., 1985), monkeys (Collins et al., 2009) and membrane feeding has been shown (Coleman et al., 2004; Junkum et al., 2005).
An. donaldi Reid	No	Yes	*P. vivax* circumsporozoite proteins have been detected in *An. donaldi* specimens sampled in Malaysia (Seng et al., 1999). *An. donaldi* was successfully infected with *P. vivax* and *P. falciparum* sporozoites in the lab and was capable of transmitting both species to man (Hardin et al., 1973; Harrison and Scanlon, 1975).
An. hodgkini Reid	No	Yes	*P. vivax* was detected in specimens collected in Thailand (VK247) (Coleman et al., 2002).
An. karwari (James)	No	Yes	*P. vivax* circumsporozoite proteins in *An. karwari* specimens collected in Thailand (Frances et al., 1996).
An. kochi Dönitz	No	Yes	Wild *P. vivax*-infected *An. kochi* mosquitoes were found in Assam, India (Prakash et al., 2004).
An. lesteri Baisas & Hu (synonymous with *An. anthropophagus* Xu & Feng)	Yes	Yes(?)	*An. lesteri* (*An. anthropophagus*) were experimentally infected with *P. vivax* from humans in Republic of Korea (Shin et al., 2002). Human laboratory feedings showed *An. lesteri* to be a highly competent vector (Joshi et al., 2009).
An. letifer Sandosham	No	Yes(?)	Specimens collected in Sarawak, Malaysia were found to be sporozoite positive and although the species of the parasite was not identified, >90% of the malaria infections in the region at the time were due to *P. vivax* (Chang et al., 1997). Implied transmission, but with no direct evidence, is described in other references (Rahman et al., 1997; Fryauff et al., 1998).

Continued

Table 1.2 Known and potential vector species of *Plasmodium vivax—cont'd*

Species, species complex* or group	Distribution and bionomics reviewed by MAP†	*P. vivax* vector (Yes/No)‡	Notes and reference(s)
An. leucosphyrus Dönitz & *An. latens* Sallum & Peyton	Yes (sister species in the Leucosphyrus Complex)	Yes	*P. vivax* circumsporozoite proteins have been detected in *An. leucosphyrus* specimens sampled in Malaysia (Seng et al., 1999). However, the species sampled was most likely *An. latens*, which was commonly misidentified as its sister species *An. leucosphyrus* (Sallum et al., 2005).
An. ludlowae (Theobald)	No	?	No references were found to confirm that *An. ludlowae* transmits *P. vivax*, but it is a potential vector in areas where vivax malaria is found (e.g. the Philippines and Indonesia) (Lien et al., 1977; Wooster and Rivera, 1985).
An. minimus complex	Yes	Yes	*P. vivax* was detected in specimens collected in Thailand (VK210 and VK247) (Coleman et al., 2002) and India (VK247) (Prakash et al., 2004). *An. minimus* has also been experimentally infected with *P. vivax* from monkeys (Rattanarithikul et al., 1996; Collins et al., 2009).
An. nigerrimus Giles	No	Yes	*P. vivax* (VK210) circumsporozoite proteins have been detected in *An. nigerrimus* sampled in China (Alam et al., 2010) and *An. nigerrimus* has been incriminated as a vector there (WHO, 1986).
An. sinensis Wiedemann	Yes (part of the Hyracanus Group)	Yes	*P. vivax* (VK210 and VK247) was detected in specimens collected in the Republic of Korea (Lee et al., 2002). Deliberate infection was also observed in laboratory conditions (Shin et al., 2002).
An. subpictus complex	Yes	Yes	*P. vivax* circumsporozoite proteins have been detected in *An. subpictus* sampled in Sri Lanka (Amerasinghe et al., 1991a).
An. sundaicus complex	Yes	Yes(?)	*An. sundaicus* is described as a vector of primarily *P. falciparum* malaria, implying that it will also transmit *P. vivax* (Am et al., 1993) and *P. vivax* oocysts and sporozoites have been found in infected laboratory colonies (Adak et al., 2005).

An. tessellatus Theobald	No	Yes	*P. vivax* circumsporozoite proteins were detected in *An. tessellatus* species sampled in Sri Lanka (Amerasinghe et al., 1991a).
An. vagus Dönitz	No	Yes	*P. vivax* sporozoites have been detected (by ELISA) in wild captured *An. aconitus* in Sri Lanka (Amerasinghe et al., 1991a) and India (Prakash et al., 2004).
Asia-Pacific			
An. balabacensis Baisas	Yes (part of the *An. leucosphyrus* complex)	Yes(?)	Sporozoites were found in specimens collected in Central Java, Indonesia (Barcus et al., 2002) and Palawan, Philippines (Schultz, 1992). Laboratory reports indicate transmission of *P. vivax* to non-human primates (Collins et al., 1980).
An. farauti complex	Yes	Yes	*P. vivax* circumsporozoite proteins were found in specimens collected in Papua New Guinea (Burkot et al., 1988).
An. flavirostris (Ludlow)	Yes	?	*An. flavirostris* has been described as the 'principal vector' in the Philippines, where *P. vivax* and *P. falciparum* malaria occur (Foley et al., 2003) and has been found with sporozoites in Malaysia (Hii et al., 1985) and the Philippines (Oberst et al., 1988), but the parasite species was not given.
An. koliensis Owen	Yes (part of the Punctulatus Group)	Yes	*P. vivax* circumsporozoite positivity of 0.76% was reported for wild-infected *An. koliensis* in Papua New Guinea (Attenborough et al., 1997).
An. maculatus Theobald/Maculatus Group	Yes	Yes	*P. vivax* (VK210 and VK247) was detected in *An. maculatus* specimens collected in Thailand (Coleman et al., 2002). *An. maculatus* has been experimentally infected with *P. vivax* from monkeys (Wirtz et al., 1985; Collins et al., 2009) with a reported infectivity of 4.8% (Collins et al., 1980).
An. punctulatus complex	Yes	Yes	*P. vivax* circumsporozoite antigens were detected in wild populations of the *An. punctulatus* complex sampled in Papua New Guinea (Burkot et al., 1988).

Continued

Table 1.2 Known and potential vector species of *Plasmodium vivax—cont'd*

Species, species complex* or group	Distribution and bionomics reviewed by MAP†	*P. vivax* vector (Yes/No)‡	Notes and reference(s)
The Americas			
An. albimanus Wiedemann	Yes	Yes	*P. vivax* sporozoites have been detected (by ELISA) in wild captured *An. albimanus* in Mexico (Ramsey et al., 1994) and in specimens experimentally infected from humans (Wirtz et al., 1985). Experimental feeding on monkeys revealed that *An. albimanus* was much more susceptible to 'New World' *P. vivax* (*P. vivax* specimens from Central and South America; 21.2% infection rate), than 'Old World' *P. vivax* (0.4% infection rate) (Li et al., 2001). Other laboratory studies using monkeys showed infectivity rates of only 0.6–0.7% (Collins et al., 1980).
An. albitarsis complex	Yes	Yes	*P. vivax* infection was detected (by dissection to find presence and ELISA to determine *Plasmodium* species) in wild captured *An. albitarsis* in Northern Brazil (de Arruda et al., 1986). Mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers; sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991). *P. vivax* infection was also seen in mosquitoes experimentally fed on infected monkeys (Collins et al., 1985).
An. aquasalis Curry	Yes	Yes	*P. vivax* sporozoites have been detected by ELISA in wild captured *An. aquasalis* in Brazil (Povoa et al., 2003). *An. aquasalis* has been experimentally infected with *P. vivax* from infected humans (da Silva et al., 2006).
An. argyritarsis Robineau-Desvoidy	No	Yes(?)	*An. argyritarsis* naturally infected with *P. vivax* have been found in Argentina and the species is identified as a principal vector in parts of Brazil. However, there is some debate whether the species sampled were *An. argyritarsis* or misidentified *An. darlingi* specimens (Linthicum, 1988).

An. bellator Dyar & Knab	No	Yes(?)	*An. bellator* has been found with *P. vivax* oocysts or both oocysts and sporozoites in Brazil (Deane, 1986) and have been experimentally infected from an infected human (Rozeboom and Laird, 1942).
An. benarrochi Gabaldón	No	Yes(?)	Naturally *P. vivax* infected *An. benarrochi* were observed in Peru (Flores-Mendoza et al., 2004), however when mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers, no sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991).
An. braziliensis (Chagas)	No	Yes(?)	Naturally *P. vivax* infected *An. braziliensis* were observed in Brazil (da Silva-Vasconcelos et al., 2002), however when mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers; no sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991).
An. cruzii Dyar & Knab	No	Yes	*P. vivax* (VK247) infectivity rates of 0.086–0.179% in wild-infected mosquitoes were reported in Brazil (Branquinho et al., 1997).
An. darlingi Root	Yes	Yes	*An. darlingi* specimens were found to be naturally infected with *P. vivax* in Amapa (Conn et al., 2002) and Para States (by dissection to find presence and ELISA to determine *Plasmodium* species) (de Arruda et al., 1986) in Brazil and in French Guiana (Girod et al., 2008). Mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers; sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991).
An. deaneorum Rosa-Freitas	No (part of *An. albitaris* complex)	Yes	*An. deaneorum* is a sibling of *An. albitarsis* complex and may be an important vector in Amazonian Brazil (Conn et al., 2002). This species can be experimentally infected by both *P. vivax* and *P. falciparum* (Klein et al., 1991; Senise et al., 2006).
An. freeborni Aitken	Yes	Yes(?)	*An. freeborni* has been experimentally infected with *P. vivax* from humans (Burgess and Young, 1950) and monkeys (Collins *et al*., 2009).

Continued

Table 1.2 Known and potential vector species of *Plasmodium vivax*—*cont'd*

Species, species complex* or group	Distribution and bionomics reviewed by MAP†	*P. vivax* vector (Yes/No)‡	Notes and reference(s)
An. hermsi Barr & Guptavanj	No	Yes(?)	*An. hermsi* has been implicated in outbreaks of *P. vivax* in California (Maldonado et al., 1990; Ginsberg, 1991) and *An. hermsi* has been experimentally infected with vivax from monkeys (Collins et al., 2009).
An. marajoara Galvão & Damasceno	Yes (part of the *An. albitarsis* complex)	Yes	Specimens of *An. marajoara* (a sibling of *An. albitarsis s.l.*) collected in Amapa State, Brazil were found to be naturally infected with vivax and may be a superior vector for this parasite over *An. darlingi* in this location (Conn et al., 2002).
An. mediopunctatus Lutz	No	Yes(?)	Mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers and *P. vivax* sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991), however other literature states that this species was present in low numbers (de Arruda et al., 1986) or was a non-vector (Vittor et al., 2006).
An. nuneztovari complex	Yes	Yes	*P. vivax* infection was detected in wild captured *An. nuneztovari* by dissection and ELISA was used to determine *Plasmodium* species in Para State, Northern Brazil (de Arruda et al., 1986).
An. oswaldoi Peryassú	No	Yes	*An. oswaldoi* was incriminated as a vector of *P. vivax* in Southern Colombia (Quinones et al., 2006) and Northern Brazil (de Arruda et al., 1986). Mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers and *P. vivax* sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991).

An. pseudopunctipennis complex	Yes	Yes(?)	*An. pseudopunctipennis* is implied as the vector of *P. vivax* in Bolivia (Lardeux et al., 2007) and has been experimentally infected with *P. vivax* with variable infectivity rates from different vivax strains (Rodriguez et al., 2000). While natural infections with *P. vivax* have been reported from Argentina and Mexico (Warren et al., 1980), there is also evidence of *Plasmodium* sporozoites being absent from *An. pseudopunctipennis* populations in Belize (Achee et al., 2000) and El Salvador (Warren et al., 1980) and experiments have shown *An. pseudopunctipennis* to be refractory to *P. vivax* (VK210) (Gonzalez-Ceron et al., 2007).
An. punctimacula Dyar & Knab	No	Yes	Experimental man-to-mosquito transmission has been confirmed, and naturally infected *An. punctimacula* have been found in Panama (Simmons, 1937; Quinones et al., 2006; Ulloa et al., 2006; Loaiza et al., 2008).
An. quadrimaculatus complex	Yes	Yes(?)	Species of the *An. quadrimaculatus* are implied to be vectors of *P. vivax* malaria (Jensen et al., 1996) and has been infected in the laboratory from *P. vivax*-infected monkeys (Collins et al., 2009).
An. triannulatus Neiva & Pinto	No	Yes	*P. vivax* was detected in wild captured *An. triannulatus* in Para State, Northern Brazil (de Arruda et al., 1986). Mosquito oocyst infection was observed in laboratory conditions after feeding on infected human volunteers and *P. vivax* sporozoites were found in the salivary glands approximately two weeks after feeding (Klein et al., 1991).
Africa+			
An. arabiensis Patton	Yes (part of the *An. gambiae* complex)	Yes	*P. vivax* sporozoites have been detected (by ELISA) in wild captured *An. arabiensis* in Southern Ethiopia (Taye et al., 2006) and Madagascar (Fontenille et al., 1990). *An. arabiensis* have been experimentally infected with *P. vivax* from infected monkeys (Collins et al., 2009).
An. funestus Giles	Yes	Yes	*P. vivax* (VK247) circumsporozoite proteins were found in *An. funestus s.s.* specimens collected in Kenya (Ryan et al., 2006).

Continued

Table 1.2 Known and potential vector species of *Plasmodium vivax*—*cont'd*

Species, species complex* or group	Distribution and bionomics reviewed by MAP†	*P. vivax* vector (Yes/No)‡	Notes and reference(s)
An. gambiae Giles/ *An. gambiae* complex	Yes	Yes	*P. vivax* (VK247) circumsporozoite proteins were found in *An. gambiae s.s.* specimens collected in Kenya (Ryan et al., 2006) and *An. gambiae* has been infected with *P. vivax* in laboratory settings from infected monkeys (Wirtz et al., 1985; Collins et al., 2009).
An. labranchiae Falleroni	Yes	Yes	*An. labranchiae* has been stated to be a vector of *P. vivax* (Lindsay and Thomas, 2001; Romi et al., 2001), and a vectorial capacity for *P. vivax* in *An. labranchiae* has been estimated (Romi et al., 1997; Romi et al., 2001).
An. melas Theobald	Yes (part of the *An. gambiae* complex)	?	No natural *P. vivax* infection has been found (Moreno et al., 2004; Bigoga et al., 2007), but that is perhaps because the indigenous populations, rather than the mosquito, are refractory to *P. vivax* (Moreno et al., 2004).
An. merus Dönitz	Yes (part of the *An. gambiae* complex)	?	No natural *P. vivax* infection has been found; *An. melas* sampled in Tanzania were found not to be infected with *P. vivax* (Temu et al., 1998).
An. moucheti Evans	Yes	?	There is no indication of *P. vivax* infection, but this may be due to lack of the parasite within its distribution (Antonio-Nkondjio et al., 2005).
An. multicolor Cambouliu	No	Yes(?)	No *P. vivax* parasites were found in *An. multicolor* collected in desert oases in Egypt (El Said et al., 1983; Kenawy et al., 1990; Morsy et al., 1995). However, the authors infer the potential for transmission and refer to experimental transmission of *P. vivax* by *An. multicolour* under laboratory conditions.

An. nili complex	Yes	?	There is no indication of *P. vivax* infection by ELISA in specimens sampled in Ethiopia (Nigatu et al., 1994), but this may be due to lack of the parasite within its distribution (Antonio-Nkondjio et al., 2005).
An. pharoensis Theobald	No	Yes	*P. vivax* circumsporozoite proteins were detected by ELISA in *An. pharoensis* sampled in southwest Ethiopia (Nigatu et al., 1994).

Vector species were identified during the preparation for a series of articles (Hay et al., 2010c; Sinka et al., 2010a, 2010b, 2011, 2012) regarding the dominant vector species of human malaria. The potential for 71 anopheline species and species complexes to transmit *P. vivax* are shown alphabetically by region.
*Harbach (2004).
†Sinka et al. (2010a, 2010b, 2011, 2012).
‡Yes = evidence of wild/naturally infected mosquitoes; Yes(?) = evidence of laboratory infected mosquitoes or secondary information implicating species as a vector; ? = no information found to confirm species can carry *P. vivax*, but nothing to state it cannot, or unconfirmed or unclear references to species *P. vivax* vector status.

(Sinka et al., 2012). In addition to its sympatric range with *An. culicifacies s.l.* and *An. fluviatilis s.l.* over India, *An. stephensi* was predicted to occur in Afghanistan and Pakistan, where it was sympatric with *An. superpictus*.

The sources for the following summary can be found in Sinka et al. (2011) and Sinka et al. (2010a).

The DVS cover a large range of varying ecological niches and the presence of a large number of species complexes in the region means that variation in behaviours are often seen. Indeed, behavioural variability within a species is also common, for example, *Anopheles annularis* has a range extending across India, down through South-East Asia, across many of the Indonesian islands and Timor Island. However, *An. annularis* only has a focal role in malaria transmission in selected areas of India. Elsewhere it is considered of little importance.

The Indian DVS have overlapping, sympatric ranges and include a number of species complexes. For example, there are five sibling species in the *An. culicifacies* complex (species A, B, C, D and E) and although species A, C, D and E are all reported to be vectors of malaria in India (Sinka et al., 2011), species E is the most important because of its highly anthropophilic and endophilic behaviour. Members of the complex are found across the plains and up into the highlands (Iyengar, 1954; Barik et al., 2009). Larval habitats include a variety of man-made and naturally occurring water bodies and tolerance of brackish water has been observed in species E (Roberts, 1996). Species C shows plasticity in its behaviour, being found in both forested and deforested areas. Varied bionomics are therefore observed both within and among members of this species complex.

The other primary Indian DVS are *An. fluviatilis s.l.* and *An. stephensi*. Variable behaviour is observed amongst the three members of the Fluviatilis Complex (species S, T and U). Within the complex, there are reports of anthropophilic, zoophilic and exo- and endophagic populations, biting at dusk and during the night. *Anopheles stephensi* is primarily zoophilic and endophagic, and both *An. fluviatilis s.l.* and *An. stephensi* rest primarily indoors. *Anopheles stephensi* is unusual amongst *Anopheles* species in that it appears to be able to use virtually any water condition/habitat as a larval site, hence, its success in urban areas. Less is known regarding the larval habitats of *An. fluviatilis s.l.*, but the complex is associated with slow-moving streams or river margins.

Amongst the northern Asia DVS, *An. sinensis s.l.* and *An. superpictus* are considered potential primary vectors of *P. vivax*. In Thailand, members of the *An. sinensis* complex are considered zoophilic and exophagic, and therefore,

of little vectorial importance. In the Korean Peninsula, on the other hand, *An. sinensis s.l.* shows a propensity for biting humans and has been identified as a primary vector (Lee et al., 2001). The larvae of *An. sinensis* complex are associated with lowland, shallow, freshwater habitats. Adult females may hibernate to overwinter from October to April when the temperatures drop below 19 °C (Chow, 1970). Little is known regarding the biting habits of *An. superpictus*, particularly regarding host preference, but they are generally considered to be exophagic. The larvae are commonly found in shallow, flowing, fresh, sun-lit water. However, the species can adapt to human influences by using man-made pits and holes.

The primary DVS in the south-eastern part of Asia are the *An. dirus* and *An. minimus* complexes, both of which contain highly competent vector species. Within the Dirus Complex, which contains seven sibling species, again there is some variability in behaviours (depending on location and sibling species) but the main vectors of the complex (*An. dirus s.s.* and *Anopheles baimaii*) are both highly anthropophilic and will feed both indoors and out. Biting activity appears highly sibling/species dependent with *Anopheles scanloni* (a focal vector within the complex) biting at dusk, *An. dirus* biting in the evening (between 2000 h and 2300 h) and *An. baimaii* biting as late as 0200 h. In general, outdoor resting both before and after feeding has been reported for this species complex. The larvae of this complex inhabit small, temporary, freshwater bodies and are therefore most abundant during the rainy (monsoon) season. *Anopheles minimus s.l.* has three siblings, only two of which (*An. minimus s.s.* and *Anopheles harrisoni*) are considered viable contemporary vectors. Variability in reported behaviour may be a consequence of misidentification or lack of differentiation between these two species, although both are also considered opportunistic and plastic in their behaviour. *Anopheles minimus s.s.* is potentially more anthropophilic, but also adaptable depending on host availability. Overall, *An. harrisoni* is considered more exophagic, exophilic and zoophilic than its sibling. Minimus Complex larvae are typically associated with small- or moderate-sized streams with slow, clear, cool and partially shaded water. *Anopheles minimus s.s.* has been observed in a wider range of habitats from canopied forests to open fields. *Anopheles harrisoni*, on the other hand, is more commonly associated with agricultural areas. Further studies that utilise molecular assays to differentiate sibling species are needed to understand the varied behaviour and habitat preference of *An. minimus s.s.* and *An. harrisoni*.

It is clear that the behaviour of the Asian vectors is complicated, variable and, as yet, not well understood. However, many of the species do exhibit

characters that are amenable to control, for example, those that bite indoors at night are conducive to control using ITNs and indoor resting facilitates the effectiveness of IRS.

Asia summary. *Plasmodium vivax* is an important public health problem across large parts of Asia. Six of the 10 largest PAR estimates of the 95 *Pv*MECs were in Asia and these countries accounted for 87% of the global PAR. Control of *P. vivax* in *Pv*MECs in Asia would, therefore, dramatically reduce the impact of this widespread disease. In Asia, in particular, densely populated areas are endemic with stable *P. vivax* transmission. This is due to environmental suitability as well as a vector species (*An. stephensi*) that is well suited to urban transmission. The limits and endemicity maps produced show that *P. vivax* has a wide geographic range in this region, across which the level of transmission is generally low, with small pockets of intense stable transmission (Fig. 1.3A1 and A2). In high-prevalence areas, such as those found in parts of India (Orissa State) and Myanmar, the clinical character of *P. vivax* has been found to behave more like *P. falciparum*, resulting in cases of severe disease and death (Mendis et al., 2001; Baird, 2007; Price et al., 2007a; Kochar et al., 2009; Mahgoub et al., 2012). Uncertainty estimates were high throughout India and much of Myanmar and population-weighted estimates emphasise the need to improve the estimates in India with more survey data (Fig. 1.4A1 and A2). Improved *P. vivax* surveillance in these regions would greatly improve prevalence predictions. By 2012, several countries in this region (Azerbaijan, Bhutan, China, Georgia, Korea DPR, Republic of Korea, Kyrgyzstan, Sri Lanka, Tajikistan, Thailand, Turkey, Uzbekistan and Viet Nam) had entered the malaria elimination phase (The Global Health Group and the Malaria Atlas Project, 2011). Improved control in neighbouring countries with intense, stable transmission, such as Myanmar, will improve the success of those nations already working towards elimination by reducing the potential for imported malaria from human migration (Tatem and Smith, 2010). As elimination efforts continue, spatially detailed mapping will be needed to capture focal areas of transmission that remain. Further and improved vector surveillance using molecular techniques to disaggregate species complexes is needed to illuminate which vectors are truly the primary vector species of *P. vivax* to allow for appropriate control measures to be deployed in this highly variable region.

4.2. Asia-Pacific

The Asia-Pacific region (defined for these purposes as the southern islands of Asia-Pacific and the Malaysian Peninsula; Fig. 1.2) has amongst the highest

endemicity estimates of *P. vivax* malaria (Fig. 1.3B) and the most complex spectrum of vectors. This region has a relatively small land mass, which means it has a much lower estimated PAR but the range in epidemiological and entomological factors here mean it presents a unique and challenging control setting.

Defining the limits of transmission. Annual parasite index data were provided for all the *P. vivax*-endemic countries in Asia-Pacific. The *Pv*MECs in this region are Indonesia, Malaysia, the Philippines, Papua New Guinea, the Solomon Islands, Timor-Leste and Vanuatu.

Transmission was estimated to span 2.74 million km^2 of land in this region, which is primarily made up of islands. The area at risk in Asia-Pacific constituted 6% of the global area at risk. Indonesia, the largest country in the region, had the greatest area at risk, consisting of 1.71 million km^2 of land, which was 90% of the country's total area (1.90 million km^2) and 60% of the total area at risk for the region.

Estimating endemicity. There were 5277 records of prevalence data collected from Asia-Pacific, which made up more than half (53%) of the global *P. vivax* PR survey database. This was primarily due to a large data contribution from Indonesia, which provided 4457 (84%) data points in the Asia-Pacific region and 45% of the global dataset (Fig. 1.3B1).

As with mainland Asia, *P. vivax* endemicity in regions with stable transmission was predicted to vary greatly across the Asia-Pacific region as shown in Panel B2 of Fig. 1.3. *Pv*$PR_{1–99}$ point estimates were predicted to exceed 7% in small parts of Indonesia and the Solomon Islands, and much of Papua New Guinea. Estimates towards the east of the region (Sumatra and Kalimantan) were generally low, while endemicity values of around 5% were predicted to the north (Malaysian Borneo and the Philippines). The vast number of surveys from Indonesia meant that the certainty of the predictions in that portion of the region was relatively high (Fig. 1.4B1). Uncertainty was greatest on New Guinea and parts of Borneo, where prevalence data were sparse. The Philippines, which also had little survey data available, was predicted with relatively low certainty. Again, the population-weighted uncertainty map was substantively different (Fig. 1.4B2). While the highly populated areas of Asia were predicted with less certainty, the high-density areas of Asia-Pacific, which also had a high number of prevalence surveys (Indonesia), showed an increase in certainty in the population-weighted estimates. Uncertainty was also greatly reduced following population-weighting in focal areas in Papua, Indonesia and Papua New Guinea, where population density is low. The Philippines had relatively low uncertainty

even when weighted by population estimates, indicating that this region would benefit from increased prevalence surveillance data.

Population at risk. Relative to the geographic size of the region, Asia-Pacific had a large PAR due to pockets of intense transmission in highly populated endemic areas. It was estimated that there were 215 million people at risk in this region. This was only 11% of the PAR in mainland Asia. However, the proportion of the population that was at any risk of *P. vivax* relative to the total population of the *Pv*MECs in Asia-Pacific was comparable to the Asia region: the PAR was 59% of the total population in Asia-Pacific and 58% in Asia. The largest PAR in Asia-Pacific was in Indonesia, with an estimated population of 129.28 million individuals at risk of *P. vivax* out of a total population of 231.60 million (56%) in 2010. The PAR of stable and unstable transmission in Indonesia was 60% of the PAR of Asia-Pacific as a whole. The next largest PAR estimates were from the Philippines and Malaysia. The Philippines had 50.34 million people at risk or 54% of the total population of the Philippines and 23% of the regional PAR. Malaysia had 27.88 million (99.8% of the Malaysian population; 13% of the regional PAR) people at risk. Papua New Guinea had a PAR of 5.64 million (81% of the national population; 3% of the regional PAR) and Timor-Leste had 1.15 million (98% of the national population; 0.5% of the regional PAR). The Solomon Islands and Vanuatu had a PAR less than one million: 0.53 and 0.24 million (both nearly 100% of the total national populations and <0.3% of the regional PAR). Seventy percent of the total PAR in the region was exposed to unstable transmission. Unstable transmission predominated in Indonesia and Malaysia, where 79% (26.78 million) and 84% (4.45 million) of the PAR, respectively, were exposed to that level of transmission. In the Philippines, 53% (26.59 million) of the PAR lived in areas of stable transmission. Nearly all of the transmission in Vanuatu was stable (99.57%). The entire PAR in both Timor-Leste and the Solomon Islands experienced stable transmission. More than half of the population (59%) in Asia-Pacific is at risk of some level of vivax malaria transmission and the 215 million people at risk was 8.7% of the global PAR.

Vectors. Of the 26 potential vector species of *P. vivax* (Table 1.2), we predict that 16 are DVS that occur in Asia-Pacific (*Anopheles aconitus*, *An. annularis*, *An. balabacensis*, *An. barbirostris* complex, *An. dirus* complex, *An. farauti* complex, *An. flavirostris*, *An. koliensis*, *An. lesteri*, *An. leucosphyrus* & *An. latens* (sister species grouped together here due to known mis-identification problems within the current literature), Maculatus Group, *An. minimus* complex, *An. punctulatus* complex, *An. sinensis* complex, *An. subpictus* complex

and *An. sundaicus* complex), six of which are found only in Asia-Pacific (*An. balabacensis*, *An. farauti* complex, *An. flavirostris*, *An. koliensis*, *An. leucosphyrus* & *An. latens*, and the Punctulatus Group) (Sinka et al., 2011). The predicted distributions of the DVS in this region are shown in Fig. 1.9B.

For the 16 Asia-Pacific DVS, 9052 unique georeferenced occurrence records were assembled. The most data-rich species/species complexes were the *An. farauti* (n = 1737), *An. subpictus* (n = 1143) and *An. barbirostris* (n = 1064) complexes; and the most data-poor were *An. leucosphyrus* & *An. latens* (n = 12). Occurrence data were reported from all seven *Pv*MECs in this region. Most of the data originated from Papua New Guinea (n = 1503 sites), followed by Indonesia (n = 890), Solomon Islands (n = 160), Malaysia (n = 145), the Philippines (n = 124), Vanuatu (n = 36) and Timor-Leste (n = 1). Given the heterogeneity in vector distribution in this region, improved data collection from many of these countries would improve the fidelity of the predicted distributions of this complicated region.

The distribution and behaviour of the DVS discussed below are those that met the criteria for having broad distributions in the region (Fig. 1.9C) and have been conclusively incriminated as vectors of *P. vivax* (Table 1.2). A complete discussion of their ranges and bionomics may be found elsewhere (Sinka et al., 2011, 2012). The species that were identified as specific vectors of *P. vivax* in Asia-Pacific are *An. barbirostris s.l.*, *An. farauti s.l.*, *An. koliensis*, *An. leucosphyrus* & *An. latens* and *An. punctulatus s. l.*

Anopheles barbirostris s.l. is predicted to occur across large areas of inland Sumatra and Borneo. It is typically a highland species, but is also found in coastal regions (e.g. western Timor). Adult females of this species complex tend to feed outdoors on animals during dusk and night time, bringing its significance as a vector of malaria into question. Further molecular analysis and identification of the sibling species is needed to determine if variations observed in biting behaviours are due to plasticity within a species or differences amongst different species in the complex. A variety of larval habitats are also observed for *An. barbirostris s.l.* Though they are typically considered swamp breeds, a variety of water depth, size, light intensity and movement has been observed. This variation may again be due to the differences among species within the complex rather than the plasticity of a single species.

Anopheles farauti s.l., *An. koliensis*, and *An. punctulatus s.l.* are in sympatry across most of the islands of New Guinea, New Britain and into the Solomon Islands. All three are found within the Punctulatus Group. There are two sibling species in the Punctulatus Complex: *An. punctulatus s.s.* and *Anopheles* sp. near *punctulatus*.

The latter of the two species is relatively uncommon, whereas *An. punctulatus s.s.* is reportedly widespread and an important vector on the island of New Guinea. The distribution of *An. punctulatus s.s.* spans eastern Indonesia, Papua New Guinea and the Solomon Islands. Adult females feed readily on humans outdoors or occasionally indoors. Those that feed indoors may rest there, but are more likely to exit to find an outdoor resting spot. Peak biting times vary based on the geographic location. The larvae of *An. punctulatus s.s.* thrive in temporary pools that result from disturbed ecology. The habitats tend to be small, shallow, sunlit pools of water that may be turbid, but never brackish.

The distribution of the eight cryptic species of the Farauti Complex is largely dependent on the species' tolerance of salinity such that some members are coastal species (e.g. *An. farauti s.s.*) while others are restricted to inland areas with freshwater larval habitats (e.g. *Anopheles hinesorum*). The feeding habits of *An. farauti s.l.* also vary; host preference changes based on the availability of hosts and resting may be dependent on where feeding occurred. Peak biting times fluctuate with location, but it should be noted that daytime biting may occur. The larval habitats of members of the complex are typically in natural rain-fed water bodies, but artificial containers such as water drums or coconut shells may also be used. The habitats range in salinity and light intensity due to the wide geographic distribution of the species complex.

The predicted distribution of *An. koliensis* is similar to that of the Punctulatus Complex. The adult females tend to be anthropophilic, but will feed on other animal hosts. Although females may go indoors in search of a host, resting typically occurs outdoors. Feeding can occur throughout the night, with variable peak times based on location. The larval habitats of *An. koliensis* are intermediate between those of *An. farauti s.l.* and *An. punctulatus s.l.* and are generally associated with permanent freshwater bodies in open grassland such as irrigation ditches.

Anopheles leucosphyrus and *An. latens* are sister species of the Leucosphyrus Complex. Both species are found in forested areas with the former on the island of Sumatra and the latter on the southern portions of the Thai and Malaysian Peninsula and parts of Borneo. Both species are considered to be important vectors of malaria and were long thought of as the same species until molecular and cross-mating studies distinguished between them (Baimai *et al.*, 1988). Little is known of the bionomics of *An. leucosphyrus s.s.* because much of the literature on '*An. leucosphyrus*' is now known to refer to *An. latens*. The species has been shown to bite humans both inside and outside, however. *Anopheles latens* bites throughout the night with peak

times that vary based on the seasonal climate and location. Its larval habitats include shaded temporary pools found in the forest such as swamps, stump ground holes or wheel tracks.

The bionomics of the DVS in Asia-Pacific are as variable as in mainland Asia. This is due to variation observed both within species complexes and among separate species. Such variations increase the challenge of developing a single universal vector control strategy across the region, however, a combination of methods could prove effective (Lindsay et al., 2004).

Asia-Pacific summary. The vivax malaria problem in Asia-Pacific is complex and presents a variety of challenges to control. Indonesia and the Philippines are densely populated over large land masses (Fig. 1.8B) and *P. vivax* is predicted to be endemic across most of their national territories. They have the fourth and fifth largest PAR estimates globally with 129 million at risk in Indonesia and 50 million in the Philippines. Some of the highest predicted *Pv*$PR_{1–99}$ values were also observed in this region (Fig. 1.3B1). As in the Asia regions, it is in these areas, where intense *P. vivax* transmission occurs (e.g. Indonesia and Papua New Guinea), where drug resistance is emerging and malaria cases of equivalent clinical severity to *P. falciparum* have been observed (Baird, 2004; Pukrittayakamee et al., 2004; Tjitra et al., 2008; Price et al., 2009). This and the large areas of unstable transmission make this region a worthy and vital candidate for research, control and eventually elimination efforts; four of the seven *Pv*MECs have already declared their status of working towards elimination: Malaysia, the Philippines, the Solomon Islands and Vanuatu (The Global Health Group and the Malaria Atlas Project, 2011). Uncertainty estimates, shown in Fig. 1.4B1 and B2, were high in areas that also displayed high *Pv*$PR_{1–99}$ estimates (Fig. 1.3B2), such as Papua New Guinea. While the population-weighted correction shows that parts of these regions are, for the most part, sparsely populated (Fig. 1.8B), improved surveillance to facilitate high-resolution mapping will be essential to tackle the final transmission hotspots of this region as elimination efforts progress. The vector situation in this region is complex and further research is needed to differentiate members of species complexes and incriminate those vectors that are vectors of *P. vivax* so that appropriate control measures may be taken.

4.3. Americas

The Americas have amongst the highest *P. vivax* endemicity values in the world but these typically occur in areas of very low population density, so the populations at risk are much lower than those found in Asia (Fig. 1.3C1 and Fig. 1.8C). The diversity of vector species is also less in this region (Fig. 1.9C).

Defining the limits of transmission. *Plasmodium vivax* annual clinical incidence data were available from all 19 countries that are considered to be endemic in the Americas. The most recent year of reporting was 2008 for the majority of the countries. Guyana, Nicaragua and Panama reported data up to 2007; Belize, El Salvador, French Guiana and Guatemala provided data up to 2006 and Colombia's last year of reporting was 2005.

Endemic areas in this region were estimated to span 9.46 million km^2, most of which were found in the Amazon basin. The majority of the areas at risk in the Americas (85%; 8.08 million km^2) were at stable transmission (Fig. 1.3C1). This indicated that, for the most part, the disease was either endemic or wholly absent in this region with limited areas at risk of unstable transmission. The transmission level was generally found to be inversely proportional to the population density: areas experiencing stable transmission were those of lower population density. For example, the largest area at risk of stable transmission in a single country was 4.40 million km^2 in Brazil, which was 54% of the stable transmission area for the Americas region. However, the population at stable risk in Brazil was 26% (13.02 million) of the region's total PAR of stable transmission. The Americas comprised 53% of the global land area at stable transmission but only 5% of the population at that level of risk.

Estimating endemicity. There were relatively few PR records and surveys for the Americas, with 388 unique records available for modelling. The three most data-rich countries in this region were Brazil (n = 175; 45%), Venezuela (n = 65; 17%) and Peru (n = 51; 13%), as shown in Fig. 1.3C1.

The predicted prevalence, or $Pv\text{PR}_{1–99}$, was highly variable throughout the Americas. Much of the areas of stable transmission had PRs between 3 and 5%, with isolated areas nearing 0% and others exceeding 7%. Regions of high prevalence (>7%) were found in Amazonia (Northeast Brazil) and Central America (Nicaragua and Honduras). The uncertainty in the $Pv\text{PR}_{1–99}$ predictions was relatively high overall due to the relatively sparse PR survey data in this region (Fig. 1.4C1). However, when population density was incorporated into the weighted uncertainty calculations, the resulting uncertainty index was greatly lowered because the areas with stable risk and high uncertainty predictions had low population densities (Fig. 1.4C2). This implies that the observed uncertainty has relatively little operational importance at the global scale, although highlights the current lack of precision for defining high-risk foci in remote Amazonian settings.

Population at risk. The PAR in the Americas was relatively low given the geographic limits of *P. vivax* due to low population densities in areas of high transmission. It was estimated that there were 137.45 million people living

at risk of *P. vivax* in the Americas in 2010. The majority of the PAR was in areas of unstable transmission (64%; 87.66 million) and 36% (49.79 million) were exposed to stable *P. vivax* transmission. Brazil had the largest PAR with 45.22 million people living at any risk, which is 26% of its total population and represents 33% of the region's PAR). The largest number of people living at stable risk (13.02 million) was in Brazil, followed by Venezuela with 4.43 million at stable risk and 22.87 million total at risk (83% of the national population and 17% of the regional PAR). Colombia was estimated to have 7.53 million at stable risk and 19.15 million at risk (56% of its total population and 14% of the region's total). El Salvador, with a PAR 3.17 million, did not have any individuals living at stable risk. French Guiana (PAR 158,000) and Suriname (PAR 30,000; the lowest PAR in the region) did not have any populations living at unstable risk. The portion of the PAR at stable transmission in Belize was also nearly 100%; only 30 individuals out of 220,000 at risk were in unstable transmission areas. Of the 1.5 billion people living at unstable and 964 million at stable transmission around the globe, 5.8% and 5.2% of those were found in the Americas, respectively.

Vectors. Of the 20 potential vectors (Table 1.2) of *P. vivax* in the Americas, nine are considered DVS. These include *Anopheles albimanus*, *An. albitarsis* complex, *An. aquasalis*, *An. darlingi*, *An. freeborni*, *An. marajoara*, *An. nuneztovari* complex, *An. pseudopunctipennis* and *An. quadrimaculatus* (Sinka et al., 2010b). The predicted distribution of these species is shown in Fig. 1.9D. There were 4141 georeferenced spatiotemporally unique occurrence points obtained for *Anopheles* DVS from 25 countries in the Americas. Twenty-two percent of the occurrence records (n = 926) referred to *An. darlingi* and 20% (n = 851) to *An. albimanus*. The remaining species had occurrence data of between 63 (*An. freeborni*) and 572 (*An. quadrimaculatus s.l.*) records. Of the total number of point locations identified (n = 1509), the greatest number came from the United States (n = 377), followed by Brazil (n = 304) and Belize (n = 228). The *Pv*MECs with the fewest point records (<10) were French Guiana (n = 7), El Salvador (n = 3), Paraguay (n = 2), Guyana (n = 1), Honduras (n = 1) and Nicaragua (n = 1).

Distribution maps indicate a relatively straightforward vector profile across the *Pv*MECs of the Americas. Of the nine DVS in this region, *An. darlingi* and *An. albitarsis s.l.* met the criteria for further description because these species are predicted to have a wide distribution in the region and conclusive supporting evidence indicating transmission of *P. vivax* in the wild, indeed, the distribution of *An. darlingi* is remarkably similar to that of *P. vivax* in the region (1.3C2 and 1.9D). Although *An. darlingi* is considered

a riverine species inhabiting forest locations, it has been observed to take advantage of deforestation, exploiting areas with reduced canopy cover compared to those more densely forested locations. The larval habitats of this species are typically clear and natural water bodies such as slow-flowing, clear rivers and streams. Adult female *An. darlingi* are both exo- and endophagic, though they tend to rest outdoors regardless of where their blood meal was taken. Host preference also varies within this species as does biting activity (time of biting), which may adapt to correspond to human behaviour (i.e. late night human activity influences or causes late night biting by *An. darlingi*).

The Albitarsis Complex is composed of five sibling species including the known vectors, *An. albitarsis*, *An. deaneorum* and *An. marajoara*. Members of this complex exhibit some behaviours that are similar to *An. darlingi*, for example, the adult females tend to rest indoors but will bite indiscriminately without much host preference both indoors and outdoors. The larval habitats tend to be clear, sunlit, freshwater.

The Americas summary. National surveillance data have shown a widespread decrease in morbidity and mortality from both major *Plasmodium* species across the American continent since 2000 (WHO/PAHO, 2008). This is due largely to the successful implementation of integrated vector management (Roberts et al., 1997; Butler and Roberts, 2000; Rojas et al., 2001; Killeen et al., 2002; Roberts et al., 2002; Shiff, 2002; WHO/PAHO, 2006) and has led eight of the 21 *Pv*MECs in the Americas (Argentina, Belize, Costa Rica, El Salvador, Mexico, Nicaragua, Panama and Paraguay) to target malaria elimination (The Global Health Group and the Malaria Atlas Project, 2011). However, the prevalence estimates of *P. vivax* in the Americas are heterogeneous with isolated areas of intense transmission (Fig. 1.3C2). Vector occurrence data coverage was uniformly low in the American *Pv*MECs. Improved vector research in the Americas will advance the maps of species-specific distributions, which is essential for the continued success of vector management to curb malaria morbidity and mortality. Pockets of high transmission (*Pv*PR > 7%) were observed in large areas of the region. However, it is important to note that these were generally in areas with low population estimates. Sparse prevalence data resulted in high uncertainty estimates in the region that were greatly reduced when the density of the population in endemic areas was taken into consideration (Fig. 1.4C). *Anopheles darlingi* appears to be the primary vector in this region; however, eight of the potential vectors of *P. vivax* in this region lack conclusive evidence regarding their potential to transmit the parasite. Therefore, further research is needed to decisively incriminate the vectors of *P. vivax* in the Americas.

Plasmodium vivax makes up the vast majority of malaria transmission that occurs in the Americas (Arevalo-Herrera et al., 2010). Survey data from this region were relatively sparse and records represent a picture of heterogeneous transmission levels. The Americas contributed a small fraction (5.5%) of global PAR of *P. vivax*, but comprised nearly a quarter (22%) of the global area at risk. Areas of high transmission in highly dispersed populations may present unique challenges to malaria control. To advance elimination efforts, high-resolution mapping will be needed to accurately illustrate the degree of heterogeneity in this region and areas that require the greatest resources. This will demand more data from areas with high predicted prevalence and uncertainty, such as Honduras and Nicaragua in Central America and northwest Brazil in South America. Although Brazil is a large and populous country, its areas of stable transmission occur in sparsely populated regions of the Amazon basin. In 2010, Brazil had the seventh largest PAR of *P. vivax* globally (45 million), the largest area at risk (4.90 million km^2) as well as the largest area at stable risk (4.40 million km^2) and, therefore, has an important role to play in the region's future of malaria control and elimination.

4.4. Africa+

The estimates for endemicity and populations at risk of *P. vivax* malaria in Africa+ (defined here as Africa, Saudi Arabia and Yemen) are mitigated by the high prevalence of Duffy negativity in these populations, and the vector situation in this region is relatively straightforward.

Defining the limits of transmission. High *P. falciparum* endemicity in Africa, coupled with high prevalences of Duffy negativity, has meant that collection of *P. vivax*–specific data has not been a priority in the past. The data available from which to estimate the limits of *P. vivax* transmission is, therefore, limited. *Plasmodium vivax* annual parasite incidence (*Pv*API) data were only available from six Africa+ countries (13%). The last year of reporting available was 2009 for four countries (Dijbouti, Namibia, South Africa and Swaziland); whilst for the remaining two countries (Saudi Arabia and Yemen) the last available reports were from 2006. Based on the limited *Pv*API and *Pv*PR data available for Africa+, and the presence of suitable vectors and climatic conditions, forty-six countries are assumed to be *P. vivax* endemic in this region.

The area at risk was estimated to span over 22 million km^2 of Africa, Yemen and Saudi Arabia. Despite the historical misconception that *P. vivax* is absent from the African continent, 84% of the 22.46 million km^2 of land in the region was at some risk of *P. vivax* transmission (Fig. 1.3D2). However, the vast majority (92%; 20.60 million km^2) of total area at risk was estimated to house unstable transmission.

Estimating endemicity. There were 1640 records of prevalence data from 97 sources obtained for Africa+, 79% of which reported an absence of *P. vivax*. Approximately 16% of the global *Pv*PR data records used in the modelling were from surveys conducted in Africa, Yemen and Saudi Arabia. Ethiopia, Zambia and Sudan were the most data rich countries, contributing 50% (n = 826), 18% (n = 295) and 18% (n = 290) of the total data, respectively.

The predicted prevalence estimates were uniformly low for areas at stable risk in Africa, Yemen and Saudi Arabia as shown in Fig. 1.3D2. The point estimates of predicted $PvPR_{1–99}$ rarely exceeded 2%. The uncertainty of the predictions in this region was also very low (Fig. 1.4D). The incorporation of information regarding the proportion of Duffy negative individuals strengthened the predictions in this area and compensated for the sparse parasite rate data available from this region (Fig. 1.3D1, Fig. 1.4D and Fig. 1.7D). Areas with higher $PvPR_{1–99}$ (Ethiopia, South Sudan and Madagascar) predictions were also found to have the highest uncertainty, likely because these areas were outside the range of high Duffy negativity and could not borrow from the certainty conferred by that restriction. As with the Americas, the uncertainty in Africa+ was reduced in the population-weighted predictions because of the low population density found in parts of the continent, particularly Madagascar (Fig. 1.4D2). Uncertainty remained relatively high in the highlands of Ethiopia, which are more densely populated and are therefore important targets for control.

Population at risk. The PAR in Africa+ was low given the large geographic coverage of the region because of the high prevalence of Duffy negativity (Fig. 1.3D1 and Fig. 1.7D). There were an estimated 86 million people living at risk of *P. vivax* in Africa, Yemen and Saudi Arabia in 2010, which was 3.5% of the global total. More than half (56%), or 48.72 million, of the PAR were those living in unstable transmission and the remaining 37.66 million (44%) lived in areas at the level of stable transmission. Nine countries had an estimated PAR of zero, all of which are located in West Africa: Côte d'Ivoire, Ghana, Guinea, The Gambia, Guinea-Bissau, Senegal, Sierra Leone and São Tomé and Príncipe. Togo, Equatorial Guinea, Gabon and Burkina Faso also had a total PAR of less than one thousand individuals. Thirty-four countries in the Africa+ region had zero individuals at risk of stable transmission. Of the 12 countries with stable transmission, Madagascar was the only country where the entire PAR (5.23 million) experienced stable transmission. Madagascar has the fourth highest PAR in the region; the highest were Ethiopia, Sudan and Yemen with 35.19 million (41% of

the PAR in the Africa+ region), 14.78 million (17%) and 13.06 million (15%) people at risk in each, respectively.

Although the high proportion of Duffy-negative individuals in Africa is thought to result in an absence of the parasite across the continent, evidence of transmission in Africa+ (Guerra et al., 2010; Ménard et al., 2010) supported the decision to assume stable transmission in all areas of Africa+ that were not excluded by *Pv*API data or biological masks. Areas were then reclassified following $PvPR_{1–99}$ predictions such that locations with endemicity levels below 1% were reclassified as unstable; this was the majority of the region. High proportions of Duffy negativity are especially common among the population of West Africa. Ghana, for example, had a PAR of 24 million individuals when accounting for environmental exclusions and before the inclusion of Duffy negativity, which subsequently reduced the PAR of Ghana to zero. The incorporation of the Duffy negativity layer reduced the PAR of the Africa+ region from 840 million to 86 million, which is 3.5% of the global PAR of *P. vivax*.

Vectors. The high impact of malaria on the African continent is largely due to the efficiency of the African DVS in transmitting *P. falciparum*. However, *P. vivax* may also play an important role. As more countries on the continent move towards elimination, there may be in shift in the endemicity of *P. vivax* as levels of *P. falciparum* decline. It is essential to know which vector species are capable of transmitting *P. vivax*.

Africa is the home to two of the most efficient vector species of human malaria: *An. gambiae* and *An. funestus* (Gillies and de Meillon, 1968; Coluzzi, 1999). Of the 10 potential vectors of *P. vivax* in Africa+ (Table 1.2), eight (*Anopheles arabiensis*, *An. funestus* complex, *An. gambiae* complex, *An. Labranchiae*, *An. melas*, *An. merus*, *An. moucheti* and *An. nili* complex (Sinka et al., 2010a)) are considered to be DVS of human malaria, but their importance in transmitting *P. vivax* is still uncertain. For example, it is unclear whether the forest vector *An. moucheti* is a poor/non-vector of *P. vivax* or that this parasite simply does not exist in areas where this species occurs (Table 1.2). The distributions of *An. arabiensis*, *An. funestus s.l.* and *An. gambiae s.l.* and the DVS that occur in Saudi Arabia and Yemen are shown in Fig. 1.9E. Only the three primary DVS are shown in Africa because of the lack of evidence of the ability for the other African DVS to transmit *P. vivax* (Table 1.2). For the African DVS there were 8338 spatially and temporally unique occurrence records. The majority of these occurrence records were obtained for *An. funestus s.l.* (n = 2692), *An. arabiensis* (n = 2301) and *An. gambiae s.l.* (n = 2291). Point data were obtained from 44 countries. The largest number

of point records per country were from Kenya (*n* = 757), followed by Tanzania and Cameroon (*n* = 383 for both nations). For the most part, the other countries on the continent had less than 100 point records, with the exception of Burkina Faso (*n* = 310), Equatorial Guinea (*n* = 113), Ghana (*n* = 106), Madagascar (*n* = 198), Mali (*n* = 166), Nigeria (*n* = 190), Senegal (*n* = 209), Sudan (*n* = 125), the Gambia (*n* = 192) and Uganda (*n* = 135). African countries classified as *P. vivax* endemic that had very few (≤5) vector occurrence point records were Central African Republic (*n* = 3), Congo (*n* = 2), Liberia (*n* = 4), Namibia (*n* = 5) and Togo (*n* = 1).

The compiled species distribution maps (Fig. 1.9E) illustrate a relatively straightforward picture of the distribution of the region's DVS: *Anopheles arabiensis*, *An. funestus s.l.*, and *An. gambiae s.l.* which all have broad distributions in the region and are confirmed vectors of *P. vivax*. These three DVS dominate in heterogeneous ranges of different pairs and combinations (of one another) in such a way that the species are present on their own in only focused locations. The co-dominant range of the *An. funestus* complex and *An. gambiae* in Central Africa is surrounded by an 'envelope' that houses all three primary DVS, that is further surrounded by *An. arabiensis* and the Funestus Complex, and then only *An. arabiensis* on the periphery. *Anopheles arabiensis* tolerates drier environments and is therefore absent from the forested areas of western Central Africa. The *An. funestus* complex distribution indicates a presence throughout all of sub-Saharan Africa, including Madagascar, but excluding much of southern Africa. *Anopheles gambiae* has a more complex distribution across a band from East (including Madagascar) to West Africa.

The bionomics of all the African DVS of malaria are summarised in full elsewhere (Sinka et al., 2010a). Here, we briefly describe the behaviours of those three DVS identified as potential *P. vivax* vectors in the region. *Anopheles arabiensis*, *An. funestus s.l.*, and *An. gambiae s.l.* are known to be primary vectors of *P. falciparum*, but have also been incriminated as vectors of *P. vivax* through the detection *P. vivax* circumsporozoite proteins in wild-caught specimens (Table 1.2). *Anopheles arabiensis* is often described as zoophilic, exophagic and exophilic, yet its behaviour appears to be quite variable, depending on location. For example, *An. arabiensis* found in West Africa are generally more anthropophilic and endophagic than those in the East. Such behavioural variability may enhance this species' ability to transmit *P. vivax* (or any human malaria) allowing it to adapt to avoid control methods such as IRS. Moreover, with peak biting times ranging from evening (1900) to early morning (0300), *An. arabiensis* may also avoid control via ITNs.

Anopheles arabiensis tends to be found in drier climates within Africa and larval habitats are typically small, temporary, clear, sunlit freshwater pools similar to those of *An. gambiae s.l.* However, *An. arabiensis* larvae have also been sampled from large or small man-made water bodies, including rice fields, as well as flowing or even brackish waters.

There is less variation observed in the bionomics of *An. funestus s.l.* relative to *An. arabiensis*. This species is known to be highly anthropophilic and endophilic and have a late biting time (after 2200), making IRS or ITNs highly effective interventions (although pyrethroid resistance has now been reported (Hargreaves et al., 2000; Coetzee and Fontenille, 2004)). The larval habitats of the Funestus Complex tend to be large, permanent (or semi-permanent) bodies of freshwater such as a pond, lake edge or rice field, where the larvae use emergent plants as protection against predation. There is some variation in the behaviour of members of the complex and further investigations using molecular identification methods are needed to determine the distribution of the different subtypes and the bionomics they exhibit.

Anopheles gambiae s.l. is perhaps the most well-known vector of human malaria and the most studied. It occupies a wide geographic range and is considered to be a highly efficient vector because of its highly anthropophilic biting behaviour and relatively long adult life stage. Females generally feed late at night indoors and also rest indoors, again making IRS and ITNs successful intervention strategies. However, in studies comparing feeding and resting location preference, *An. gambiae s.s.* exhibited both indoor and outdoor feeding and resting habits (Sinka et al., 2010a). This is likely ascribed to different chromosomal and molecular forms now identified within the species (Bockarie et al., 1993). The larval habitats of *An. gambiae s.l.* were long thought to be restricted to clear sunlit, temporary pools such as puddles or hoof prints, however, larvae have been reported from turbid and even polluted water and from large semi-permanent water sources such as rice fields. The variation of larval habitats is again attributed to divergences of the chromosomal or molecular forms.

The vector species of *P. vivax* in Africa occupy wide geographic ranges across the region and exhibit variation in behaviours both among and within individual species or species complexes. This may present challenges for control, yet, the tendency for these species to bite and rest indoors at night makes control methods such as ITNs effective control strategies. Vivax malaria is not currently the main focus of malaria research in this region and incrimination studies on wild populations have yet to be performed on half

of the potential DVS in this region (Table 1.2). Further research identifying which anopheline species are the most efficient vectors of this parasite will be beneficial to the Africa+ region as goals of control and elimination of *P. falciparum* are realised and the focus moves to address *P. vivax*.

Africa+ summary. Estimates of the PAR of *P. vivax* in Africa, which is climatically well suited for malaria transmission, were very low, with the exception of countries around the Horn of Africa. This was because of the assumption that Duffy negative individuals, found in high frequencies on the continent (Howes et al., 2011), are refractory to *P. vivax* infection. This supposition was incorporated into the model, despite observations of *P. vivax* malaria infections in Duffy-negative individuals in Madagascar (Menard et al., 2010) and on the mainland of the continent (Ryan et al., 2006; Mendes et al., 2011; Wurtz et al., 2011). While this information contradicts our working assumption of complete protection, there is insufficient evidence to determine if these are more than just rare occurrences on the continent that would have a significant effect on the epidemiology of *P. vivax* in Africa.

Prevalence estimates for Africa are characterised by low predicted values and high levels of uncertainty (Fig. 1.4D). The PR data from Africa that served as the input data for the endemicity predictions were sparse, with the exception of a few regions (Fig. 1.3D1). This is largely because *P. vivax* is not the main priority for Africa. There were 753 million people at risk of stable *P. falciparum* in Africa in 2010 (Gething et al., 2011a), compared to the 38 million at stable risk *P. vivax*. However, 30% (228 million) of the 753 million were living in regions of low stable transmission ($Pf\mathrm{PR}_{2-10} \leq 5\%$); the prevalence of *P. falciparum* in these areas falls to a point where elimination is viable, the situation of *P. vivax* will increase relatively, a consequence of its tendency to be the last parasite standing during elimination efforts (Garnham, 1951; Yekutiel, 1960; Pampana, 1969; Wernsdorfer et al., 2009; Tatem et al., 2010). This reinforces the need for increased vector surveillance and incrimination of species specifically for *P. vivax*. The endemicity map (Fig. 1.3D2) presented indicates that while *P. vivax* is present at very low levels in Africa, it is circulating. Consideration for how those prevalence estimates may be affected by decreasing *P. falciparum* levels must be considered.

4.5. Areas Where Lack of Geographical Data is Acute

Asia. Although there was thorough coverage of annual clinical incidence data and a large number of prevalence data records, given the large area of *Pv*MECs in Central Asia (20.5 million km^2), there were still large regions with a dearth of data (Fig. 1.3A1). These regions are highlighted by the

uncertainty maps (Fig. 1.4A). India, the nation with the largest PAR of *P. vivax* globally, had disproportionately little prevalence data available. Given that parts of India are predicted to experience intense transmission, high-resolution mapping is needed to identify foci of transmission and this will only be achieved through improved surveillance coverage or the release of data that have not been previously shared. The uncertainty map also highlights lack of data in Myanmar. Regions of high uncertainty, which were also present in the population-weighted estimates, correspond with areas plotted to be highly endemic ($Pv\text{PR}_{1–99} > 7\%$). Improved certainty in predictions in this country would be beneficial as neighbouring Thailand and China work towards malaria elimination. There would also be utility in the provision of contemporary data as progress towards elimination is made (e.g. in China) and transmission dynamics are therefore altered.

Asia-Pacific. Coverage of annual clinical incidence and prevalence survey data was relatively strong for this region (Fig. 1.3B1). However, the vast majority of the prevalence data originated from Indonesia. Sites on the island of New Guinea and parts of Borneo had very sparse prevalence data and were therefore predicted with greater uncertainty (Fig. 1.4B1 and B2). These regions have some of the most intense *P. vivax* transmission settings in the world. Improved surveillance would help localise hotspots and generate prime targets for control. The Philippines should also be noted as being a region with high uncertainty that was retained even after accounting for population density in the predictions. The smaller islands of this region would benefit from improved vector data collection since many had very little vector data available. This would improve the fidelity of the species distribution models in this region to better inform control decisions.

The Americas. Relative to other regions with stable *P. vivax* transmission, there were very little prevalence data available for this region (Fig. 1.3C1). Improved PR data from this part of the world, particularly those countries with intense transmission (Brazil, Honduras and Nicaragua), would benefit *P. vivax* mapping efforts. High-resolution mapping is needed to accurately map these highly focalized transmission settings, and this will require broad coverage survey data. There was good vector data coverage in this region, which is likely linked to the research and implementation of successful vector control efforts in the Americas. However, data collection on vector occurrence could be further improved. Some of the countries with the highest transmission (Honduras and Nicaragua) had the fewest occurrence records available.

Africa+. The *P. vivax* data available from Africa+ were also sparse, a reflection of the perception that the parasite is largely absent from the continent

(Fig. 1.3D1). In this region, the importance of *P. vivax* is overshadowed by *P. falciparum* and as such is not recorded in many countries. However, *P. vivax* infection has been observed in Duffy-positive and -negative individuals in the region, hence, it would be prudent for countries to monitor the prevalence of the parasite. While prevalence data were lacking from the region as a whole, the areas that would benefit most from improved surveillance are those with higher intensity transmission: Ethiopia, Madagascar and Somalia. This would reduce the uncertainty of the predictions and allow for better monitoring of control efforts. Lastly, the vector data available for the DVS in the region was relatively poor. Although the vector profile of Africa+ is straightforward, increased occurrence records may capture intricacies in distribution patterns and help distinguish vector and non-vector species of species complexes.

5. DISCUSSION

Plasmodium vivax malaria imposes serious public health burdens and is the most widespread of all the human malarias, particularly in women and children in poorly resourced communities (Poespoprodjo et al., 2008, 2009). Robust evidence demonstrates that *P. vivax*, despite long-held convention, is not a 'benign' infection (Baird, 2007; Price et al., 2007b). Moreover, studies show that 60-year-old drugs used to treat vivax malaria are failing throughout much of the *P. vivax*-endemic world (Baird, 2009; Price et al., 2009). The hypnozoitocidal component of that treatment, primaquine, is a deeply flawed therapy due to the threat it imposes to G6PD deficient patients (Baird, 2007). The reader is also referred to the chapter by Howes et al. that appears elsewhere in this thematic issue of *Advances in Parasitology* (Chapter 4, Volume 81). These issues are brought into focus as international targets, such as the Millennium Development Goals (http://www.mdgmonitor.org/), are developed to halt or mitigate malaria incidence and further goals are set for the elimination of the disease (WHO, 2007; Feachem and Sabot, 2008; The Global Health Group and the Malaria Atlas Project, 2011). Neglect of the impact and study of *P. vivax* is incompatible with these expressed goals.

The *P. vivax* PR estimates reviewed here are, across the American, African and Asian regions, uniformly low in comparison with PR estimates derived for *P. falciparum* (Gething et al., 2011a). The PR spectrum for *P. vivax* ranges from 0% to 7%, whereas for *P. falciparum*, it is 0 to 70%. This seems to imply *P. falciparum* transmission is an order-of-magnitude greater,

but this is potentially misleading. The absolute prevalence of *P. vivax* in heavily endemic zones may also reach or exceed 70%. Relatively low parasite densities in blood (arising from its strict preference for reticulocytes) may lead to high rates of false-negative diagnoses by microscopy or rapid diagnostic tests (RDTs) (Mueller et al., 2009a). Microscopic diagnoses very often underestimate the true prevalence of *P. vivax* in blood in both high- and low-transmission settings (Mueller et al., 2009b; da Silva et al., 2010; Harris et al., 2010; Katsuragawa et al., 2010; Steenkeste et al., 2010). Further details regarding the diagnosis of *P. vivax* may be found in a review elsewhere in this thematic volume of *Advances in Parasitology* (Chapter 4, Volume 80). Missed diagnosis is important in mixed species infections where *P. vivax* may be a minor contributor to parasitaemia and often overlooked (Mayxay et al., 2004). There is no proven evidence that this is due to cross-species immunity, rather that plasmodia species are mutually suppressive in mixed infections (Richie, 1988). While *P. falciparum* tends to dominate *P. vivax*, infection with *P. vivax* reduces the intensity of falciparum infections (Snounou and White, 2004). Mixed infections are complex and often under-diagnosed (Mayxay et al., 2004) and it is difficult to interpret the effect they may have on prevalence estimates. The reader is referred to a review of acquired immunity to *P. vivax* provided in this volume (Chapter 3, Volume 81).

The difference in PR spectrums between *P. vivax* and *P. falciparum* could also be a reflection of the age range used as the input data for the endemicity model. To model endemicity, *P. vivax* predictions were standardised across the 1–99 age range, rather than the 2–10 years of age range, which was used for *P. falciparum* (Gething et al., 2011a). Malaria transmission peaks in 2–10-year olds and, therefore, the use of the 1–99-year age range 'dilutes' the prevalence estimates. Regardless, it is evident that the association between the risk of disease and parasite prevalence is markedly different for *P. vivax* than *P. falciparum*, with significant risk of *P. vivax* disease at lower parasite densities.

The maps reviewed here represent progress towards an improved understanding of the epidemiology of this unique malaria parasite. Included in this work is the first ever *P. vivax*-specific global endemicity map and updated limits of transmission. These maps are intended to aid control strategy formulation and operational decision-making. Stratification by transmission intensity has been supported by mathematical modelling in the context of *P. falciparum* management (Smith et al., 2006, 2008; Okell et al., 2008; Smith and Hay, 2009; Chitnis et al., 2010a, 2010b; Griffin et al., 2010; Ross et al., 2011), but *P. vivax* is rarely differentiated by endemicity level.

Consensus has yet to be reached in defining suitable control-oriented strata for *P. vivax*. Without such stratified control, there has been little impetus to fill that knowledge gap and generate reliable stratified risk maps. These *P. vivax* mapping efforts may also be unified with those for *P. falciparum* (Hay et al., 2009; Gething et al., 2011a) to help rectify this and elucidate where control efforts of the two parasites can be amalgamated. An example of such is the potential use of artemisinin-combination therapy (ACT) as presumptive treatment for diagnosed malaria in co-endemic areas discussed in the review of the use of anti-malarial drugs to reduce *P. vivax* transmission that is provided elsewhere in this volume (Chapter 5).

Primaquine, the only drug currently licensed to treat the liver stage of the parasite is contraindicated in individuals with G6PD deficiency. The map of *P. vivax* endemicity is an important complement to the recently developed map of G6PD deficiency (Howes et al., 2012) and will help to identify areas with high prevalence of both *P. vivax* and G6PD deficiency. Overlaying these is essential for estimating the potential risk of adverse outcomes that could occur from treatment with primaquine in areas where G6PD deficiency testing cannot be guaranteed (Ruwende and Hill, 1998; Cappellini and Fiorelli, 2008). The reader is again referred to a review of G6PD deficiency that details the geographic distribution, genetic variants and the implication of primaquine therapy provided elsewhere in this thematic issue of *Advances in Parasitology* (Chapter 4, Volume 81).

Plasmodium vivax endemicity maps provide bench-marks for progress in control and elimination. This information is increasingly needed by international organizations and groups that are once again assessing the prospect of eradication of all species of human malaria (Lines et al., 2007; Feachem and Sabot, 2008; Greenwood, 2008; RBMP, 2008; Mendis et al., 2009; Malaria Eradication Research Agenda, 2011b). At present, these maps are of particular importance outside of Africa, where *P. vivax* is the primary threat. Thirty of the 95 *Pv*MECs are in Asia and their populations comprise 91% of the global PAR of *P. vivax*. Knowledge of the global distribution and impact of *P. vivax* is also important to estimate market size and investment priorities for those developing targets for drug (Malaria Eradication Research Agenda, 2011a) and vaccine (Brown et al., 2009; Malaria Eradication Research Agenda, 2011c) research and development. The maps also facilitate national priority setting and advocacy.

The maps of dominant vector species of human malaria presented here (Fig. 1.9) and information regarding the potential vector species of *P. vivax* (Table 1.2) highlight a further knowledge gap in our understanding of *P.*

vivax compared to *P. falciparum*. Classification of anopheline species as vectors of malaria has typically meant vectors of *P. falciparum* malaria. This is most evident in the Africa+ region, but conclusive information is also missing from the Asian and American regions, where *P. vivax* is the dominant parasite. In many countries, the status of vector species is determined by absence of evidence rather than decisive evidence of absence of the parasite. The research into transmission of *P. vivax* is lagging behind that of *P. falciparum*. Improved vector surveillance would benefit all regions, but not without incrimination studies that focus specifically on *P. vivax*.

Currently, the cartographic foundations for estimating the public health burden of *P. vivax* do not exist and this is perhaps the highest priority for moving forward. The link between *P. vivax* prevalence and clinical burden must be established for regions where the disease is monoendemic as well as where it is sympatric with *P. falciparum*. National estimates of *P. vivax* rely on routine case reporting sources that vary in their fidelity and are often crudely distinguished from *P. falciparum* (WHO, 2011). Cartographic estimates of *P. vivax* burden would provide valuable information regarding the impact of the disease independent of inherent biases that may accompany health system data (Gupta et al., 2009; Rowe et al., 2009; Hay et al., 2010a, 2010b; Malaria Eradication Research Agenda, 2011b; Mueller et al., 2011). There is also a need to define the burden of *P. vivax* in clinically vulnerable groups such as pregnant women (Nosten et al., 1999; Mueller et al., 2009a) and, most notably, children. To estimate the burden of *P. vivax*, it will first be necessary to gain a better understanding of the impact of relapsing infections on prevalence. The degree to which relapses contribute to clinical disease is known to vary geographically, but the cause and pattern of the variation is still poorly understood (Battle et al., 2011; Betuela et al., 2011; White, 2011) and calls for further investigation. The extent to which *P. falciparum* affects the burden of *P. vivax* remains unknown, and now emerges as an important question in elimination strategy (Maitland et al., 1996, 1997; Snounou and White, 2004; Genton et al., 2008).

This chapter described the global public health significance of *P. vivax* in light of new cartographic technology and evidence, along with the fundamentally important emerging understanding of the infection as threatens to life. The parasite occurs across a broader geographic range, in more diverse habitats, in more anopheline vector species, and threatens more people than *P. falciparum*. Vivax malaria is an overwhelmingly Asian and Asia-Pacific problem, home to 91% of the 2.49 billion PAR, dozens of mosquito vector species, and endemic habitats as distinct as the temperate Korean peninsula

and tropical New Guinea. This is the same region where treatments are failing due to drug resistance or reluctance to use threatening and impractical treatments like primaquine. Vivax malaria today threatens many people with very serious illness and is especially difficult to prevent and treat. The evidence presented in this chapter and others in this thematic issue disclaim the long-held presumption of inconsequence with infection by *P. vivax*.

6. METHODS

The full methods that would allow the reader to reproduce these maps are given in Gething et al. (2012) and Sinka et al. (2010b). Here, we bring together the full suite of methodologies and present a set of summaries that highlight the data used and the assumptions made to allow the value of the resulting estimates to be assessed.

6.1. Defining the Limits of *P. vivax* Transmission

To adequately assess the global burden of *P. vivax*, the limits of the infection and the global distribution of risk must first be identified. Knowledge of the spatial distribution, and the clinical incidence within those limits, provide a foundation on which control efforts and measures of progress can be based. MAP began these efforts through defining the spatial limits of infection and the PAR, informed by ecological variables and the distribution of Duffy negative individuals, and then applied those limits to mapping the transmission and varying levels of *P. vivax* endemicity globally.

6.1.1. International Limits of P. vivax

The global spatial limits of *P. vivax* malaria were first defined for 2009 (Guerra et al., 2010) and the methods and results have since been updated for 2010 (Gething et al., 2012). A list of 95 *P. vivax* malaria endemic countries (*Pv*MECs), illustrated in Fig. 1.10A1–D1, was identified using previous methods based on international health and travel guidelines (Centers for Disease Control and Prevention, 2009; Guerra et al., 2010; WHO, 2010b; Gething et al., 2011a). The *Pv*MECs were grouped into three regions: the Americas; Africa, Saudi Arabia and Yemen (Africa+); and Central and South East Asia (CSE Asia), which was further divided into Asia and Asia-Pacific in order to resolve PAR estimates (referred to here as Asia and Asia-Pacific) (Fig. 1.2). The borders of these countries, along with national survey information and relevant, published sources and personal communications, defined the first version of the *P. vivax* spatial limits map (Guerra et al., 2010).

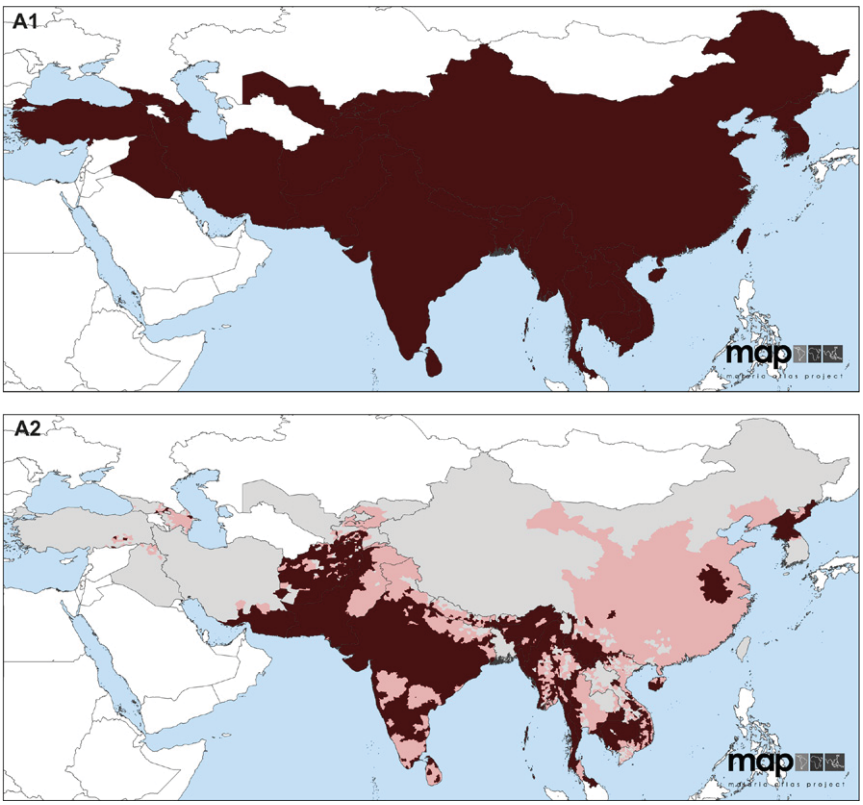


Figure 1.10 Map sequence illustrating the different exclusion layers applied by region. Maps are shown by region: Asia (A), Asia-Pacific (B), the Americas (C) and Africa+. Panel 1 = all regions of the regional *P. vivax* endemic countries; 2 = downgrading or exclusion of risk informed by annual parasite incidence data (*Pv*API); 3 = additional exclusion of risk informed by the biological temperature mask; 4 = additional downgrading or exclusion of risk informed by the aridity mask; 5 = additional downgrading or exclusion of risk informed by medical intelligence and international travel and health guidelines; 6 = the final limits definition after additionally downgrading risk in stable areas predicted to have very low prevalence by the model-based geostatistics (MBG) model. Stable transmission is shown in red, unstable transmission in pink, *P. vivax* malaria free areas in grey and countries non-endemic for *P. vivax* or outside of the region in white. For interpretation of the references to colour in this figure legend, the reader is referred to the online version of this book.

*Pv*API data, which report the number of confirmed *P. vivax* malaria cases per administrative unit per 1000 people per annum (p.a.) from *Pv*MECs, were used to further constrain the spatial limits of infection (Fig. 1.10A2–D2). *Pv*API data were obtained from a variety of sources which are provided in detail elsewhere (Gething et al., 2012). Data were unavailable for 41 of

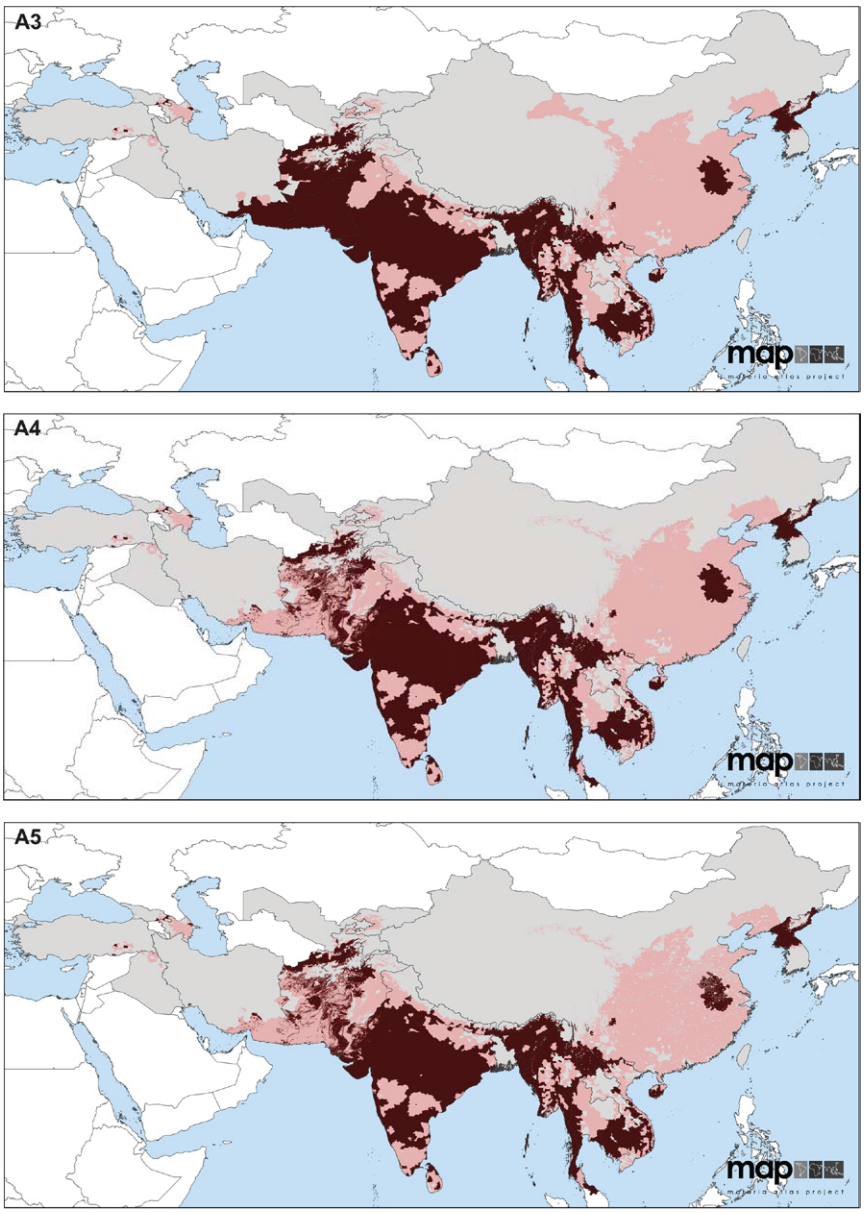


Figure 1.10, cont'd

the 95 *Pv*MECs, which were all within the Africa+ region, with the exception of Uzbekistan. Ideally, the *Pv*API data were available per administrative unit per year with each record containing information regarding the number of people in each administrative unit and the number of *P. falciparum*

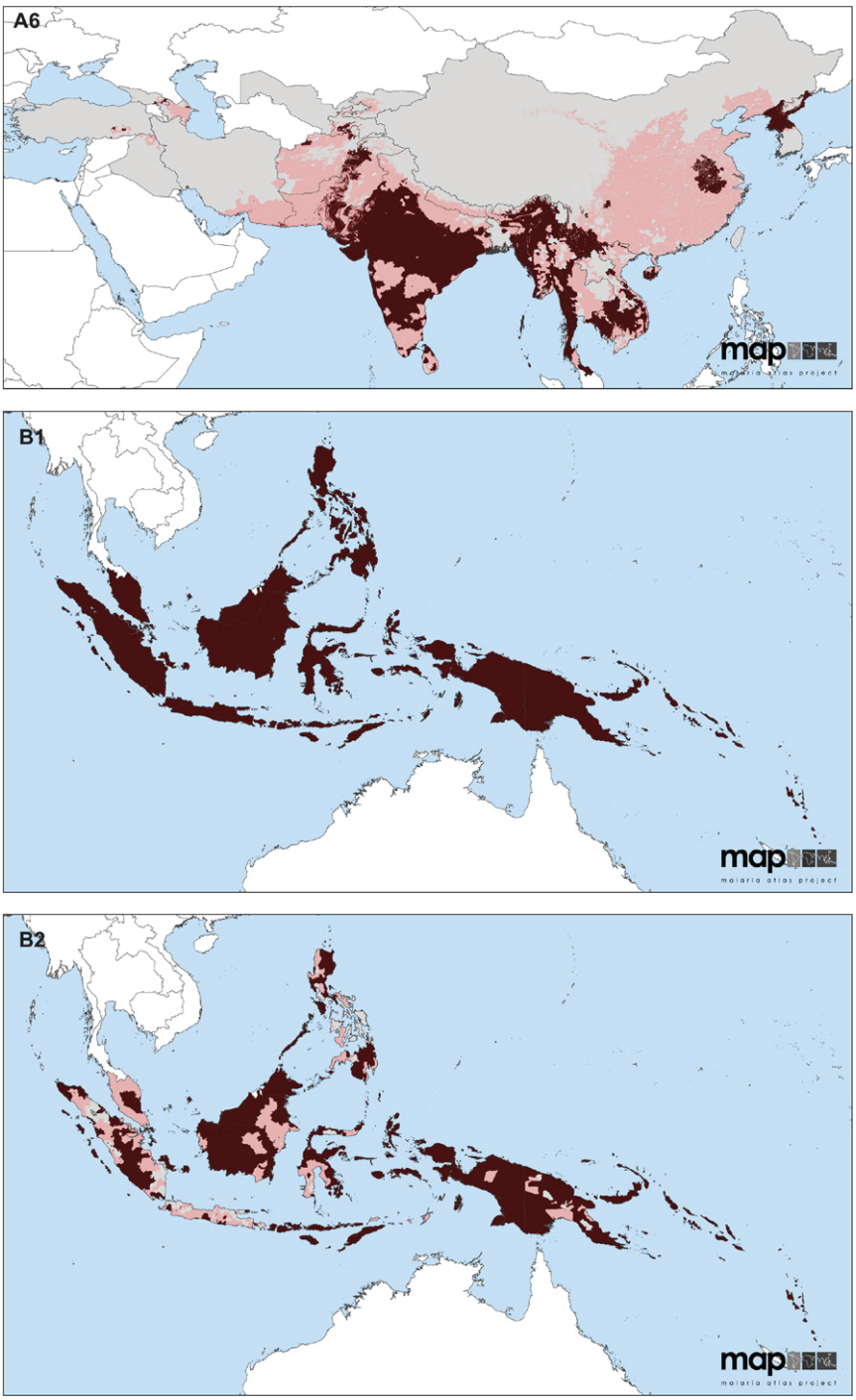


Figure 1.10, cont'd

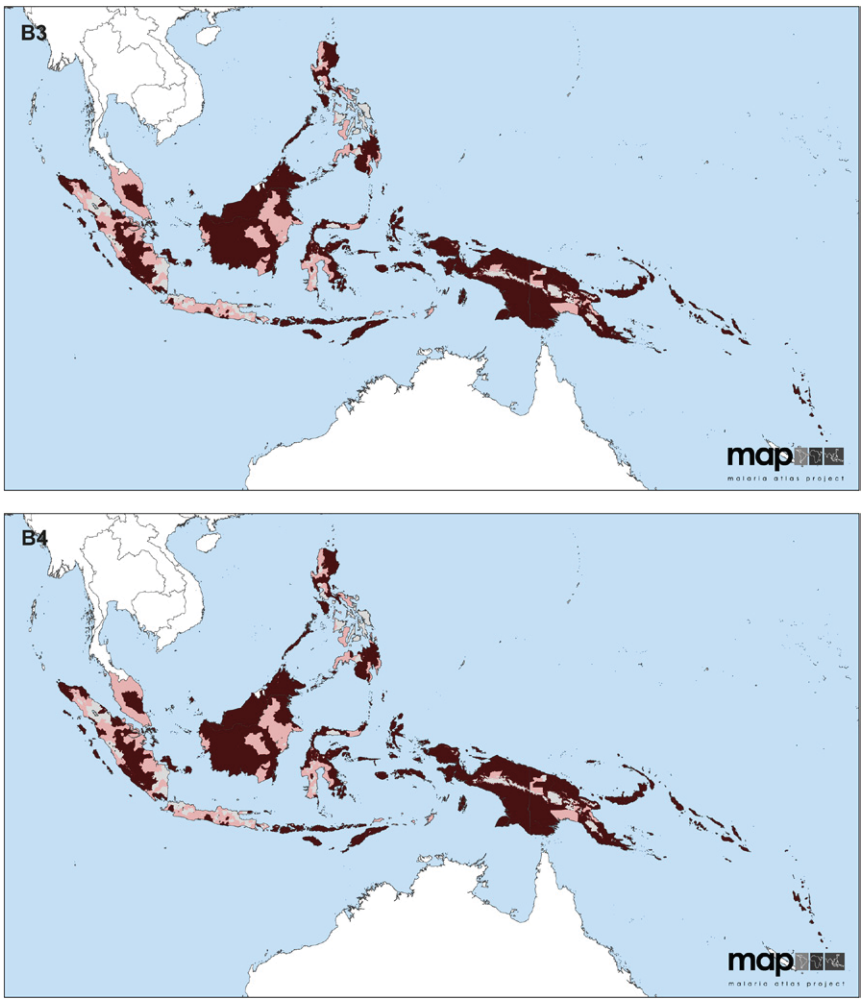


Figure 1.10, cont'd

and *P. vivax* cases. When necessary, the missing data were extrapolated from *Pv*API data estimates from preceding years or parasite species ratios confirmed by alternative sources. The aim was to obtain data for the four most recent previous years (up to 2010) at the second administrative (ADMIN2) level (or third, ADMIN3, if data were available). To map the *Pv*API data, the information was reconciled to the digital administrative boundaries available from the 2009 Global Administrative Unit Layers (GAUL) from the Food and Agriculture Association of the United Nations (FAO) within the Food Security for Action Programme (FAO, 2008). *Pv*API data were

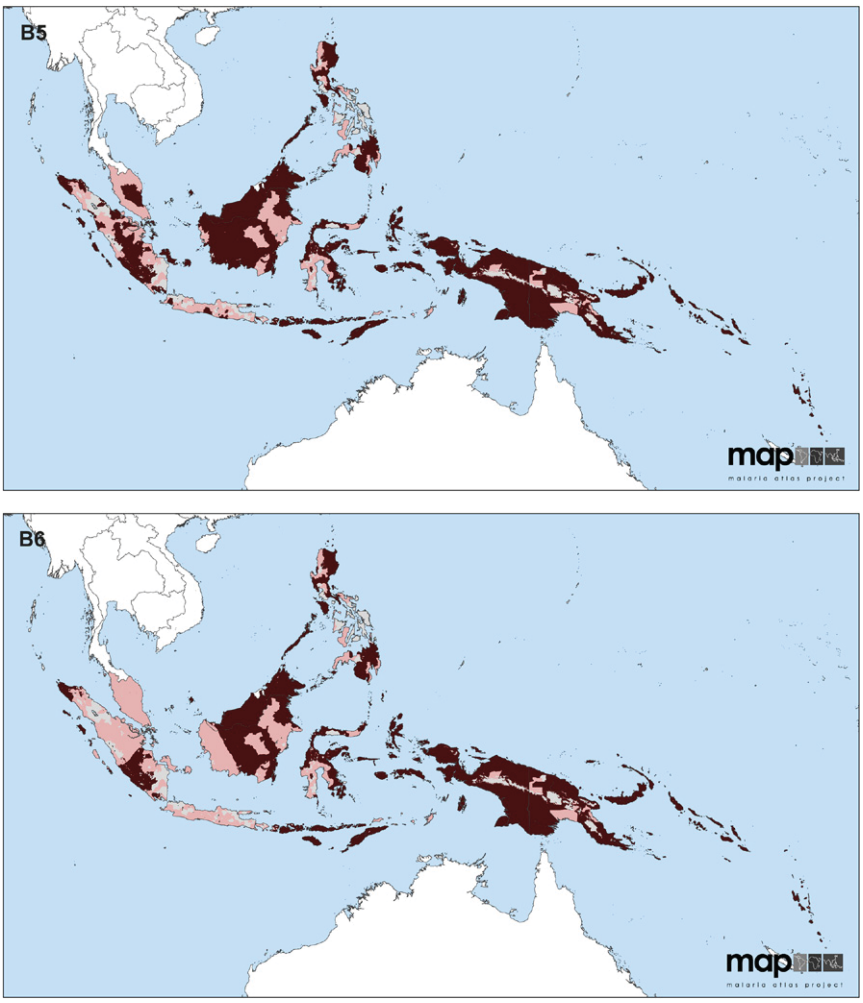


Figure 1.10, cont'd

arranged to classify areas at risk into three categories: malaria free, unstable (<0.1 cases per 1000 p.a.) or stable (≥0.1 cases per 1000 p.a.) transmission. These values were based on the Global Malaria Eradication Programme classifications; API values of <0.1‰ were considered a reliable indication to cease IRS and to shift to the consolidation phase of eradication (Pampana, 1969; Yekutiel, 1980; Guerra et al., 2008; Hay et al., 2008). This followed a transition from a cut off of 0.5 cases per 1000 p.a., which was deemed less reliable because malaria would at times return after the cessation for spraying at an API of 0.5‰ (Guerra et al., 2008). This was likely because

Figure 1.10, cont'd

surveillance, whether passive or active, was not wholly reliable or reflective of true endemicity levels. The more conservative categorization that is also applied here, of unstable transmission equating to less than 0.1 cases per 1000 p.a. allowed for less confidence in the fidelity of the prevalence source information and accounted for inaccuracies in district or provincial level reporting (Snow et al., 2005; Erhart et al., 2007; Sharma, 2007).

Biological masks were applied to further constrain the risk in the *P. vivax* endemic regions. Environmental conditions can suppress malaria transmission by limiting various components of the transmission cycle. Temperature has been shown to affect vector survival, emergence and feeding rates (Ross, 1911; Detinova, 1962; Mahmood and Reisen, 1981; Ahumada et al., 2004). The most limiting effect of temperature on malaria transmission is the

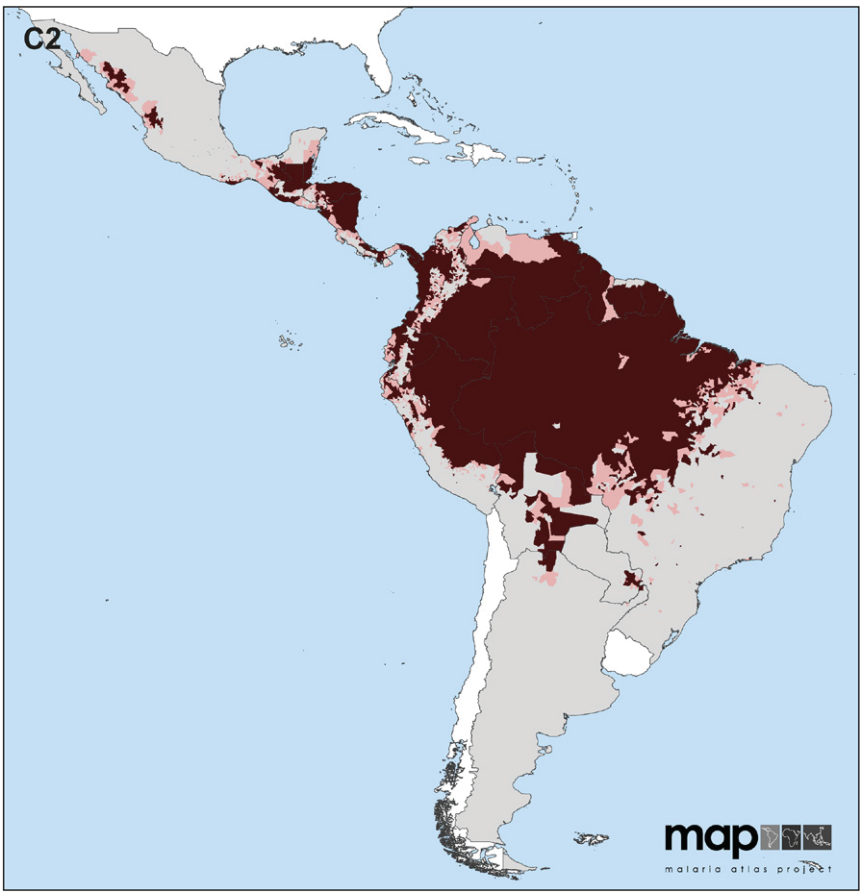


Figure 1.10, cont'd

interaction between vector lifespan and the length of the sporogonic cycle; as temperature varies throughout the year, as does the length of the extrinsic incubation period during which the parasite matures to the sporozoite stage within the gut of the mosquito (Nikolaev, 1935). For transmission to occur, the anopheline population must survive long enough for sporogony of the parasite to occur. A model that assesses the effects of temperature on *P. vivax* over time (Gething et al., 2011b) was used to generate a grid of temperature suitability proportional to vectorial capacity, a measure of transmission potential (Garrett-Jones, 1964; Smith and McKenzie, 2004). Daily vector survival was calculated as a function of local temperature and vector lifespan was assumed to be 1 month (Kiszewski et al., 2004), with exceptions being made for areas with longer living species (*Anopheles sergentii* and

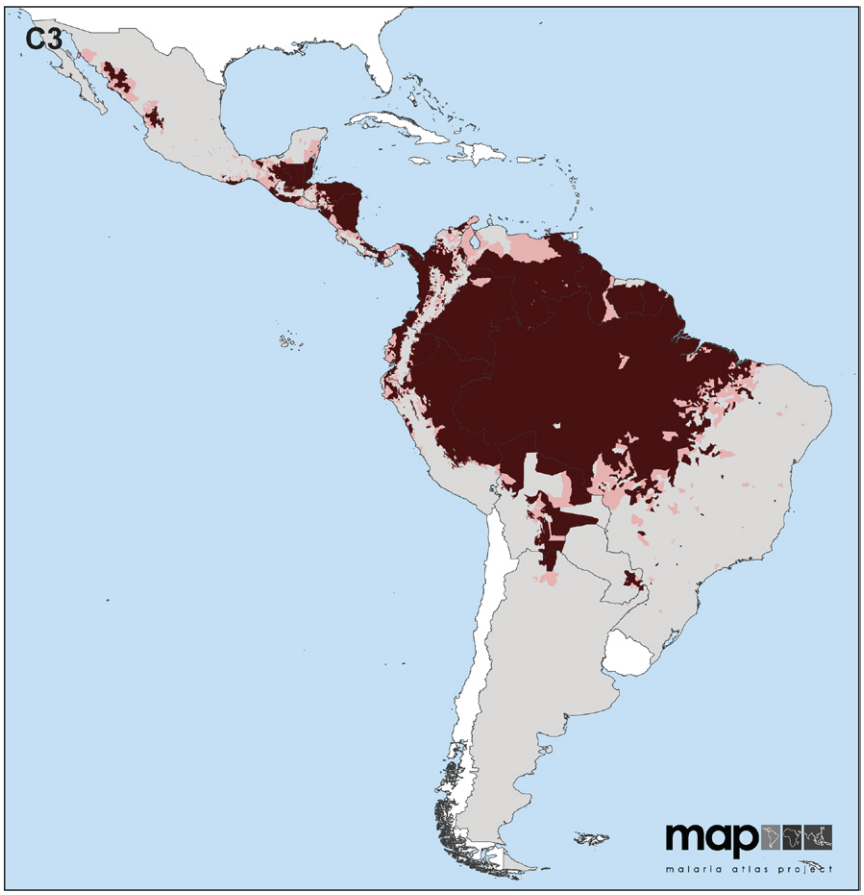


Figure 1.10, cont'd

An. superpictus) (Guerra et al., 2008; Sinka et al., 2010a). The interaction between vector lifespan and the length of the sporogonic cycle was modelled for each pixel of the temperature suitability grid (Fig. 1.5), so that pixels in which there was no time during the year that sporogony could be completed were classified as being at no risk for *P. vivax* transmission (Fig. 1.10A3–D3).

A second environmental driver of suitability for *P. vivax* transmission is availability of sufficient moisture. A mask of aridity was used to exclude areas where arid conditions would preclude anopheline survival at all life stages (Shililu et al., 2004; Gray and Bradley, 2005). Arid areas were identified using the global GlobCover Land Cover product (ESA/ESA GlobCover Project, led by MEDIAS-France/POSTEL) (Bicheron et al., 2008). The aridity mask (Fig. 1.6) was treated differently from the temperature mask

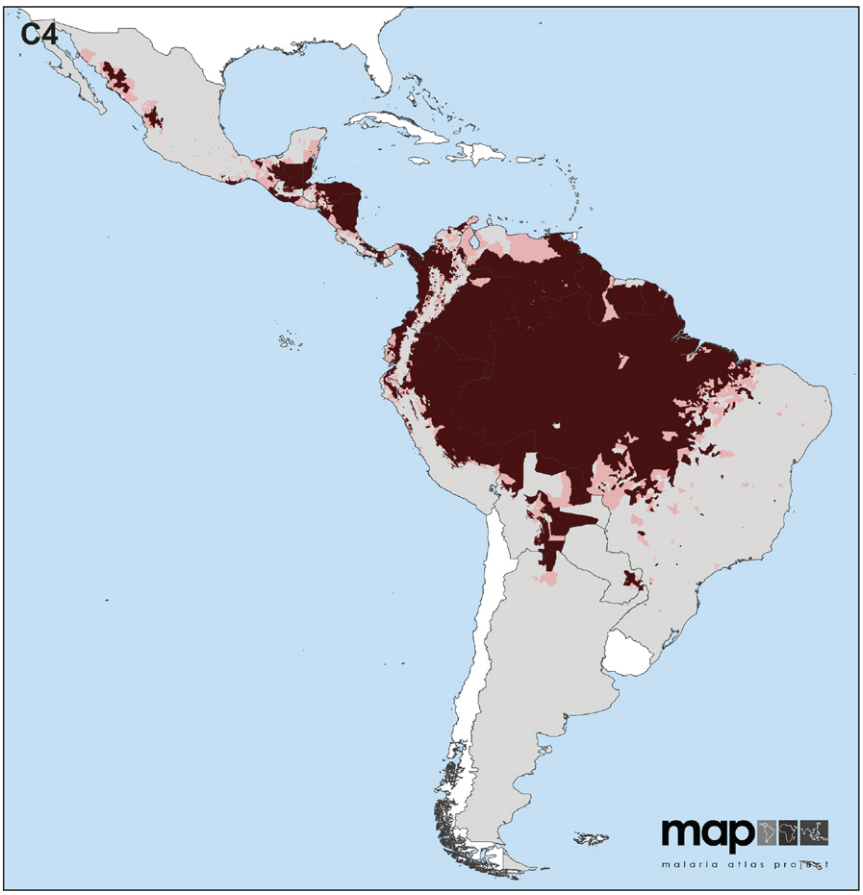


Figure 1.10, cont'd

to accommodate for the possibility of human and vector species adapting to arid conditions so that risk classes were downgraded depending on the land cover. GlobCover bare areas classified as stable risk according to the *Pv*API were downgraded to unstable, and unstable risk was stepped down to malaria free (Fig. 1.10A4–D4).

Medical intelligence was also used to refine the limits of *P. vivax* risk (Fig. 1.10A5–D5). International travel guidelines (Centers for Disease Control and Prevention, 2009; WHO, 2010b) and data obtained through personal communications were applied to identify urban and sub-national malaria free-areas. Urban areas have less malaria transmission than rural areas because of the distinct ecological conditions that result from a man-made environment (Hay et al., 2005; Tatem et al., 2008). Different species

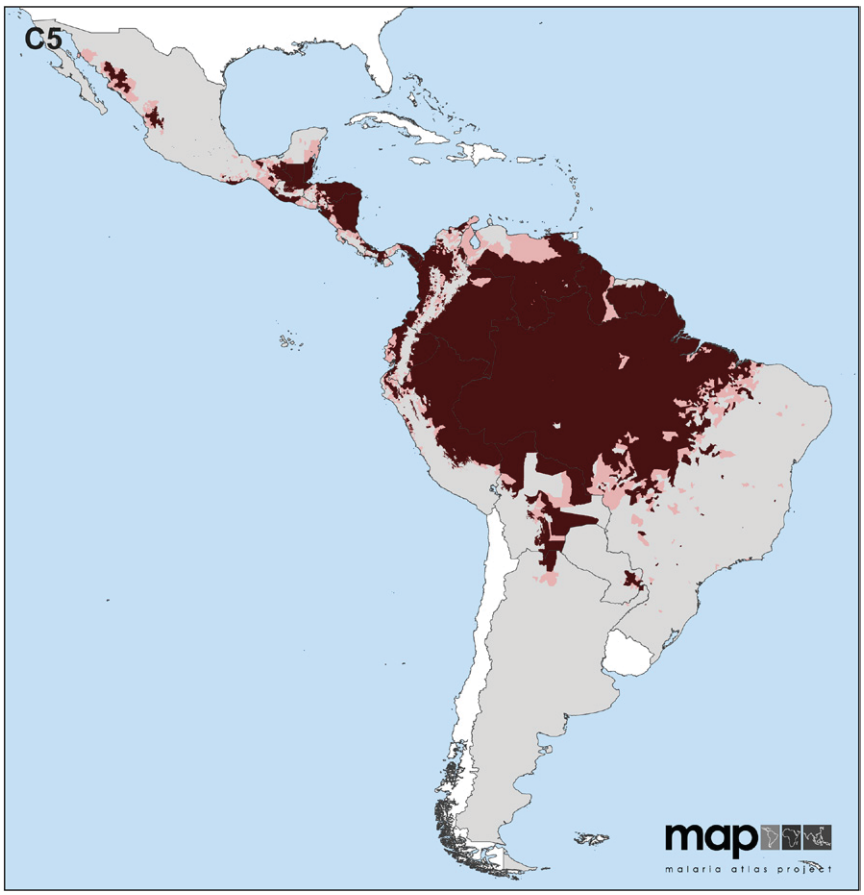


Figure 1.10, cont'd

of *Anopheles* mosquitoes respond at varying degrees, but urbanisation has been shown to reduce malaria transmission across Africa and in parts of the Americas (Hay et al., 2005; de Castro et al., 2006). In Asia, a principal vector, *An. stephensi*, has adapted to urban environments by breeding in artificial water collections (Sharma et al., 1993; Sinka et al., 2011) to the extent that stable transmission has been observed in the majority of cities (70 out of 86 cities examined) in India reporting annual parasite index data (Akhtar et al., 2009). *Anopheles culicifacies s.l.* has also been shown to transmit malaria in urban settings, but population densities and sporozoite rates indicate that this species is more affected by the environmental changes of urban areas (Nalin et al., 1985; Sharma et al., 1993; Sharma, 1995; Sinka et al., 2011). Therefore, it was assumed that urban transmission was maintained by only

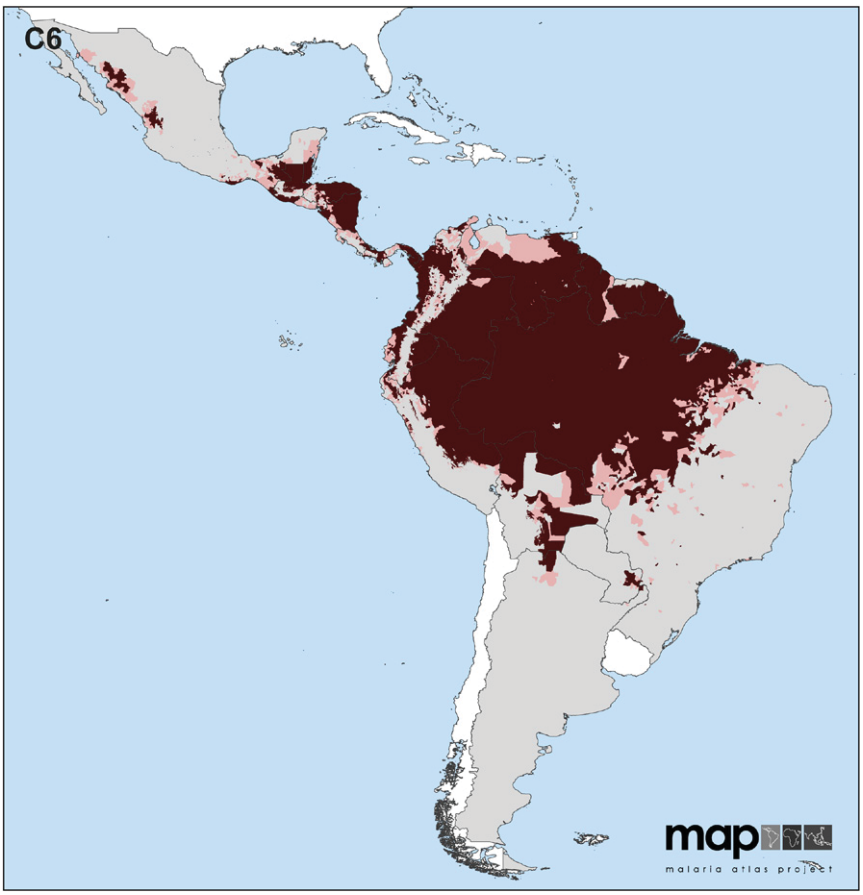


Figure 1.10, cont'd

An. stephensi. Cities specified as being malaria free by the international travel and health guidelines (Centers for Disease Control and Prevention, 2009; WHO, 2010b) or other sources were mapped using the GRUMP urban extent layer (Balk et al., 2006). Urban areas outside of the range of *An. stephensi* were classified as being at no risk of *P. vivax* transmission and areas within the vector species' range were down-graded by one risk level (stable to unstable; unstable to malaria free) accordingly. In addition to urban areas, some administrative areas, such as islands, were also declared malaria free by the international travel and health guidelines and a few specific personal communications. These areas were also classified as no risk if they were not already specified as such by the *Pv*API layer. The final limits for each region are shown in Fig. 1.10A6–D6.

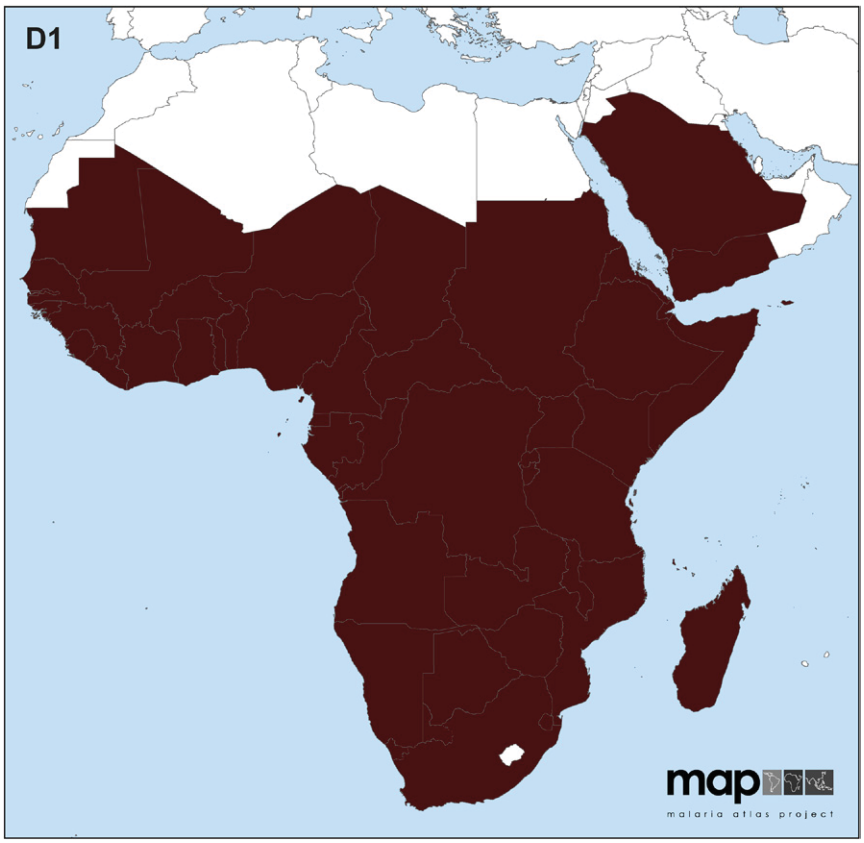


Figure 1.10, cont'd

6.1.2. The Availability of PvAPI Data by Region

Plasmodium vivax API data were available from 53 countries at a variety of administrative levels from years ranging between 2002 and 2010.

Asia. Annual parasite index data were aggregated at a variety of administrative levels. Azerbaijan, Bhutan, Cambodia, Kyrgyzstan and Thailand provided data at the first administrative (ADMIN1) level. China reported data from areas known to have little to no transmission regions at ADMIN1, and at ADMIN3 in the remaining areas. Myanmar and Nepal reported all data from ADMIN3 units. The remaining countries (Afghanistan, Bangladesh, Georgia, India, Iran, Iraq, Lao PDR, Pakistan, Republic of Korea, Sri Lanka, Tajikistan, Turkey and Viet Nam) collected *Pv*API data from ADMIN2 units. *Pv*API data were available from 4443 risk units in Asia.

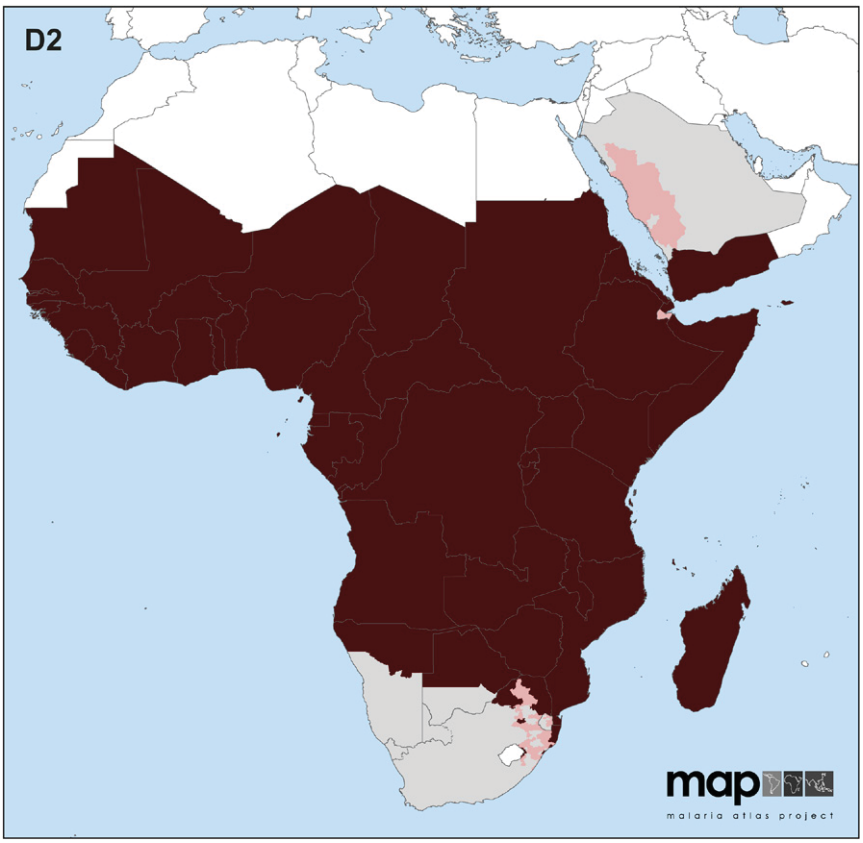


Figure 1.10, cont'd

Asia-Pacific. Annual parasite index data were provided for all of the *P. vivax* endemic countries in Asia-Pacific. Only Malaysia provided data up until 2010. Indonesia and Timor-Leste had data until 2008 available and 2007 was the last reporting year for the remaining countries. Malaysia, the Solomon Islands, Timor-Leste and Vanuatu provided data at the ADMIN1 level and Indonesia, Papua New Guinea and the Philippines reported data from ADMIN2 units. *Pv*API data were obtained from a total of 559 risk units in Asia-Pacific.

The Americas. *Pv*API data were available from all *Pv*MECs in the Americas. For the years reported, which varied by country, six countries (Belize, El Salvador, Guatemala, Guyana, Nicaragua and Suriname) reported data at the first administrative (ADMIN1) level and Venezuela reported data at the ADMIN1 and ADMIN2 level. Eleven countries (Argentina, Bolivia, Brazil, Colombia, Costa Rica, Ecuador, French Guiana, Honduras, Mexico,

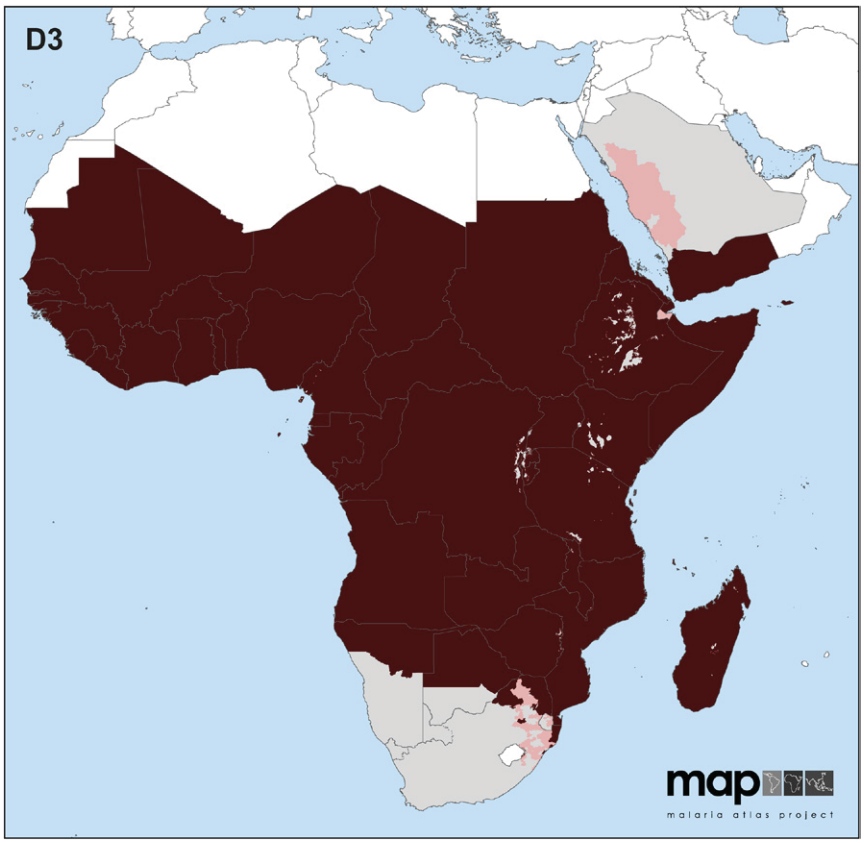


Figure 1.10, cont'd

Panama and Paraguay) reported ADMIN2 data and Peru reported data from ADMIN3 units. Data were available for a total of 12,514 administrative (or risk) units in the Americas.

Africa+. *Pv*API data were aggregated at a variety of levels by the six (out of 46) *Pv*MECs that reported data in the region. For the years of available data, Dijbouti, Saudi Arabia and Yemen reported ADMIN1 level data, Namibia reported from a mix of ADMIN1 and ADMIN2 units, and South Africa and Swaziland had data available at the ADMIN2 level. Data were provided for a total of 377 risk units for the Africa+ region.

6.1.3. The Global Distribution of the Duffy Blood Group

The distribution of Duffy-negative populations was used as another exclusion layer to constrain the limits of *P. vivax* transmission (Fig. 1.7). The reader

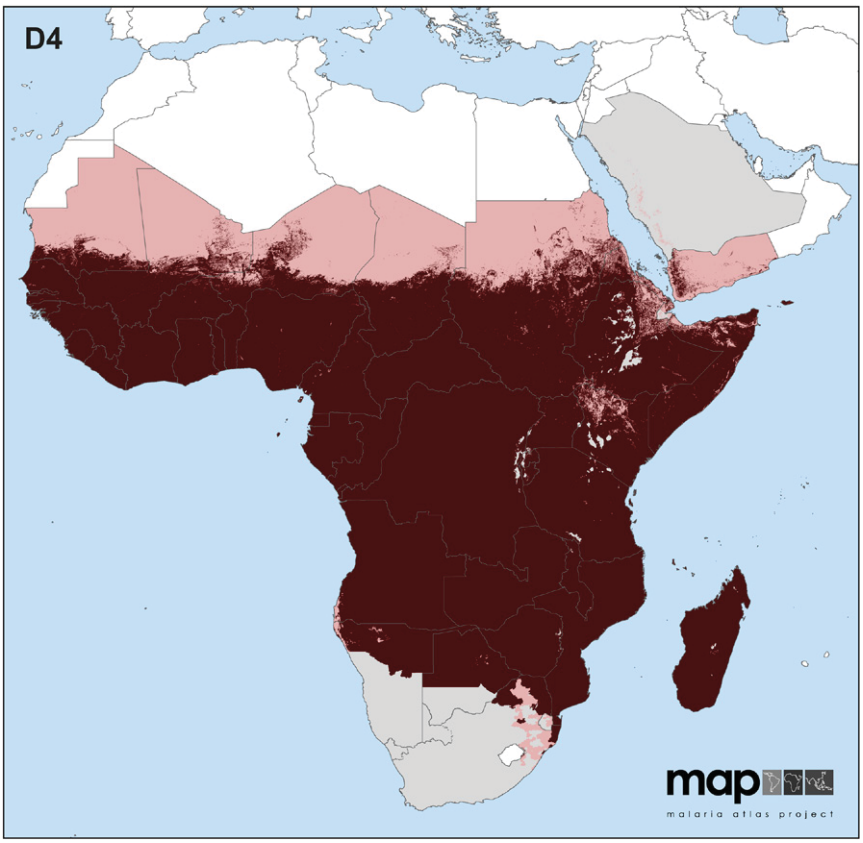


Figure 1.10, cont'd

is again referred to a detailed review of the effect of the lacking Duffy antigen on resistance to *P. vivax* provided in this thematic issue of *Advances in Parasitology* (Chapter 2, Volume 81). The Duffy antigen refers to a receptor expressed on the surface of red blood cells, which *P. vivax* has been shown to be dependent upon for erythrocytic invasion (Miller et al., 1976; Barnwell et al., 1989; Wertheimer and Barnwell, 1989). Duffy-negative individuals, who lack the antigen, are therefore largely refractory to *P. vivax* infection and high frequencies of the phenotype are presumed to suppress *P. vivax* endemicity in areas that would otherwise be well suited for transmission. A continuous map of the Duffy-negative phenotype, described in detail elsewhere (Howes et al., 2011) and briefly here, was used as the exclusion surface.

To model the global distribution of the Duffy-negative phenotype, a database of Duffy blood group surveys was assembled. Surveys of Duffy-variant

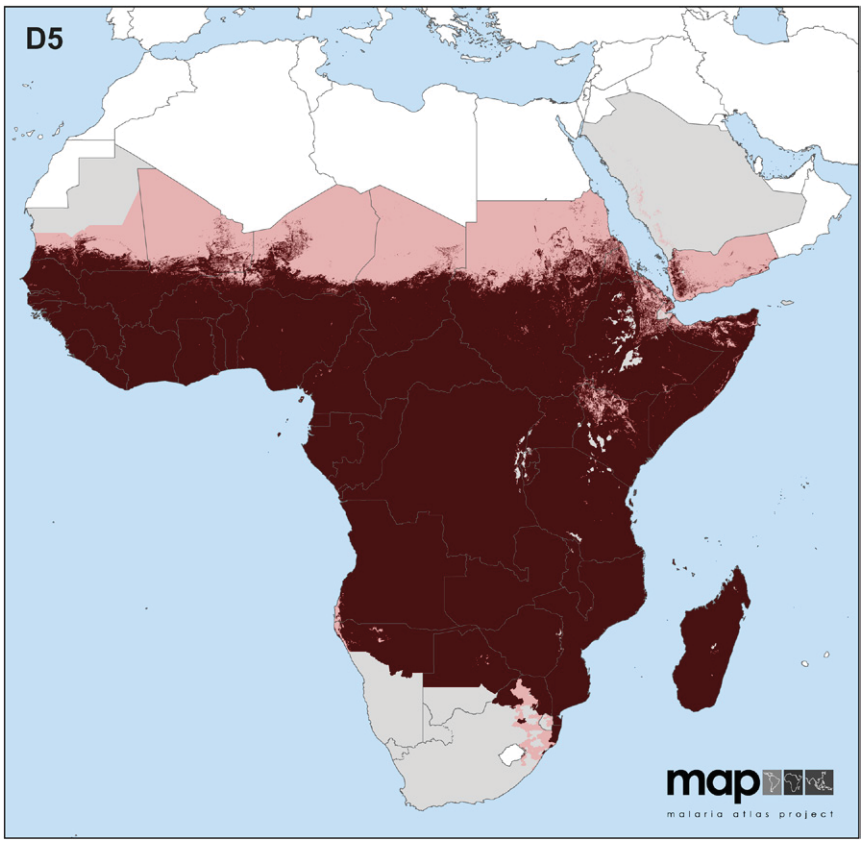


Figure 1.10, cont'd

frequencies were compiled from systematic searches of published literature, personal communications and sources obtained through previously published databases (Mourant et al., 1976; Cavalli-Sforza et al., 1994). Surveys dating back to 1950, when the blood group was first described (Cutbush and Mollison, 1950), were included. Results were refined so that only population-based surveys were used and potentially biased samples were removed. Survey data were geopositioned following guidelines described previously by MAP (Guerra et al., 2007).

In addition to the frequency of the blood group variants, the diagnostic method used in each survey was recorded to classify the type of information provided to ultimately inform the model. The Duffy antigen has two main variant forms, Fy^a and Fy^b. These differ by a single amino acid substitution (Gly42Asp), which is encoded in two alleles, *FY*A* and

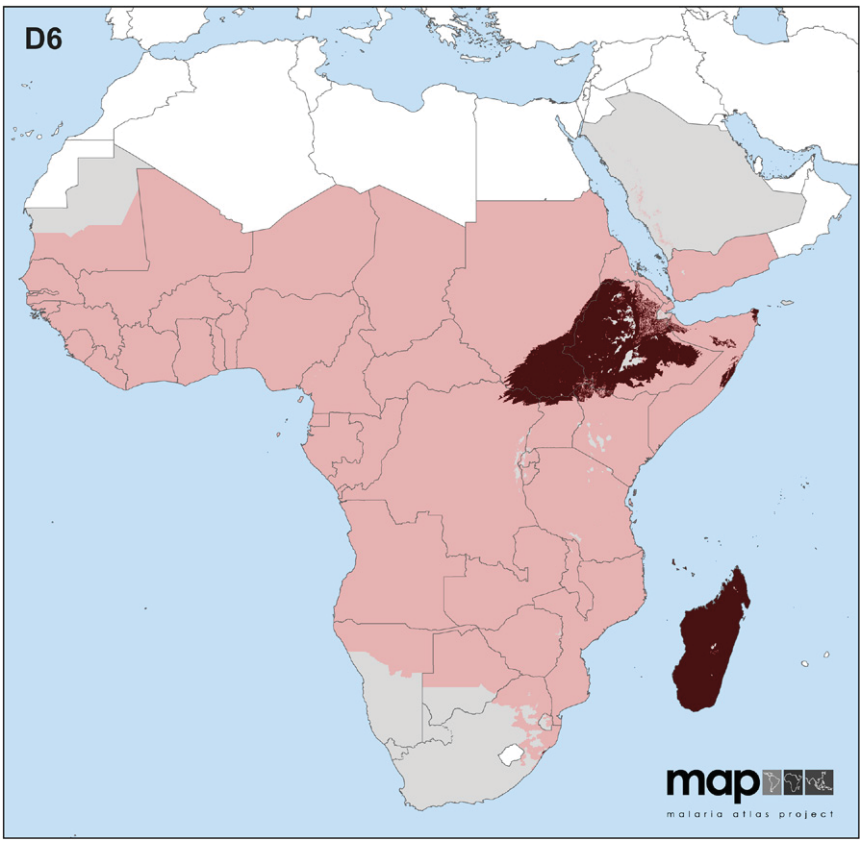


Figure 1.10, cont'd

*FY*B*, that vary based on a single base substitution (G125A) (Zimmerman, 2004; Langhi and Bordin, 2006). A separate single base substitution in the gene's promoter region (T-33C) may block the expression of the gene and result in a null 'erythrocyte silent' (ES) phenotype. This mutation is most commonly associated with the *FY*B* (*FY*B*^{ES}) allele (the *FY*A*^{ES} allele is extremely rare) (Langhi and Bordin, 2006; Sellami et al., 2008). These four alleles (*FY*A*, *FY*B*, *FY*B*^{ES}, *FY*A*^{ES}) result in 10 possible genotypes and four possible phenotypes: Fy(a+b+), Fy(a+b−), Fy(a−b−) and Fy(a−b−) (Howes et al., 2011). Methods to diagnose Duffy types can focus on different aspects of the system: the individual polymorphic sites, the genotype, the presence or absence of the antigen, and which particular type of antigen is being expressed. The assembled data were therefore grouped into five types according to which diagnostic approach was used:

i) *Genotype*, where full genotypes were reported; ii) *Phenotype*, where full serological diagnoses (anti-Fy^a and anti-Fy^b sera) were used; iii) *Promoter*, if results only stated antigen expression or non-expression (no distinction between Fy^a and Fy^b); iv) *Phenotype-a*, if only Fy^a was tested for (no distinction between Fy^b and the negative phenotype); and v) *Phenotype-b*, if only Fy^b was tested for (no distinction between Fy^a and the negative phenotype).

This dataset formed the evidence-base for a geostatistical model used to predict a continuous map of the prevalence of the Duffy-negative phenotype [Fy(a−b−)], the phenotype encoded by the $FY{\star}B^{ES}/FY{\star}B^{ES}$ genotype. The $FY{\star}A^{ES}$ is highly infrequent and so was modelled as a small constant rather than a spatially variable allele. Although not all of the five data types directly informed the prevalence of Duffy negativity, the model was able to infer useful information directly from each (by ruling out certain genotypes), so all were included as model inputs. To allow this, the model structure was based on the frequencies of the two loci involved in the system: the promoter type (determining expression *versus* non-expression at the T-33C site) and the coding region (determining the G125A polymorphism). The Bayesian framework considered the two loci as spatially independent random fields, and used the survey data to map the frequency of the expression of each variant. Each of the data types informed different aspects of these loci: some informed both loci and others excluded possible variants. The model also incorporated a land cover variable to distinguish sub-Saharan African populations from others on the continent. This was to inform the model of the high probability of association between silencing mutation of the Fy^b variant and the $FY{\star}B^{ES}$ allele, which was observed in high frequencies across sub-Saharan Africa.

The Bayesian model-based geostatistical framework predicted Duffy group expression frequencies in all geographic locations across a 5 × 5 km grid to generate a continuous global surface of each variant. The Duffy negativity phenotype was expressed by the squared frequency of the $FY{\star}B^{ES}$ allele. The map reveals that the homozygous null phenotype is highly constrained to sub-Saharan African populations, with localized high-frequency areas in the Americas. Historical perceptions have supported the assumption that *P. vivax* is absent from much of the African continent (Rosenberg, 2007). However, evidence of autochthonous transmission within Africa indicated that areas of the continent should not be excluded a priori. This map, therefore, provided an evidence-based exclusion layer for Duffy-negative populations resistant to *P. vivax* infection in Africa.

6.2. Global Endemicity of *P. vivax*

6.2.1. P. vivax *Parasite Rate Data*

To map *P. vivax* malaria endemicity within the limits of stable transmission, in addition to the updated geographic boundaries and corresponding populations at risk of *P. vivax*, an updated georeferenced database of *P. vivax* parasite rate (*Pv*PR) survey data were compiled. PR was used here because it is the most ubiquitous of the malariometric measures of risk (Hay et al., 2009). *Pv*PR data represent the proportion of a randomly sampled population to have detectable *P. vivax* parasitaemia when screened via microscopy or RDTs and is the most consistently measured index of malaria endemicity. The final database consisted of unique surveys obtained from published and unpublished literature sources spanning the period 1985–2010. The database included information on the survey origin, how the location was determined (georeferencing method), time period, age group, sample size and diagnostic method used.

The *Pv*PR database was made up of 9970 spatiotemporally unique records from 432 different sources. Data were available from 53 countries, 12 of which were in the Americas, 19 in the Africa+ region, 15 in Asia and 7 in Asia-Pacific. There were 44 *Pv*MECs not represented in the database, most of which were in Africa, with the exception of Argentina, Azerbaijan, Belize, Bhutan, Korea DPR, El Salvador, Georgia, Guyana, Iran, Kyrgyzstan, Panama, Paraguay, Republic of Korea and Uzbekistan. Details of the PR data that were input into the model from each region are given below.

6.2.2. Modelling P. vivax *Endemicity*

To generate a continuous surface of *P. vivax* endemicity using PR data, a flexible modelling framework based on model-based geostatistics (MBG) (Diggle et al., 1998; Diggle and Ribeiro, 2007) was used. With areas of stable transmission converted into a 5 × 5 km grid, MBG models allow for endemicity values to be predicted at each pixel as a function of the geographically varying mean of survey values and a weighted average of neighbouring data values. MBG models are well suited for predicting endemicity values, in this case PRs, for a number of reasons. First, the mean PR values may be defined as a function of multiple environmental covariates that influence malaria transmission. Second, a covariance function may be employed to define the spatial heterogeneity of the PR data and, in turn, define the appropriate weight for each data point when generating a prediction. Third, uncertainty can be based on the nature and density of data surrounding a pixel. Fitting MBG models with Bayesian

inference and a Markov chain Monte Carlo (MCMC) algorithm produces uncertainty metrics around the final predictions as well as the model inputs in the form of predictive posterior distributions (Patil et al., 2011). Areas with the least uncertainty are those with a large number of recent surveys with relatively homogenous results, whereas greater uncertainty would be found in places with sparse or old surveys with ranges of different observed PRs.

To model a global endemicity surface of *P. vivax* malaria, an MBG framework that had been successfully employed for falciparum malaria (Hay et al., 2009; Gething et al., 2011a) was used following modifications made to accommodate biological features unique to *P. vivax*. The modelling methods incorporate surveys from a wide time period such that older surveys are given less weight than recent ones. The environmental covariates included were those that had an a priori expectation to affect malaria transmission intensity. These were urban areas defined by the GRUMP urban extent product (Balk et al., 2006; CIESIN/IFPRI/WB/CIAT, 2007), a long-term average vegetation index product used as a proxy for available moisture for vector reproduction and survival (Hay et al., 2006; Scharlemann et al., 2008), and the temperature suitability index derived from the model described above, which identifies areas suitable for transmission based on the requirements of vector survival and sporogony (Gething et al., 2011b).

PR data were standardised by age because of variation in infection rates observed in different age groups. It is often observed that malaria prevalence rises rapidly in infancy before reaching a plateau in early childhood and declining through adolescence and adulthood. This phenomenon was modelled using a previously described framework (Smith et al., 2007) to standardise for the prevalence variability among age groups. The model was informed with finely age-stratified *Pv*PR surveys to represent vivax-specific age profiles (Mueller et al., 2009b; Lin et al., 2010) and was used to convert all the observed survey prevalence values to standardised age-independent values for use in the MBG modelling. Predictions were made for all-age prevalence estimates for individuals aged one to 99 years ($PvPR_{1\text{-}99}$). Children aged <1 year were not included because of the potential confounding effect of maternal antibodies, but all other ages were included. This deviated from the method of using the 2- to 10-year cohort for falciparum malaria (Guerra et al., 2007; Gething et al., 2011a), because all age ranges are typically sampled for *P. vivax*, which is found at relatively lower prevalence rates.

To determine the endemicity of *P. vivax*, it was also important to incorporate Duffy negativity into the modelling framework because of the refractory nature of the phenotype to the parasite. The map of Duffy

negativity described above informed the model of the fraction of Duffy-negative individuals in the population at each pixel of the predicted surface. The parasite endemicity prediction could be made, therefore, from the vivax-susceptible or Duffy positive portion of the population. This meant that the proportion of *Pv*PR could not exceed the percentage of the population who were Duffy positive and that predictions in data-sparse portions of Africa could borrow strength from the Duffy negativity surface because estimates were limited to a more restricted range of potential outcomes.

The endemicity surface that results from the MBG modelling framework described is a 5 × 5 km grid of predictions for $PvPR_{1–99}$ within the limits of stable *P. vivax* transmission. For practical reasons, and because areas endemic for *P. vivax* have distinct ecological, entomological and epidemiological characteristics, the *P. vivax* endemic world was divided into four regions: the Americas, Africa+, Asia and Asia-Pacific. Separate models were fitted to each region so that a $PvPR_{1–99}$ estimate averaged across the 12 months of 2010 was found. The endemicity map was created by using the mean of each posterior distribution as a point estimate and uncertainty was shown as the ratio of the posterior distribution IQR to its mean. The IQR was found to express the precision with which the $PvPR_{1–99}$ values were predicted. Calculation of the ratio of the IQR of each posterior distribution to its mean generated an index that demonstrated how the model performance varied with data density in different locations. This index was also weighted by population density to generate a map to show where high levels of uncertainty may be operationally significant.

6.3. The Refined Population at Risk of *P. vivax*

The various data sources described above were combined to produce the final spatial limits map (Fig. 1.10A6–D6). After the implementation of MBG modelling, some of the regions that were estimated as being within the limits of stable transmission were downgraded to unstable transmission. If the model outputs of *Pv*PR were extremely low due to a large abundance of surveys reporting zero infections in that area or, in Africa, because of high Duffy negativity, a decision rule was applied such that pixels that were predicted with high certainty (probability >0.9) to be less than 1% *Pv*PR were reassigned as unstable.

The PAR of *P. vivax* malaria was estimated using the constrained limits of infection and population values for 2010 projected from the year 2000 GRUMP *beta* version population counts. The result was a 1 × 1 km spatial grid of population surface of the number of people living at stable or

unstable risk in each country as well as surface area of the regions at risk. The population surface, along with uncertainty maps, was also used to calculate a population-weighted index of uncertainty. The initial PAR based on the *Pv*MECs was 5.36 billion individuals in a land area of 69 million km^2. After applying the *Pv*API data (1.3 billion PAR excluded), temperature suitability (sporogony duration; 61 million PAR excluded), aridity (32 million PAR excluded), medical intelligence (713 million PAR excluded) and the Duffy negativity (768 million PAR excluded) layers, the remaining PAR was 2.49 billion in an area of 44 million km^2. This indicates that *P. vivax* was endemic across approximately a third of the world's surface. Half of that area was found in Africa (51%) and a quarter was in both America (22%) and Asia (27%). However, high population density in parts of Asia and the large proportion of the protective Duffy-negative phenotype found in African populations, meant that 82% of the 2.49 billion people at risk were in Asia with the remaining 17% being spread across Asia-Pacific (9%), the Americas (6%) and Africa (3%). Well over half of the PAR lived in areas of unstable transmission (62%; 1.52 billion) where transmission was very low and unlikely to exceed one case per 10,000 people per annum. A total of 965 million people were estimated to be living at risk of stable transmission. The endemicity maps (Fig. 1.3A2–D2) demonstrate that the transmission potential in and among stable transmission areas may vary greatly, even within relatively small geographic areas.

6.4. Mapping the Range of Dominant Vector Species

Vector distribution maps have long been used as a tool to aid malaria control globally. Examples of these maps date back to the 1950s and include vector species maps (May, 1951) and the ecological zones of malaria epidemiology determined by Macdonald based on climatology and known vector species ranges (Macdonald, 1957). More recent ecological (Mouchet et al., 2004) and global vector distribution (Kiszewski et al., 2004) maps have been widely adopted by the malaria research community. A recent series of publications by MAP has attempted to update malaria vector distribution maps with a comprehensive and extensive evidence base (Hay et al., 2010c; Sinka et al., 2010a, 2010b, 2011, 2012). There are 465 formally described species of *Anopheles* mosquitoes and more than 50 unnamed species and species complexes (Harbach, 2011). Approximately 70 species and species complexes have been incriminated to transmit malaria parasites (Service and Townson, 2002) and of those, 41 have been identified as DVS (Hay et al., 2010c). Determination of vector dominance is generally based

on factors that increase overall vectorial capacity (Takken and Lindsay, 2003), including abundance, propensity for feeding on humans and the mean adult longevity (to determine if the species lives long enough to transmit the parasite) (Hay et al., 2010c). In the process of determining vector dominance, it was noted which anophelines may have the potential to transmit *P. vivax*. A literature search of 'vivax' and the *Anopheles* species name was performed to identify evidence supporting wild transmission of the vivax parasite. The list of the *Anopheles* species, species complexes and groups as well as the evidence gathered regarding the potential for transmission of *P. vivax* are shown in Table 1.2.

To predict the geographic range of the 41 DVS of malaria, known occurrence points, expert opinion maps, ecological data and modelling techniques were applied. To begin, 15,837 occurrence records from 4800 sources were acquired from systematic searches of formal and informal literature sources and compiled into a comprehensive database (Hay et al., 2010c; Sinka et al., 2012). Expert opinion (EO) maps were then digitized from exhaustive searches of published maps, which are referenced in detail elsewhere (Sinka et al., 2010a, 2010b, 2011), and refined by consultation with a TAG. A suite of environmental and climatic variables known to shape vector distribution landscapes (such as elevation, land surface temperature and precipitation) were also included in the database (Sinka et al., 2010b). BRT modelling methodology (Elith et al., 2008) was applied to generate a predicted distribution map for each DVS. Distributions were generated for nine species/species complexes in the Americas; 13 for Africa, Europe and the Middle East; and 19 in Asia, 14 of which were in Asia (five were only in Asia), and 16 in Asia-Pacific (three were only in Asia-Pacific). Information was also gathered regarding the bionomics of the DVS, which greatly affects the potential impacts of common malaria interventions such as ITNs and IRS. Behaviours that were searched for and catalogued by species included larval site and habitat types, and adult resting and biting behaviours.

The predicted ranges of the DVS varied greatly across the regions, with relatively straightforward vector profiles in the Americas and Africa and very complex vector distributions across Asia. The distribution maps illustrate a probability of occurrence, but do not indicate the predicted prevalence. A positive (coloured) pixel does indicates that the probability of occurrence is >0.5 (>0.5 and ≤ 1.0) and that a negative pixel (not coloured/grey) represents a probability of occurrence <0.5 (0–0.5). The regional maps shown here are generated from an amalgamation of individual DVS distribution maps. The TAG identified the top three DVS per country (if the country

had more than three DVS) and ranked the DVS by their importance. Where there was great variation within countries (e.g. Indonesia), more detailed spatial information regarding the DVS was gathered. The rankings were then used to determine the order in which the species-specific distribution maps were layered to generate global-scale maps for the *Anopheles* vectors of greatest public health importance.

LIST OF ABBREVIATIONS

ACT Artemisinin-based combination therapy
ADMIN Administrative unit
Africa+ Africa, Yemen and Saudi Arabia
Asia Mainland Asia excluding the Malaysian Peninsula
Asia-Pacific The Malaysian Peninsula and the islands of Asia and the Pacific
CSE Asia Central and South East Asia
DVS Dominant vector species or species complexes
ELISA Enzyme-linked immunosorbent assay
G6PD Glucose-6-phosphate dehydrogenase
GRUMP Global Rural-Urban Mapping Project
IQR Inter-quartile range
Lao PDR Lao People's Democratic Republic
Korea DPR Democratic People's Republic of Korea
MAP Malaria Atlas Project
MBG Model-based geostatistics
MCMC Markov chain Monte Carlo
PAR Populations at risk
***Pv*API** *P. vivax* annual parasite incidence
***Pv*MEC** *P. vivax* malaria endemic country
***Pv*PR** *P. vivax* parasite rate
***Pv*PR$_{1–99}$** *P. vivax* parasite rate in 1–99-year olds
RDT Rapid diagnostic test
s.l. *sensu lato*
s.s. *sensu stricto*
TAG Technical advisory group

ACKNOWLEDGEMENTS

The large global assembly of parasite prevalence data was critically dependent on the generous contributions of data made by a large number of people in the malaria research and control communities and these individuals are listed on the MAP website (http://www.map.ac.uk/acknowledgements). We thank David Pigott and Jennie Charlton for proofreading the manuscript. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (#095066), which also supports PWG, CAG and KEB. IRFE is funded by grants from the University of Oxford—Li Ka Shing Foundation Global Health Program and the Oxford Tropical Network. CLM is funded by a Biomedical Resources Grant from the Wellcome

Trust (#091835). The initial vector database and species distribution maps were created while MES was funded by a project grant from the Wellcome Trust (#083534) to SIH. The compilation of the global distributions was conducted with support and funding from the Bill and Melinda Gates Foundation via the VECNet consortium (http://www.vecnet.org/). REH is funded by a Biomedical Resources Grant from the Wellcome Trust (#085406). RNP is a Wellcome Trust Senior Fellow in Clinical Science (#091625). JKB is supported by a Wellcome Trust grant (#B9RJIXO). PWG and SIH also acknowledge support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health (http://www.fic.nih.gov). This work reviewed here forms part of the output of the Malaria Atlas Project (MAP, http://www.map.ox.ac.uk), principally funded by the Wellcome Trust, UK (http://www.wellcome.ac.uk). MAP also acknowledges the support of the Global Fund to Fight AIDS, Tuberculosis, and Malaria (http://www.theglobalfund.org). The funders had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Achee, N.L., Korves, C.T., Bangs, M.J., Rejmankova, E., Lege, M., Curtin, D., Lenares, H., Alonzo, Y., Andre, R.G., Roberts, D.R., 2000. *Plasmodium vivax* polymorphs and *Plasmodium falciparum* circumsporozoite proteins in *Anopheles* (Diptera: Culicidae) from Belize, Central America. J. Vector Ecol. 25, 203–211.

Adak, T., Kaur, S., Singh, O.P., 1999. Comparative susceptibility of different members of the *Anopheles culicifacies* complex to *Plasmodium vivax*. Trans. R. Soc. Trop. Med. Hyg. 93, 573–577.

Adak, T., Sharma, V.P., Orlov, V.S., 1998. Studies on the *Plasmodium vivax* relapse pattern in Delhi, India. Am. J. Trop. Med. Hyg. 59, 175–179.

Adak, T., Singh, O.P., Das, M.K., Wattal, S., Nanda, N., 2005. Comparative susceptibility of three important malaria vectors *Anopheles stephensi, Anopheles fluviatilis*, and *Anopheles sundaicus* to *Plasmodium vivax*. J. Parasitol. 91, 79–82.

Adak, T., Singh, O.P., Nanda, N., Sharma, V.P., Subbarao, S.K., 2006. Isolation of a *Plasmodium vivax* refractory *Anopheles culicifacies* strain from India. Trop. Med. Int. Health 11, 197–203.

Ahumada, J.A., Lapointe, D., Samuel, M.D., 2004. Modeling the population dynamics of *Culex quinquefasciatus* (Diptera: Culicidae), along an elevational gradient in Hawaii. J. Med. Entomol. 41, 1157–1170.

Akhtar, R., Dutt, A., Wadhwa, V., 2009. Malaria resurgence in urban India: lessons from health planning strategies. In: Akhtar, R., Dutt, A., Wadhwa, V. (Eds.), Malaria in South Asia: Eradication and Resurgence During the Second Half of the Twentieth Century (Advances in Asian Human-Environmental Research), Spring, London, pp. 141–155.

Alam, M.S., Khan, M.G., Chaudhury, N., Deloer, S., Nazib, F., Bangali, A.M., Haque, R., 2010. Prevalence of anopheline species and their *Plasmodium* infection status in epidemic-prone border areas of Bangladesh. Malar. J. 9, 15.

Alonso, P.L., Brown, G., Arevalo-Herrera, M., Binka, F., Chitnis, C., Collins, F., Doumbo, O.K., Greenwood, B., Hall, B.F., Levine, M.M., Mendis, K., Newman, R.D., Plowe, C.V., Rodriguez, M.H., Sinden, R., Slutsker, L., Tanner, M., 2011. A research agenda to underpin malaria eradication. PLoS Med. 8, e1000406.

Am, N.T., Riec, L.Q., Huyen, V.T., Lan, N.B., 1993. Etudes entomo-épidémiologiques du paludisme dans la zone côtière de Ho Chi Minh-Ville, 1990–1992. Cahiers Santé 3, 464–473.

Amerasinghe, F.P., Amerasinghe, P.H., Peiris, J.S., Wirtz, R.A., 1991a. Anopheline ecology and malaria infection during the irrigation development of an area of the Mahaweli Project, Sri Lanka. Am. J. Trop. Med. Hyg. 45, 226–235.

Amerasinghe, F.P., Amerasinghe, P.H., Peiris, J.S., Wirtz, R.A., 1991b. Anopheline ecology and malaria infection during the irrigation development of an area of the Mahaweli Project, Sri Lanka. Am. J. Trop. Med. Hyg. 45, 226–235.

Amerasinghe, F.P., Indrajith, N.G., 1994. Postirrigation breeding patterns of surface water mosquitoes in the Mahaweli Project, Sri Lanka, and comparisons with preceding developmental phases. J. Med. Entomol. 31, 516–523.

Anonymous, 2009. Kenya Malaria Monitoring and Evaluation Plan 2009–2017.

Anstey, N.M., Russell, B., Yeo, T.W., Price, R.N., 2009. The pathophysiology of vivax malaria. Trends Parasitol. 25, 220–227.

Antonio-Nkondjio, C., Simard, F., Awono-Ambene, P., Ngassam, P., Toto, J.C., Tchuinkam, T., Fontenille, D., 2005. Malaria vectors and urbanization in the equatorial forest region of south Cameroon. Trans. R. Soc. Trop. Med. Hyg. 99, 347–354.

Arevalo-Herrera, M., Chitnis, C., Herrera, S., 2010. Current status of *Plasmodium vivax* vaccine. Hum. Vaccin. 6, 124–132.

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

Awab, G.R., Pukrittayakamee, S., Imwong, M., Dondorp, A.M., Woodrow, C.J., Lee, S.J., Day, N.P., Singhasivanon, P., White, N.J., Kaker, F., 2010. Dihydroartemisinin-piperaquine versus chloroquine to treat vivax malaria in Afghanistan: an open randomized, non-inferiority, trial. Malar. J. 9, 105.

Baimai, V., Harbach, R.E., Sukowati, S., 1988. Cytogenetic evidence for two species within the current concept of the malaria vector *Anopheles leucosphyrus* in Southeast Asia. J. Am. Mosq. Control Assoc. 4, 44–50.

Baird, J.K., 2004. Chloroquine resistance in *Plasmodium vivax*. Antimicrob. Agents Chemother. 48, 4075–4083.

Baird, J.K., 2007. Neglect of *Plasmodium vivax* malaria. Trends Parasitol. 23, 533–539.

Baird, J.K., 2009. Resistance to therapies for infection by *Plasmodium vivax*. Clin. Microbiol. Rev. 22, 508–534.

Baird, J.K., 2010. Eliminating malaria – all of them. Lancet 376, 1883–1885.

Baird, J.K., Hoffman, S.L., 2004. Primaquine therapy for malaria. Clin. Infect. Dis. 39, 1336–1345.

Baird, J.K., Rieckmann, K.H., 2003. Can primaquine therapy for vivax malaria be improved?. Trends Parasitol. 19, 115–120.

Baird, J.K., Schwartz, E., Hoffman, S.L., 2007. Prevention and treatment of vivax malaria. Curr. Infect. Dis. Rep. 9, 39–46.

Baker, E.Z., Beier, J.C., Meek, S.R., Wirtz, R.A., 1987. Detection and quantification of *Plasmodium falciparum* and *P. vivax* infections in Thai-Kampuchean Anopheles (Diptera: Culicidae) by enzyme-linked immunosorbent assay. J. Med. Entomol. 24, 536–541.

Balk, D.L., Deichmann, U., Yetman, G., Pozzi, F., Hay, S.I., Nelson, A., 2006. Determining global population distribution: methods, applications and data. Adv. Parasitol. 62, 119–156.

Barcus, M.J., Laihad, F., Sururi, M., Sismadi, P., Marwoto, H., Bangs, M.J., Baird, J.K., 2002. Epidemic malaria in the Menoreh Hills of Central Java. Am. J. Trop. Med. Hyg. 66, 287–292.

Barik, T.K., Sahu, B., Swain, V., 2009. A review on *Anopheles culicifacies*: from bionomics to control with special reference to Indian subcontinent. Acta Trop. 109, 87–97.

Barnwell, J.W., Nichols, M.E., Rubinstein, P., 1989. *In vitro* evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. J. Exp. Med. 169, 1795–1802.

Basseri, H.R., Doosti, S., Akbarzadeh, K., Nategpour, M., Whitten, M.M., Ladoni, H., 2008. Competency of *Anopheles stephensi mysorensis* strain for *Plasmodium vivax* and the role of inhibitory carbohydrates to block its sporogonic cycle. Malar. J. 7, 131.

Battle, K.E., Van Boeckel, T., Gething, P.W., Baird, J.K., Hay, S.I., 2011. A review of the geographical variations in *Plasmodium vivax* relapse rate. Am. J. Trop. Med. Hyg. 85 (Suppl. 451-A-516), 470.

Betuela, I., Rosanas, A., Kiniboro, B., Stanisic, D., Samol, L., Lazzari, E., Bassat, Q., del Portillo, H., Siba, P., Alonso, P., Muller, I., 2011. Relapses are contributing significantly to risk of *P. vivax* infection and disease in Papua New Guinean children 1-5 years of age. Trop. Med. Int. Health 16, 60–61.

Beutler, E., 1994. G6PD deficiency. Blood 84, 3613–3636.

Bicheron, P., Defourny, P., Brockmann, C., Schouten, L., Vancutsem, C., Huc, M., Bontemps, S., Leroy, M., Achard, F., Herold, M., Ranera, F. and Arino, O. (2008). GLOBCOVER: Products Description and Validation Report.

Bigoga, J.D., Manga, L., Titanji, V.P., Coetzee, M., Leke, R.G., 2007. Malaria vectors and transmission dynamics in coastal south-western Cameroon. Malar. J. 6, 5.

Black, R.E., Cousens, S., Johnson, H.L., Lawn, J.E., Rudan, I., Bassani, D.G., Jha, P., Campbell, H., Walker, C.F., Cibulskis, R., Eisele, T., Liu, L., Mathers, C., 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 375, 1969–1987.

Bockarie, M.J., Dagoro, H., 2006. Are insecticide-treated bednets more protective against *Plasmodium falciparum* than *Plasmodium vivax*-infected mosquitoes?. Malar. J. 5, 15.

Bockarie, M.J., Service, M.W., Toure, Y.T., Traore, S., Barnish, G., Greenwood, B.M., 1993. The ecology and behaviour of the forest form of Anopheles gambiae s.s. Parassitologia 35 (Suppl.), 5–8.

Bousema, T., Drakeley, C., 2011. Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. Clin. Microbiol. Rev. 24, 377–410.

Boyd, M.F., Kitchen, S.F., 1937. On the infectiousness of patients infected with *Plasmodium vivax* and *Plasmodium falciparum*. Am. J. Trop. Med. Hyg. s1–17, 253–262.

Branquinho, M.S., Marrelli, M.T., Curado, I., Natal, D., Barata, J.M., Tubaki, R., Carreri-Bruno, G.C., de Menezes, R.T., Kloetzel, J.K., 1997. Infection of *Anopheles (Kerteszia) cruzii* by *Plasmodium vivax* and *Plasmodium vivax* variant VK247 in the municipalities of Sao Vicente and Juquitiba, Sao Paulo. Rev. Panam. Salud Publica 2, 189–193.

Brooker, S., Leslie, T., Kolaczinski, K., Mohsen, E., Mehboob, N., Saleheen, S., Khudonazarov, J., Freeman, T., Clements, A., Rowland, M., Kolaczinski, J., 2006. Spatial epidemiology of *Plasmodium vivax*, Afghanistan. Emerg. Infect. Dis. 12, 1600–1602.

Brown, G.V., Moorthy, V.S., Reed, Z., Mendis, K., Arevalo-Herrera, M., Alonso, P., 2009. Priorities in research and development of vaccines against *Plasmodium vivax* malaria. Vaccine 27, 7228–7235.

Burgess, R.W., Young, M.D., 1950. The comparative susceptibility of *Anopheles quadrimaculatus* and *Anopheles freeborni* to infection by *Plasmodium vivax* (St. Elizabeth strain). J. Natl. Malar. Soc. 9, 218–221.

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

Butler, W.P., Roberts, D.R., 2000. Malaria in the Americas: a model of reemergence. Mil. Med. 165, 897–902.

Cappellini, M.D., Fiorelli, G., 2008. Glucose-6-phosphate dehydrogenase deficiency. Lancet 371, 64–74.

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. The History and Geography of Human Genes. Princeton University Press, Princeton.

Centers for Disease Control and Prevention, 2009. CDC Health Information for International Travel 2010.

Chang, M.S., Hii, J., Buttner, P., Mansoor, F., 1997. Changes in abundance and behaviour of vector mosquitoes induced by land use during the development of an oil palm plantation in Sarawak. Trans. R. Soc. Trop. Med. Hyg. 91, 382–386.

Chen, N., Auliff, A., Rieckmann, K., Gatton, M., Cheng, Q., 2007. Relapses of *Plasmodium vivax* infection result from clonal hypnozoites activated at predetermined intervals. J. Infect. Dis. 195, 934–941.

Chitnis, N., Schapira, A., Smith, D.L., Hay, S.I., Smith, T., Steketee, R.W., Global Partnership to Roll Back Malaria, 2010a. Mathematical modeling to support malaria control and elimination. Roll Back Malaria Progress and Impact Series 5, 48.

Chitnis, N., Schapira, A., Smith, T., Steketee, R., 2010b. Comparing the effectiveness of malaria vector-control interventions through a mathematical model. Am. J. Trop. Med. Hyg. 83, 230–240.

Chow, C.Y., 1970. Bionomics of malaria vectors in the Western Pacific Region. Southeast Asian J. Trop. Med. Public Health 1, 40–57.

CIESIN/IFPRI/WB/CIAT, 2007. Global Rural Urban Mapping Project (GRUMP) Alpha: Gridded Population of the World, Version 2, with Urban Reallocation (GPW-UR). URL: http://sedac.ciesin.columbia.edu/gpw.

Coatney, G.R., 1976. Relapse in malaria – an enigma. J. Parasitol. 62, 3–9.

Coatney, G.R., Cooper, W.C., 1948. Recrudescence and relapse in vivax malaria. In: Proceedings of the 4th International Congress on Tropical Medicine.

Coetzee, M., Fontenille, D., 2004. Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa. Insect Biochem. Mol. Biol. 34, 599–605.

Cogswell, F.B., 1992. The hypnozoite and relapse in primate malaria. Clin. Microbiol. Rev. 5, 26–35.

Coleman, R.E., Kumpitak, C., Ponlawat, A., Maneechai, N., Phunkitchar, V., Rachapaew, N., Zollner, G., Sattabongkot, J., 2004. Infectivity of asymptomatic *Plasmodium*-infected human populations to *Anopheles dirus* mosquitoes in western Thailand. J. Med. Entomol. 41, 201–208.

Coleman, R.E., Sithiprasasna, R., Kankaew, P., Kiaattiut, C., Ratanawong, S., Khuntirat, B., Sattabongkot, J., 2002. Naturally occurring mixed infection of *Plasmodium vivax* VK210 and *P. vivax* VK247 in *Anopheles* mosquitoes (Diptera: Culicidae) in western Thailand. J. Med. Entomol. 39, 556–559.

Collins, W.E., Jeffery, G.M., 1996. Primaquine resistance in *Plasmodium vivax*. Am. J. Trop. Med. Hyg. 55, 243–249.

Collins, W.E., Sullivan, J.S., Nace, D., Williams, T., Sullivan, J.J., Galland, G.G., Grady, K.K., Bounngaseng, A., 2002. Experimental infection of *Anopheles farauti* with different species of Plasmodium. J. Parasitol. 88, 295–298.

Collins, W.E., Sullivan, J.S., Strobert, E., Galland, G.G., Williams, A., Nace, D., Williams, T., Barnwell, J.W., 2009. Studies on the Salvador I strain of *Plasmodium vivax* in non-human primates and anopheline mosquitoes. Am. J. Trop. Med. Hyg. 80, 228–235.

Collins, W.E., Warren, M., Contacos, P.G., Skinner, J.C., Richardson, B.B., Kearse, T.S., 1980. The Chesson strain of *Plasmodium vivax* in *Aotus* monkeys and anopheline mosquitoes. J. Parasitol. 66, 488–497.

Collins, W.E., Warren, M., Skinner, J.C., Sutton, B.B., 1985. Infectivity of two strains of *Plasmodium vivax* to *Anopheles albitarsis* mosquitoes from Colombia. J. Parasitol. 71, 771–773.

Coluzzi, M., 1999. The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. Parassitologia 41, 277–283.

Conn, J.E., Wilkerson, R.C., Segura, M.N.O., De Souza, R.T.L., Schlichting, C.D., Wirtz, R.A., Povoa, M.M., 2002. Emergence of a new neotropical malaria vector facilitated by human migration and changes in land use. Am. J. Trop. Med. Hyg. 66, 18–22.

Cutbush, M., Mollison, P.L., 1950. The Duffy blood group system. Heredity (Edinb) 4, 383–389.

da Silva-Vasconcelos, A., Kato, M.Y., Mourao, E.N., de Souza, R.T., Lacerda, R.N., Sibajev, A., Tsouris, P., Povoa, M.M., Momen, H., Rosa-Freitas, M.G., 2002. Biting indices, host-seeking activity and natural infection rates of anopheline species in Boa Vista, Roraima, Brazil from 1996 to 1998. Mem. Inst. Oswaldo Cruz 97, 151–161.

da Silva, A.N., Santos, C.C., Lacerda, R.N., Machado, R.L., Povoa, M.M., 2006. Susceptibility of *Anopheles aquasalis* and *An. darlingi* to *Plasmodium vivax* VK210 and VK247. Mem. Inst. Oswaldo Cruz 101, 547–550.

da Silva, N.S., da Silva-Nunes, M., Malafronte, R.S., Menezes, M.J., D'Arcadia, R.R., Komatsu, N.T., Scopel, K.K., Braga, E.M., Cavasini, C.E., Cordeiro, J.A., Ferreira, M.U., 2010. Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia. Trans. R. Soc. Trop. Med. Hyg. 104, 343–350.

Daskova, N.G., Rasnicyn, S.P., 1982. Review of data on susceptibility of mosquitoes in the USSR to imported strains of malaria parasites. Bull. World Health Organ. 60, 893–897.

de Arruda, M., Carvalho, M.B., Nussenzweig, R.S., Maracic, M., Ferreira, A.W., Cochrane, A.H., 1986. Potential vectors of malaria and their different susceptibility to *Plasmodium falciparum* and *Plasmodium vivax* in northern Brazil identified by immunoassay. Am. J. Trop. Med. Hyg. 35, 873–881.

de Castro, M.C., Monte-Mor, R.L., Sawyer, D.O., Singer, B.H., 2006. Malaria risk on the Amazon frontier. Proc. Natl. Acad. Sci. U.S.A. 103, 2452–2457.

de Zulueta, J., Ramsdale, C.D., Coluzzi, M., 1975. Receptivity to malaria in Europe. Bull. World Health Organ. 52, 109–111.

Deane, L.M., 1986. Malaria vectors in Brazil. Mem. Inst. Oswaldo Cruz 81, 5–14.

Detinova, T.S., 1962. Age-grouping methods in Diptera of medical importance, with special reference to some vectors of malaria.

DFID, 2010. Breaking the Cycle: Saving Lives and Protecting the Future. vol. 65, The UK's Framework for Results for Malaria in the Developing World.

Diggle, P.J., Ribeiro, P.J., 2007. Model-based Geostatistics. Springer, New York.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. R. Stat. Soc. Ser. C Appl. Stat. 47, 299–326.

Douglas, N.M., Anstey, N.M., Angus, B.J., Nosten, F., Price, R.N., 2010. Artemisinin combination therapy for vivax malaria. Lancet Infect. Dis. 10, 405–416.

El Said, S.E., El Sawaf, B., Gebril, M., Kenawy, M.A., Azzab, A., 1983. Bionomics of anopheline mosquitoes in the Faiyoum Governate, Egypt, in relation to transmission and control of malaria. J. Egypt Public Health Assoc. 58, 189–241.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.

Erhart, A., Thang, N.D., Xa, N.X., Thieu, N.Q., Hung, L.X., Hung, N.Q., Nam, N.V., Toi, L.V., Tung, N.M., Bien, T.H., Tuy, T.Q., Cong, L.D., Thuan, L.K., Coosemans, M., D'Alessandro, U., 2007. Accuracy of the health information system on malaria surveillance in Vietnam. Trans. R. Soc. Trop. Med. Hyg. 101, 216–225.

Eshghi, N., Motabar, M., Javadian, E., Manoutcheri, A.V., 1976. Biological features of *Anopheles fluviatilis* and its role in the transmission of malaria in Iran. Trop. Geogr. Med. 28, 41–44.

FAO, 2008. The Global Administrative Unit Layers (GAUL): Technical Aspects.

Faulde, M.K., Hoffmann, R., Fazilat, K.M., Hoerauf, A., 2007. Malaria reemergence in northern Afghanistan. Emerg. Infect. Dis. 13, 1402–1404.

Feachem, R., Sabot, O., 2008. A new global malaria eradication strategy. Lancet 371, 1633–1635.

Feachem, R.G.A., Phillips, A.A., Targett, G.A., Malaria Elimination Group (Eds.), 2009. Shrinking the Malaria Map: A Prospectus on Malaria Elimination, The Global Health Group, University of California - Santa Cruz Global Health Sciences, San Francisco, U.S.A.

Flores-Mendoza, C., Fernandez, R., Escobedo-Vargas, K.S., Vela-Perez, Q., Schoeler, G.B., 2004. Natural *Plasmodium* infections in *Anopheles darlingi* and *Anopheles benarrochi* (Diptera: Culicidae) from eastern Peru. J. Med. Entomol. 41, 489–494.

Foley, D.H., Torres, E.P., Mueller, I., Bryan, J.H., Bell, D., 2003. Host-dependent *Anopheles flavirostris* larval distribution reinforces the risk of malaria near water. Trans. R. Soc. Trop. Med. Hyg. 97, 283–287.

Fontenille, D., Lepers, J.P., Campbell, G.H., Coluzzi, M., Rakotoarivony, I., Coulanges, P., 1990. Malaria transmission and vector biology in Manarintsoa, high plateaux of Madagascar. Am. J. Trop. Med. Hyg. 43, 107–115.

Frances, S.P., Klein, T.A., Wirtz, R.A., Eamsila, C., Pilakasiri, C., Linthicum, K.J., 1996. *Plasmodium falciparum* and *P. vivax* circumsporozoite proteins in anophelines (Diptera: Culicidae) collected in eastern Thailand. J. Med. Entomol. 33, 990–991.

Fryauff, D.J., Tuti, S., Mardi, A., Masbar, S., Patipelohi, R., Leksana, B., Kain, K.C., Bangs, M.J., Richie, T.L., Baird, J.K., 1998. Chloroquine-resistant *Plasmodium vivax* in transmigration settlements of West Kalimantan, Indonesia. Am. J. Trop. Med. Hyg. 59, 513–518.

Galappaththy, G.N., Omari, A.A., Tharyan, P., 2007. Primaquine for preventing relapses in people with *Plasmodium vivax* malaria. Cochrane Database Syst. Rev. (Online), CD004389.

Garnham, P.C., 1951. Some effects on the community of malaria eradication with special reference to the relapse phenomenon. East Afr. Med. J. 28, 6–10.

Garnham, P.C., 1989. Malaria Parasites of Man: Life-cycles and Morphology Excluding Ultrastructure). Malaria: Principles and Practice of Malariology. Churchill Livingstone, Edinburgh, pp. 61–96.

Garnham, P.C., Bray, R.S., Bruce-Chwatt, L.J., Draper, C.C., Killick-Kendrick, R., Sergiev, P.G., Tiburskaja, N.A., Shute, P.G., Maryon, M., 1975. A strain of *Plasmodium vivax* characterized by prolonged incubation: morphological and biological characteristics. Bull. World Health Organ. 52, 21–32.

Garrett-Jones, C., 1964. Prognosis for interruption of malaria transmission through assessment of the mosquito's vectorial capacity. Nature 204, 1173–1175.

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

Gething, P.W., Elyazar, I.R.F., Moyes, C.M., Smith, D.L., Battle, K.E., Guerra, C.A., Patil, A.P., Tatem, A.J., Howes, R.E., Myers, M.F., George, D.B., Horby, P., Wertheim, H.F.L., Price, R.N., Mueller, I., Baird, J.K., Hay, S.I., 2012. A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. PLoS Negl. Trop. Dis. 6 e1814.

Gething, P.W., Patil, A.P., Hay, S.I., 2010a. Quantifying aggregated uncertainty in *Plasmodium falciparum* malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. PLoS Comput. Biol. 6 e1000724.

Gething, P.W., Patil, A.P., Smith, D.L., Guerra, C.A., Elyazar, I.R., Johnston, G.L., Tatem, A.J., Hay, S.I., 2011a. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. Malar. J. 10, 378.

Gething, P.W., Smith, D.L., Patil, A.P., Tatem, A.J., Snow, R.W., Hay, S.I., 2010b. Climate change and the global malaria recession. Nature 465, 342–345.

Gething, P.W., Van Boeckel, T.P., Smith, D.L., Guerra, C.A., Patil, A.P., Snow, R.W., Hay, S.I., 2011b. Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*. Parasit. Vectors 4, 92.

Gillies, M.T., de Meillon, B. (Eds.), 1968. The Anophelinae of Africa South of the Sahara (Ethiopian Zoogeographical Region), The South African Institute for Medical Research, Johannesburg.

Ginsberg, M.M., 1991. Transmission of malaria in San Diego County, California. West J. Med. 154, 465–466.

Girod, R., Gaborit, P., Carinci, R., Issaly, J., Fouque, F., 2008. *Anopheles darlingi* bionomics and transmission of *Plasmodium falciparum, Plasmodium vivax* and *Plasmodium malariae* in Amerindian villages of the Upper-Maroni Amazonian forest, French Guiana. Mem. Inst. Oswaldo Cruz 103, 702–710.

Global Partnership to Roll Back Malaria, Johansson, E.W., Cibulskis, R.E., Steketee, R.W, 2010. Malaria funding and resource utilization: the first decade of Roll Back Malaria. 95.

Gonzalez-Ceron, L., Rodriguez, M.H., Chavez-Munguia, B., Santillan, F., Nettel, J.A., Hernandez-Avila, J.E., 2007. *Plasmodium vivax*: impaired escape of Vk210 phenotype ookinetes from the midgut blood bolus of *Anopheles pseudopunctipennis*. Exp. Parasitol. 115, 59–67.

Gray, E.M., Bradley, T.J., 2005. Physiology of desiccation resistance in *Anopheles gambiae* and *Anopheles arabiensis*. Am. J. Trop. Med. Hyg. 73, 553–559.

Greenwood, B.M., 2008. Control to elimination: implications for malaria research. Trends Parasitol. 24, 449–454.

Griffin, J.T., Hollingsworth, T.D., Okell, L.C., Churcher, T.S., White, M., Hinsley, W., Bousema, T., Drakeley, C.J., Ferguson, N.M., Basanez, M.G., Ghani, A.C., 2010. Reducing *Plasmodium falciparum* malaria transmission in Africa: a model-based evaluation of intervention strategies. PLoS Med. 7, e1000324.

Guerra, C.A., Gikandi, P.W., Tatem, A.J., Noor, A.M., Smith, D.L., Hay, S.I., Snow, R.W., 2008. The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. PLoS Med. 5, e38.

Guerra, C.A., Hay, S.I., Lucioparedes, L.S., Gikandi, P.W., Tatem, A.J., Noor, A.M., Snow, R.W., 2007. Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. Malar. J. 6, 17.

Guerra, C.A., Howes, R.E., Patil, A.P., Gething, P.W., Van Boeckel, T.P., Temperely, W.H., Kabaria, C.W., Tatem, A.J., Manh, B.H., Elyazar, I., Baird, J.K., Snow, R.W., Hay, S.I., 2010. The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Negl. Trop. Dis. 4 e774.

Guerra, C.A., Snow, R.W., Hay, S.I., 2006a. Defining the global spatial limits of malaria transmission in 2005. Adv. Parasitol. 62, 157–179.

Guerra, C.A., Snow, R.W., Hay, S.I., 2006b. Mapping the global extent of malaria in 2005. Trends Parasitol. 22, 353–358.

Gupta, S., Gunter, J.T., Novak, R.J., Regens, J.L., 2009. Patterns of *Plasmodium vivax* and *Plasmodium falciparum* malaria underscore importance of data collection from private health care facilities in India. Malar. J. 8, 227.

Harbach, R.E., 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bull. Entomol. Res. 94, 537–553.

Harbach, R.E., 2011. Genus *ANOPHELES* Meigen, 1818. Mosquito Taxonomic Inventory. URL: http://mosquito-taxonomic-inventory.info/genus-anopheles-meigen-1818.

Hardin, S., Santa Maria, M.E. and Liaw, C.F., 1973. Experimental Infection of *Anopheles donaldi* Reid with *Plasmodium falciparum* and *Plasmodium vivax*.

Hargreaves, K., Koekemoer, L.L., Brooke, B.D., Hunt, R.H., Mthembu, J., Coetzee, M., 2000. *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. Med. Vet. Entomol. 14, 181–189.

Harris, I., Sharrock, W.W., Bain, L.M., Gray, K.A., Bobogare, A., Boaz, L., Lilley, K., Krause, D., Vallely, A., Johnson, M.L., Gatton, M.L., Shanks, G.D., Cheng, Q., 2010. A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting. Malar. J. 9, 254.

Harrison, B.A., Scanlon, J.E., 1975. Medical entomology studies II: the subgenus *Anopheles* in Thailand. Contrib. Am. Entomol. Inst. 12, 1–308.

Hay, S.I., Gething, P.W., Snow, R.W., 2010a. India's invisible malaria burden. Lancet 376, 1716–1717.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W., 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med. 6, e1000048.

Hay, S.I., Guerra, C.A., Tatem, A.J., Atkinson, P.M., Snow, R.W., 2005. Urbanization, malaria transmission and disease burden in Africa. Nat. Rev. Microbiol. 3, 81–90.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M., Snow, R.W., 2004. The global distribution and population at risk of malaria: past, present, and future. Lancet Infect. Dis. 4, 327–336.

Hay, S.I., Okiro, E.A., Gething, P.W., Patil, A.P., Tatem, A.J., Guerra, C.A., Snow, R.W., 2010b. Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLoS Med. 7, e1000290.

Hay, S.I., Sinka, M.E., Okara, R.M., Kabaria, C.W., Mbithi, P.M., Tago, C.C., Benz, D., Gething, P.W., Howes, R.E., Patil, A.P., Temperley, W.H., Bangs, M.J., Chareonviriyaphap, T., Elyazar, I.R., Harbach, R.E., Hemingway, J., Manguin, S., Mbogo, C.M., Rubio-Palis, Y., Godfray, H.C., 2010c. Developing global maps of the dominant *Anopheles* vectors of human malaria. PLoS Med. 7, e1000209.

Hay, S.I., Smith, D.L., Snow, R.W., 2008. Measuring malaria endemicity from intense to interrupted transmission. Lancet Infect. Dis. 8, 369–378.

Hay, S.I., Snow, R.W., 2006. The Malaria Atlas Project: developing global maps of malaria risk. PLoS Med. 3, e473.

Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J., Rogers, D.J., 2006. Global environmental data for mapping infectious disease distribution. Adv. Parasitol. 62, 37–77.

Hii, J., Kan, S., Pereira, M., Parmar, S.S., Campos, R.L., Chan, M.K., 1985. Bancroftian filariasis and malaria in island and hinterland populations in Sabah, Malaysia. Trop. Geogr. Med. 37, 93–101.

Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W., Zimmerman, P.A., Barnadas, C., Beall, C.M., Gebremedhin, A., Menard, D., Williams, T.N., Weatherall, D.J., Hay, S.I., 2011. The global distribution of the Duffy blood group. Nat. Commun. 2, 266.

Howes, R.E., Piel, F.B., Patil, A.P., Nyangiri, O.A., Gething, P.W., Dewi, M., Hogg, M.M., Battle, K.E., Padilla, C.D., Baird, J.K., Hay, S.I., 2012. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: geostatistical model-based map. PLoS Med., 9, e1001339

Hulden, L., 2009. The decline of malaria in Finland–the impact of the vector and social variables. Malar. J. 8, 94.

Hulden, L., Hulden, L., Heliovaara, K., 2008. Natural relapses in vivax malaria induced by Anopheles mosquitoes. Malar. J. 7, 64.

Imwong, M., Snounou, G., Pukrittayakamee, S., Tanomsing, N., Kim, J.R., Nandy, A., Guthmann, J.P., Nosten, F., Carlton, J., Looareesuwan, S., Nair, S., Sudimack, D., Day, N.P., Anderson, T.J., White, N.J., 2007. Relapses of *Plasmodium vivax* infection usually result from activation of heterologous hypnozoites. J. Infect. Dis. 195, 927–933.

Iyengar, M.O., 1954. Vector of malaria in Kabul, Afghanistan. Trans. R. Soc. Trop. Med. Hyg. 48, 319–324.

James, S.P., 1931. Some general results of a study of induced malaria in England. Trans. R. Soc. Trop. Med. Hyg. 24, 478–525.

Jensen, T., Cockburn, A.F., Kaiser, P.E., Barnard, D.R., 1996. Human blood-feeding rates among sympatric sibling species of *Anopheles quadrimaculatus* mosquitoes in northkern Florida. Am. J. Trop. Med. Hyg. 54, 523–525.

Joshi, D., Choochote, W., Park, M.H., Kim, J.Y., Kim, T.S., Suwonkerd, W., Min, G.S., 2009. The susceptibility of *Anopheles lesteri* to infection with Korean strain of *Plasmodium vivax*. Malar. J. 8, 42.

Junkum, A., Jitpakdi, A., Jariyapan, N., Komalamisra, N., Somboon, P., Suwonkerd, W., Saejeng, A., Bates, P.A., Choochote, W., 2005. Susceptibility of two karyotypic forms of *Anopheles aconitus* (Diptera: Culicidae) to *Plasmodium falciparum* and *P. vivax*. Rev. Inst. Med. Trop. Sao Paulo 47, 333–338.

Kaneko, A., Taleo, G., Kalkoa, M., Yaviong, J., Reeve, P.A., Ganczakowski, M., Shirakawa, C., Palmer, K., Kobayakawa, T., Bjorkman, A., 1998. Malaria epidemiology, glucose 6-phosphate dehydrogenase deficiency and human settlement in the Vanuatu Archipelago. Acta Trop. 70, 285–302.

Kasap, H., 1990. Comparison of experimental infectivity and development of *Plasmodium vivax* in *Anopheles sacharovi* and *An. superpictus* in Turkey. Am. J. Trop. Med. Hyg. 42, 111–117.

Katsuragawa, T.H., Gil, L.H., Tada, M.S., de Almeida e Silva, A., Costa, J.D., Araujo Mda, S., Escobar, A.L., da Silva, L.H., 2010. The dynamics of transmission and spatial distribution of malaria in riverside areas of Porto Velho, Rondonia, in the Amazon region of Brazil. PLoS One 5, e9245.

Kenawy, M.A., Beier, J.C., Asiago, C.M., el Said, S.E., Roberts, C.R., 1990. Interpretation of low-level *Plasmodium* infection rates determined by ELISA for anophelines (Diptera: Culicidae) from Egyptian oases. J. Med. Entomol. 27, 681–685.

Killeen, G.F., Fillinger, U., Kiche, I., Gouagna, L.C., Knols, B.G., 2002. Eradication of *Anopheles gambiae* from Brazil: lessons for malaria control in Africa?. Lancet Infect. Dis. 2, 618–627.

Kiszewski, A., Mellinger, A., Spielman, A., Malaney, P., Sachs, S.E., Sachs, J., 2004. A global index representing the stability of malaria transmission. Am. J. Trop. Med. Hyg. 70, 486–498.

Klein, T.A., Lima, J.B., Tada, M.S., Miller, R., 1991. Comparative susceptibility of anopheline mosquitoes in Rondonia, Brazil to infection by *Plasmodium vivax*. Am. J. Trop. Med. Hyg. 45, 463–470.

Kochar, D.K., Das, A., Kochar, S.K., Saxena, V., Sirohi, P., Garg, S., Kochar, A., Khatri, M.P., Gupta, V., 2009. Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India. Am. J. Trop. Med. Hyg. 80, 194–198.

Langhi Jr., D.M., Bordin, J.O., 2006. Duffy blood group and malaria. Hematology 11, 389–398.

Lardeux, F., Loayza, P., Bouchite, B., Chavez, T., 2007. Host choice and human blood index of *Anopheles pseudopunctipennis* in a village of the Andean valleys of Bolivia. Malar. J. 6, 8.

Lee, D.K., 1998. Effect of two rice culture methods on the seasonal occurrence of mosquito larvae and other aquatic animals in rice fields of southwestern Korea. J. Vector Ecol. 23, 161–170.

Lee, H.I., Lee, J.S., Shin, E.H., Lee, W.J., Kim, Y.Y., Lee, K.R., 2001. Malaria transmission potential by *Anopheles sinensis* in the Republic of Korea. Korean J. Parasitol. 39, 185–192.

Lee, H.W., Shin, E.H., Cho, S.H., Lee, H.I., Kim, C.L., Lee, W.G., Moon, S.U., Lee, J.S., Lee, W.J., Kim, T.S., (2002). Detection of vivax malaria sporozoites naturally infected in anopheline mosquitoes from endemic areas of northern parts of Gyeonggi-do (Province) in Korea. Korean J. Parasitol. 40, 75–81.

Li, J., Collins, W.E., Wirtz, R.A., Rathore, D., Lal, A., McCutchan, T.F., 2001. Geographic subdivision of the range of the malaria parasite *Plasmodium vivax*. Emerg. Infect. Dis. 7, 35–42.

Lien, J.C., Kawengian, B.A., Partono, F., Lami, B., Cross, J.H., 1977. A brief survey of the mosquitoes of South Sulawesi, Indonesia, with special reference to the identity of *Anopheles barbirostris* (Diptera: Culicidae) from the Margolembo area. J. Med. Entomol. 13, 719–727.

Lin, E., Kiniboro, B., Gray, L., Dobbie, S., Robinson, L., Laumaea, A., Schopflin, S., Stanisic, D., Betuela, I., Blood-Zikursh, M., Siba, P., Felger, I., Schofield, L., Zimmerman, P., Mueller, I., 2010. Differential patterns of infection and disease with *P. falciparum* and *P. vivax* in young Papua New Guinean children. PLoS One 5, e9047.

Lindsay, S., Kirby, M., Baris, E., Bos, R., 2004. Environmental management for malaria control in the East Asia and Pacific region.

Lindsay, S.W., Thomas, C.J., 2001. Global warming and risk of vivax malaria in Great Britain. EcoHealth 2, 80–84.

Lines, J., Whitty, C.J., Hanson, K., 2007. Prospects for eradication and elimination of malaria: a technical briefing for DFID.

Linthicum, K.J., 1988. A revision of the Argyritarsis section of the subgenus *Nyssorhynchus* of *Anopheles* (Diptera: Culicidae). Mosq. Syst. 20, 98–271.
Livingstone, F.B., 1984. The Duffy blood groups, vivax malaria, and malaria selection in human populations: a review. Hum. Biol. 56, 413–425.
Loaiza, J.R., Bermingham, E., Scott, M.E., Rovira, J.R., Conn, J.E., 2008. Species composition and distribution of adult *Anopheles* (Diptera: Culicidae) in Panama. J. Med. Entomol. 45, 841–851.
Luxemburger, C., Perea, W.A., Delmas, G., Pruja, C., Pecoul, B., Moren, A., 1994. Permethrin-impregnated bed nets for the prevention of malaria in schoolchildren on the Thai-Burmese border. Trans. R. Soc. Trop. Med. Hyg. 88, 155–159.
Lysenko, A.J., Beljaev, A.E., 1969. An analysis of the geographical distribution of *Plasmodium ovale*. Bull. World Health Organ. 40, 383–394.
Lysenko, A.J., Semashko, I.N., 1968. Geography of malaria. A medico-geographic profile of an ancient disease. In: Lebedew, A.W. (Ed.), Itogi Nauki: Medicinskaja Geografija, Academy of Sciences, USSR, Moscow, pp. 25–146.
Macdonald, G., 1957. Local Features of Malaria. The Epidemiology and Control of Malaria. Oxford University Press, London, pp. 63–99.
Mahgoub, H., Gasim, G.I., Musa, I.R., Adam, I., 2012. Severe *Plasmodium vivax* malaria among Sudanese children at New Halfa Hospital, Eastern Sudan. Parasit. Vectors 5, 154.
Mahmood, F., Reisen, W.K., 1981. Duration of the gonotrophic cycle of *Anopheles culicifacies* Giles and *Anopheles stephensi* Liston, with observations on reproductive activity and survivorship during winter in Punjab province, Pakistan. Mosq. News. 41, 41–50.
Maitland, K., Williams, T.N., Bennett, S., Newbold, C.I., Peto, T.E., Viji, J., Timothy, R., Clegg, J.B., Weatherall, D.J., Bowden, D.K., 1996. The interaction between *Plasmodium falciparum* and *P. vivax* in children on Espiritu Santo island, Vanuatu. Trans. R. Soc. Trop. Med. Hyg. 90, 614–620.
Maitland, K., Williams, T.N., Newbold, C.I., 1997. *Plasmodium vivax* and *P. falciparum*: biological interactions and the possibility of cross-species immunity. Parasitol. Today 13, 227–231.
Malaria Eradication Research Agenda, 2011a. A research agenda for malaria eradication: drugs. PLoS Med. 8, e1000402.
Malaria Eradication Research Agenda, 2011b. A research agenda for malaria eradication: monitoring, evaluation, and surveillance. PLoS Med. 8, e1000400.
Malaria Eradication Research Agenda, 2011c. A research agenda for malaria eradication: vaccines. PLoS Med. 8, e1000398.
Maldonado, Y.A., Nahlen, B.L., Roberto, R.R., Ginsberg, M., Orellana, E., Mizrahi, M., McBarron, K., Lobel, H.O., Campbell, C.C., 1990. Transmission of *Plasmodium vivax* malaria in San Diego County, California, 1986. Am. J. Trop. Med. Hyg. 42, 3–9.
Manguin, S., Garros, C., Dusfour, I., Harbach, R.E., Coosemans, M., 2008. Bionomics, taxonomy, and distribution of the major malaria vector taxa of *Anopheles* subgenus *Cellia* in Southeast Asia: an updated review. Infect. Genet. Evol. 8, 489–503.
May, J.M., 1951. Map of the world distribution of malaria vectors. Geogr. Rev. 41, 638–639.
Mayxay, M., Pukrittayakamee, S., Newton, P.N., White, N.J., 2004. Mixed-species malaria infections in humans. Trends Parasitol. 20, 233–240.
McKenzie, F.E., Jeffery, G.M., Collins, W.E., 2002. *Plasmodium vivax* blood-stage dynamics. J. Parasitol. 88, 521–535.
McLaughlin, C., Levy, J., Noonan, K., Rosqueta, K., 2009. Lifting the burden of malaria: an investment guide for impact-driven philanthropy, 79.
Meek, S.R., 1995. Vector control in some countries of Southeast Asia: comparing the vectors and the strategies. Ann. Trop. Med. Parasitol. 89, 135–147.
Ménard, D., Barnadas, C., Bouchier, C., Henry-Halldin, C., Gray, L.R., Ratsimbasoa, A., Thonier, V., Carod, J.F., Domarle, O., Colin, Y., Bertrand, O., Picot, J., King, C.L., Grimberg,

B.T., Mercereau-Puijalon, O., Zimmerman, P.A., 2010. *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. Proc. Natl. Acad. Sci. U.S.A. 107, 5967–5971.

Mendes, C., Dias, F., Figueiredo, J., Mora, V.G., Cano, J., de Sousa, B., do Rosario, V.E., Benito, A., Berzosa, P., Arez, A.P., 2011. Duffy negative antigen is no longer a barrier to *Plasmodium vivax* – molecular evidences from the African West Coast (Angola and Equatorial Guinea). PLoS Negl. Trop. Dis. 5, e1192.

Mendis, K., Rietveld, A., Warsame, M., Bosman, A., Greenwood, B., Wernsdorfer, W.H., 2009. From malaria control to eradication: the WHO perspective. Trop. Med. Int. Health 14, 802–809.

Mendis, K., Sina, B.J., Marchesini, P., Carter, R., 2001. The neglected burden of *Plasmodium vivax* malaria. Am. J. Trop. Med. Hyg. 64, 97–106.

Miller, L.H., Mason, S.J., Clyde, D.F., McGinniss, M.H., 1976. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy. N. Engl. J. Med. 295, 302–304.

Moonen, B., Cohen, J.M., Snow, R.W., Slutsker, L., Drakeley, C., Smith, D.L., Abeyasinghe, R.R., Rodriguez, M.H., Maharaj, R., Tanner, M., Targett, G., 2010. Operational strategies to achieve and maintain malaria elimination. Lancet 376, 1592–1603.

Moreno, M., Cano, J., Nzambo, S., Bobuakasi, L., Buatiche, J.N., Ondo, M., Micha, F., Benito, A., 2004. Malaria Panel Assay versus PCR: detection of naturally infected *Anopheles melas* in a coastal village of Equatorial Guinea. Malar. J. 3, 20.

Morsy, T.A., el Kadry, A.A., Salama, M.M., Sabry, A.H., el Sharkawy, I.M., 1995. Studies on the bionomics and vector competence of adult anopheline mosquitoes in El Faiyum Governorate, Egypt. J. Egypt Soc. Parasitol. 25, 213–244.

Mouchet, J., Carnevale, P., Coosemans, M., Julvez, J., Manguin, S., Richard-Lenoble, D., Sircoulon, J., 2004. Biodiversité du paludisme dans le monde. John Libbey Eurotext, Montrouge, France.

Mourant, A.E., Kopec, A.C., Domaniewska-Sobczak, K., 1976. The Distribution of the Human Blood Groups and Other Polymorphisms. Oxford University Press, London.

Mueller, I., Galinski, M.R., Baird, J.K., Carlton, J.M., Kochar, D.K., Alonso, P.L., del Portillo, H.A., 2009a. Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. Lancet Infect. Dis. 9, 555–566.

Mueller, I., Slutsker, L., Tanner, M., 2011. Estimating the burden of malaria: the need for improved surveillance. PLoS Med. 8, e1001144.

Mueller, I., Widmer, S., Michel, D., Maraga, S., McNamara, D.T., Kiniboro, B., Sie, A., Smith, T.A., Zimmerman, P.A., 2009b. High sensitivity detection of *Plasmodium* species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea. Malar. J. 8, 41.

Nalin, D.R., Mahood, F., Rathor, H., Muttalib, A., Sakai, R., Chowdhry, M.A., Safdar, G., ul Haq, I., Munir, M., Suleiman, M., et al., 1985. A point survey of periurban and urban malaria in Karachi. J. Trop. Med. Hyg. 88, 7–15.

Nguyen, D.M., Tran, D.H., Harbach, R.E., Elphick, J., Linton, Y.M., 2000. A new species of the Hyrcanus Group of *Anopheles,* subgenus *Anopheles*, a secondary vector of malaria in coastal areas of southern Vietnam. J. Am. Mosq. Control Assoc. 16, 189–198.

Nigatu, W., Petros, B., Lulu, M., Adugna, N., Wirtz, R., 1994. Species composition, feeding and resting behaviour of the common anthropophilic anopheline mosquitoes in relation to malaria transmission in Gambella, south west Ethiopia. Insect Sci. Appl. 15, 371–377.

Nikolaev, B.P., 1935. On the influence of temperature on the development of malaria *plasmodia* inside the mosquito.

Nosten, F., McGready, R., Simpson, J.A., Thwai, K.L., Balkan, S., Cho, T., Hkirijaroen, L., Looareesuwan, S., White, N.J., 1999. Effects of *Plasmodium vivax* malaria in pregnancy. Lancet 354, 546–549.

Oberst, R.B., Schultz, G.W., Laughlin, L.W., Sy, N.E., Stantos, M.N., Casimiro, C., 1988. Epidemiological study of malaria in Palawan. Philippine J. Microbiol. Infect. Dis. 17, 41–48.

Obsomer, V., Defourny, P., Coosemans, M., 2007. The *Anopheles dirus* complex: spatial distribution and environmental drivers. Malar. J. 6, 26.

Okell, L.C., Drakeley, C.J., Bousema, T., Whitty, C.J., Ghani, A.C., 2008. Modelling the impact of artemisinin combination therapy and long-acting treatments on malaria transmission intensity. PLoS Med. 5, e226 discussion e226.

Pampana, E., 1969. A Textbook of Malaria Eradication. Oxford University Press, London.

PATH, 2011. Staying the Course? Malaria Research and Development in a Time of Economic Uncertainty.

Patil, A.P., Gething, P.W., Piel, F.B., Hay, S.I., 2011. Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol. 27, 246–253.

Patil, A.P., Okiro, E.A., Gething, P.W., Guerra, C.A., Sharma, S.K., Snow, R.W., Hay, S.I., 2009. Defining the relationship between *Plasmodium falciparum* parasite rate and clinical disease: statistical models for disease burden estimation. Malar. J. 8, 186.

Pigott, D.M., Atun, R., Moyes, C.M., Hay, S.I., Gething, P.W., 2012. Funding for malaria control 2006–2010: a comprehensive global assessment. Malar. J. 11, 246.

Poespoprodjo, J.R., Fobia, W., Kenangalem, E., Lampah, D.A., Hasanuddin, A., Warikar, N., Sugiarto, P., Tjitra, E., Anstey, N.M., Price, R.N., 2009. Vivax malaria: a major cause of morbidity in early infancy. Clin. Infect. Dis. 48, 1704–1712.

Poespoprodjo, J.R., Fobia, W., Kenangalem, E., Lampah, D.A., Warikar, N., Seal, A., McGready, R., Sugiarto, P., Tjitra, E., Anstey, N.M., Price, R.N., 2008. Adverse pregnancy outcomes in an area where multidrug-resistant *Plasmodium vivax* and *Plasmodium falciparum* infections are endemic. Clin. Infect. Dis. 46, 1374–1381.

Povoa, M.M., Conn, J.E., Schlichting, C.D., Amaral, J.C., Segura, M.N., Da Silva, A.N., Dos Santos, C.C., Lacerda, R.N., De Souza, R.T., Galiza, D., Santa Rosa, E.P., Wirtz, R.A., 2003. Malaria vectors, epidemiology, and the re-emergence of *Anopheles darlingi* in Belem, Para, Brazil. J. Med. Entomol. 40, 379–386.

Prakash, A., Bhattacharyya, D.R., Mohapatra, P.K., Mahanta, J., 2001. Estimation of vectorial capacity of *Anopheles dirus* (Diptera: Culicidae) in a forest-fringed village of Assam (India). Vector Borne Zoonotic Dis. 1, 231–237.

Prakash, A., Bhattacharyya, D.R., Mohapatra, P.K., Mahanta, J., 2004. Role of the prevalent *Anopheles* species in the transmission of *Plasmodium falciparum* and *P. vivax* in Assam state, north-eastern India. Ann. Trop. Med. Parasitol. 98, 559–568.

Price, L., Planche, T., Rayner, C., Krishna, S., 2007a. Acute respiratory distress syndrome in *Plasmodium vivax* malaria: case report and review of the literature. Trans. R. Soc. Trop. Med. Hyg. 101, 655–659.

Price, R.N., Douglas, N.M., Anstey, N.M., 2009. New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance. Curr. Opin. Infect. Dis. 22, 430–435.

Price, R.N., Tjitra, E., Guerra, C.A., Yeung, S., White, N.J., Anstey, N.M., 2007b. Vivax malaria: neglected and not benign. Am. J. Trop. Med. Hyg. 77, 79–87.

Prudencio, M., Rodriguez, A., Mota, M.M., 2006. The silent path to thousands of merozoites: the *Plasmodium* liver stage. Nat. Rev. Microbiol. 4, 849–856.

Pukrittayakamee, S., Imwong, M., Looareesuwan, S., White, N.J., 2004. Therapeutic responses to antimalarial and antibacterial drugs in vivax malaria. Acta Trop. 89, 351–356.

Quinones, M.L., Ruiz, F., Calle, D.A., Harbach, R.E., Erazo, H.F., Linton, Y.M., 2006. Incrimination of *Anopheles* (*Nyssorhynchus*) *rangeli* and *An.* (*Nys.*) *oswaldoi* as natural vectors of *Plasmodium vivax* in Southern Colombia. Mem. Inst. Oswaldo Cruz 101, 617–623.

Rahman, W.A., Che'Rus, A., Ahmad, A.H., 1997. Malaria and *Anopheles* mosquitos in Malaysia. Southeast Asian J. Trop. Med. Public Health 28, 599–605.

Ramsey, J.M., Salinas, E., Bown, D.N., Rodriguez, M.H., 1994. Plasmodium vivax sporozoite rates from *Anopheles albimanus* in southern Chiapas, Mexico. J. Parasitol. 80, 489–493.

Ratcliff, A., Siswantoro, H., Kenangalem, E., Maristela, R., Wuwung, R.M., Laihad, F., Ebsworth, E.P., Anstey, N.M., Tjitra, E., Price, R.N., 2007. Two fixed-dose artemisinin combinations for drug-resistant falciparum and vivax malaria in Papua, Indonesia: an open-label randomised comparison. Lancet 369, 757–765.

Rattanarithikul, R., Konishi, E., Linthicum, K.J., 1996. Detection of *Plasmodium vivax* and *Plasmodium falciparum* circumsporozoite antigen in anopheline mosquitoes collected in southern Thailand. Am. J. Trop. Med. Hyg. 54, 114–121.

RBMP, 2008. The global malaria action plan for a malaria free world.

Richie, T.L., 1988. Interactions between malaria parasites infecting the same vertebrate host. Parasitology 96 (Pt 3), 607–639.

Roberts, D., 1996. Mosquitoes (Diptera:Culicidae) breeding in brackish water: female ovipositional preferences or larval survival?. J. Med. Entomol. 33, 525–530.

Roberts, D.R., Laughlin, L.L., Hsheih, P., Legters, L.J., 1997. DDT, global strategies, and a malaria control crisis in South America. Emerg. Infect. Dis. 3, 295–302.

Roberts, D.R., Masuoka, P., Yu, A.Y., 2002. Determinants of malaria in the Americas. In: Casman, E.A., Dowlatabadi, H. (Eds.), The Contextual Determinants of Malaria, Resources for the Future Press, Washington, D.C., pp. 35–58.

Rodriguez, M.H., Gonzalez-Ceron, L., Hernandez, J.E., Nettel, J.A., Villarreal, C., Kain, K.C., Wirtz, R.A., 2000. Different prevalences of *Plasmodium vivax* phenotypes VK210 and VK247 associated with the distribution of *Anopheles albimanus* and *Anopheles pseudopunctipennis* in Mexico. Am. J. Trop. Med. Hyg. 62, 122–127.

Rojas, W., Botero, S., Garcia, H.I., 2001. An integrated malaria control program with community participation on the Pacific Coast of Colombia. Cad. Saude Publica 17 (Suppl), 103–113.

Romi, R., Pierdominici, G., Severini, C., Tamburro, A., Cocchi, M., Menichetti, D., Pili, E., Marchi, A., 1997. Status of malaria vectors in Italy. J. Med. Entomol. 34, 263–271.

Romi, R., Sabatinelli, G., Majori, G., 2001. Could malaria reappear in Italy? Emerg. Infect. Dis. 7, 915–919.

Rosenberg, R., 2007. *Plasmodium vivax* in Africa: hidden in plain sight? Trends Parasitol. 23, 193–196.

Ross, A., Maire, N., Sicuri, E., Smith, T., Conteh, L., 2011 Determinants of the cost-effectiveness of intermittent preventive treatment for malaria in infants and children. PLoS One 6, e18391.

Ross, R., 1911. The Prevention of Malaria. John Murray, London.

Rowe, A.K., Kachur, S.P., Yoon, S.S., Lynch, M., Slutsker, L., Steketee, R.W., 2009. Caution is required when using health facility-based data to evaluate the health impact of malaria control efforts in Africa. Malar. J. 8, 209.

Rowland, M., Mohammed, N., Rehman, H., Hewitt, S., Mendis, C., Ahmad, M., Kamal, M., Wirtz, R., 2002. Anopheline vectors and malaria transmission in eastern Afghanistan. Trans. R. Soc. Trop. Med. Hyg. 96, 620–626.

Rozeboom, L.E., Laird, R.L., 1942. *Anopheles* (Kerteszia) *bellator* Dyar and Knab as vector of malaria in Trinidad, British West Indies. Am. J. Trop. Med. Hyg. 22, 83–91.

Ruwende, C., Hill, A., 1998. Glucose-6-phosphate dehydrogenase deficiency and malaria. J. Mol. Med. (Berl.) 76, 581–588.

Ryan, J.R., Stoute, J.A., Amon, J., Dunton, R.F., Mtalib, R., Koros, J., Owour, B., Luckhart, S., Wirtz, R.A., Barnwell, J.W., Rosenberg, R., 2006. Evidence for transmission of *Plasmodium vivax* among a Duffy antigen negative population in Western Kenya. Am. J. Trop. Med. Hyg. 75, 575–581.

Sachs, J., Malaney, P., 2002. The economic and social burden of malaria. Nature 415, 680–685.

Sallum, M.A.M., Peyton, E.L., Wilkerson, R.C., 2005. Six new species of the *Anopheles leucosphyrus* group, reinterpretation of *An. elegans* and vector implications. Med.Vet. Entomol. 19, 158–199.

Sattabongkot, J., Tsuboi, T., Zollner, G.E., Sirichaisinthop, J., Cui, L., 2004. *Plasmodium vivax* transmission: chances for control?. Trends Parasitol. 20, 192–198.

Scharlemann, J.P., Benz, D., Hay, S.I., Purse, B.V., Tatem, A.J., Wint, G.R., Rogers, D.J., 2008. Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data. PLoS One 3, e1408.

Schultz, G.W., 1992. Biting activity of mosquitos (Diptera: Culicidae) at a malarious site in Palawan, Republic of The Philippines. Southeast Asian J. Trop. Med. Public Health 23, 464–469.

Sellami, M.H., Kaabi, H., Midouni, B., Dridi, A., Mojaat, N., Boukef, M.K., Hmida, S., 2008. Duffy blood group system genotyping in an urban Tunisian population. Ann. Hum. Biol. 35, 406–415.

Seng, C.M., Matusop, A., Sen, F.K., 1999. Differences in *Anopheles* composition and malaria transmission in the village settlements and cultivated farming zone in Sarawak, Malaysia. Southeast Asian J. Trop. Med. Public Health 30, 454–459.

Senise, L., Motoki, M.T., Sallum, M.A.M., 2006. Notes on the holotype of *Anopheles marajoara* Galvao & Damasceno (Diptera, Culicidae). Rev. Bras. Entomol. 50, 453–457.

Service, M.W., Townson, H., 2002. The *Anopheles* vector. In: Gilles, H.M. (Ed.), Essential Malariology, Fourth ed.. Arnold, London, pp. 59–84.

Sharma, R.S., 1995. Urban malaria and its vectors *Anopheles stephensi* and *Anopheles culicifacies* (Diptera: Culicidae) in Gurgaon, India. Southeast Asian J. Trop. Med. Public Health 26, 172–176.

Sharma, S.N., Subbarao, S.K., Choudhury, D.S., Pandey, K.C., 1993. Role of *An. culicifacies* and *An. stephensi* in malaria transmission in urban Delhi. Indian J. Malariol. 30, 155–168.

Sharma, V.P., 2007. Battling the malaria iceberg with chloroquine in India. Malar. J. 6, 105.

Shiff, C., 2002. Integrated approach to malaria control. Clin. Microbiol. Rev. 15, 278–293.

Shililu, J.I., Grueber, W.B., Mbogo, C.M., Githure, J.I., Riddiford, L.M., Beier, J.C., 2004. Development and survival of *Anopheles gambiae* eggs in drying soil: influence of the rate of drying, egg age, and soil type. J. Am. Mosq. Control Assoc. 20, 243–247.

Shin, E.H., Kim, T.S., Lee, H.W., Lee, J.S., Lee, W.J., 2002. Vector competence of *Anopheles lesteri* Baisas and Hu (Diptera: Culicidae) to *Plasmodium vivax* in Korea. Korean J. Parasitol. 40, 41–44.

Shute, P.G., Lupascu, G., Branzei, P., Maryon, M., Constantinescu, P., Bruce-Chwatt, L.J., Draper, C.C., Killick-Kendrick, R., Garnham, P.C., 1976. A strain of *Plasmodium vivax* characterized by prolonged incubation: the effect of numbers of sporozoites on the length of the prepatent period. Trans. R. Soc. Trop. Med. Hyg. 70, 474–481.

Simmons, J.S., 1937. Observations on the importance of *Anopheles punctimacula* as a malaria vector in Panama, and report of experimental infections in *An. neomaculipalpis, An. apicimacula*, and *An. eiseni*. Am. J. Trop. Med. Hyg. 17, 191–212.

Simsek, F.M., Kayans, S., Ozensoy Toz, S., Ozbel, Y., Alten, B., S.T. Chan, A., 2010. Evaluation of the VecTest™ malaria antigen panel assay using *Anopheles sacharovi* specimens in an endemic area, Sanliurfa Province, Turkey. Kafkas. Univ. Vet. Fak. Derg. 16, S231–S234.

Singh, N., Mishra, A.K., 2000. Anopheline ecology and malaria transmission at a new irrigation project area (Bargi Dam) in Jabalpur (Central India). J. Am. Mosq. Control Assoc. 16, 279–287.

Sinka, M.E., Bangs, M.J., Manguin, S., Chareonviriyaphap, T., Patil, A.P., Temperley, W.H., Gething, P.W., Elyazar, I.R., Kabaria, C.W., Harbach, R.E., Hay, S.I., 2011. The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis. Parasit. Vectors 4, 89.

Sinka, M.E., Bangs, M.J., Manguin, S., Coetzee, M., Mbogo, C.M., Hemingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W., Okara, R.M., Van Boeckel, T., Godfray,

H.C., Harbach, R.E., Hay, S.I., 2010a. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis. Parasit. Vectors 3, 117.

Sinka, M.E., Bangs, M.J., Manguin, S., Rubio-Palis, Y., Chareonviriyaphap, T., Coetzee, M., Mbogo, C.M., Hemmingway, J., Patil, A.P., Temperley, W.H., Gething, P.W., Kabaria, C.W., Burkot, T.R., Harbach, R.E., Hay, S.I., 2012. A global map of dominant malaria vectors. Parasit. Vectors 5, 69.

Sinka, M.E., Rubio-Palis, Y., Manguin, S., Patil, A.P., Temperley, W.H., Gething, P.W., Van Boeckel, T., Kabaria, C.W., Harbach, R.E., Hay, S.I., 2010b. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic precis. Parasit. Vectors 3, 72.

Smith, D.L., Guerra, C.A., Snow, R.W., Hay, S.I., 2007. Standardizing estimates of the *Plasmodium falciparum* parasite rate. Malar. J. 6, 131.

Smith, D.L., Hay, S.I., 2009. Endemicity response timelines for *Plasmodium falciparum* elimination. Malar. J. 8, 87.

Smith, D.L., McKenzie, F.E., 2004. Statics and dynamics of malaria infection in *Anopheles* mosquitoes. Malar. J. 3, 13.

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

Smith, T., Maire, N., Ross, A., Penny, M., Chitnis, N., Schapira, A., Studer, A., Genton, B., Lengeler, C., Tediosi, F., de Savigny, D., Tanner, M., 2008. Towards a comprehensive simulation model of malaria epidemiology and control. Parasitology 135, 1507–1516.

Snounou, G., White, N.J., 2004. The co-existence of *Plasmodium*: sidelights from falciparum and vivax malaria in Thailand. Trends Parasitol. 20, 333–339.

Snow, R.W., Guerra, C.A., Noor, A.M., Myint, H.Y., Hay, S.I., 2005. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature 434, 214–217.

Sokolova, M.I., Snow, K.R., 2002. Malaria vectors in European Russia. J. Eur. Mosq. Control Assoc. 12, 1–6.

Steenkeste, N., Rogers, W.O., Okell, L., Jeanne, I., Incardona, S., Duval, L., Chy, S., Hewitt, S., Chou, M., Socheat, D., Babin, F.X., Ariey, F., Rogier, C., 2010. Sub-microscopic malaria cases and mixed malaria infection in a remote area of high malaria endemicity in Rattanakiri province, Cambodia: implication for malaria elimination. Malar. J. 9, 108.

Subbarao, S.K., Adak, T., Vasantha, K., Joshi, H., Raghvendra, K., Cochrane, A.H., Nussenzweig, R.S., Sharma, V.P., 1988. Susceptibility of *Anopheles culicifacies* species A and B to *Plasmodium vivax* and *Plasmodium falciparum* as determined by immunoradiometric assay. Trans. R. Soc. Trop. Med. Hyg. 82, 394–397.

Subbarao, S.K., Vasantha, K., Joshi, H., Raghavendra, K., Usha Devi, C., Sathyanarayan, T.S., Cochrane, A.H., Nussenzweig, R.S., Sharma, V.P., 1992. Role of *Anopheles culicifacies* sibling species in malaria transmission in Madhya Pradesh state, India. Trans. R. Soc. Trop. Med. Hyg. 86, 613–614.

Takken, W., Lindsay, S.W., 2003. Factors affecting the vectorial competence of *Anopheles gambiae*: a question of scale. In: Takken, W., Scott, T.W. (Eds.), Ecological Aspects for Application of Genetically Modified Mosquitoes, Kluwer Academic Publishers, Dordrecht, pp. 75–90.

Tanner, M., Hommel, M., 2010. Towards malaria elimination–a new thematic series. Malar. J. 9, 24.

Tatem, A.J., Guerra, C.A., Kabaria, C.W., Noor, A.M., Hay, S.I., 2008. Human population, urban settlement patterns and their impact on *Plasmodium falciparum* malaria endemicity. Malar. J. 7, 218.

Tatem, A.J., Smith, D.L., 2010. International population movements and regional *Plasmodium falciparum* malaria elimination strategies. Proc. Natl. Acad. Sci. U.S.A. 107, 12222–12227.

Tatem, A.J., Smith, D.L., Gething, P.W., Kabaria, C.W., Snow, R.W., Hay, S.I., 2010. Ranking of elimination feasibility between malaria-endemic countries. Lancet 376, 1579–1591.

Taye, A., Hadis, M., Adugna, N., Tilahun, D., Wirtz, R.A., 2006. Biting behavior and *Plasmodium* infection rates of *Anopheles arabiensis* from Sille, Ethiopia. Acta Trop. 97, 50–54.

Temu, E.A., Minjas, J.N., Coetzee, M., Hunt, R.H., Shift, C.J., 1998. The role of four anopheline species (Diptera: Culicidae) in malaria transmission in coastal Tanzania. Trans. R. Soc. Trop. Med. Hyg. 92, 152–158.

The Global Health Group and the Malaria Atlas Project, 2011. Atlas of Malaria-eliminating Countries, 2011. The Global Health Group, Global Health Sciences, University of California, San Francisco, U.S.A.

Tjitra, E., Anstey, N.M., Sugiarto, P., Warikar, N., Kenangalem, E., Karyana, M., Lampah, D.A., Price, R.N., 2008. Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. PLoS Med. 5, e128.

Ulloa, A., Gonzalez-Ceron, L., Rodriguez, M.H., 2006. Host selection and gonotrophic cycle length of *Anopheles punctimacula* in southern Mexico. J. Am. Mosq. Control Assoc. 22, 648–653.

Vijay, S., Rawat, M., Adak, T., Dixit, R., Nanda, N., Srivastava, H., Sharma, J.K., Prasad, G.B., Sharma, A., 2011. Parasite killing in malaria non-vector mosquito *Anopheles culicifacies* species B: implication of nitric oxide synthase upregulation. PLoS One 6, e18400.

Vittor, A.Y., Gilman, R.H., Tielsch, J., Glass, G., Shields, T., Lozano, W.S., Pinedo-Cancino, V., Patz, J.A., 2006. The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. Am. J. Trop. Med. Hyg. 74, 3–11.

Vythilingam, I., Chan, S.T., Shanmugratnam, C., Tanrang, H., Chooi, K.H., 2005. The impact of development and malaria control activities on its vectors in the Kinabatangan area of Sabah, East Malaysia. Acta Trop. 96, 24–30.

Warren, M., Collins, W.E., Jeffery, G.M., Richardson, B.B., 1980. *Anopheles pseudopunctipennis*: laboratory maintenance and malaria susceptibility of a strain from El Salvador. Am. J. Trop. Med. Hyg. 29, 503–506.

Wells, T.N., Burrows, J.N., Baird, J.K., 2010. Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination. Trends Parasitol. 26, 145–151.

Wernsdorfer, W., Hay, S.I., Shanks, G.D., 2009. Learning from History. Shrinking the Malaria Map: a Prospectus on Malaria Elimination, pp. 95–107.

Wertheimer, S.P., Barnwell, J.W., 1989. *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. Exp. Parasitol. 69, 340–350.

White, N.J., 2011. Determinants of relapse periodicity in *Plasmodium vivax* malaria. Malar. J. 10, 297.

WHO, 1986. Abstracts of Recent Chinese Publications on Malaria, 17, World Health Organization, Geneva, Switzerland.

WHO, 2007. Malaria Elimination: a Field Manual for Low and Moderate Endemic Countries, 85, World Health Organization, Geneva, Switzerland.

WHO, 2008. The Global Burden of Disease 2004, 215, World Health Organization, Geneva, Switzerland.

WHO, 2010a. Guidlines for the Treatment of Malaria, second ed.. World Health Organization, Geneva, Switzerland, 194.

WHO, 2010b. International Travel and Health: Situation as on 1 January 2010, World Health Organization, Geneva, Switzerland.

WHO, 2011. World Malaria Report 2011, 246, World Health Organization, Geneva, Switzerland.

WHO/PAHO, 2006. Regional Strategic Plan for Malaria in the Americas 2006–2010, 71, Pan American Health Organization, Office for the Americas, Washington, DC.

WHO/PAHO, 2008. Malaria in the Americas: Time Series Epidemiological Data from 2000 to 2007, Pan American Health Organization, Office for the Americas, Washington, DC.

Winckel, C.W.F., 1955. Long latency in *Plasmodium vivax* infections in a temperate zone. Doc. Med. Geogr. Trop. 7.

Wirtz, R.A., Burkot, T.R., Andre, R.G., Rosenberg, R., Collins, W.E., Roberts, D.R., 1985. Identification of *Plasmodium vivax* sporozoites in mosquitoes using an enzyme-linked immunosorbent assay. Am. J. Trop. Med. Hyg. 34, 1048–1054.

Wooster, M.T., Rivera, D., 1985. Breeding point and larval association of anopheline mosquitoes of northwest Mindoro, Philippines. Southeast Asian J. Trop. Med. Public Health 16, 59–65.

World Bank, 2009. World Development Report 2009: Reshaping Economic Geography.

Wurtz, N., Mint Lekweiry, K., Bogreau, H., Pradines, B., Rogier, C., Ould Mohamed Salem Boukhary, A., Hafid, J.E., Ould Ahmedou Salem, M.S., Trape, J.F., Basco, L.K., Briolant, S., 2011. Vivax malaria in Mauritania includes infection of a Duffy-negative individual. Malar. J. 10, 336.

Yekutiel, P., 1960. Problems of epidemiology in malaria eradication. Bull. World Health Organ. 22, 669–683.

Yekutiel, P., 1980. III The Global Malaria Eradication Campaign. In: Klingberg, M.A. (Ed.), Eradication of Infectious Diseases: a Critical Study, Karger, Basel, pp. 34–88.

Zanzibar Malaria Control Program, 2009. Malaria Elimination in Zanzibar: a Feasibility Assessment.

Zimmerman, P.A., 2004. The enigma of *Plasmodium vivax* malaria and erythrocyte Duffy negativity. In: Dronamraju, K.R. (Ed.), Infectious Disease and Host-pathogen Evolution, Cambridge University Press, Cambridge, pp. 141–172.

Chapter 3 – Geographical variation in *Plasmodium vivax* relapse

---

This chapter describes analyses done to characterize the observed geographic variation of *Plasmodium vivax* relapse periodicity. This work has been published in the *Malaria Journal* and is included here in its final form. The Additional files referred to in the text are provided in the Appendix of the thesis and online. Additional files 1 and 4, which were published in document form, appear in the appendix. Additional files 2 and 3 include over 230 and 30,000 records respectively, and were therefore not included here, but can be accessed through the article doi:10.1186/1475-2875-13-144.

As described in Chapter 1, the ability of *P. vivax* to cause relapsing infections presents challenges to understanding the epidemiology of *P. vivax* and its control. The geographic variation in relapse patterns described here are applied to the model in Chapter 4 which describes the relationship between prevalence of infection and incidence of clinical disease, both of which would be influenced by relapsing infections.

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

MALARIA JOURNAL

RESEARCH Open Access

Geographical variation in *Plasmodium vivax* relapse

Katherine E Battle[1*], Markku S Karhunen[1], Samir Bhatt[1], Peter W Gething[1], Rosalind E Howes[1], Nick Golding[1], Thomas P Van Boeckel[2], Jane P Messina[1], G Dennis Shanks[3], David L Smith[4,5], J Kevin Baird[6,7] and Simon I Hay[1,5*]

Abstract

Background: *Plasmodium vivax* has the widest geographic distribution of the human malaria parasites and nearly 2.5 billion people live at risk of infection. The control of *P. vivax* in individuals and populations is complicated by its ability to relapse weeks to months after initial infection. Strains of *P. vivax* from different geographical areas are thought to exhibit varied relapse timings. In tropical regions strains relapse quickly (three to six weeks), whereas those in temperate regions do so more slowly (six to twelve months), but no comprehensive assessment of evidence has been conducted. Here observed patterns of relapse periodicity are used to generate predictions of relapse incidence within geographic regions representative of varying parasite transmission.

Methods: A global review of reports of *P. vivax* relapse in patients not treated with a radical cure was conducted. Records of time to first *P. vivax* relapse were positioned by geographic origin relative to expert opinion regions of relapse behaviour and epidemiological zones. Mixed-effects meta-analysis was conducted to determine which geographic classification best described the data, such that a description of the pattern of relapse periodicity within each region could be described. Model outputs of incidence and mean time to relapse were mapped to illustrate the global variation in relapse.

Results: Differences in relapse periodicity were best described by a historical geographic classification system used to describe malaria transmission zones based on areas sharing zoological and ecological features. Maps of incidence and time to relapse showed high relapse frequency to be predominant in tropical regions and prolonged relapse in temperate areas.

Conclusions: The results indicate that relapse periodicity varies systematically by geographic region and are categorized by nine global regions characterized by similar malaria transmission dynamics. This indicates that relapse may be an adaptation evolved to exploit seasonal changes in vector survival and therefore optimize transmission. Geographic patterns in *P. vivax* relapse are important to clinicians treating individual infections, epidemiologists trying to infer *P. vivax* burden, and public health officials trying to control and eliminate the disease in human populations.

Keywords: Malaria, *Plasmodium vivax*, Map, Relapse, Periodicity, Recurrence, Recrudescence, Strain

Background

Malaria is a significant global public health problem and the greatest burden of disease is found in the world's poorest countries [1]. The majority of malaria morbidity and mortality is caused by two of the five species of *Plasmodium* that naturally infect humans, *Plasmodium falciparum* and *Plasmodium vivax*. The broader global distribution of *P. vivax* relative to *P. falciparum* puts an estimated 2.5 billion people at risk for endemic vivax malaria [2,3]. An increasing body of evidence has shown that *P. vivax* should no longer be thought of as a benign and rarely fatal disease [4-9], but instead as being capable of causing severe disease and death, particularly in pregnant women and small children [9-12].

Plasmodium vivax is epidemiologically and biologically different to *P. falciparum* and it is not, therefore, appropriate to assume that control methods developed for falciparum malaria are directly transferable to *P. vivax* [13-16]. Biological features of *P. vivax* that distinguish it from *P. falciparum* also present unique challenges to the control of the parasite [17-19]; in elimination settings, *P. vivax* is often the "last parasite standing" following

* Correspondence: katherine.battle@zoo.ox.ac.uk; simon.hay@zoo.ox.ac.uk
[1]Department of Zoology, Spatial Ecology and Epidemiology Group, Tinbergen Building, University of Oxford, South Parks Road, Oxford, UK
[5]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
Full list of author information is available at the end of the article

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

P. falciparum elimination [20,21]. *Plasmodium vivax* gametocytes are present earlier in the progression of a primary or recrudescent infection compared to *P. falciparum* [17,22], such that the majority of patients have sufficient gametocytaemia to allow for transmission before diagnosis or treatment may occur [23-25]. In addition, *P. vivax* gametocytes are transmitted more efficiently to *Anopheles* mosquito vectors than those of *P. falciparum* and are transmissible at lower parasite densities [18,26,27]. Within the mosquito, *P. vivax* sporozoites develop faster than *P. falciparum* at equivalent temperatures, which contributes to its exploitation of a wider geographic range [28].

Perhaps the most epidemiologically important feature of *P. vivax* biology is its ability to relapse in the weeks and months following a primary parasitaemia via a dormant liver stage known as the hypnozoite [29-31]. This potential for long-term latency provides the obvious advantage of safe harbour during cold winter months when circulation in blood creates potential host immune system dangers without the benefit or opportunity for onward transmission. Therefore the term "infection" has various meanings for *P. vivax*. Infection may refer to the introduction and presence of parasites in the body, but with *P. vivax*, unlike *P. falciparum*, this can also refer to a symptom-less latent infection. The origin of renewed parasitaemia following a primary vivax infection or a "recurrence" is also ambiguous; it could be due to a hypnozoite-triggered relapse, a resurgence of erythrocytic parasites as a recrudescence, or an entirely new reinfection. See Figure 1 and Table 1 for a description of the pathways between types of infection and attack, and distinctions in terminology. The hypnozoite fundamentally distinguishes *P. vivax* from *P. falciparum* in almost every important biological, epidemiological, clinical, and public health respect.

The hypnozoite stage in the life cycle of *P. vivax* and the potential for relapse makes chemical therapies that target only the blood stage of infection ineffective as a radical cure. The 8-aminoquinolines are the only class of drugs known to have activity on the hypnozoite parasite [32-34]. Primaquine therapy, the only currently licensed radical cure, comes with caveats that add to the challenge of controlling the parasite to the point of elimination. Primaquine is associated with potentially fatal haemolysis in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency [32,35,36] and is contraindicated in pregnant women because of the risk of acute haemolytic anaemia in the foetus of unknown G6PD status [37]. The hypnozoite stage and the paucity of therapy for safe and effective treatment render vivax malaria an exceedingly difficult challenge for clinicians and those responsible for the control of endemic malaria. Relapse also has critical implications for understanding epidemiological metrics such as the basic reproduction number and force of infection, obtained from prevalence rates derived from malariometric surveys and cartographic studies that form a central part in elimination scenario planning [38,39].

It has long been known that there is significant geographical variation in the rate at which a "strain" of *P. vivax* relapses [40-42]. Temperate and subtropical strains often exhibit either a long incubation or latent period (Figure 1) of around eight to ten months. Tropical strains are characterized by short incubation times and short latency (approximately three to six weeks) [43]. Incubation period refers to the time from sporozoite

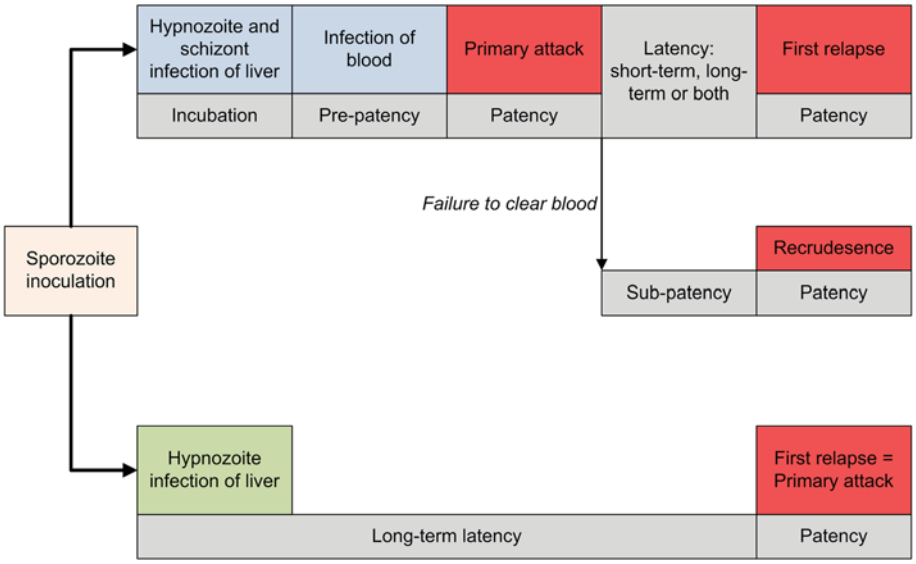


Figure 1 Pathways to infection of blood and clinical attacks in *Plasmodium vivax* malaria.

Battle et al. Malaria Journal 2014, 13:144
http://www.malariajournal.com/content/13/1/144

Table 1 Glossary of terms relevant to *Plasmodium vivax* relapse

Term	Definition
Infection	Presence of parasites in any of its forms of incubation, prepatency, patency, subpatency or latency.
Incubation	The period between inoculation of sporozoites and release of merozoites into the blood stream (primary exo-erythrocytic cycle).
Prepatency	The period prior to a primary attack where asexual parasites in the blood are both not detectable and asymptomatic, though present.
Patency	The period of clinical attack with demonstration of asexual parasites in blood as the cause of illness.
Subpatency	The period following a primary attack where asexual parasites are both not detectable and asymptomatic, though present.
Recrudescence	An event following sub-patency when parasites are both demonstrated to be present and the cause of another clinical attack or asymptomatic patency.
Latency	The period between a primary attack and a relapse; in some strains also the period between inoculation of sporozoites and the occurrence of a patent primary attack, typically six months or more.
Relapse	Patent asexual parasitaemia originating from activation of latent hypnozoites.
Re-infection	Patency by asexual blood stage parasites deriving from a new inoculation of sporozoites.
Recurrence	A newly patent parasitaemia occurring at any point after clearance of sub-patency of a primary parasitaemia where the origin is not known with certainty as being a reinfection, recrudescence or relapse.

Definitions are presented in chronological order of events.

inoculation (the mosquito bite) to the primary blood-stage infection. The latent period describes the time from the primary attack to relapse. How hypnozoite relapse is triggered, and the source of this phenotypic variation, is unresolved [44]. One theory is that the mechanism is an adaptive trait of the parasite to sequester or "hibernate" during times when climatic conditions would be inhospitable to the parasite's anopheline mosquito vectors [45-47]. Another is that latent hypnozoites are activated by a systemic febrile illness, explaining the large number of *P. vivax* relapses that follow *P. falciparum* infections [47-49]. These hypotheses need not be mutually exclusive.

Regardless of the triggering mechanism and aetiology of relapse, evidence from both controlled experimental and natural settings indicate considerable geographical variation in the timing of relapse. Although the historical perception has been that frequent relapsing strains originate from the tropics and long-latency strains from temperate regions [31], it does not sufficiently describe the observed variation in relapsing phenotypes. This binary classification conflicts with evidence of long-latency strains in tropical regions in the Americas, for instance. Coatney *et al*. [43] described in 1971 that there were three patterns of relapse. These included the Chesson strain of New Guinea-South Pacific which exhibits a short incubation period (seven to fourteen days), followed by regular re-invasions of the bloodstream within approximately three weeks after the primary infection and may continue to relapse for more than 18 months without a radical cure of the hypnozoite stage. The St Elizabeth strain from southern USA has a similar incubation period to the Chesson strain, but hypnozoites remain quiescent for several months following the primary infection before relapsing at regular intervals of three to four weeks for up to two years [43]. A third variety, such as the strain once found in the Netherlands [42], has a long period of incubation (around eight months) before the primary clinical episode followed by frequent relapses (the lower arm of Figure 1). This three-type classification likely oversimplifies the degree of variation in *P. vivax* relapse periodicity and offers limited information regarding the geographic origin of the described phenotypes. Furthermore, the majority of *P. vivax* strains have short incubation periods and the greatest difference lies in the latency period from primary attack to first relapse.

Geographic zones have been proposed for distinguishing areas with similar timing and frequency of *P. vivax* relapse following sporozoite inoculation. Figure 2A, adapted from White [47], and modified by the boundaries of the malaria endemicity map proposed by Lysenko and Semashko (for the maximum range of malaria *circa* 1900) [50], illustrates a proposed distribution of relapse phenotypes. White noted the historical perception that strains that relapse quickly originate from Southeast Asia. Temperate and subtropical areas are characterized by long-latency strains. The Indian subcontinent and South America are thought to contain both long-latency and frequent relapsing strains. A second geographic description of variation in time to relapse is described in a recent study by Lover and Coker [51]. The authors analysed the time to relapse in experimentally infected individuals relative to the geographic origin of the strain. They found that, overall, temperate strains relapse more slowly than tropical ones. However, they also found that New World tropical strains relapse more slowly than Old World tropical strains, and Eurasian temperate strains relapse more slowly than temperate strains from the Western Hemisphere [51]. Figure 2B illustrates this implied classification system.

While the maps in Figure 2 were derived from observations of relapse phenotype, others have grouped regions based on similar ecological and epidemiological

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

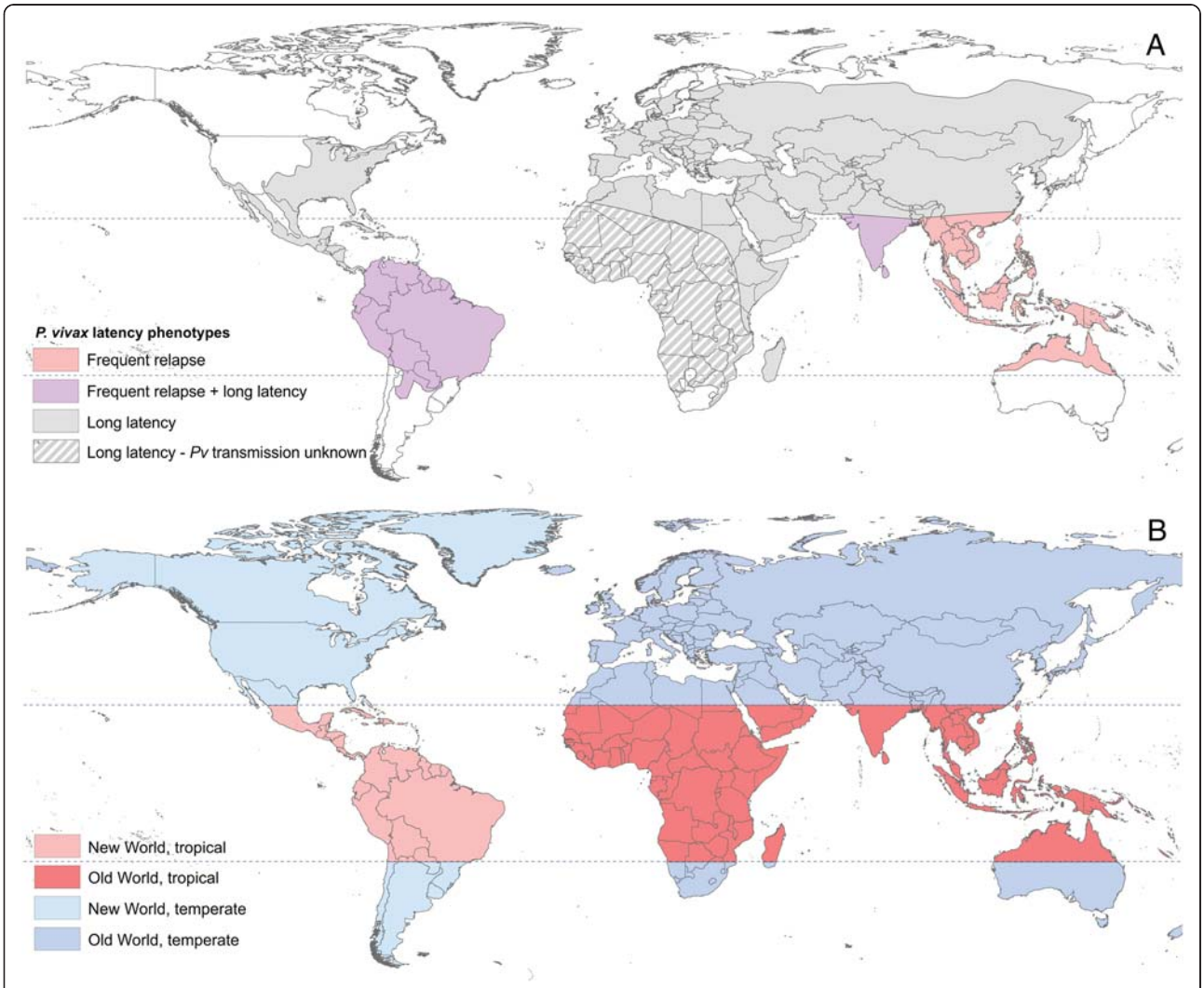


Figure 2 Proposed distributions of *Plasmodium vivax* relapse latency phenotypes. Panel **A**, adapted from White [47], shows the historical distribution of frequent relapsing and long-latency relapse strains. The geographic limits were modified using a historical malaria endemicity map from Lysenko and Shemashko [50]. Tropical frequent relapsing strains are in pink and long-latency strains in grey. Much of Africa is shown with grey hatching because the influence of Duffy negativity and its effect on vivax transmission in this part of the world is not yet understood. Purple areas are thought to have both long-latency and frequent relapsing strains. Panel **B** shows the Old and New World classification system based on the analysis and findings from Lover and Coker [51]. Tropical zones (red and pink) harbour strains that relapse more quickly than those in temperate zones (light and dark blue). New World tropical strains (pink) relapse more slowly than Old World tropical strains (red) and Old World temperate strains (dark blue) relapse slower than New World temperate strains (light blue). The dotted lines indicate the ±23.5° latitude lines to delineate temperate and tropical areas. Old World refers to Africa, Eurasia and the Pacific and New World to the Americas and Caribbean regions.

characteristics in the absence of relapse observations. These regions may reveal patterns of relapse and help elucidate determinants of the periodicity observed. The zones proposed by Macdonald [52] are shown in Figure 3. They are described by Macdonald as "zoogeographical" malaria regions and share commonalities with many historical biogeographical zonations [53-55]. The approximate boundaries of the zones are delineated by climatic variables that influence malaria transmission rates, such as temperature and rainfall, the intensity of transmission observed, as well as the abundance and behaviours of the locally dominant vector species [52].

A systematic review of *P. vivax* relapse events was conducted with the aim of revealing systematic geographical patterns of relapse frequency and a quantitative description of the potential time to relapse in different regions of the *P. vivax*-endemic world.

Methods

Data assembly

A formal literature search for data was conducted on PubMed [56] with the keywords: "vivax AND relapse" on 15 November, 2012 and updated on 24 October, 2013. The search returned 449 references. This list was

Battle et al. Malaria Journal 2014, 13:144
http://www.malariajournal.com/content/13/1/144

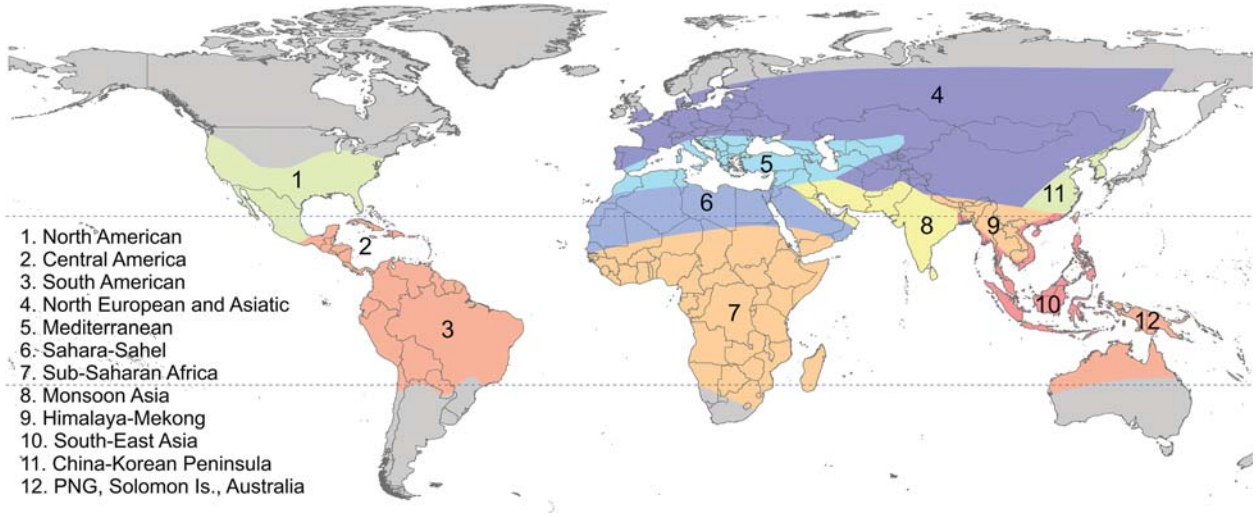


Figure 3 Zoogeographical zones of malaria transmission. Zones defined by Macdonald (1957) based on regions with similar ecological and epidemiological characteristics [52].

augmented with the reports of *P. vivax* relapse cited by Baird and Hoffman [32] and from the reference list of two recent reviews of variation in relapse periodicity [47,51]. Additional studies were obtained by correspondence with colleagues active in this research area. Malaria in the military was also examined using references from the US Army and medical records from British soldiers who contracted malaria during duty in World War II. Medical records were obtained from contacts and the Malaria Research Library, now the Malaria Reference Library, kept at the London School of Hygiene and Tropical Medicine. The aim was to obtain as much data as possible regarding *P. vivax* recurrence in all regions where vivax malaria is or has been endemic. The literature sources ostensibly refer to the recurrence events as relapses; however treatment trials and studies conducted in endemic areas may include recrudescences and reinfections in measures of relapse. Because much of the data on *P. vivax* relapse in patients not treated with primaquine originated before use of the drug became common following World War II [57], no restrictions were applied on study dates.

The exclusion criteria for the studies were minimal. Data were not used from patients who had been treated with a sufficient dose (15 mg per kg for 7 or 14 days) of primaquine, or any 8-aminoquinoline drug (pamaquine, plasmochin or pentaquine), due to the effect of the drug on hypnozoites, and therefore patterns of relapse. For example, a series of clinical trials demonstrated that 8-aminoquinoline drug failures (relapses) occurred 60–90 days post-patency, whereas untreated relapses almost always occurred between day 17 and day 35 post-patency [58]. Studies that had treated patients with a five-day course of 15 mg base of primaquine or less, which was shown to be ineffective in preventing relapse [59], were permitted. Relapse in patients treated with a seven-day course of primaquine in South America, where the treatment schedule has been shown to be inadequate [60], were also considered. Blood-stage treatments were not exclusion criteria, but were noted for analysis purposes. Mefloquine prophylaxis and treatments such as mepacrine (quinacrine, atabrine) or chloroquine may cause a delay in relapse because the drugs retain suppressive levels in the patient for long periods after treatment [61,62], making it difficult to distinguish the observed relapse as a first or second relapse. A 14-day cut-off was applied to data abstracted from drug treatment trials. Any re-appearance of infection before the 14th day was categorized as a treatment failure, while infection after day 14 was listed as a relapse. This conservative cut-off was applied to maximize sensitivity. A primary relapse is unlikely to occur before two weeks, even in fast relapsing strains [63]. However, a 14-day cut-off may result in some recrudescence events being classed as a relapse.

When possible, data on time to recurrence from the start of treatment of the primary infection were abstracted to the individual level. Start of treatment was almost invariably the first day of patency, and we considered it most probable that the vast majority of recurrences represented relapses. The majority of studies reported by the day, but those that reported the week or month of relapse were also included (with the last day of the month or week given as the time to relapse). Studies that aggregated months together were excluded. The period of follow-up was recorded for all individuals in each study, including those that did not experience a relapse. Data on the type of patient (prison volunteer, malaria therapy patients, soldiers, etc.) were also collected

Battle et al. Malaria Journal 2014, 13:144
http://www.malariajournal.com/content/13/1/144

as it is likely that this would have influenced the time to first relapse. Data on duration of prepatent period from studies performed in experimental settings were recorded where this information was available.

Georeferencing

Geopositioning of relapse studies was implemented using established methods [64] for those references that did not provide specific coordinates for the study site. The latitude and longitude of entries that provided cities or towns were located as points (<10 sq km) using Google Maps [65]. The centroid of small and large polygons (>25 to <100 sq km and >100 sq km, respectively), such as islands, regions or countries were obtained using ArcMAP 10.0 [66] to determine the latitude and longitude of those areas. The latitude and longitude values recorded corresponded to the origin of infection. Therefore, infections in returning travellers were geopositioned to the place of travel and malaria therapy or experimental trials were positioned to the origin of the strain used.

Statistical analysis

The number of cases, total person time observed, and mean and standard deviation of time to first relapse were calculated among individuals who experienced at least one relapse in each study. The incidence rate of relapse was calculated from the number of relapse events and total person time. The factors affecting this rate were modelled using mixed-effects meta-analysis in the package metafor [67,68] within the statistical programming language R [69]. As data must be normalized for use in metafor, how to best accomplish this was tested by applying different transformations to the data from each study and assessing deviation from normality by the Shapiro-Wilk test.

The geographic zones described above (Figures 2 and 3) were included in the meta-analysis of relapse rate as categorical moderators. These included the three phenotypic zones shown in White (hereafter referred to as the White-3 system), which were also shown differentiated by Old World and New World (White-5), the four zones described by Lover and Coker, and lastly the 12 zoogeographical regions as described by Macdonald [47,51,52]. Model choice was performed between these geographic systems using information criteria given by metafor (see Table 2). For the best geographic system obtained from this, a meta-analysis of mean time to first relapse among patients with observed relapse events was performed. See Additional file 1 for more information regarding model choice and the meta-analysis. Kaplan-Meier survival curves were generated from pooled individual data from each zone. These curves are intended to illustrate the observed qualitative patterns of relapse in each zone.

Table 2 Comparison of geographic classification systems

System	τ^2	I^2	H^2	R^2	AIC	AICc
White-3	1.3	97.1	34.0	33.6%	707.7	707.9
Lover	1.6	97.5	40.8	19.4%	728.4	729.0
White-5	0.9	95.4	21.8	57.3%	623.6	624.0
Macdonald	0.8	95.0	20.0	59.9%	612.7	614.1
Modified Macdonald	0.8	95.0	19.9	60.1%	612.9	614.0

Mixed-effects meta-analysis has been performed for 214 different studies or study arms by using the R package metafor [68]. The statistics are: τ^2, amount of residual heterogeneity; I^2, residual heterogeneity/unaccounted variability; H^2, unaccounted variability/sampling variability; R^2, calculated from the residual heterogeneity (τ^2) and the residual heterogeneity of an empty model as suggested by Raudenbush [70]; AIC, Akaike information criterion; and AICc, corrected AIC. The values of AIC, BIC and AICc are based on the restricted maximum likelihood, as this corresponds to the REML (restricted maximum likelihood) estimator recommended by experts [67].

Relapse maps

To visualize geographic variation in relapse, maps were generated by plotting points of median time to relapse in individuals who experienced a relapse from each study included in the final dataset. Regional maps were produced to illustrate the relapse incidence and mean time to relapse modelled within the geographic system chosen in the meta-analysis.

Results

Data assembly

Following the literature search and collection of unpublished sources, 121 references were found to contain data on time to first recurrence in patients not treated with a sufficient radical cure. Further details regarding the results of the literature review are found in Additional file 1. Details and summary statistics for the 121 references showing time to first recurrence are shown in Additional file 2. Of those, 87 references reported data at the individual level with time to relapse reported in days or values less than or equal to one month (Figure 4). The resulting dataset contained information on 30,049 individuals, of whom 5,731 experienced at least one recurrence. These data are provided in Additional file 3. The observed recurrences are most likely to be relapses, but, in probably rare instances, recrudescences may also be represented among data classified as early relapse (<60 days). The list of references included in the final database is available in Additional file 4.

Relapse rate data were available from 29 different countries and regions. The vast majority of the data were from India (78%, 23537/30049); although of the patients to experience a relapse, only one third (34%, 1931/5731) originated from the subcontinent (Additional file 1). There were data from 23 known strains in experimental infections, but the majority of individuals contracted wild *P. vivax* (94%, 28149/30049). Many of the subjects were not residents of endemic areas. For example, over one third of the patients to relapse (37%, 2731/5731) were

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

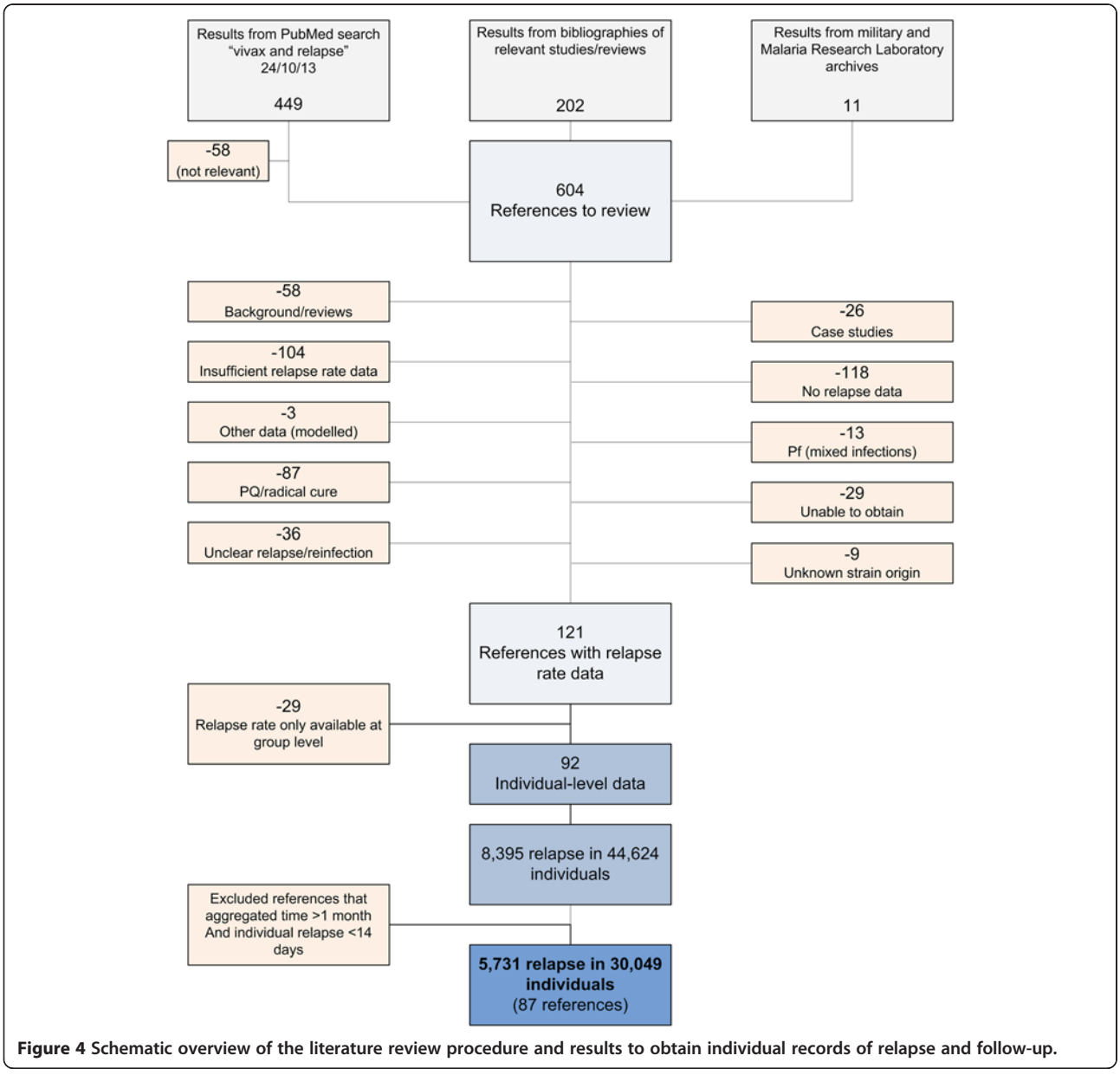


Figure 4 Schematic overview of the literature review procedure and results to obtain individual records of relapse and follow-up.

military personnel deployed from non-endemic regions. A summary of key aspects of the dataset, such as treatment and patient type, is available in Additional file 1.

Statistical analysis

Regardless of the transformation applied to the raw data for each study (identity, log, square root or Freeman-Tukey [71]), the time to first relapse deviated significantly from normal (plots shown in Additional file 1). The deviation was smallest under the log-transformation, which is common for incidence-rate data [72]. The model choice analysis was therefore carried out using log-transformed data. The Macdonald classification system yielded the best description of the data, judging by pseudo-R^2, the Akaike information criterion (AIC) and corrected AIC (AICc, which accounts for small sample sizes, see Table 2). Possible ways to simplify the Macdonald system such that geographically contiguous zones with similar transmission suitability would be combined were assessed. Further details are provided in Additional file 1. Combining zones 4 and 11 slightly improved the model fit. Zones 5 and 6, and 9 and 10, respectively, were also combined as there were no relapse data available from zones 6 and 9. The revised zones are shown in Figure 5A.

Table 3 presents two estimates of relapse incidence rate for the modified Macdonald system as the number of first relapses per 100,000 person days. One is the crude estimate based on raw data, and the other is the

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

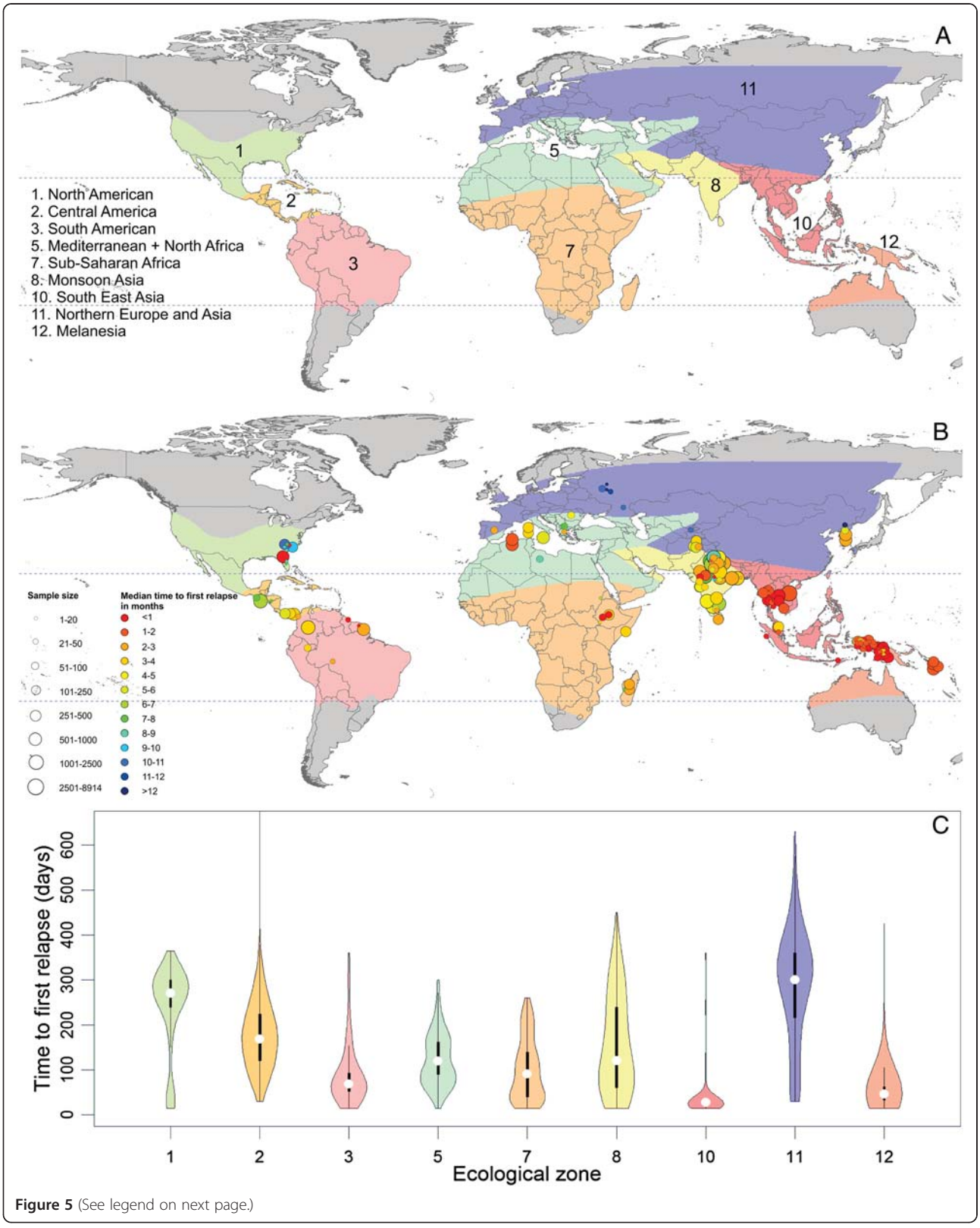


Figure 5 (See legend on next page.)

Battle et al. Malaria Journal 2014, 13:144
http://www.malariajournal.com/content/13/1/144

(See figure on previous page.)
Figure 5 Revised zoogeographical zones and observed time to first relapse. Panel **A** illustrates the revised zoogeographical zones used to describe the time to first relapse. Panel **B** shows the median observed time to relapse in each study used to obtain individual data. The size of each point varies by sample size and the time to first relapse is shown on a spectrum of red (less than one month) to dark blue (>12 months). Violin plots in Panel **C** show the observed time to first relapse in individuals from each zone in Panels **A** and **B**. The coloured areas correspond to each zone and to a smoothed approximation of the frequency distribution (a kernel density plot) of the time relapse within each geographic region. The black central bar represents the interquartile range and the white circles indicate the median values. Note that the maximum value for zone 2 extends beyond the plot.

result obtained from the meta-analysis. The highest observed and predicted incidence values are found in zones 9 + 10 and 12, corresponding to Southeast Asia and Papua New Guinea (PNG) plus the Solomon Islands (Melanesia). It is predicted that there will be approximately 800–1,200 relapses per 100,000 person days in this part of the world, compared with the estimated 130 relapses in northern Asia and Europe (zones 4 and 11). The crude and predicted estimates are very different for some zones (namely 3 and 8). This is because the mixed-effect meta-analysis attributes unusually low and high case numbers to inter-study variation and these do not contribute substantively to the rate estimate. A relevant incidence measure could not be calculated for zone 8 (India) because the observed data included several large studies (>2,000 patients) in which the majority of patients did not experience a relapse (Additional file 1).

Table 4 shows the raw and modelled estimates of mean relapse time for each of the geographic zones in the modified Macdonald system. These were obtained by running a meta-analysis for mean time to relapse and its standard deviation within each study (in a separate analysis from the original run for the incidence rates), and these are calculated from only the observed relapse events. This produces a modelled estimate of relapse time that is based on the least variable data sources. The Indian zone figures are therefore more plausible in these results. The modelled results show the fastest times to relapse are found in Southeast Asia (zone 9 + 10) and Melanesia (zone 12), around 45 days. South America (zone 3) also had a rapid time to relapse (65 days). However, there were relatively few records from zone 3 (Figure 5B) and some of the heterogeneity in relapse patterns may have been missed. Zone 1, North America, is predicted to have relapse times of about six months. Central America, zone 2, is estimated to have a relapse time of five to six months, driven by a few studies with long relapse intervals observed in Mexico [74]. The Mediterranean zone (5 + 6), a region of seasonal transmission, has a modelled mean time to relapse of five months. Based on the raw data, the mean time to relapse from the few data points in the sub-Saharan Africa zone (7) was only one month. Finally, the northern Europe and Asia zone (11 + 4) has by far the longest modelled mean time to relapse of ten months.

Finally, Figure 6 presents the survival curves for the modified Macdonald system. Note that the meta-analysis models described above do not yield survival curves. The curves in Figure 6 are based on the Kaplan-Meier survival function estimator, and they are calculated from pooled raw data within the geographic zones to provide a quantitative comparison of relapse patterns among zones. For

Table 3 Relapse incidence rates for the modified Macdonald system

Ecological zone	Based on raw data, ML with 95% CI	Model-based, REML with 95% CI
1	357 (CI: 318–400)	455 (CI: 313–662)
2	217 (CI: 198–238)	259 (CI: 120–557)
3	419 (CI: 328–528)	1093 (CI: 535–2,233)
5 + 6	214 (CI: 186–245)	262 (CI: 82–839)
7	221 (CI: 191–255)	213 (CI: 95–477)
8	25 (CI: 24–26)	62 (CI: 33–116)
9 + 10	975 (CI: 811–1163)	836 (CI: 351–1,995)
11 + 4	138 (CI: 120–159)	134 (CI: 64–278)
12	1023 (CI: 981–1,067)	1224 (CI: 689–2,174)

The numbers are presented as first relapses per 100,000 person days. The estimate based on raw data is obtained by dividing the number of relapses by follow-up time and using Ulm's exact formula [73] for confidence intervals. The model-based estimates are calculated from the results obtained from meta-analysis performed by the R package metafor. ML refers to the maximum likelihood and REML to restricted maximum likelihood.

Table 4 Mean time to relapse among geographic zones

Ecological zone	Based on raw data, ML with 95% CI	Model-based, REML with 95% CI
1	100 (CI: 99–101)	185 (CI: 162–208)
2	239 (CI: 239–240)	164 (CI: 117–212)
3	53 (CI: 53–53)	65 (CI: 18–113)
5 + 6	153 (CI: 153–153)	151 (CI: 80–221)
7	31 (CI: 30–31)	107 (CI: 57–158)
8	181 (CI: 180–181)	120 (CI: 82–159)
9 + 10	289 (CI: 288–290)	41 (CI: −11–92)
11 + 4	89 (CI: 87–90)	299 (CI: 254–345)
12	122 (CI: 121–123)	47 (CI: 12–81)

The model-based estimates have been calculated by using the R package metafor, which acknowledges interstudy variation. Thus the numbers differ from raw means calculated from the data. ML refers to the maximum likelihood and REML to restricted maximum likelihood. Please note that sample means by the very definition concern only observed events, and consequently this table ignores person time from censored observations.

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

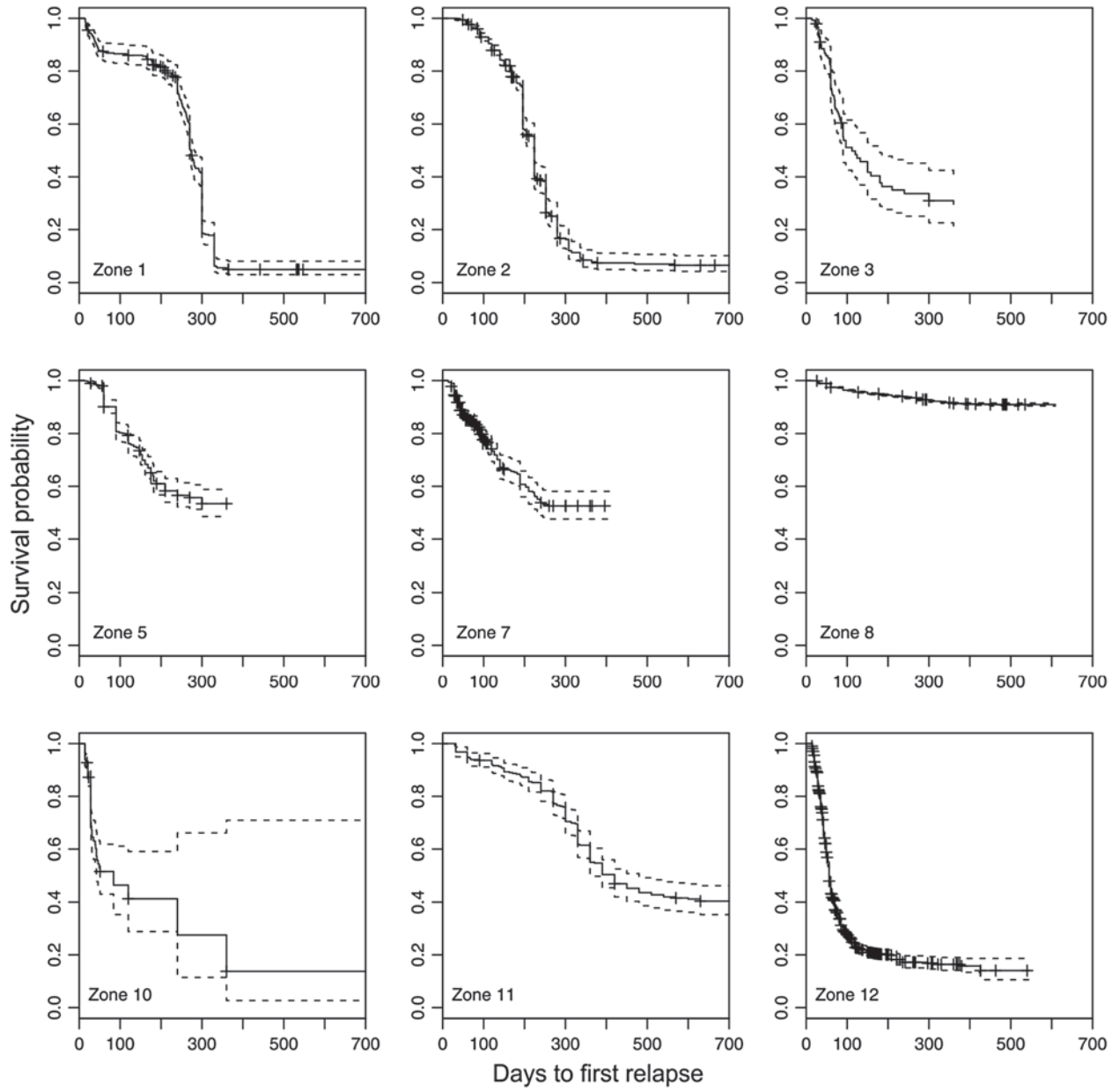


Figure 6 Survival curves for the modified Macdonald system. Shown are the Kaplan-Meier estimates (solid lines) with 95% confidence intervals (dotted lines). For each curve, all individual-level data from the respective zone have been pooled. The curves terminate at the longest follow-up day in each zone. The ticks indicate censoring events, i.e., losses to follow-up or the completion of a study without relapse. Zone 1: North America, zone 2: Central America, zone 3: South America, zone 5: Mediterranean and North Africa, zone 7: sub-Saharan Africa, zone 8: Monsoon Asia, zone 10: Southeast Asia, zone 11: northern Europe and Asia, zone 12: Melanesia.

example, in zone 1, the majority of the patients observed had relapsed by day 300, whereas in zone 12, most patients had relapsed before day 100. The curve in zone 8 does not reveal much regarding the time to relapse because of the small number of relapsing patients.

Relapse maps

The revised zoogeographical zones used in the analyses described above are shown in Figure 5A. In Figure 5B the median time to relapse for study locations is specified with points inside the geographic zones. The map illustrates a concentration of fast-relapsing strains in Southeast Asia and Melanesia. The heterogeneity in relapse periodicity observed in zone 8 is also apparent in this map. The variation in the North American zone is due to the behaviour of strains occasionally relapsing quickly after a long incubation period between inoculation and primary attack (data available from experimental

Battle *et al. Malaria Journal* 2014, **13**:144
http://www.malariajournal.com/content/13/1/144

inoculations only), but for the most part relapses followed a long period of latency after a short incubation period. Summary statistics of the time to relapse by zone are shown in Additional file 1. Panel C in Figure 5 is a violin plot of observed time to relapse in each zone. This illustrates that those shorter relapses in North America are fairly rare. The violin plots also show that the longer periods to relapse in Central America and Southeast Asia are rare. Heterogeneity in other zones, such as 8 and 11 + 4, is also demonstrated by the violin plots. Figure 7 illustrates the results of the predicted relapse incidence and mean time to relapse by region. In Figure 7A, zone 8 is hatched out as the resulting estimate is biologically implausible and is most likely caused by the handful of studies with large numbers of patients not reporting relapse during the observation period.

Discussion

The aims of this paper were to review the timing and frequency of *P. vivax* relapse of known origin in patients not treated with a hypnozoitocide to characterize variance in these patterns within geographic dimension, to identify a system to classify the variation in relapse observed and to describe the pattern of relapse in each area. A modified classification of the zoogeographical zones of malaria transmission outlined by Macdonald [52] was found to best describe the observed variation in relapse incidence. The rate of relapse and mean time from primary infection to relapse was predicted in each of the nine zones. These quantitative estimates of the contribution of relapses to *P. vivax* case incidence are crucial in informing estimates of disease burden and the origin of acute attack, i.e., from biting mosquitoes or

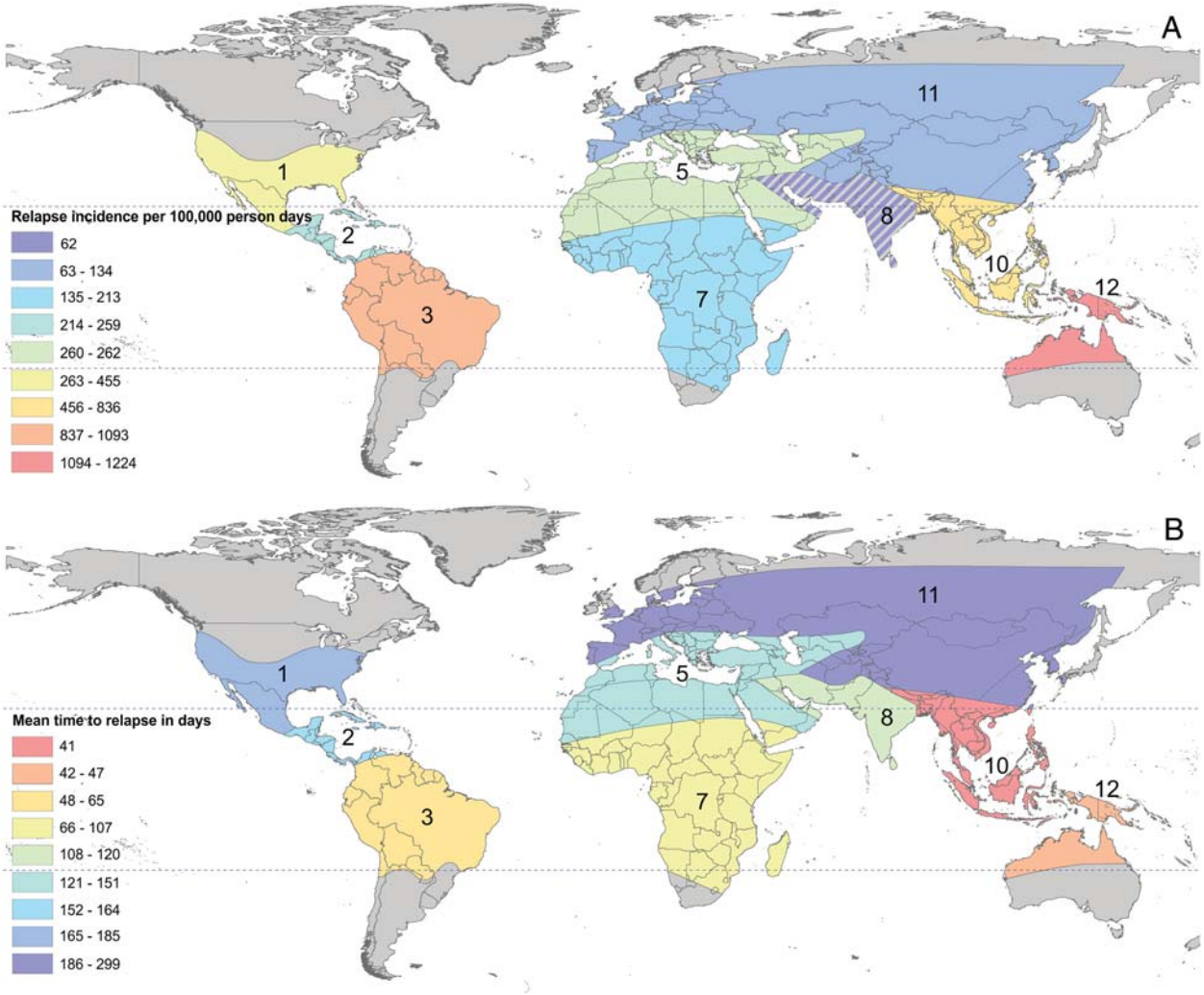


Figure 7 Modelled relapse incidence and mean time to relapse. Panel **A** illustrates the relapse incidence per 100,000 person days on a spectrum of blue to red, with red being the highest incidence of relapse. Zone 8 is hatched to indicate that the predication is particularly uncertain. Panel **B** illustrates the predicted mean time to relapse on a spectrum from blue to red, with red being most frequent relapse. The numbers of the zones correspond to those shown in Figure 5A.

Battle et al. Malaria Journal 2014, 13:144
http://www.malariajournal.com/content/13/1/144

emergent hypnozoites. This understanding, in turn, informs critical decision-making in control strategies that effectively weigh the benefit of anti-mosquito *versus* hypnozoitocidal interventions. They also help identify regions in which strains have long-term latency and are therefore undetectable to standard diagnostic methods (rapid diagnostic tests and microscopy) for long periods of time.

The results presented here further refine historical interpretations and recent analyses of the geographic variation in relapse periodicity. As shown in Figure 2, tropical strains relapse more rapidly than temperate strains and New World strains vary from those in the Old World. White's illustration of the variation in relapse phenotype shown in Figure 2A shows that Southeast Asia and Asia-Pacific is the only region having exclusively frequent relapse behaviour. The results also showed infections from this region to relapse quickly, with a few rare exceptions. White showed that both frequent relapsing and long-latency strains are present in India and South America. The data from India appear to affirm this, with relapse patterns from the subcontinent and surrounding areas so heterogeneous that it was impossible to generate logical model predictions of relapse incidence for the region. The low incidence in region 8 shown in Table 3 is not believed to be a reflection of the presence of long relapsing strains, but rather a result of natural infections that either did not result in a relapse or the resulting parasitaemia was too low to be detected by the study. There is not presently an explanation for the lack of relapses, but this phenomenon has also been observed in recent tafenoquine trials in India [75]. The variation in relapse timing in the raw data observed in India (Figure 5B) is likely a result of the wide variation in transmission settings found within this zone. There are tropical forest areas, similar to zone 12, dry habitats like those in zone 5 and highland areas that border zone 11. In addition, the presence of *Anopheles stephensi*, adapted to breeding in artificial water collections [76,77], has extended transmission into urban areas. Therefore, in addition to issues of data availability and study design, the range of ecological settings in the zone, and likely some of the other zones, may also contribute to the variance in relapse behaviours observed.

The results presented for South America were different from what was shown in White's phenotype map (Figure 2A). South America was predicted to have a high relapse incidence, comparable to Southeast Asia, and a two-month mean time from primary attack to relapse. The data available from this region were limited and determining the cause for the observed difference is therefore difficult. There has been renewed interest in the origin of the American strains of *P. vivax*, whether they originate from somewhere in Asia or were sent west from Africa by migration and the slave trade, as has been proposed for *P. falciparum* [78]. This could influence the nature of the relapse periodicity observed. Improved understanding of the phylogeny of *P. vivax* may reveal information about the pattern of relapse in this region.

The analysis by Lover and Coker [51] revealed that tropical regions relapse more quickly than temperate strains. However, Central America and sub-Saharan Africa had relapse patterns similar to the Mediterranean with moderate relapse incidence (around 250 relapses per 100,000 person days) and five to six months between primary infection and relapse. These regions seem to be better described as an "intermediate" relapsing phenotype between the frequent relapsing strains in Southeast Asia and South America and the long-latency temperate strains in North America and Europe/Asia.

The results of the northern temperate regions concur with the findings of Lover and Coker [51]. The authors noted that while in general, temperate strains relapse more slowly than tropical strains and that New World tropical strains were slower than Old World strains, the opposite was true of the temperate strains. Based on this analysis, the New World temperate strains relapsed more rapidly than the Old World temperate strains (Figure 7). The modelled results showed that the relapse incidence was 455 per 100,000 person days and mean time to relapse was six months in North America. However, again, this high incidence compared to that in northern Asia and Europe (134 relapses per 100,000 person days) could be due to a few exceptional experimental subjects who received large sporozoite inoculations [79-81].

The utility of the predictions made is limited by the nature of the data available. There are few contemporary data on *P. vivax* infections in patients not treated with a hypnozoitocide. Therefore, much of the data used were from drug trials on adult workers, military personnel and prison "volunteers", as well as data from when malaria was used to treat neurosyphilis patients. The age and immunity of the patients would perhaps not be representative of relapse as it would occur among residents of the strain region of origin. This could be important because the children in many endemic settings likely carry the greatest burden of relapsing infections [82]. Some experimental challenge subjects were inoculated with relatively heavy sporozoite dosages, a factor that greatly influences the time from primary attack to relapse [47] and is likely to differ from relapses following more modest numbers of sporozoites acquired from wild anophelines. It is straightforward to attribute an infection as a relapse or re-infection in experimental settings, but this could not be distinguished for *P. vivax* infections acquired in the wild. Effort was made to obtain studies where the follow-up period was conducted in a non-endemic area (for example, in a hospital in a city).

Of the 5731 records of relapse 88% (n = 5030) were obtained from experimental or non-local populations (such as military personnel). Lastly, the strains used in therapy and drug trials in the first half of the twentieth century were of "known" origin and infections were geopositioned to those sites. However, it is not certain if the strains used were of their named origin. For example, the "Madagascar" strain was obtained from an Indian seaman whose last port of call was in Madagascar [83] and may conceivably, therefore, originate from elsewhere.

A principal limitation of this work is the inability to conclusively identify a recurrent infection as a relapse, recrudesence or reinfection. This is particularly an issue given the rising rate of resistance to standard treatments such as chloroquine [84]. In cases of chloroquine failure, recurrence can occur one to two months after initial treatment [63], making differentiation between relapse and recrudescence a challenge. Chloroquine and chloroquine-combination therapies were by far the most common treatment regimens (82%, 24787/30049). Of those patients to receive chloroquine, 2519 patients experienced a relapse or recurrence, 523 of which occurred before 60 days. This is equivalent to 9% (523/5731) of the relapse records. In addition, the effect of resistance on recurrence is likely abated by the historical nature of the dataset. The first cases of chloroquine-resistant *P. vivax* were reported in 1989 [85]. Of the 5731 records of relapse, 2080 occurred before 60 days and 82% of those (n = 1701) were observed before 1989. Therefore, increased resistance is unlikely to have a large effect on the results and instances of recrudesence being classed as a relapse would have been rare. While the relapse signal represented in these data certainly contains some noise due to reinfection or recrudescence, we considered these other sources of recurrences improbable relative to relapse. Figures 5C and 6 seem consistent with this assumption because recurrence due to reinfection or recrudescence would have been far more stochastic than relapse, obscuring or effacing the patterns shown by the randomness of timing of those events relative to primary parasitaemia.

There are aspects of the data that were not incorporated into the analysis performed here that could be addressed in future work. First, it was difficult to account for strains with long incubation periods before primary infection (information only available for a subset of data from experimental settings) followed by relatively short time to relapse. This was occasionally exhibited by the North American St Elizabeth strain and *P. vivax multinucleatum* from China (see Additional file 3). The link between the sporozoite dose and latency, mentioned above, was shown in the literature to be an important factor in determining relapse patterns [46,86,87], but was not incorporated into the analysis as it cannot be known for wild infections. This may be a possible explanation why the incidence of relapse in North America was greater than that predicted for Central America. However, this was likely not a common problem with the experimental studies used. The majority of studies aimed to induce patency and the sporozoite inoculations were large, but not extreme. There were only a handful of studies included that used particularly large inoculations in order to study the effect of dosage on relapse pattern [79-81].

The type of patient varied among studies (prison volunteers, military personnel, malaria therapy patients, outpatients in an endemic area, etc.) and the drug type and dosage varied within and among studies. In some studies, primary attacks were treated with insufficient doses of 8-aminoquinolines or drugs that have long half-lives. Mepacrine has a half-life of up to a month and can delay relapses by about 30 days [88,89] and chloroquine, the most common drug in the dataset, can delay parasite reappearance by anywhere from two to six weeks [47]. Inclusion of the subject-type and treatment as explanatory variables was tested; however, the results were similar to the simpler model used (see Additional file 1). Finally, the analysis only addressed the periodicity between primary attack and first observed recurrence (relapse). The modelled estimates of incidence do not account for multiple relapses. Both the frequency and number of relapses will vary based on a variety of factors including inoculums, age

Table 5 Strategies for modelling survival data obtained from many dissimilar sources

Statistical method	Accounts for individual-level variation	Accounts for between-study variation	R packages
Fixed-effects meta-analysis	No; operates on summary statistics	No	Many software packages, e.g., meta and metafor
Mixed-effects meta-analysis	No; operates on summary statistics	Yes	Many packages, e.g., meta and metafor; also general-purpose software such as lme4 may be used
Survival analysis for pooled data	Yes	No	A number of packages, e.g., survival, eha and flexsurv
Survival analysis with mixed effects	Yes	Yes	Most notably R/coxme; flexible software seems to be hard to find

Data were analysed using mixed-effects meta-analysis, which is common for this type of study. All of the methods have strengths and weaknesses.

Battle et al. Malaria Journal 2014, 13:144
http://www.malariajournal.com/content/13/1/144

of patient and origin of infection. While many studies did not follow patients long enough to report multiple relapses, further work in this area will be essential to obtain measures of the *P. vivax* force of infection.

In addition to the limitations posed by the data and survey study designs, the analysis is limited by the types of statistical methods available for this kind of task (see Table 5). It would be preferable to use a statistical model that is both hierarchical (to account for between-study variation) and employs a suitable survival-analysis likelihood. Unfortunately, software used to fit this type of model was numerically unstable and hence the mixed-effects meta-analysis was employed.

The mechanism of hypnozoite activation to cause an acute attack (relapse) remains unknown. There is clearly variation in relapse "phenotype". Based on the results presented here, timing of relapse appears to vary geographically in conjunction with areas of similar ecology and malaria transmission patterns. However, it is difficult to determine whether long latency occurs in regions of frequent relapse (tropical areas such as zones 9 + 10 and 12). A long-latency relapse may be thought to be another short-term relapse in a succession of rapid relapses, and the genotype cannot reveal if it is in fact a separate relapse "event" [47]. Nonetheless, understanding broad patterns of relapse is of use epidemiologically. There tends to be fewer overall relapses in the long-latency strains because hepatocytes that host the hypnozoites may die before the relapse event occurs. The resulting burden of hypnozoites from different strains or regions has implications for sensitivity to primaquine and therefore the dosage of primaquine that should be used [47]. This was observed in treatment of soldiers returning from Korea (fewer hypnozoites) [90] relative to those returning from the Pacific (high hypnozoite burden) [58] and will have implications for future control strategies.

Conclusions

Frequency of relapse varies geographically. The association between relapse rate and the geographic regions does not clarify causation. Geographic variation does not directly imply environmental cues as triggers for relapse, even if the revised Macdonald system resembles the distribution of transmission suitability. Relapse frequency may result from evolved responses to average transmission season duration or arise from proximate cues, such as triggers from other infections, correlated with *P. vivax* transmission and/or vector suitability periods. There is likely an interaction between activation of latent hypnozoites from infection and an evolved trait for strains from areas of seasonal transmission to remain dormant during periods of low mosquito abundance. Regardless of the cause, these patterns are important for the treatment of individual infections, measures of *P. vivax* burden and the prospects for control and eventual elimination of the disease from endemic areas.

Additional files

Additional file 1: Additional results and statistical analyses.

Additional file 2: Table S1. Summary relapse dataset.

Additional file 3: Table S2. Individual-level relapse dataset used in analysis.

Additional file 4: Individual-level data reference list.

Abbreviations

95% CI: 95% confidence interval; AIC: Akaike information criterion; AICc: Corrected Akaike information criterion; BIC: Bayesian information criterion; G6PD: Glucose-6-phosphate dehydrogenase; ML: Maximum likelihood; PNG: Papua New Guinea; REML: Restricted maximum likelihood.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KEB and SIH conceived the study and oversaw its implementation with assistance from MSK, PWG and JKB. KEB wrote the first draft of the manuscript and assembled data with assistance from REH, JPM and GDS. Environmental data were provided by TPVB. MSK led the design of the modelling framework with input from NG, SB and DLS. All authors participated in the interpretation of results and in the writing and editing of the manuscript. All authors read and approved the final manuscript. KEB and SIH will act as guarantors for the paper.

Acknowledgements

We would like to thank Nick White, David Warhust, Ellis McKenzie, Wallace Peters, Bill Collins, Andrei Beljaev, John Barnwell and David Bradley for their input and suggestions at the early stages of this work. We acknowledge the contribution of references from Ric Price, Socrates Herrera, unpublished data contribution from François Nosten and Cindy Chu and data support from Lilia Gonazalez-Ceron and Manjoit Roy. We also thank David Pigott for proofreading.

This work was supported by a Wellcome Trust Senior Research Fellowship (#095066) held by SIH, which also supports KEB, MSK and REH. PWG is a Medical Research Council Career Development Fellow (#K00669X) and receives support from the Bill and Melinda Gates Foundation (#OPP1068048) that also supports SB. NG is funded by a grant from the Bill & Melinda Gates Foundation (#OPP1053338). JPM is funded by the International Research Consortium on Dengue Risk Assessment Management and Surveillance (IDAMS, European Commission 7th Framework Programme (#21803)). TPVB was funded by a grant from the Belgian Fond National pour la Recherche Scientifique and the Fondation Wiener-Anspach. DLS and SIH acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. JKB is supported by Wellcome Trust grant #B9RJIXO.

Author details

[1]Department of Zoology, Spatial Ecology and Epidemiology Group, Tinbergen Building, University of Oxford, South Parks Road, Oxford, UK. [2]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [3]Australian Army Malaria Institute, Enoggera, Queensland, Australia. [4]Department of Epidemiology and Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [5]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. [6]Eijkman-Oxford Clinical Research Unit, Jalan Diponegoro No 69, Jakarta, Indonesia. [7]Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, Oxford, UK.

Received: 7 February 2014 Accepted: 31 March 2014
Published: 15 April 2014

References

1. Sachs J, Malaney P: **The economic and social burden of malaria.** *Nature* 2002, **415:**680–685.
2. Gething PW, Elyazar IRF, Moyes CM, Smith DL, Battle KE, Guerra CA, Patil AP, Tatem AJ, Howes RE, Myers MF, George DB, Horby P, Wertheim HFL, Price RN, Mueller I, Baird JK, Hay SI: **A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010.** *PLoS Negl Trop Dis* 2012, **6:**e1814.
3. Battle KE, Gething PW, Elyazar IR, Moyes CL, Sinka ME, Howes RE, Guerra CA, Price RN, Baird KJ, Hay SI: **The global public health significance of *Plasmodium vivax*.** *Adv Parasitol* 2012, **80:**1–111.
4. Anstey NM, Russell B, Yeo TW, Price RN: **The pathophysiology of vivax malaria.** *Trends Parasitol* 2009, **25:**220–227.
5. Barcus MJ, Basri H, Picarima H, Manyakori C, Sekartuti, Elyazar I, Bangs MJ, Maguire JD, Baird JK: **Demographic risk factors for severe and fatal vivax and falciparum malaria among hospital admissions in northeastern Indonesian Papua.** *Am J Trop Med Hyg* 2007, **77:**984–991.
6. Karyana M, Burdarm L, Yeung S, Kenangalem E, Wariker N, Maristela R, Umana KG, Vemuri R, Okoseray MJ, Penttinen PM, Ebsworth P, Sugiarto P, Anstey NM, Tjitra E, Price RN: **Malaria morbidity in Papua Indonesia, an area with multidrug resistant *Plasmodium vivax* and *Plasmodium falciparum*.** *Malar J* 2008, **7:**148.
7. Kochar DK, Saxena V, Singh N, Kochar SK, Kumar SV, Das A: ***Plasmodium vivax* malaria.** *Emerg Infect Dis* 2005, **11:**132–134.
8. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM: **Vivax malaria: neglected and not benign.** *Am J Trop Med Hyg* 2007, **77:**79–87.
9. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M, Lampah DA, Price RN: **Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia.** *PLoS Med* 2008, **5:**e128.
10. Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, Alpers MP, Muller I: ***Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea.** *PLoS Med* 2008, **5:**e127.
11. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Hasanuddin A, Warikar N, Sugiarto P, Tjitra E, Anstey NM, Price RN: **Vivax malaria: a major cause of morbidity in early infancy.** *Clin Infect Dis* 2009, **48:**1704–1712.
12. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Warikar N, Seal A, McGready R, Sugiarto P, Tjitra E, Anstey NM, Price RN: **Adverse pregnancy outcomes in an area where multidrug-resistant *Plasmodium vivax* and *Plasmodium falciparum* infections are endemic.** *Clin Infect Dis* 2008, **46:**1374–1381.
13. Baird JK: **Eliminating malaria - all of them.** *Lancet* 2010, **376:**1883–1885.
14. Bockarie MJ, Dagoro H: **Are insecticide-treated bednets more protective against *Plasmodium falciparum* than *Plasmodium vivax*-infected mosquitoes?** *Malar J* 2006, **5:**15.
15. Bousema T, Drakeley C: **Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination.** *Clin Microbiol Rev* 2011, **24:**377–410.
16. Luxemburger C, Perea WA, Delmas G, Pruja C, Pecoul B, Moren A: **Permethrin-impregnated bed nets for the prevention of malaria in schoolchildren on the Thai-Burmese border.** *Trans R Soc Trop Med Hyg* 1994, **88:**155–159.
17. Mendis K, Sina BJ, Marchesini P, Carter R: **The neglected burden of *Plasmodium vivax* malaria.** *Am J Trop Med Hyg* 2001, **64:**97–106.
18. Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J, Cui L: ***Plasmodium vivax* transmission: chances for control?** *Trends Parasitol* 2004, **20:**192–198.
19. Wells TN, Burrows JN, Baird JK: **Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination.** *Trends Parasitol* 2010, **26:**145–151.
20. Pampana E: *A textbook of malaria eradication.* London: Oxford University Press; 1969.
21. Tatem AJ, Smith DL, Gething PW, Kabaria CW, Snow RW, Hay SI: **Ranking of elimination feasibility between malaria-endemic countries.** *Lancet* 2010, **376:**1579–1591.
22. McKenzie FE, Jeffery GM, Collins WE: ***Plasmodium vivax* blood-stage dynamics.** *J Parasitol* 2002, **88:**521–535.
23. Awab GR, Pukrittayakamee S, Imwong M, Dondorp AM, Woodrow CJ, Lee SJ, Day NP, Singhasivanon P, White NJ, Kaker F: **Dihydroartemisinin-piperaquine versus chloroquine to treat vivax malaria in Afghanistan: an open randomized, non-inferiority, trial.** *Malar J* 2010, **9:**105.
24. Douglas NM, Anstey NM, Angus BJ, Nosten F, Price RN: **Artemisinin combination therapy for vivax malaria.** *Lancet Infect Dis* 2010, **10:**405–416.
25. Ratcliff A, Siswantoro H, Kenangalem E, Maristela R, Wuwung RM, Laihad F, Ebsworth EP, Anstey NM, Tjitra E, Price RN: **Two fixed-dose artemisinin combinations for drug-resistant falciparum and vivax malaria in Papua, Indonesia: an open-label randomised comparison.** *Lancet* 2007, **369:**757–765.
26. Boyd MF, Kitchen SF: **On the infectiousness of patients infected with *Plasmodium vivax* and *Plasmodium falciparum*.** *Am J Trop Med Hyg* 1937, **s1-17:**253–262.
27. Collins WE, Sullivan JS, Nace D, Williams T, Sullivan JJ, Galland GG, Grady KK, Bounngaseng A: **Experimental infection of *Anopheles farauti* with different species of *Plasmodium*.** *J Parasitol* 2002, **88:**295–298.
28. Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, Hay SI: **Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*.** *Parasit Vectors* 2011, **4:**92.
29. Coatney GR: **Relapse in malaria - an enigma.** *J Parasitol* 1976, **62:**3–9.
30. Garnham PC: **Malaria parasites of man: life-cycles and morphology (excluding ultrastructure)**. In *Malaria: Principles and practice of malariology.* Edinburgh, UK: Churchill Livingstone; 1989:61–96.
31. Krotoski WA: **The hypnozoite and malarial relapse.** *Prog Clin Parasitol* 1989, **1:**1–19.
32. Baird JK, Hoffman SL: **Primaquine therapy for malaria.** *Clin Infect Dis* 2004, **39:**1336–1345.
33. Galappaththy GN, Omari AA, Tharyan P: **Primaquine for preventing relapses in people with *Plasmodium vivax* malaria.** *Cochrane Database Syst Rev* 2007, **1:**CD004389.
34. Anthony MP, Burrows JN, Duparc S, Moehrle JJ, Wells TN: **The global pipeline of new medicines for the control and elimination of malaria.** *Malar J* 2012, **11:**316.
35. Beutler E: **G6PD deficiency.** *Blood* 1994, **84:**3613–3636.
36. Cappellini MD, Fiorelli G: **Glucose-6-phosphate dehydrogenase deficiency.** *Lancet* 2008, **371:**64–74.
37. WHO: *Guidlines for the treatment of malaria, Second Edition.* Geneva, Switzerland: World Health Organization; 2010.
38. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, Snow RW: **Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007.** *PLoS Med* 2010, **7:**e1000290.
39. Patil AP, Okiro EA, Gething PW, Guerra CA, Sharma SK, Snow RW, Hay SI: **Defining the relationship between *Plasmodium falciparum* parasite rate and clinical disease: statistical models for disease burden estimation.** *Malar J* 2009, **8:**186.
40. Coatney GR, Cooper WC: **Recrudescence and relapse in vivax malaria.** *Proceedings of the 4th International Congress on Tropical Medicine* 1948, **1:**629–639.
41. Garnham PC, Bray RS, Bruce-Chwatt LJ, Draper CC, Killick-Kendrick R, Sergiev PG, Tiburskaja NA, Shute PG, Maryon M: **A strain of *Plasmodium vivax* characterized by prolonged incubation: morphological and biological characteristics.** *Bull World Health Organ* 1975, **52:**21–32.
42. Winckel CWF: **Long latency in *Plasmodium vivax* infections in a temperate zone.** *Doc Med Geog Trop* 1955, **7:**292–298.
43. Coatney GR, Collins WE, Contacos PG: ***Plasmodium vivax***. In *The Primate Malarias.* Bethesda, MD: U.S. National Institute of Allergy and Infectious Diseases; 1971:43–67.
44. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, Del Portillo HA: **Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite.** *Lancet Infect Dis* 2009, **9:**555–566.
45. Baird JK, Rieckmann KH: **Can primaquine therapy for vivax malaria be improved?** *Trends Parasitol* 2003, **19:**115–120.
46. Shute PG, Lupascu G, Branzei P, Maryon M, Constantinescu P, Bruce-Chwatt LJ, Draper CC, Killick-Kendrick R, Garnham PC: **A strain of *Plasmodium vivax* characterized by prolonged incubation: the effect of numbers of sporozoites on the length of the prepatent period.** *Trans R Soc Trop Med Hyg* 1976, **70:**474–481.
47. White NJ: **Determinants of relapse periodicity in *Plasmodium vivax* malaria.** *Malar J* 2011, **10:**297.
48. Douglas NM, Nosten F, Ashley EA, Phaiphun L, Van Vugt M, Singhasivanon P, White NJ, Price RN: ***Plasmodium vivax* recurrence following falciparum and mixed species malaria: risk factors and effect of antimalarial kinetics.** *Clin Infect Dis* 2011, **52:**612–620.
49. Shanks GD, White NJ: **The activation of vivax malaria hypnozoites by infectious diseases.** *Lancet Infect Dis* 2013, **13:**900–906.
50. Lysenko AJ, Semashko IN: **Geography of malaria. A medico-geographic profile of an ancient disease**. In *Itogi Nauki: Medicinskaja Geografija.* Edited by Lebedew AW. Moscow: Academy of Sciences, USSR; 1968:25–146.

51. Lover AA, Coker RJ: **Quantifying effect of geographic location on epidemiology of *Plasmodium vivax* malaria.** *Emerg Infect Dis* 2013, **19:**1058–1065.
52. Macdonald G: *The Epidemiology and Control of Malaria.* London: Oxford University Press; 1957.
53. Holt B, Lessard JP, Borregaard MK, Fritz SA, Araujo MB, Dimitrov D, Fabre PH, Graham CH, Graves GR, Jonsson KA, Nogues-Bravo D, Wang ZH, Whittaker RJ, Fjeldsa J, Rahbek C: **An update of Wallace's zoogeographic regions of the world.** *Science* 2013, **339:**74–78.
54. Peel MC, Finlayson BL, McMahon TA: **Updated world map of the Koppen-Geiger climate classification.** *Hydrol Earth Sys Sci* 2007, **11:**1633–1644.
55. Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'Amico JA, Itoua I, Strand HE, Morrison JC, Loucks CJ, Allnutt TF, Ricketts TH, Kura Y, Lamoreux JF, Wettengel WW, Hedao P, Kassem KR: **Terrestrial ecoregions of the worlds: a new map of life on Earth.** *Bioscience* 2001, **51:**933–938.
56. *PubMed.* http://www.ncbi.nlm.nih.gov/pubmed/.
57. Howes RE, Battle KE, Satyagraha AW, Baird JK, Hay SI: **G6PD deficiency: global distribution, genetic variants and primaquine therapy.** *Adv Parasitol* 2013, **81:**133–201.
58. Cooper WC, Myatt AV, Hernandez T, Jeffery GM, Coatney GR: **Studies in human malaria. XXXI. Comparison of primaquine, isopentaquine, SN-3883, and pamaquine as curative agents against Chesson strain vivax malaria.** *Am J Trop Med Hyg* 1953, **2:**949–957.
59. Gogtay NJ, Desai S, Kamtekar KD, Kadam VS, Dalvi SS, Kshirsagar NA: **Efficacies of 5- and 14-day primaquine regimens in the prevention of relapses in *Plasmodium vivax* infections.** *Ann Trop Med Parasitol* 1999, **93:**809–812.
60. Alvarez G, Pineros JG, Tobon A, Rios A, Maestre A, Blair S, Carmona-Fonseca J: **Efficacy of three chloroquine-primaquine regimens for treatment of *Plasmodium vivax* malaria in Colombia.** *Am J Trop Med Hyg* 2006, **75:**605–609.
61. Looareesuwan S, White NJ, Chittamas S, Bunnag D, Harinasuta T: **High rate of *Plasmodium vivax* relapse following treatment of falciparum malaria in Thailand.** *Lancet* 1987, **2:**1052–1055.
62. Taggart JV, Earle DP Jr, Berliner RW, Welch WJ, Zubrod CG, Jailer JW, Kuhn BH, Norwood J, Shannon JA: **Studies on the chemotherapy of the human malarias; the antimalarial activity of quinacrine.** *J Clin Invest* 1948, **27:**93–97.
63. Baird JK: **Chloroquine resistance in *Plasmodium vivax*.** *Antimicrob Agents Chemother* 2004, **48:**4075–4083.
64. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, Noor AM, Snow RW: **Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project.** *Malar J* 2007, **6:**17.
65. *Google Maps.* https://maps.google.com/.
66. ESRI: *ArcGIS Desktop 10.0.* Redlands, CA: Environmental Systems Resource Institute; 2010.
67. Viechtbauer W: **Bias and efficiency of meta-analytic variance estimators in the random-effects model.** *J Educ Behav Stat* 2005, **30:**261–293.
68. Viechtbauer W: **Conducting meta-analyses in R with the metafor package.** *Journal of Statistical Software* 2010, **36:**1–48.
69. R Core Team: *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2013.
70. Raudenbush SW: **Analyzing effect sizes: random effects models**. In *The handbook of research synthesis and meta-analysis.* 2nd edition. Edited by Cooper H, Hedges LV, Valentine JC. New York: Russell Sage Foundation; 2009:295–315.
71. Freeman MF, Tukey JW: **Transformations related to the angular and the square root.** *Ann Math Stat* 1950, **21:**607–611.
72. Clayton D, Hills M: *Statistical Models in Epidemiology.* Oxford, UK: Oxford University Press; 2013.
73. Ulm K: **A simple method to calculate the confidence interval of a standardized mortality ratio (SMR).** *Am J Epidemiol* 1990, **131:**373–375.
74. Gonzalez-Ceron L, Mu J, Santillan F, Joy D, Sandoval MA, Camas G, Su X, Choy EV, Torreblanca R: **Molecular and epidemiological characterization of *Plasmodium vivax* recurrent infections in southern Mexico.** *Parasit Vectors* 2013, **6:**109.
75. Llanos-Cuentas A, Lacerda MV, Rueangweerayut R, Krudsood S, Gupta SK, Kochar SK, Arthur P, Chuenchom N, Mohrle JJ, Duparc S, Ugwuegbulam C, Kleim JP, Carter N, Green JA, Kellam L: **Tafenoquine plus chloroquine for the treatment and relapse prevention of Plasmodium vivax malaria (DETECTIVE): a multicentre, double-blind, randomised, phase 2b dose-selection study.** *Lancet* 2014, **383:**1049–1058.
76. Sharma SN, Subbarao SK, Choudhury DS, Pandey KC: **Role of *An. culicifacies* and *An. stephensi* in malaria transmission in urban Delhi.** *Indian J Malariol* 1993, **30:**155–168.
77. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, Gething PW, Elyazar IR, Kabaria CW, Harbach RE, Hay SI: **The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis.** *Parasit Vectors* 2011, **4:**89.
78. Taylor JE, Pacheco MA, Bacon DJ, Beg MA, Machado RL, Fairhurst RM, Herrera S, Kim JY, Menard D, Povoa MM, Villegas L, Mulyanto, Snounou G, Cui L, Zeyrek FY, Escalante AA: **The evolutionary history of *Plasmodium vivax* as inferred from mitochondrial genomes: parasite genetic diversity in the Americas.** *Mol Biol Evol* 2013, **30:**2050–2064.
79. Arnold J, Alving AS, Hockwald RS, Clayman CB, Dern RJ, Beutler E: **Natural history of Korean vivax malaria after deliberate inoculation of human volunteers.** *J Lab Clin Med* 1954, **44:**723–726.
80. Cooper WC, Ruhe DS, Coatney CR, Josephson ES, Young MD: **Studies in human malaria. VIII. The protective and therapeutic action of quinacrine against St. Elizabeth strain vivax malaria** *Am J Hyg* 1949, **49:**25–40.
81. Tiburskaja NA, Vrublevskaja OS: **The course of infection caused by the North Korean strain of *Plasmodium vivax*.** *In WHO Malaria Report 1977* 1977, **895:**1–19. 1–19.
82. Mueller I, Galinski MR, Tsuboi T, Arevalo-Herrera M, Collins WE, King CL: **Natural acquisition of immunity to *Plasmodium vivax*: epidemiological observations and potential targets.** *Adv Parasitol* 2013, **81:**77–131.
83. Bignami A: **Concerning the pathogenesis of relapses in malarial fevers.** *South Med J* 1913, **6:**79–89.
84. Price RN, Douglas NM, Anstey NM: **New developments in *Plasmodium vivax* malaria: severe disease and the rise of chloroquine resistance.** *Curr Opin Infect Dis* 2009, **22:**430–435.
85. Rieckmann KH, Davis DR, Hutton DC: ***Plasmodium vivax* resistance to chloroquine?** *Lancet* 1989, **2:**1183–1184.
86. Boyd MF: **The influence of sporozoite dosage in vivax malaria.** *Am J Trop M* 1940, **20:**279–286.
87. Coatney GR, Cooper WC, YM D, Burgess RW: **Studies in human malaria. XVIII. The life pattern of sporozoite-induced St Elizabeth strain vivax malaria.** *Am J Epidemiol* 1950, **51:**200–215.
88. Findlay GM: *Recent advances in chemotherapy.* Philadelphia, USA: The Blakiston Company; 1951.
89. Most H: **Clinical trials of anti-malarial drugs**. In *Internal Medicine in World War II, Infectious Diseases, vol 2. Volume 2.* Washington, D.C.: Office of the Surgeon General, Department of the Army; 1963:479–492.
90. Coatney GR, Alving AS, Jones R Jr, Hankey DD, Robinson DH, Garrison PL, Coker WG, Donovan WN, Di Lorenzo A, Marx RL, Simmons IH: **Korean vivax malaria. V. Cure of the infection by primaquine administered during long-term latency.** *Am J Trop Med Hyg* 1953, **2:**985–988.

doi:10.1186/1475-2875-13-144
Cite this article as: Battle *et al.*: Geographical variation in *Plasmodium vivax* relapse. *Malaria Journal* 2014 **13**:144.

Chapter 4 – Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease

To generate a global estimate of *Plasmodium vivax* clinical burden, it was necessary be able to translate readily available estimates of *P. vivax* parasite into estimates of clinical incidence rates. This chapter has been published and is shown here in its final form. Supplementary figures to the publication are shown in the Appendix. In addition, a separate data descriptor detailing the database assembled for the purpose of the analysis has been published and is also included in the Appendix. The analysis utilizes the geographic zones identified in the previous chapter to allow for the relationship between the prevalence of infection and incidence of clinical disease to differ based on the regional relapse epidemiology.

Battle et al. Malaria Journal (2015) 14:191
DOI 10.1186/s12936-015-0706-3

MALARIA JOURNAL

RESEARCH Open Access

Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease

Katherine E Battle[1*], Ewan Cameron[1], Carlos A Guerra[2], Nick Golding[3], Kirsten A Duda[1], Rosalind E Howes[1], Iqbal RF Elyazar[4], Ric N Price[5,6], J Kevin Baird[4,6], Robert C Reiner Jr[7,8], David L Smith[1,2,8], Peter W Gething[1] and Simon I Hay[3,8,9]

Abstract

Background: Though essential to the development and evaluation of national malaria control programmes, precise enumeration of the clinical illness burden of malaria in endemic countries remains challenging where local surveillance systems are incomplete. Strategies to infer annual incidence rates from parasite prevalence survey compilations have proven effective in the specific case of *Plasmodium falciparum*, but have yet to be developed for *Plasmodium vivax*. Moreover, defining the relationship between *P. vivax* prevalence and clinical incidence may also allow levels of endemicity to be inferred for areas where the information balance is reversed, that is, incident case numbers are more widely gathered than parasite surveys; both applications ultimately facilitating cartographic estimates of *P. vivax* transmission intensity and its ensuring disease burden.

Methods: A search for active case detection surveys was conducted and the recorded incidence values were matched to local, contemporary parasite rate measures and classified to geographic zones of differing relapse phenotypes. A hierarchical Bayesian model was fitted to these data to quantify the relationship between prevalence and incidence while accounting for variation among relapse zones.

Results: The model, fitted with 176 concurrently measured *P. vivax* incidence and prevalence records, was a linear regression of the logarithm of incidence against the logarithm of age-standardized prevalence. Specific relationships for the six relapse zones where data were available were drawn, as well as a pooled overall relationship. The slope of the curves varied among relapse zones; zones with short predicted time to relapse had steeper slopes than those observed to contain long-latency relapse phenotypes.

Conclusions: The fitted relationships, along with appropriate uncertainty metrics, allow for estimates of clinical incidence of known confidence to be made from wherever *P. vivax* prevalence data are available. This is a prerequisite for cartographic-based inferences about the global burden of morbidity due to *P. vivax*, which will be used to inform control efforts.

Keywords: Malaria, *Plasmodium vivax*, Epidemiology, Incidence, Prevalence, Model

Background

Reliable estimates of clinical incidence of malaria have been an enduring challenge for epidemiologists working to measure the impact of the disease, define targets for control, and evaluate progress towards elimination [1-10]. Direct clinical incidence surveys are costly and time-consuming; as a result, many published large-scale estimates of incidence rely on passive reporting of cases to routine health information systems. These data are often incomplete or inaccurate [11,12] and must be adjusted using relationships between variables of unknown certainty [10]. Prevalence, or parasite rate (PR), on the other hand, is a more easily measured and widely available malaria metric [13]. A species-specific modelled relationship between *Plasmodium vivax* PR (*Pv*PR) and the rate of clinical illness, similar to that developed for *Plasmodium falciparum* [5,14], would be an important step towards the generation of a continuous global map of *P. vivax*

* Correspondence: katherine.battle@zoo.ox.ac.uk
[1]Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford, UK
Full list of author information is available at the end of the article

burden from which national and sub-national aggregate estimates of annual incidence can be compiled with known uncertainties.

Enumeration of the global disease burden attributable to *P. vivax* malaria has been identified as a key knowledge gap [15-17]. Large discrepancies exist in the currently available burden estimates [18], which have been calculated using a variety of methods. Figures based on cases reported to health systems estimate *P. vivax* incidence to be 15.8 million cases per year [2,18]. However, estimates derived from the 'cartographic' approach using mapped endemicity classes and populations at risk suppose that these values would be far greater: 132-391 million cases annually [4,17]. The cartographic method bypasses some of the challenges inherent in the surveillance-based approach, in which the numbers of cases reported are adjusted to account for incompleteness in reporting, usage of health facilities, and diagnostic confirmation, and it is difficult to quantify the precision of these adjustments. The cartographic approach, on the other hand, estimates cases through a geostatistical model of endemicity constrained by the input data, with strength borrowed implicitly from observations at neighbouring sites, such that the resulting case estimates carry a formal uncertainty metric testable via cross-validation. Both techniques have their limitations and reconciling them is a long-term goal; the first step towards which is a fuller implementation of the cartographic approach for *P. vivax*.

A global map of *P. vivax* prevalence from which cartographic incidence estimates may be generated has been published for 2010 [19] and efforts to update this map are underway. The 2010 map, shown in Figure 1, displays the stable and unstable limits of transmission as defined according to annual parasite incidence (API) data, as well as the predicted *Pv*PR (as a population average over the one to 99 year-old age range) at a 5 × 5 km pixel scale within the stable limits of *P. vivax* transmission (API ≥0.1 per 1,000 per annum) [20]. As this map illustrates, large swaths of densely populated areas are exposed to stable transmission, though it remains unclear how many clinical infections arise from the 2.5 billion people who live within the limits of *P. vivax* transmission [21] because the relationship between *Pv*PR and incident morbidity has not yet been reliably established for *P. vivax*.

It is necessary to model *P. vivax* separately from *P. falciparum* because of the biological and epidemiological differences that affect their observed prevalence of infection and patterns of clinical incidence [22]. Vivax malaria circulates in the blood at much lower parasite densities than *P. falciparum*, making it less likely to be detected by diagnostic techniques commonly used to measure prevalent infections: light microscopy and rapid diagnostic tests (RDTs) [22]. Nevertheless, low blood-parasite densities are still able to elicit symptomatic disease [23]. Cartographic estimates of *Pv*PR are an approximate order of magnitude lower than those for *P. falciparum* [19,24]. Although prevalence values exceeding those shown on the scale in Figure 1 are observed, particularly among children (see associated dataset [25]), the community prevalence of *P. vivax* is consistently low relative to *P. falciparum*, as illustrated in Figure 2. A possible explanation for this effect is natural immunity, which is acquired more rapidly against *P. vivax* than *P. falciparum*, such that infection prevalence peaks in young children, with *Pv*PR in adults significantly lower [22]. Prevalence of *P. vivax* starts to decline after the second year of age, whereas *P. falciparum*

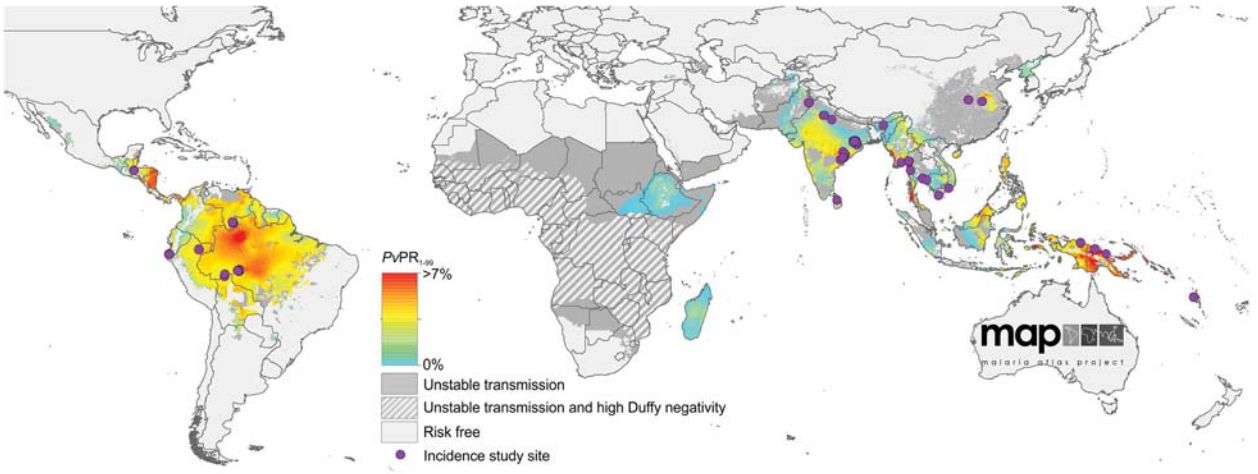


Figure 1 The spatial distribution of *Plasmodium vivax* endemicity in 2010 overlaid by ACD study sites. The spatial distribution of *P. vivax* [19] is shown using the MBG point estimates of the annual mean *Pv*PR (1 to 99 year-olds) within the spatial limits of stable transmission, displayed on a continuum of blue (low prevalence) to red (high prevalence). Areas within the stable limits that were predicted with high certainty (>0.9) to have a *Pv*PR less than 1% were classed as unstable. Regions where Duffy negativity gene frequency is predicted to exceed 90% [42] are shown in hatching for additional context. The location of study sites of the incidence records used in the final analysis are shown as purple points.

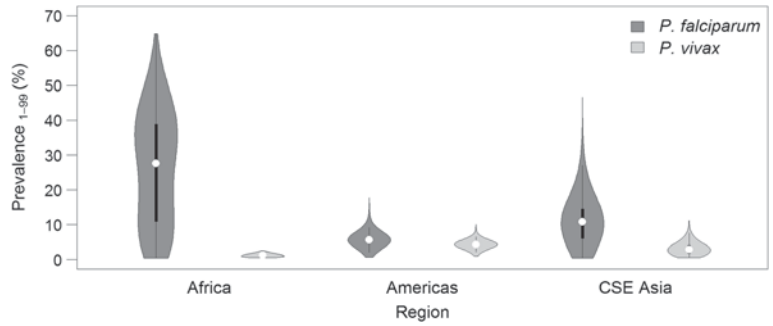


Figure 2 Comparison of *Plasmodium falciparum* and *Plasmodium vivax* prevalence. Prevalence values, obtained from the mapped *P. falciparum* and *P. vivax* endemicity surfaces [19,24]. Data for *P. falciparum* has been standardized to the 1 to 99 years age range to reflect *P. vivax* data [36]. The shaded areas correspond to each species and show a smoothed approximation of the frequency distribution (a kernel density plot) of parasite prevalence within each geographic region. The black central bar represents the interquartile range and the white circles indicate the median values.

prevalence continues to rise until later in life in all but the most intense transmission settings [26,27].

The most significant biological difference between *P. vivax* and *P. falciparum* is the ability of *P. vivax* to form liver stages capable of causing relapsing infections weeks to months after the initial inoculation [28]. Hence, in contrast to *P. falciparum*, which has only sporozoite-induced infections, blood-stage parasitaemia in *P. vivax* can arise from either mosquito-borne sporozoites or liver-borne hypnozoites. This has significant consequences for measuring the force of infection of blood-stage *P. vivax*. However, for the purpose of estimating the burden of clinical disease, the origin of the infection is not of utmost significance. A primary mosquito-borne infection and a relapsing hypnozoite-borne infection are both capable of causing symptomatic illness as well as onward transmission, and incidence from both are correlated with parasite prevalence. Hence, this study did not attempt to differentiate the incidence of relapse from the incidence of new infections, but rather examined the relationship between prevalent parasitaemia and incidence of clinical disease by geographic regions stratified by differing relapse patterns [28].

Issues inherent in estimating the burden of *P. vivax* malaria are addressed here by defining the relationship between published symptomatic *P. vivax* incidence rates derived from active case detection (ACD), matched with age-standardized measures of infection prevalence.

Methods

Data assembly

The aim of the data assembly was to build a comprehensive database of reports of clinical (symptomatic) incidence of *P. vivax* measured by ACD since 1 January, 1985, to be consistent with the *Pv*PR data used to develop global endemicity maps. A formal literature search was conducted in PubMed [29] on 27 November, 2013 using the search terms: ((malaria[MeSH Terms]) AND ("Incidence" [Mesh] OR "Epidemiology" [Mesh] OR "epidemiology" [Subheading])) AND ("1985/01/01"[Date - Publication] : "3000"[Date - Publication]). This returned 11,272 references.

Abstracts of all references returned were reviewed to determine if clinical incidence data could potentially be included in the paper. Reviews, case studies, and reports on imported malaria, animal studies, vector-only studies, and technical analyses (such as genetic mapping or transmission models) were excluded at this stage. Studies that did not explicitly report *P. vivax* incidence data collection in the abstract were not excluded in case it was reported in the main body.

The full texts of the 898 selected references, plus 78 publications flagged from previous studies [14,30], were then checked for the following criteria: (i) they contained longitudinal survey data involving ACD of symptomatic cases (typically defined by presence or recent history of fever); (ii) they were conducted in the general community (i.e., not patient sub-groups); (iii) malaria was diagnosed using microscopy or RDTs; and, (iv) results were presented in such a way that the number of cases and person-time observed could be determined. Due to diagnostic limitations, cases could not be distinguished as hypnozoite-borne and sporozoite-borne infections. There were no restrictions placed on age of the study population. For the initial data extraction, no limit was placed on the length or regularity of ACD, as long as the case detection methods were specifically reported. Studies that used passive case detection (PCD) only or were cross-sectional surveys were excluded.

Studies were geopositioned to a region (Africa, Americas or Central and Southeast (CSE) Asia), country and place name, and mapped to a specific latitude and longitude using location information from the source and gazetteers such as Encarta [31] and Google Maps [32], as described

previously [33]. The studies were also classified to a geographic zone of relapse phenotype as defined by Battle *et al.* [28]. Patterns in the timing of the first relapse event are thought to vary geographically among the zones illustrated in Additional file 1: Figure S1. The size and age range of a study cohort was extracted, and a single study reporting on multiple age ranges was disaggregated into separate records. Likewise, if a study contained different treatment or intervention arms, these were entered as separate records, and any control methods in place separate from the study were noted accordingly. Details regarding the time and length of the survey were recorded, along with type (ACD only or ACD + PCD) and frequency of detection. The number of cases and the person-years observed were recorded to determine incidence, as well as the diagnostic method, case definition, and any parasite density threshold applied to that definition.

If the number of person-years observed was not reported, it was estimated by multiplying the population of the study cohort by the length of the study. As this method of estimation may over-estimate person-time due to study members being lost to follow-up (and therefore under-estimate incidence), it was recorded whether person-time was explicitly reported in the study or if it had to be estimated.

Matching incidence to prevalence

Where possible, *Pv*PR data were extracted from the same publication as the incidence data to provide a temporally matched measure of prevalence in the same community. If *Pv*PR data were not reported, the Malaria Atlas Project (MAP) database [33,34] was searched for a prevalence study conducted in the same community in the same time period as the incidence study. For the records without a matched *Pv*PR value, a predicted prevalence was extracted from the *P. vivax* MAP endemicity surface using ArcGIS [35]. The methodology used to generate this surface is described in detail elsewhere [19,20], but briefly: the predicted *Pv*PR values represent an annualized mean prevalence in all ages (1-99 years) drawn from a species-specific model-based geostatistical (MBG) framework using 9,970 *Pv*PR surveys collected from 1985 to 2010 plus a suite of environmental covariates to estimate the prevalence in every 5 × 5 km square within the limits of stable transmission.

To facilitate modelling of the prevalence – incidence relationship, each inferred prevalence was standardized to a common age range of 0 to 85 years using the age-standardization model developed by Smith *et al.* initially for *P. falciparum* [36] and later updated for *P. vivax* [19]. The age-standardization was implemented using a freely available software package developed by the authors for the R statistical programming environment [37,38]. The full dataset used in this study and further details regarding its assembly are available in a dedicated publication [25]. A schematic of the data assembly stages is shown in Figure 3.

Model development

A Bayesian hierarchical model was developed to describe the relationship between the population prevalence and clinical incidence of vivax malaria. The model included a composite likelihood function to account for various aspects of the data: (i) the inherent randomness of the standard sampling distributions for both the parasite positive count (binomial) and the clinical case count (Poisson) at each site; (ii) a potential over-dispersion (extra-Poissonian variance) in the incidence observations attributable to site and study-specific random effects; (iii) a dependence of observed clinical incidence on the frequency of ACD [39]; and (iv) the impact of variation in the range of ages targeted by each study design given the importance of exposure-based, and hence age-dependent, immunity to clinical illness. While the asynchronous sampling of incidence and prevalence in different transmission seasons evinced by some surveys was not modelled explicitly, its contribution to the observational variance was effectively allowed for by these study- and site-specific random effects terms.

To account for (iii), in the absence of a single widely accepted parametric model of the effect of ACD occurrence, a non-parametric approach was used to infer this relationship. A modular statistical distribution was defined over the space of monotonically decreasing functions evaluated at the seven unique regularities of detection used in the ACD studies in the database: daily, every other day, every third day, five times per week, weekly, fortnightly, and monthly. The generative model for this distribution (denoted in Figure 4 as non-parametric gamma) was defined with respect to the joint order statistic of seven random variables; each gamma-distributed with a shape parameter of two and a rate parameter of one.

To account for (iv) a pair of scale coefficients were introduced to the likelihood function and fit simultaneously with all other random effects: the first acting to scale down the expected incidence for studies excluding children below 5 years of age, the second acting to scale up that for studies excluding older children and adults above 15 years of age. The age standardization of contemporary prevalence estimates described above was propagated to the input data as an adjustment of the numerator (observed parasite positives) in the binomial likelihood.

The fitted model for the prevalence – incidence relationship adopted here was a linear regression of the logarithm of the incidence rate against the logarithm of prevalence with zone-specific clustering effects. Note: as standard terminology in the statistical literature, the term

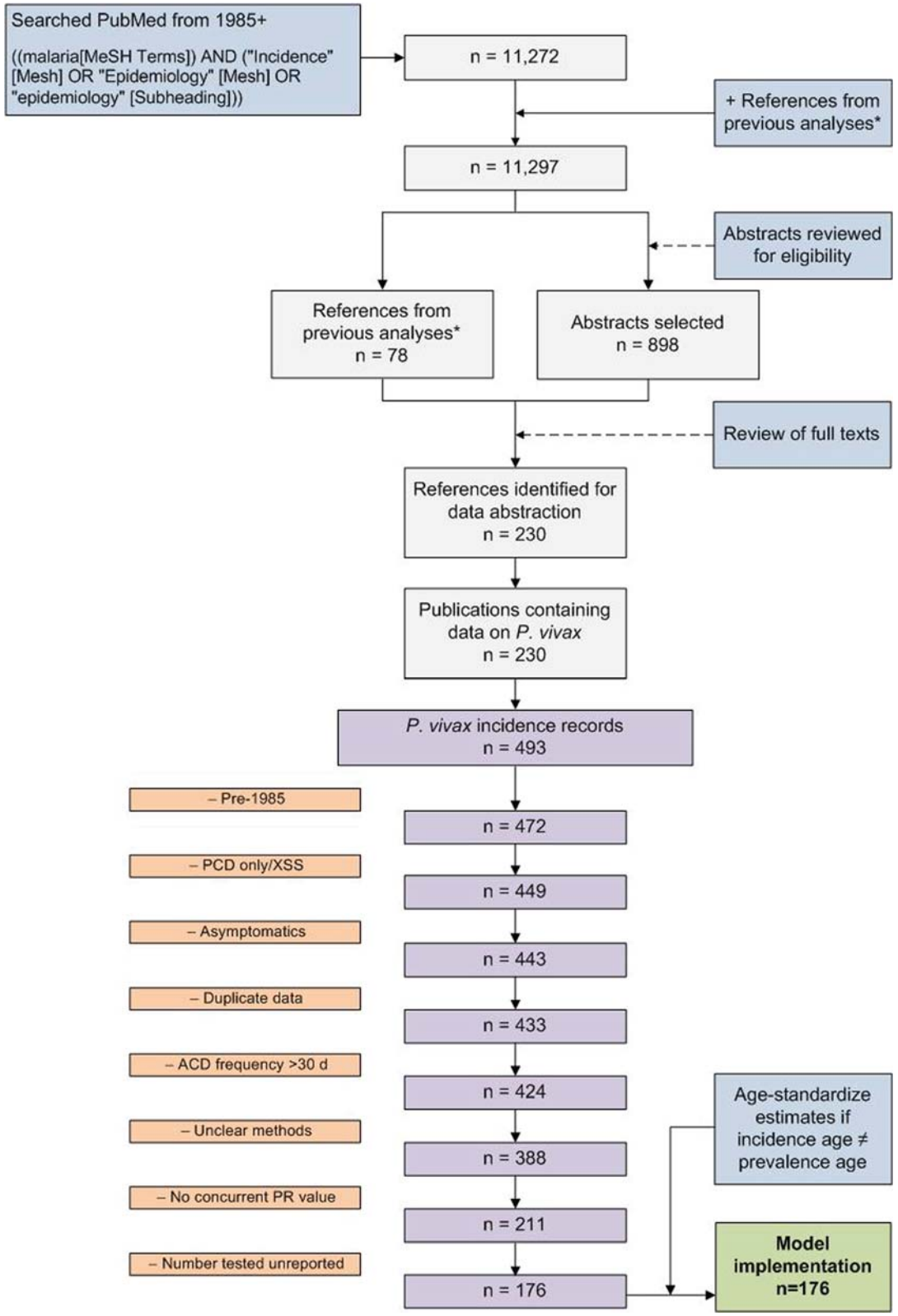


Figure 3 Schematic overview of the literature search procedure, results, and data exclusions to obtain clinical incidence records of use for model implementation. References from previous analyses* include those used by Patil *et al.* [14] and Griffin *et al.* [30].

Likelihood function

$$\mathrm{inc}_i^{\mathrm{obs}} \sim \mathrm{Poisson}(\mathrm{mean} = \exp(\mathrm{rate}_i) \times \mathrm{PYO}_i)$$

$$\mathrm{rate}_i = \exp(\alpha_{\mathrm{zone}[i]}^{\text{zone-specific}} + \Delta\alpha_i^{\text{site-specific random effect}} + c_1 \mathrm{I}_{\mathrm{LAR}[i]>5} + c_2 \mathrm{I}_{\mathrm{UAR}[i]<15} +$$

$$\beta_{\mathrm{zoe}[i]}^{\text{zone-specific}} \times (\log(\mathrm{p}_i^{\mathrm{true}}) - \log(0.025)) - \gamma_{\mathrm{ACD\ freq.}[i]}^{\text{ACD frequency effect}})$$

$$\mathrm{p}_i^{\mathrm{obs}} \sim \mathrm{Bin}(\mathrm{p}_i^{\mathrm{true}}, \mathrm{N}_i^{\mathrm{obs}})$$

Priors

$$\mathrm{p}_i^{\mathrm{true}} \sim \mathrm{Beta}(\eta_1, \eta_2);\ \eta_1 \sim \mathrm{Unif}(0.5,100);\ \eta_2 \sim \mathrm{Unif}(0.5,100)$$

$$\Delta\alpha_i^{\text{site-specific random effect}} \sim \mathrm{Norm}(\Delta\alpha_{\mathrm{study}[i]}^{\text{mean study-specific effect}}, \chi^2_{\mathrm{study}[i]});\ c_1, c_2 \sim \mathrm{Norm}(0,1)$$

$$\text{for } k \text{ in } 1,\ldots,\mathrm{N}_{\mathrm{studies}},\ \Delta\alpha_k^{\text{mean study-specific effect}} \sim Norm(0,\omega^2);\ \omega \sim \mathrm{Gamma}(0.25,0.25);\ \chi_k \sim \mathrm{Gamma}(0.25,0.25)$$

$$\alpha_{\mathrm{zone}[i]}^{\text{zone-specific}} = \alpha_{\mathrm{label}[\mathrm{zone}[i]]}^{\text{discrete}};\ \beta = \beta_{\mathrm{label}[\mathrm{zone}[i]]}^{\text{discrete}};\ \gamma_{\mathrm{ACD\ freq.}}^{\text{ACD frequency effect}} \sim \mathrm{NPG}(2,1)$$

$$\text{for } j = 1,\ldots,\mathrm{N}_{\mathrm{zones}},\ \mathrm{label}[j] \sim \mathrm{Cat}(\xi);\ \xi \sim \mathrm{Dirchlet}(\mathrm{I}_{\mathrm{N[zones]}})$$

$$\text{for } j = 1,\ldots,\mathrm{N}_{\mathrm{zones}},\ \alpha_j^{\mathrm{discrete}} \sim \mathrm{Norm}(\mu,\sigma^2);\ \mu \sim \mathrm{Norm}(-2,1);\ \sigma \sim \mathrm{Gamma}(1,1)$$

$$\text{for } j = 1,\ldots,\mathrm{N}_{\mathrm{zones}},\ \beta_j^{\mathrm{discrete}} \sim \mathrm{Gamma}(\nu_1,\nu_2);\ \nu_1 \sim \mathrm{Gamma}(2,1);\ \nu_2 \sim \mathrm{Gamma}(1,1)$$

Figure 4 The mathematical form of the model summarized in standard hierarchical Bayesian notation.

'linear' here denotes linearity in the coefficients, not necessarily the explanatory variables on which they act. Up to six unique intercepts and slopes were allowed to represent the six geographically bounded relapse phenotype zones represented in the dataset. Since variation in the epidemiology of *P. vivax* is not strictly defined by the geographic divisions proposed, the model was allowed to fit multiple zones with a common prevalence – incidence relationship by labelling these six possible slope-intercept pairs in order of increasing slope and treating label assignment for each zone as a categorical variable with proportions assigned a Dirichlet prior. Hence, multiple zones may share the same intercept and slope, and thereby share power for their inference, where this 'clustering' scenario is consistent with the observed data.

The mathematical form of this model is summarized in standard hierarchical Bayesian notation in Figure 4. Posterior simulation for this model was achieved via rejection Gibbs sampling with the JAGS (Just Another Gibbs Sampler) software package [40], with data entry and graphical summary achieved via the R statistical computing environment [37].

Results

Data assembly

Plasmodium vivax clinical incidence data were identified in 99 publications. Following checks that the studies met the inclusion criteria described above, these data were abstracted into 388 reports of incidence. The majority of the data came from CSE Asia (80%, 311/388), as shown in Table 1, with ten records from Africa and 67 from the Americas. Data originated from 18 countries in total: ten from CSE Asia, five from the Americas and three from Africa. The incidence measures observed ranged from zero to 1.6 per person year observed. The highest incidence values observed were in CSE Asia in Papua New Guinea (PNG). Summary statistics of the incidence observed by MAP region are shown in Table 2, and the violin plots in Figure 5.

Matching incidence to prevalence

Slightly less than half of the records (46%, 180/388) had a prevalence value available from the same reference. An additional 31 prevalence values were added to records using entries in the MAP database that were collected in the same site during the same year. This provided a space-time matched PR for approximately half (54%, n = 211) of the incidence records. A *Pv*PR value for each of the remaining 177 incidence records was obtained from the MAP *P. vivax* endemicity surface [19]. The MAP-based *Pv*PR values represent all-age estimates, and 123 (69%) of the incidence records without concurrent *Pv*PR were also measured in all ages. Of the incidence records with a concurrent *Pv*PR estimate, 154 (73%) of *Pv*PR surveys

Table 1 Data records by MAP region

Region	All *P. vivax* data	Data used in model
Africa+	10	0
America	67	43
CSE Asia	311	133
Total	**388**	**176**

Table 2 Incidence summary statistics

All data - incidence per 1,000 person-years observed							
Zone	**Zone name**	**N**	**Minimum**	**Mean**	**Median**	**Maximum**	**IQR**
2	**Central America**	3	72.07	103.93	80.00	159.71	43.82 (76.04, 119.86)
3	**South America**	64	0.00	227.47	161.52	977.31	281.06 (40.33, 321.39)
7	**Sub-Saharan Africa**	10	0.00	4.99	3.75	22.19	1.45 (2.48, 3.93)
8	**Monsoon Asia**	265	0.00	42.49	20.24	412.87	49.29 (8.05, 57.34)
10	**Southeast Asia**	24	0.00	291.56	290.87	710.50	497.62 (28.17, 525.79)
11	**N. Europe and Asia**	4	20.32	33.30	35.05	42.78	6.61 (30.87, 37.48)
12	**Melanesia**	18	56.81	709.63	758.19	1586.07	368.75 (531.25, 900.00)
All	**Total**	388	0	118.8	29.82	1586.07	99.38 (9.82, 109.20)
Data with concurrent *Pv*PR values used in analysis - incidence per 1,000 person-years observed							
	Region		**Minimum**	**Mean**	**Median**	**Maximum**	**IQR**
2	**Central America**	3	72.07	103.93	80.00	159.71	43.82 (76.04, 119.86)
3	**South America**	40	0.00	236.51	138.47	977.31	329.76 (22.16, 351.92)
8	**Monsoon Asia**	100	0.00	26.51	18.58	194.59	27.56 (7.36, 34.92)
10	**Southeast Asia**	18	4.48	250.68	89.77	692.31	472.71 (25.92, 498.63)
11	**N. Europe and Asia**	4	20.32	33.30	35.05	42.78	6.61 (30.87, 37.48)
12	**Melanesia**	11	56.81	674.07	658.76	1586.07	492.43(400.00, 892.43)
All	**Total**	176	0	139.10	29.10	1586.07	87.18 (11.74, 98.92)

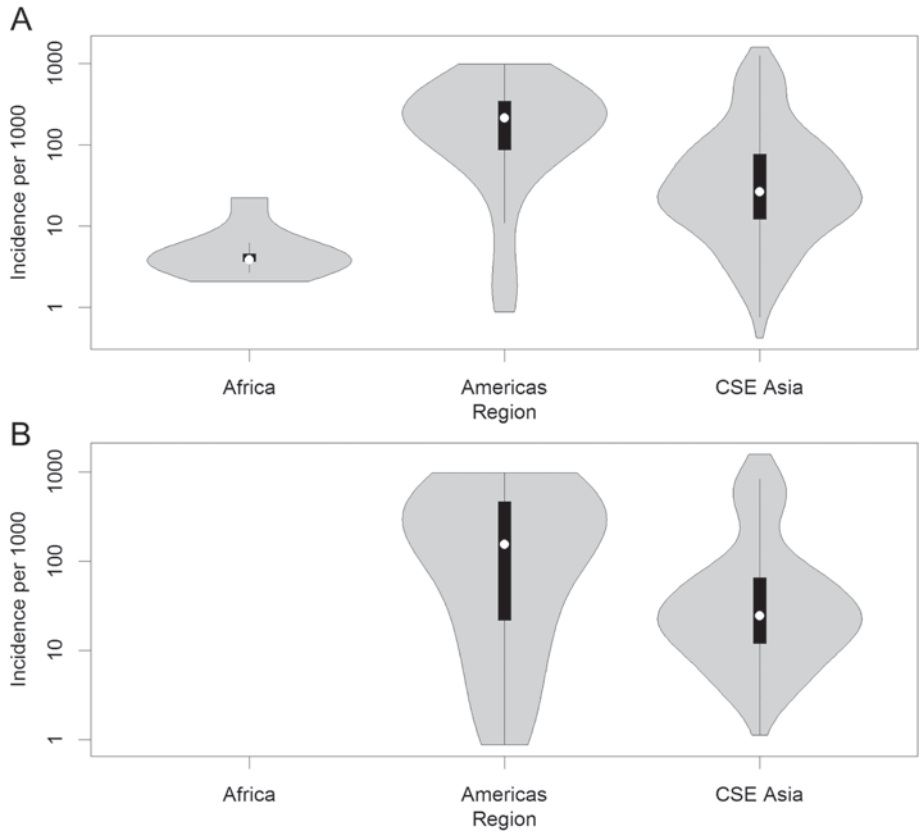


Figure 5 Violin plot of incidence (per 1,000 person-years observed). **A)** all data (n = 388) by region and **B)** data used in the analysis (n = 176) by region are shown with incidence on the logarithmic scale. The grey areas correspond to a smoothed approximation of the frequency distribution (a kernel density plot) of the incidence observed in each geographic region. The black central bar represents the interquartile range and the white circles indicate the median values.

were conducted in the same age group as the ACD cohort. The 110 *Pv*PR values that were not age-matched to the incidence data were age-standardized to the same age-range as the incidence data [36,38]. The *Pv*PR estimates for all records ranged from zero to just over 30%. The highest estimates were again observed in PNG. *Pv*PR data summary statistics are shown in Table 3.

The prevalence values extracted from the *P. vivax* endemicity map had a similar range to the *Pv*PR estimates measured alongside incidence (from close to zero to ~25%), but less variation (Additional file 2: Figure S2). This was because multiple incidence records that came from the same or nearby locations were matched to a *Pv*PR from the same or similar pixels in the predicted *Pv*PR map. Statistically, incidence records with only *Pv*PR values derived from the map were excluded because their uncertainties (in part due to the mismatch between the scale of MAP pixels and the scale of *Pv*PR variation within a pixel) were so large that the inclusion of these points did not add to the model fits. That is, only concurrently measured *Pv*PR values – reported from the same reference or another paper from the MAP database – were used. This also facilitated development of the statistical model as the selected studies all presented counts of the number examined and positive, and thus the same type of uncertainty was manifest for both the incidence and *Pv*PR estimates used in the analysis, whereas this would not be true for excluded *Pv*PR surveys that did not report the numerators or denominators (n = 35). Following all exclusions (Figure 3), 176 records from 75 sources remained to be used in the analysis. The temporal distribution and study size, based on person-time observed, of these records are shown in Figure 6.

The approximation of person-time in the majority of the selected records (76%, 133/176) was determined not to be an exclusion criterion. As illustrated in Additional file 3: Figure S3, there was a comparable level of noise in both those records with exact and approximate reported person-time. The data are plotted on both standard and logarithmic scales to also demonstrate that using the logarithm of incidence against the logarithm of prevalence better represents the distribution of the data. Note, however, that only 151 points appear in those panels using logarithmic scales because 25 records had a value of zero (four incidence and 21 prevalence) and could not be readily plotted on these axes.

A specific parasite density threshold in the case definition used in ACD studies was also set aside as an exclusion criterion. The majority of studies specified a case of *P. vivax* as a symptomatic episode at any detectable level of parasitaemia (≥1 asexual stage parasite per μl of blood), but nine of the 176 records included in the analysis specified a parasite density cut-off. Seven records were from studies applying a parasite density threshold of 500 parasites/μl to limit their Type II error rate (false attribution of vivax causality to a background fever) and an additional two applied a cut-off 1,000 parasites/μl. These studies were expected to return lower incidence estimates, but in fact were not observed to be outliers here, as seen in Additional file 4: Figure S4. That the application

Table 3 Parasite rate (%) summary statistics

All data - using concurrent *Pv*PR or MAP-based *Pv*PR from the *P. vivax* endemicity map							
Zone	**Zone name**	**N**	**Minimum**	**Mean**	**Median**	**Maximum**	**IQR**
2	**Central America**	3	1.1.6	1.28	1.27	1.40	0.12 (1.22, 1.34)
3	**South America**	64	0.00	2.51	1.80	7.65	3.41 (0.89, 3.36)
7	**Sub-Saharan Africa**	10	0.47	0.66	0.45	1.67	0.00 (2.48, 3.93)
8	**Monsoon Asia**	265	0.00	2.98	2.68	30.88	1.48 (1.26, 3.90)
10	**Southeast Asia**	24	0.00	3.39	3.43	6.98	3.47 (1.33, 4.42)
11	**N. Europe and Asia**	4	0.45	1.50	1.73	2.09	0.53 (2.11 2.94)
12	**Melanesia**	18	2.92	11.81	10.92	28.41	6.93 (9.82, 16.61)
All	**Total**	388	0.00	3.25	2.61	30.88	2.35 (1.27, 3.62)
Data with age-matched concurrent *Pv*PR used in the analysis							
	Region		**Minimum**	**Mean**	**Median**	**Maximum**	**IQR**
2	**Central America**	3	1.16	1.27	1.27	1.40	0.12 (1.22, 1.34)
3	**South America**	40	0.00	1.41	0.89	7.52	1.94 (0.00, 1.94)
8	**Monsoon Asia**	100	0.00	2.92	2.14	12.59	2.98 (0.90, 3.88)
10	**Southeast Asia**	18	0.79	2.98	2.14	6.98	3.09 (1.33, 4.42)
11	**N. Europe and Asia**	4	0.71	2.35	2.71	3.27	0.83 (2.11, 2.94)
12	**Melanesia**	11	8.25	14.52	14.77	28.41	6.04 (8.25, 15.95)
All	**Total**	176	0.00	3.27	1.87	28.41	3.05 (0.84, 3.89)

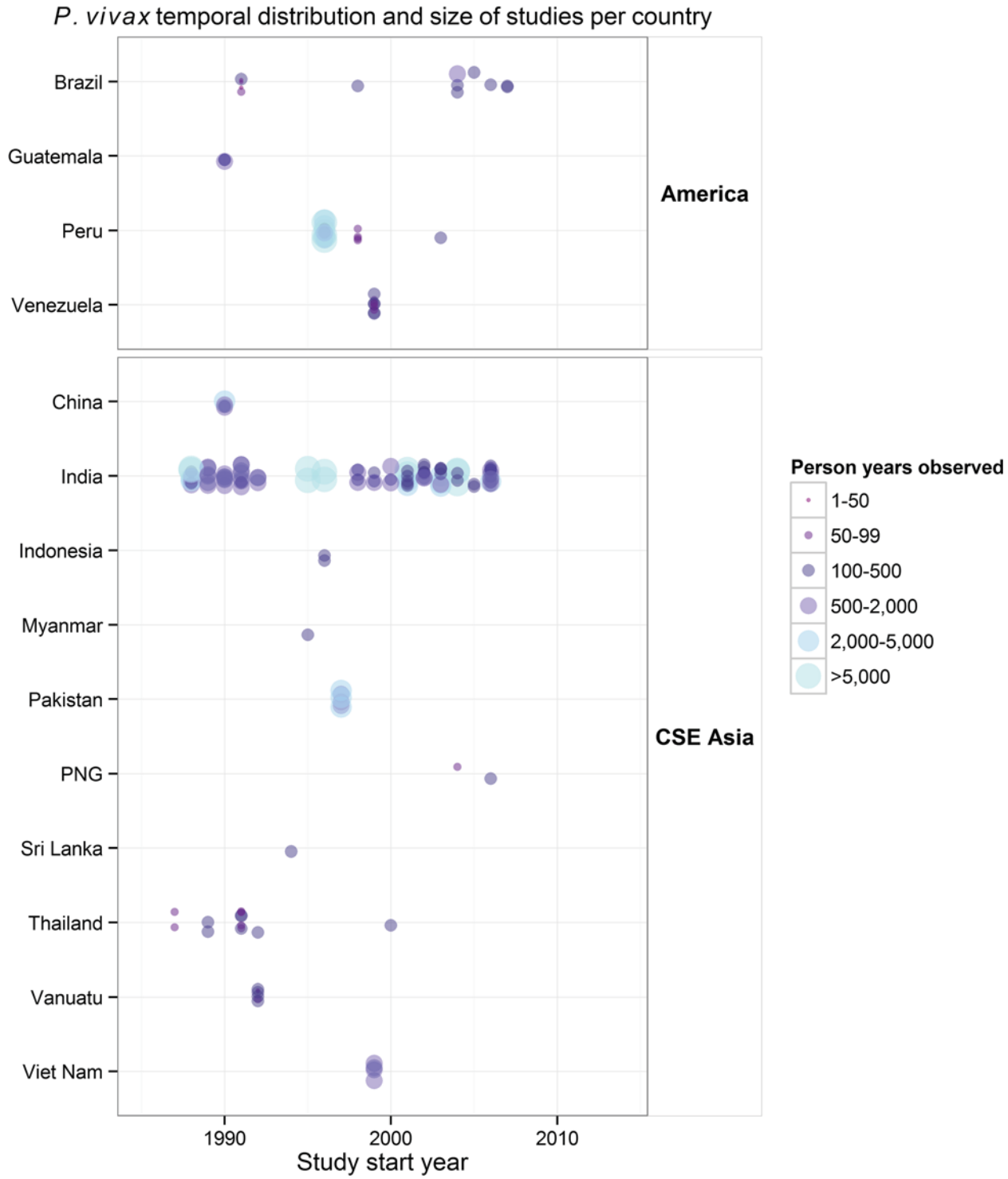


Figure 6 Temporal distribution of records used in the analysis. The size of the point reflects the number of person-years observed included in the 176 records that had an age-matched concurrent *Pv*PR measure with the incidence record.

of the cut-off did not result in lower estimates suggests that vivax-targeting ACD studies are less sensitive to case definition than is the experience for falciparum [41].

Model development

The posterior for the non-parametric fitted function modelling the impact of ACD regularity on the rate of

detected clinical incidence cases is illustrated in Additional file 5: Figure S5. In the subsequent Figures 7 and 8 the (point-wise) mean of this function was used to correct all observed incidence counts to a benchmark of fortnightly ACD. It was estimated that daily ACD studies report on average 12.2 (2.7,42) times (median and 95% credible interval, CrI) the number of fevers identified in studies with fortnightly ACD, whereas the scaling from fortnightly to monthly ACD is less marked at 0.81 (0.34,0.99). Some degree of variation in the dependence of observed incidence on ACD regularity among the geographic zones was expected, such that frequency of ACD would have a greater effect in areas with high risk of recurrence. The model of a shared effect was deemed sufficient, however, because a re-fit of the model allowing each zone to be assigned to one of two separate relationships failed to identify any significant difference in the resulting prevalence-incidence relationship.

The broad posterior credible intervals for the pair of scaling coefficients used here to account for age-dependence of the clinical incidence rate (namely, -0.28 [-0.83,0.27] for c_1 and 0.08 [-0.40,0.66] for c_2) suggest that these terms do not play a crucial role in these fits, a conclusion supported by visual inspection of the zone-specific prevalence – incidence relationships inferred upon exclusion of these terms from the model. However, the consequent inference that exposure-based immunity is unimportant for vivax malaria should be taken with caution: rather the present dataset is underpowered to investigate this effect since over 75% of the studies included here effectively report an all-ages incidence estimate.

The geographic origin of the studies was, however, of importance in the prevalence – incidence model. Figure 9 illustrates the distribution of the matched incidence and prevalence records shaded by the mean time to first relapse in each geographic zone. A significant degree of clustering between zones was identified through the fitted model. In particular, zones 8 and 11 (Monsoon Asia, and northern Asia and Europe) were found to share a common relationship in 50% of the posterior samples. These zones are characterized by long-latency relapse phenotypes (zone 11) or a combination of short and long latency (zone 8). At least three of the four remaining zones (2 - Central America, 3 - South America, 10 - Southeast Asia, and 12 - Melanesia) share a common relationship at a comparable rate. The zone-specific *Pv*PR and clinical incidence relationships thus recovered are illustrated as point-wise 68 and 95% CrIs in Figure 7 and their parameter estimates are summarized in Table 4. In the Table, α is the natural logarithm of incidence per person-year observed at a prevalence of 2.5%; in the model this is the intercept of the (logarithm of) prevalence – (logarithm of) incidence curve, such that the exponent of α is the intercept in cases per person year. Accordingly, β in Table 4 is the slope of the curve, such that if the prevalence were to increase from 2.5 to 7%, the incidence would increase by a factor of exp(β). By weighting the posterior for each zone by the proportion of observations from that zone in the dataset, a pooled relationship was produced for the entire dataset, as illustrated in Figure 8. For reference, the corresponding aggregate parameter estimates of the pooled relationship are α = -3.0 (-3.5,-2.4) and β = 0.71 (0.41,1.10). In other words, based on the pooled relationship a prevalence of 2.5% would correspond to an incidence of 49.8 cases per 1,000 person years (see Table 4).

The results benefit from the model structure by producing associated measures of uncertainty. As shown in Figures 7 and 8, the point-wise CrIs are narrowest around the axis of the regression model at 2.5% prevalence. Zone-specific relationships informed by few data points (zones 2, 11 and 12) have wider CrIs. Based on these wide uncertainty bands, the predicted incidence can change by a factor of 100. While this appears to be a large range, it is representative of the data; Figure 8 illustrates the wide range of incidence measures that were observed in communities with nearly the same *Pv*PR.

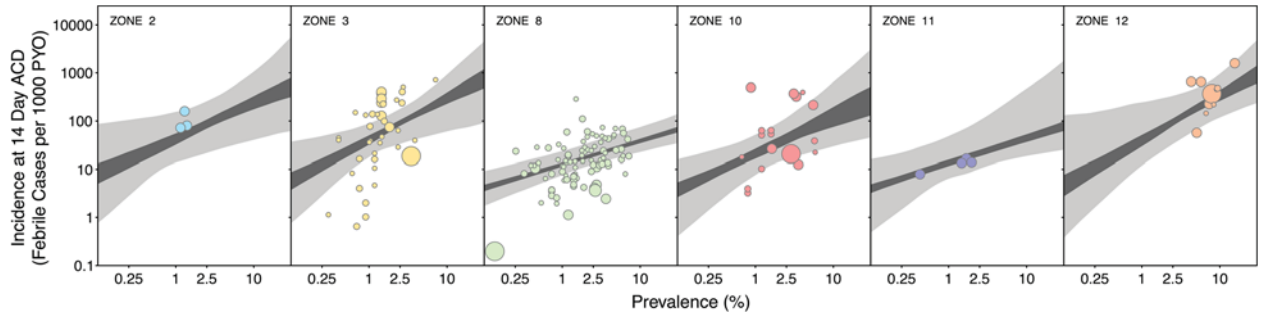


Figure 7 The zone-specific prevalence-incidence relationships shown as point-wise 68% and 95% credible intervals. Zone 2 is Central America, zone 3 is South America, zone 8 is Monsoon Asia (India), zone 10 is Southeast Asia, zone 11 is northern Asia and Europe and Zone 12 is Melanesia. The 95% CrIs are shown in light grey and the 68% CrIs are shown in dark grey. The size of the point corresponds to the time period between each ACD visit (see Figure 8) and the colours of the zones correspond to those shown in Figure 9.

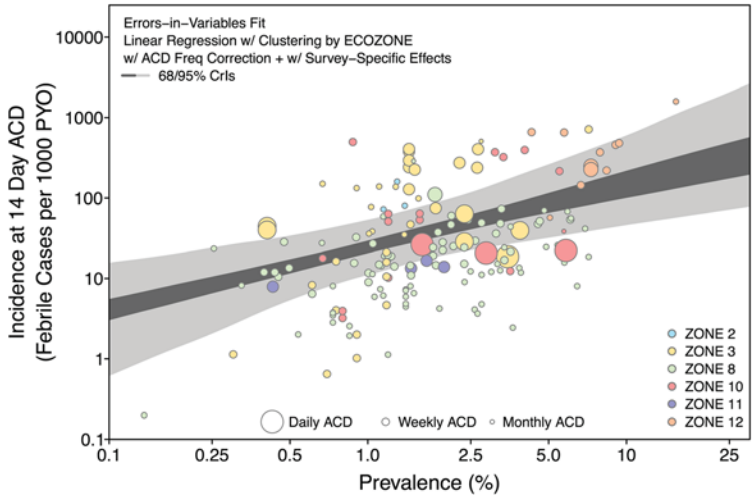


Figure 8 Pooled prevalence-incidence relationship for the entire dataset. To produce a pooled fit, the posterior of each zone was weighted by the number of observations from that zone. An errors-in-variables fit was used to allow for uncertainty in the independent variable as well as the dependent variable (ordinary linear regression would assume no uncertainty in the former). Point-wise 95% CrIs are shown in light grey and 68% CrIs are shown in dark grey. The colours of the zones match those shown in Figure 9.

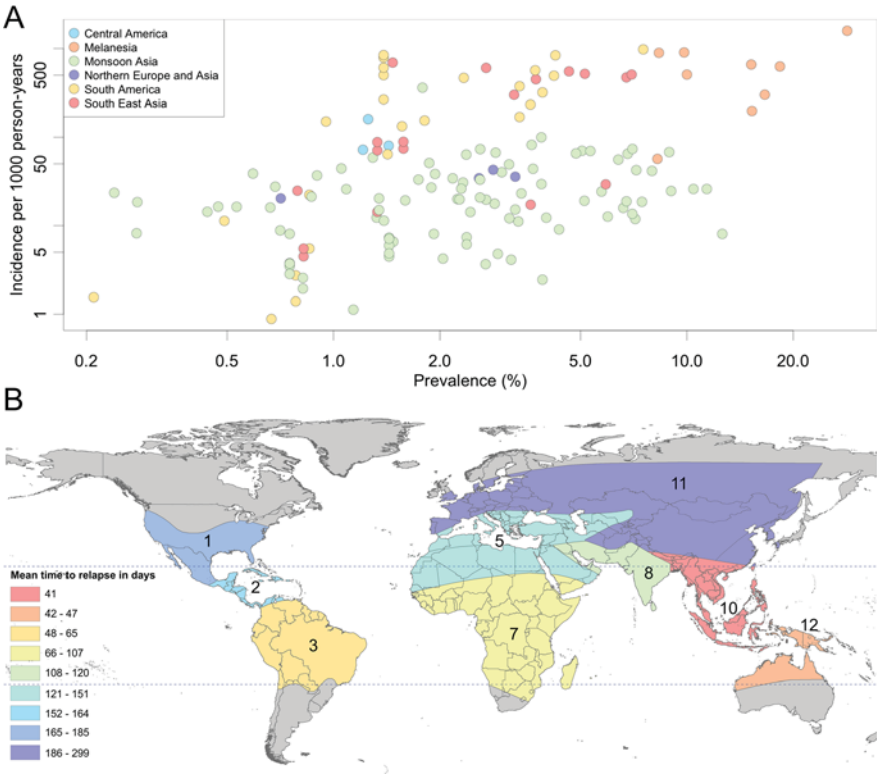


Figure 9 Scatter plot of data used in analysis coloured by relapse zones. Panel **A** plots the data used in the analysis by the relapse zones on log scales. The points are coloured by the mean time to relapse predicted in each zone shown in panel **B**.

Table 4 Parameter estimates by zone

Zone	Name	α Median [95% CrI]	exp(α)*1000 (cases per 1000 person years at 2.5%*Pv*PR)	β Median [95% CrI]
2	Central America	-2.4 [-3.8,-1.4]	90.7	0.68 [0.13,1.53]
3	South America	-2.4 [-3.4,-1.7]	90.7	0.85 [0.29,1.51]
8	Monsoon Asia	-3.9 [-4.4,-3.3]	20.2	0.49 [0.30,0.70]
10	Southeast Asia	-3.1 [-4.1,-2.1]	45.0	0.71 [0.24,1.49]
11	N. Europe and Asia	-3.8 [-4.6,-2.3]	22.4	0.51 [0.18,1.25]
12	Melanesia	-2.4 [-3.4,-1.1]	90.7	0.91 [0.17,1.55]
All	Pooled relationship	-3.0 [-3.5, -2.4]	49.8	0.71 [0.41,1.10]

Discussion

This study provides a fundamental component for calculation of the *P. vivax* clinical burden. The result of the work presented here is a model of the relationship between incidence of symptomatic vivax malaria and prevalence of detectable blood-stage *P. vivax* infection. This relationship will allow for the burden of *P. vivax* to be estimated using an updated map of *P. vivax* endemicity. Estimates of burden from maps of prevalence allow for measures of incidence to be made with associated measures of uncertainty.

Two key aspects of the analysis presented are the spatial and temporal components of the data. All incidence data were matched to *Pv*PR data that were measured in the same community at the same time. However, there were subtle differences in how *Pv*PR was measured among the various studies. In many studies, *Pv*PR was measured at the start of the ACD observation period as a baseline measure of endemicity. In other studies, there was more than one cross-sectional survey done during the incidence follow-up period. In those records, the *Pv*PR value is a pooled estimate, which was deemed acceptable because none of the studies administered a radical cure following the initial prevalence survey. This would have contributed to the noise observed in the data, but it is accounted for in the resulting models within the study-specific random effects as well as the uncertainty reflected in the CrIs.

Modelling the relationship between prevalence and incidence specifically for *P. vivax* presented new challenges not encountered in similar work for *P. falciparum* (Ewan Cameron, personal communication, 2015) [14]. There were far less incidence data available for *P. vivax* relative to *P. falciparum* [25]. The majority of the published *P. vivax* incidence data was from CSE Asia. This signals the need for improved active surveillance coverage in the Americas and implementation of RDTs that test for non-falciparum species in areas previously thought to be non-endemic for *P. vivax*, such as East Africa.

There were not age-stratified data available that would have allowed for age-specific burden modelling as done recently for *P. falciparum* (Ewan Cameron, personal communication, 2015). Age-dependent immunity causes high incidence of infection in very young children in high transmission settings with lower incidence in older children and adults [22]. Over 75% of the studies used in the analysis were conducted in whole populations, but the differing age groups in the remainder of the dataset was dealt with through a statistical correction designed to scale down the expected incidence in populations that did not include young children (under five years) and scale up the incidence in populations that did not include children and adults over 15 years of age. Further work involving this model will be improved as *P. vivax* transmission models are developed and the dependence of infection on age in different transmission settings can be explicitly derived.

Aside from issues of data availability, biological features of *P. vivax*, including its ability to cause relapsing infections following an initial mosquito-borne infection, were by necessity treated somewhat pragmatically in this modelling exercise. That is, relapse was not explicitly incorporated into the model since clinical cases due to relapse are captured by both the incidence and prevalence data. Rather, zone-specific relationships were developed to account for varying geographic patterns of relapse [28]. The slope of the prevalence-incidence relationship curve was steeper in regions where relapse is observed to occur rapidly following the primary infection. Zones with long latency relapse phenotypes, and therefore reduced annual relapse incidence (Figure 9 and Table 4), such as Monsoon Asia and northern Europe and Asia, show shallower slopes. These regions, as shown in Figure 7, reach an incidence of one case per 100 people per year at around 1% prevalence, whereas the other regions shown reach a similar incidence level at even lower prevalence values.

Conclusion

The modelling outputs presented here inform the understanding of the nature of prevalence and incidence relationships, but more importantly, the zone-specific relationships will facilitate global predictions of clinical burden to be made that account for regional differences

in *P. vivax* epidemiology. Because of its ability to relapse, *P. vivax* will be the final hurdle as regions move towards elimination in much of the malaria-endemic world. Burden estimates of known accuracy will enable assessments to be made of the impact of *P. vivax* malaria on health systems and economies within and among endemic regions, which will be essential to strategic planning for the control and ultimate elimination of *P. vivax*. The extremes of current estimates – 15.8 million *versus* 391 million clinical cases [17,18] – emphasize the need for a validated approach to measuring the burden imposed by this important and threatening parasite.

Additional files

Additional file 1: Geographic zones of relapse phenotype. Description: Relapse patterns of strains of *P. vivax* are proposed to differ among the nine ecological zones shown above [28].

Additional file 2: Incidence records plotted versus the predicted MAP-based *Pv*PR values and observed concurrent *Pv*PR values. Description: Incidence points versus MAP *Pv*PR values are shown in black and those points using concurrently measured *Pv*PR values are shown in blue.

Additional file 3: Approximate and exact person-time shown in plots of incidence per 1,000 person-years versus parasite rate. Description: The incidence records with concurrent *Pv*PR estimates are plotted below on linear (A) and log scales (B) below. The blue points are those with approximated person time and those in grey had exact person-time reported.

Additional file 4: Case parasite density threshold shown in scatter plots of incidence per 1,000 person-years versus parasite rate. Description: The incidence records are plotted below on linear (A) and log scales (B) below. The grey points are studies that used any parasitaemia in the case definition. Blue points are studies that defined a case as ≥500 parasites/μl of blood and red points, 1000 parasites/μl.

Additional file 5: The posterior for the non-parametric fitted function giving the impact of ACD frequency on the rate of detected clinical incidence cases. Description: A non-parametric statistical distribution the frequency of ACD was fit under a monotonicity restriction, which forces the posterior to preserve a strict ordering of the observed incidence scaling with respect to ACD frequency.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
KEB, PWG and SIH conceived the study and oversaw its implementation, with input from JKB, RNP and DLS. KEB wrote the first draft of the manuscript and assembled data with assistance from CAG, KAD, REH, and IRFE. NG wrote statistical software to process and standardise data. EC led the design of the modelling framework with input from NG, RCR and DLS. All authors participated in the interpretation of results and in the writing and editing of the manuscript. All authors read and approved the final manuscript. KEB and SIH will act as guarantors for the paper.

Acknowledgements
SIH is funded by a Senior Research Fellowship from the Wellcome Trust (095066), which also supports KEB, KAD and REH. NG is funded by a grant from the Bill & Melinda Gates Foundation (OPP1053338). IRFE is funded by a Public Health and Tropical Medicine Fellowship of the Wellcome Trust (B9RZGS0). RNP is a Wellcome Trust Senior Fellow in Clinical Science (091625). JKB is supported by Wellcome Trust grant (B9RJIXO). PWG is a Career Development Fellow (#K00669X) jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and receives support from the Bill & Melinda Gates Foundation (#OPP1068048, #OPP1106023), which also supports EC. DLS is funded by a grant from the Bill & Melinda Gates Foundation (OPP1110495), which also supports RCR. RCR, DLS and SIH also acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health.

Author details
[1]Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford, UK. [2]Sanaria Institute for Global Health and Tropical Medicine, Rockville, MD, USA. [3]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, UK. [4]Eijkman-Oxford Clinical Research Unit, Jalan Diponegoro No 69, Jakarta, Indonesia. [5]Global and Tropical Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia. [6]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [7]Indiana University School of Public Health, Bloomington, IN, USA. [8]Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. [9]Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA 98121, USA.

Received: 25 January 2015 Accepted: 22 April 2015
Published online: 07 May 2015

References
1. Carter R, Mendis KN. Evolutionary and historical aspects of the burden of malaria. Clin Microbiol Rev. 2002;15:564–94.
2. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C. Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. PLoS Med. 2011;8, e1001142.
3. Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW. Urbanization, malaria transmission and disease burden in Africa. Nat Rev Microbiol. 2005;3:81–90.
4. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. Lancet Infect Dis. 2004;4:327–36.
5. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, et al. Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. PLoS Med. 2010;7, e1000290.
6. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA, et al. Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2014;384:1005–70.
7. Snow RW. Sixty years trying to define the malaria burden in Africa: have we made any progress? BMC Med. 2014;12:227.
8. Snow RW, Craig M, Deichmann U, Marsh K. Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. Bull World Health Organ. 1999;77:624–40.
9. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. Nature. 2005;434:214–7.
10. WHO. World Malaria Report 2013. Geneva, Switzerland: World Health Organization; 2013. p. 199.
11. Gething PW, Noor AM, Gikandi PW, Ogara EA, Hay SI, Nixon MS, et al. Improving imperfect data from health management information systems in Africa using space-time geostatistics. PLoS Med. 2006;3, e271.
12. Moyes CL, Temperley WH, Henry AJ, Burgert CR, Hay SI. Providing open access data online to advance malaria research and control. Malar J. 2013;12:161.
13. Hay SI, Smith DL, Snow RW. Measuring malaria endemicity from intense to interrupted transmission. Lancet Infect Dis. 2008;8:369–78.
14. Patil AP, Okiro EA, Gething PW, Guerra CA, Sharma SK, Snow RW, et al. Defining the relationship between *Plasmodium falciparum* parasite rate and clinical disease: statistical models for disease burden estimation. Malar J. 2009;8:186.
15. Mendis K, Sina BJ, Marchesini P, Carter R. The neglected burden of *Plasmodium vivax* malaria. Am J Trop Med Hyg. 2001;64:97–106.
16. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL, et al. Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. Lancet Infect Dis. 2009;9:555–66.
17. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. Am J Trop Med Hyg. 2007;77:79–87.

18. WHO. World Malaria Report 2014. Geneva, Switzerland: World Health Organization; 2014. p. 226.
19. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA, et al. A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. PLoS Negl Trop Dis. 2012;6, e1814.
20. Battle KE, Gething PW, Elyazar IR, Moyes CL, Sinka ME, Howes RE, et al. The global public health significance of *Plasmodium vivax*. Adv Parasitol. 2012;80:1–111.
21. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperely WH, et al. The international limits and population at risk of *Plasmodium vivax* transmission in 2009. PLoS Negl Trop Dis. 2010;4:e774.
22. Mueller I, Galinski MR, Tsuboi T, Arevalo-Herrera M, Collins WE, King CL. Natural acquisition of immunity to *Plasmodium vivax*: epidemiological observations and potential targets. Adv Parasitol. 2013;81:77–131.
23. Baird JK. Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. Clin Microbiol Rev. 2013;26:36–57.
24. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, et al. A new world malaria map: *Plasmodium falciparum* endemicity in 2010. Malar J. 2011;10:378.
25. Battle KE, Guerra CA, Golding N, Duda KA, Cameron E, Howes RE, et al. Global database of Plasmodium falciparum and P. vivax incidence records, 1985-2013. Scientific Data 2015. In press.
26. Lin E, Kiniboro B, Gray L, Dobbie S, Robinson L, Laumaea A, et al. Differential patterns of infection and disease with *P. falciparum* and *P. vivax* in young Papua New Guinean children. PLoS ONE. 2010;5:e9047.
27. Mueller I, Widmer S, Michel D, Maraga S, McNamara DT, Kiniboro B, et al. High sensitivity detection of *Plasmodium* species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea. Malar J. 2009;8:41.
28. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, et al. Geographical variation in *Plasmodium vivax* relapse. Malar J. 2014;13:144.
29. PubMed. www.ncbi.nlm.nih.gov/pubmed/.
30. Griffin JT, Ferguson NM, Ghani AC. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. Nat Commun. 2014;5:3136.
31. Microsoft. Encarta Reference Library. Seattle: Microsoft Corporation; 2007.
32. Google. https://www.google.co.uk/maps.
33. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, Noor AM, et al. Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. Malar J. 2007;6:17.
34. Malaria Atlas Project. www.map.ox.ac.uk/.
35. ESRI. ArcGIS Desktop 10.0. Redlands, CA: Environmental Systems Resource Institute; 2010.
36. Smith DL, Guerra CA, Snow RW, Hay SI. Standardizing estimates of the *Plasmodium falciparum* parasite rate. Malar J. 2007;6:131.
37. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
38. Golding N: ageStand R package. https://github.com/SEEG-Oxford/ageStand: GitHub; 2014.
39. Snow RW, Menon A, Greenwood BM. Measuring morbidity from malaria. Ann Trop Med Parasitol. 1989;83:321–3.
40. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: 3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria. 2003.
41. Schellenberg DM, Aponte JJ, Kahigwa EA, Mshinda H, Tanner M, Menendez C, et al. The incidence of clinical malaria detected by active case detection in children in Ifakara, southern Tanzania. Trans R Soc Trop Med Hyg. 2003;97:647–54.
42. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, Gething PW, et al. The global distribution of the Duffy blood group. Nat Commun. 2011;2:266.

Chapter 5 – Modelling national treatment-seeking rates to improve interpretation of malaria case reporting

To accurately represent the burden of *Plasmodium vivax* globally, cases were estimated using both the modelling approaches described in the previous chapter and routine case-reporting data. Surveillance reports submitted to the World Health Organization are adjusted for underreporting, observed slide positivity rates and the proportion of fevers that seek treatment. This chapter aimed to improve upon the treatment-seeking data used to generate case estimates reported in the World Malaria Report, while at the same providing insight into the varying care-seeking behaviours in *P. vivax* endemic countries. Increased spatial-resolution in the adjustment parameter estimates will ideally improve the accuracy of the case estimates produced.

This chapter has been submitted for publication and is shown here in its pre-submission format with figures integrated with the text. There were two Additional Files submitted with this manuscript that are shown in the Appendix. Additional File 1 represents supplementary figures and tables and Additional File 2 contains the input and output data. Although the text is small to accommodate the page size, it has been included to show the type of data that will be made available upon publication.

5.1. Abstract

Background

The proportion of individuals who seek treatment for fever is an important quantity in understanding access to and use of health systems as well as for interpreting data on disease incidence from routine surveillance systems. Passively-detected malaria case numbers reported by malaria-endemic countries (MECs) are generally adjusted by a treatment-seeking factor before being used in malaria burden estimation. For many MECs, treatment-seeking information is available from national surveys. The aim of this paper was to predict these treatment-seeking measures for all MECs lacking national survey data.

Methods

Data on treatment-seeking for fever were obtained from Demographic and Health Surveys (DHS) and Malaria Indicator Surveys (MIS) for every MEC and year that data were available. National-level social, economic and health-related variables were gathered from the World Bank as putative covariates of treatment-seeking rates. A generalized additive mixed model (GAMM) was used to estimate treatment-seeking behaviours for countries where DHS or MIS data were unavailable. Two separate models were developed to predict the proportion of fever cases that would seek treatment at a public health facility or from any kind of treatment provider.

Results

Treatment-seeking data were available for 58 MECs and modelled for the remaining 38. The GAMMs showed that the percentage of pregnant women receiving prenatal care, education level, government health expenditure and GDP growth were important

predictors for both treatment-seeking outcomes. Using these variables, the percentages of fever cases which sought public or treatment of any kind were derived.

Conclusions

Estimates of treatment-seeking behaviours are essential to correct reported malaria case numbers to obtain more accurate measures of disease burden. Using data available from national survey and straightforward modelling techniques, this can now be done for all endemic countries, and has the potential to provide additional insight into the health systems working to control and eliminate malaria.

5.2. Background

Estimates of the burden of malaria in different parts of the endemic world are essential to inform malaria control and elimination strategies [1-3]. Work of this nature is particularly timely as the progress of Millennium Development Goals is assessed and the post-2015 future is shaped [4]. There exist two primary approaches to estimating malaria burden, which is typically reported as the number of cases per unit population (e.g. 1,000 people) per year, also called the annual parasite incidence (API). One method, referred to as the surveillance-based approach, estimates cases from routine case reports from ministries of health (MoHs) and national malaria control programs (NMCPs). The second, known as the cartographic approach, models API using global mapped predictions of infection prevalence derived from parasite rate surveys [5-7], and a modelled relationship between infection prevalence and the incidence of clinical disease [8, 9]. Different parts of the malaria endemic world tend to be better suited to one method or the other depending on the quality, quantity and type of malaria data available. In many high-endemic settings, routine surveillance data are unreliable but

the availability of one or more cross-sectional parasite rate surveys facilitates cartographic burden estimation [3, 10]. Conversely, in low endemicity countries and those working toward elimination, it is less common to perform cross-sectional prevalence surveys, but routine reporting systems are often much stronger and can form the basis of reliable burden estimates [1, 3, 11-13].

Even in strong surveillance systems, however, reports of passively detected malaria cases will capture only a certain fraction of all malaria cases in a given country [2, 14-16] and so must be adjusted by a number of parameters before use in official burden estimates [1, 3]. Such adjustment formulae have been presented in detail elsewhere [1, 3, 17] but, briefly, reported cases are adjusted to account for (i) treatment-seeking behaviour (which leads to some cases not attending health facilities and thus being omitted from reports [18-25]); (ii) malaria diagnoses made without parasitological confirmation (which leads to reported case numbers containing non-malaria illnesses [26, 27]); and (iii) incomplete reporting (which leads to cases being lost from reported data). To estimate the upper and lower bounds of cases in a country in a given year (M), the following formulae have been applied [1]:

$$M_{upper} = (C+sU)/rp$$

$$M_{lower} = ((C + sU)(1-n))/(rp)$$

where C is the number of confirmed cases; U the number of unconfirmed cases, s to the slide positivity rate and r to reporting completeness. The treatment-seeking parameters, p and n, are determined from household survey questions regarding treatment of fever, the most common symptom of malaria in children [28, 29]. The variable p is the proportion of children with fever that sought treatment from a health facility within the public or government reporting system and n is the proportion of fever cases that did

not seek treatment of any kind. In this article, *1-n* will be referred to as the proportion of children with fever that sought treatment from within or outside the public sector.

The upper and lower bounds of adjusted case estimates are therefore determined by two treatment-seeking parameters. Treatment-seeking rates vary widely between countries, and so these parameters can greatly affect final burden estimates. Where available, these parameters are drawn from nationally-representative cross-sectional household surveys such as Demographic and Health Surveys (DHS) [30] and Malaria Indicator Surveys (MIS) [31]. However, not all MECs have such survey data available, meaning these important parameters must either be assumed or omitted – leading to considerable uncertainty in final burden estimates.

The aim of this study was to build a predictive model allowing treatment-seeking rates to be estimated for all MECs, including measurements of uncertainty, allowing improved understanding of routine surveillance data and more accurate burden estimation.

5.3. Methods

5.3.1. Data assembly

Many DHS and MIS surveys ask questions to determine the prevalence and treatment of fever in children under five years of age. It was assumed that the treatment-seeking rates observed in the children would be similar in older age groups, as supported by evidence from India, Indonesia and Ethiopia [17, 19]. For a handful of surveys, data on treatment-seeking for fever were not available, so data on treatment-seeking for acute respiratory infection (ARI) were used (n=4), which has previously been shown to correlate strongly with treatment for fever [17]. The survey codes for questions regarding where treatment was sought were reviewed from each survey containing this

type of data and categorized into public/government facilities likely to have been captured by reporting systems or "any" medical treatment, which included private or NGO facilities, but excluded non-medical categories of care such as homeopathic doctors or 'healers.'

From each survey containing data on fever (or ARI) treatment-seeking, the total number of children reported to have fever was summarized by cluster and region (or only one of the two, depending on the highest resolution of geographic information available from the survey). The numbers seeking treatment at either category of facility were also totalled to obtain proportion of those seeking public/government treatment or any treatment. These data extractions from the DHS website (www.dhsprogram.com) and summaries were all automated using FME, version 2015, by Safe Software [30, 32]. The total number of children, fever cases and cases that sought treatment were finally summarized nationally in order to generate predictions for those countries without survey data available.

5.3.2. Covariate data

We identified potential covariates to test for inclusion in a predictive model using a review of the literature for "treatment-seeking" and "care-seeking" for both "malaria" and "fever" in PubMed [33]. The following have been reported as determinants of care-seeking rates: household wealth [18, 20, 22, 23, 25, 34-40], care-giver education [18, 22, 34, 36, 41] and household location (rural or urban) or access to health facilities [18, 23-25, 34, 35, 42-47]. The World Bank provides freely available national-level indicator data [48] and several indicators that were in keeping with the themes identified by the literature were downloaded: access to electricity (as a proxy for wealth and access), gross domestic product (GDP; current US$), GDP per capita, GDP growth (annual %), gross national income (GNI) per capita (current US$), total health

expenditure (% of GDP), public health expenditure (% of total), primary education completion rate (% of relevant age group), and rural population (% of total population). There also is a covariate for the percentage of pregnant women receiving pre-natal care. This was included as it would likely be indicative of child treatment-seeking and health care access.

Matching covariate data to national-level survey data was also achieved using FME. Covariate data were used from the same year as the survey when available and were otherwise matched to the closest year. Covariate data were also matched to those countries without treatment data available. For these countries, covariate data was matched to 2013, which was the most recent year of indicator data available, or the year closest to that.

Figure 1. Treatment-seeking data in malaria endemic countries by WHO region. MECs with treatment-seeking data available are shaded yellow and those missing data are shown in white. Country borders are coloured based on WHO region: Central Africa (AFRO-C, red), East Africa and high-transmission areas in Southern Africa (AFRO-E, blue), low-transmission Southern African countries (AFRO-S, green) West Africa (AFRO-W, purple), Americas (PAHO, brown), Eastern Mediterranean (EMRO, orange), Europe (EURO, yellow), South-East Asia (SEARO, pink) and Western Pacific (WPRO, grey). Areas shaded grey outside the coloured borders have no malaria risk.

In addition to the social and economic covariate data, countries were grouped geographically based on their WHO regional offices: Region of the Americas (PAHO), Eastern Mediterranean Region (EMRO), European Region (EURO), South-East Asia Region (SEARO) and Western Pacific Region (WPRO). Countries in the African region were separated into the sub-African regions reported in the World Malaria Report: West Africa (AFRO-W), Central Africa (AFRO-C), East Africa and high-transmission areas in Southern Africa (AFRO-E), and low-transmission Southern African countries (AFRO-S) [3]. These regions then formed strata within the model, as explained below. A map of the MECs coloured by region is shown in Figure 1. Countries with treatment-seeking data available are shown in yellow and those missing data are shown in white.

5.3.3. Statistical modelling

Two generalized additive mixed models (GAMMs) were developed to predict the proportion of people that sought treatment from (i) a facility covered by the government reporting system (MOD1) and (ii) any medical care provider (MOD2) [49, 50]. Where multiple surveys were used from the same country, this was accounted for by adding a country-level random effect term. Both year and global region were included as fixed effect terms, along with the suite of country-level covariates. All statistical analyses were performed in the R statistical computing environment [51].

The Spearman's rank correlation test was used to investigate the collinearity among the covariates and several indicators were excluded. First, access to electricity, which was correlated with most other variables except for pregnant women receiving care and total health expenditure, was dropped. There was correlation among all of the economic variables, so all of the GDP and GNI variables except GDP growth were excluded, as the latter was the least correlated with the other variables. Literacy and primary

education completion rates were highly correlated and the latter was retained since it had been directly referenced in the literature search of care-seeking indicators.

Following this covariate reduction, the final set of covariates in the full model was as follows: year, region, GDP growth, health expenditure, pre-natal care rates, primary education, and proportion of rural population. For predicting treatment-seeking at government facilities, public health expenditure data was used, whereas for any treatment, total health expenditure data was applied. Otherwise, the initial covariate selections were the same between the two models. Formally, our GAMM was defined as:

$$y_{care} = \beta_0 + \beta_1 * f(Year) + \beta_2 * Region + \beta_3 * GDP\ Growth + \beta_4 * Health\ Expenditure + \beta_5 * Pregnant\ Women\ Care + \beta_6 * Primary\ Completion + \beta_7 * Rural + f_{rand}(Country)$$

Where y_{care} is the percentage of the country's population seeking health care assistance, $\beta_1 * f(Year)$ is the non-linear effect of the survey year, $\beta_2 * Region$ is effect of the country's WHO region, $\beta_3 * GDP\ Growth$ is the effect of the percentage annual growth of GDP, $\beta_4 * Health\ Expenditure$ is percentage of the total country GDP that was spent public health sector (for MOD1) or any health sector (for MOD2), $\beta_5 * Pregnant\ Women\ Care$ is the percentage of pregnant women who receive prenatal care, $\beta_6 * Primary\ Completion$ is the country intake ratio to the last grade of primary education (the number of individuals starting last grade of primary education regardless of age, divided by the population of the entrance age for the last grade of primary education), $\beta_7 * Rural$ is the percentage of the total population living in rural areas and $f_{rand}(Country)$ is the country random effect. MOD1 and MOD2 were performed adopting the Gaussian distribution as a link family.

5.3.4. Model selection

Model selection was performed using a multi-model selection approach [52, 53]. Candidate models with different combinations of the selected covariates were compared based on their model fit, assessed by the Akaike Information Criterion (AIC). The model with the lowest AIC was considered the best model and the other candidate models that had a difference in AIC (ΔAIC) <2 were also selected to be included in the list of best models successively used to obtain an average model [53, 54]. The Akaike weight (ω_i) from each model was used to assess relative variable importance [53, 54]. β_1 *s*Year* and β_2**Region* were included in every candidate model because both variables were considered essential to address the spatial-temporal component of the model. Predicted treatment-seeking outcomes for each country with missing survey data were then obtained from the average models.

5.3.5. Model validation

To assess the predictive accuracy of the average models, an out-of-sample model validation was performed. The data used to fit the candidate models were randomly split into 70% and 30%, to serve as the training and test data set respectively. The average models for both treatment-seeking outcomes were then fitted to the training set and used to predict the test data. The predicted and observed treatment-seeking values of the test data set were then used to calculate the root mean square error (RMSE) for each model.

5.3.6. Mapping treatment-seeking

Existing treatment-seeking data were mapped at the regional level (boundaries shown in Figure 1) for the 58 countries with available survey data. Although some of these data were available at a point level, only regional maps were generated for consistency with the data available for other adjustment parameters in the formulae described in the

Background. For the few countries with cluster data and no regional data available, cluster level data were summarized and mapped to the first- (Colombia) or second-level (Zambia) administrative units [55]. Predicted and observed values were then combined into geographically complete maps of treatment-seeking for all MECs.

5.4. Results

5.4.1. Data assembly

Treatment-seeking data were collected from DHS (n=183) and MIS (n=13) surveys for 60 countries. The number and year of surveys available from MECs are shown in Figure S1 in Additional File 1. Indicator data were available for all but four surveys (there was no health expenditure data available for Zimbabwe). The DHS from Mali in 1987 was dropped because it revealed that no individuals sought treatment, and was therefore deemed likely to be erroneous. The GAMMs were thus fitted to 191 records from 58 countries. This left 36 MECs for which predictions of treatment-seeking outcomes were required. However, Brazil and Paraguay were also added to the prediction list since Brazil only had national level data available (survey year 1996) and Paraguay had large areas of missing data (survey year 1990). This increased the number of prediction countries to 38.

5.4.2. Model results

Five models were chosen as the best candidate models for both treatment-seeking outcomes. The fitted coefficients, AIC, ΔAIC and ω_i values for each of the best models are shown in Table 1. Model average coefficient values for each geographic region are shown in Table S1 in Additional File 1. The parameter with greatest model influence was the percentage of pregnant women receiving pre-natal care. Though not statistically significant, education (primary completion) was the next most important variable,

followed by public health expenditure for those seeking treatment at public facilities and GDP growth for those seeking any treatment.

The mean predicted values for the proportion seeking treatment (government-based or other) in 33 of the 38 countries without survey data available are shown alongside the post-2010 observed treatment-seeking values in Figure 2. To show that the predictions are representative of the observed data, a subset of treatment-seeking proportions from countries with survey data from 2010 onwards was plotted along with the corresponding predicted values (Figure 2). For the most part, the observed values are within the confidence interval (CI) ranges of the predicted values. All observed and predicted values, as well as covariate data, are shown in Additional File 2. Model validation showed good prediction performance (Figures S2-S4). The RMSE for the percentage seeking government treatment was 7.8% and for any treatment it was 12%.

Five countries could not be predicted due to insufficient covariate data. These countries were assigned the predicted values of a neighbouring or, in the case of the Democratic People's Republic of Korea (DPRK), the nearest country that was also in the same WHO region. Therefore, Eritrea was given the same value as Ethiopia, French Guiana was given Suriname's values, the DPRK was matched to Myanmar, Republic of Korea to China and Somalia to Sudan.

5.4.3. Mapping treatment-seeking

The 33 predicted and five assigned national treatment-seeking estimates were combined with those countries that already had data available. Figure 3 shows maps for both the proportion seeking government treatment (Panel A) and those seeking any treatment (Panel B). In both figures, regions in red have the lowest proportions seeking treatment and blue have high treatment-seeking values. Public treatment-seeking outcomes were

highly heterogeneous with both blue and red areas in all WHO regions. Because treatment-seeking of any kind includes government treatment, it was inherently higher globally, with patches of low proportions in AFRO-C, SEARO and PAHO.

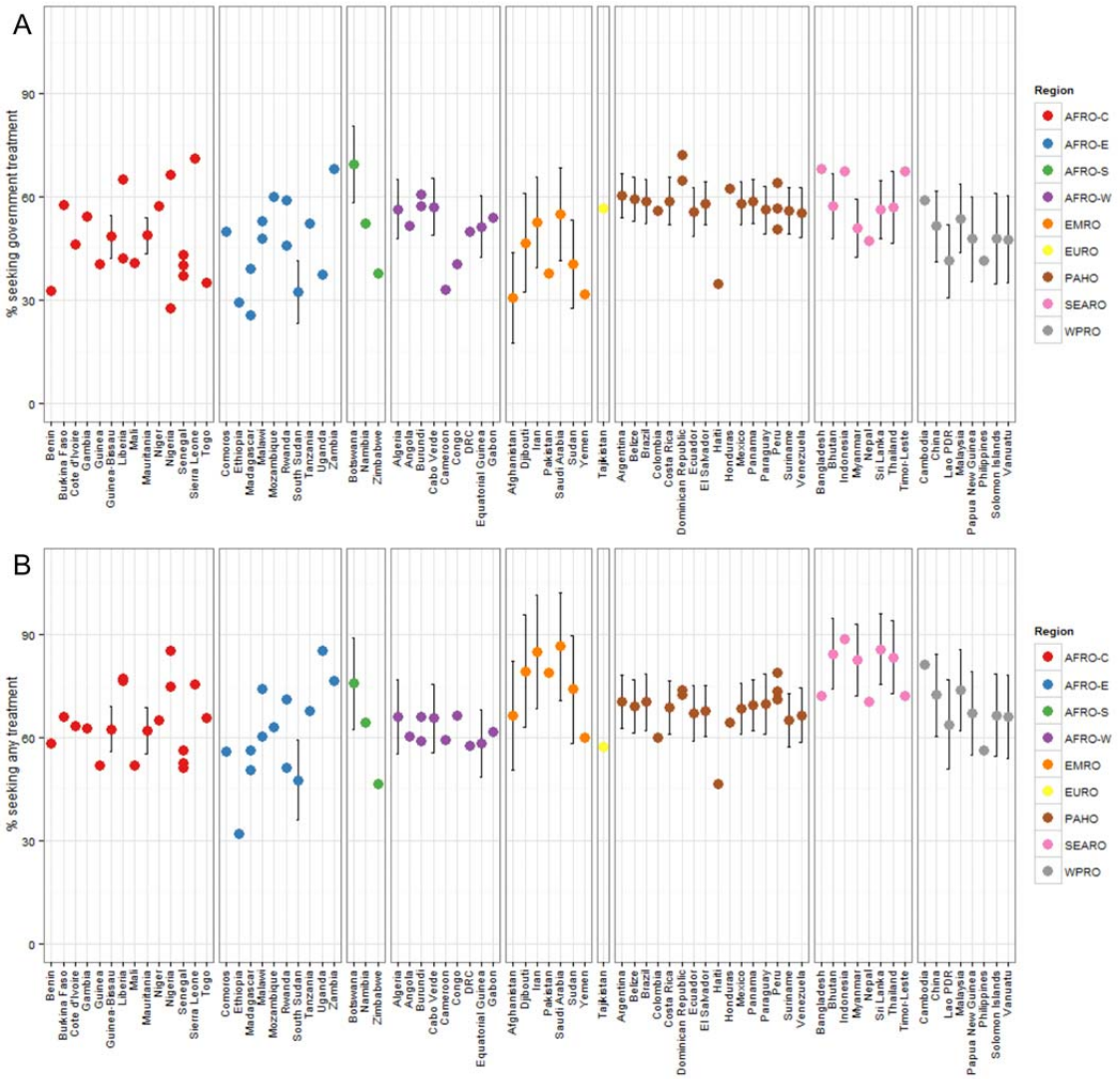


Figure 2. Predicted and observed treatment-seeking proportions. The predicted proportions estimated for 2013 are shown with 95% CI error bars alongside values observed from 2010 onward. The points are coloured by WHO region.

Table 1. Average generalized additive mixed model (GAMM) coefficients, 95% CIs, selection criteria and relative variable contributions. The intercepts and coefficients for each best candidate model included in the model average are shown below. The Aikake information criteria (AIC) and change in AIC (ΔAIC) used to assess the best models is also shown. Finally, the Akaike weights (ω_i) from each model are also shown.

Model	Intercept		GDP Growth		Health expenditure		Pregnant women care		Primary completion		Year		Region		AIC		ΔAIC		ω_i	
	Public	Any	Public	Any	Public	Any	Public	Any	Public	Any	Public	Any	Public	Any	Public	Any	Public	Any	Public	Any
1	3.04	24.83	—	0.34	—	—	0.27	0.24	0.12	0.14	+	+	+	+	-257.8	-227.5	0.00	0.00	0.29	0.27
2	-0.52	28.32	—	—	0.10	—	0.27	0.22	0.11	0.13	+	+	+	+	-257.6	-227.2	0.23	0.32	0.26	0.24
3	1.27	28.92	—	0.32	0.12	—	0.32	0.29	—	—	+	+	+	+	-256.9	-226.5	0.87	1.00	0.19	0.17
4	1.35	22.87	0.18	—	—	0.79	0.27	0.22	0.13	0.14	+	+	+	+	-256.2	-226.5	1.57	1.03	0.13	0.16
5	-2.37	20.31	0.49	0.32	0.10	0.69	0.28	0.24	0.12	0.15	+	+	+	+	-256.1	-226.4	1.72	1.11	0.12	0.16
Average	0.09	25.30	0.05	0.20	0.06	0.24	0.28	0.24	0.10	0.12	+	+	+	+						
Lower 95% CI	-10.38	11.87	-0.02	-0.08	-0.03	-0.45	0.14	0.08	-0.01	-0.01	+	+	+	+						
Upper 95% CI	13.06	40.60	0.57	0.76	0.24	2.02	0.42	0.40	0.25	0.29	+	+	+	+						
Relative variable importance ($\Sigma \omega_i$)			0.26	0.60	0.57	0.32	1.00	1.00	0.81	0.83	1.00	1.00	1.00	1.00						
N models			2	3	3	2	5	5	4	4	5	5	5	5						
p-value	0.89	0.001	0.51	0.39	0.20	0.63	<0.001	0.003	0.19	0.08										

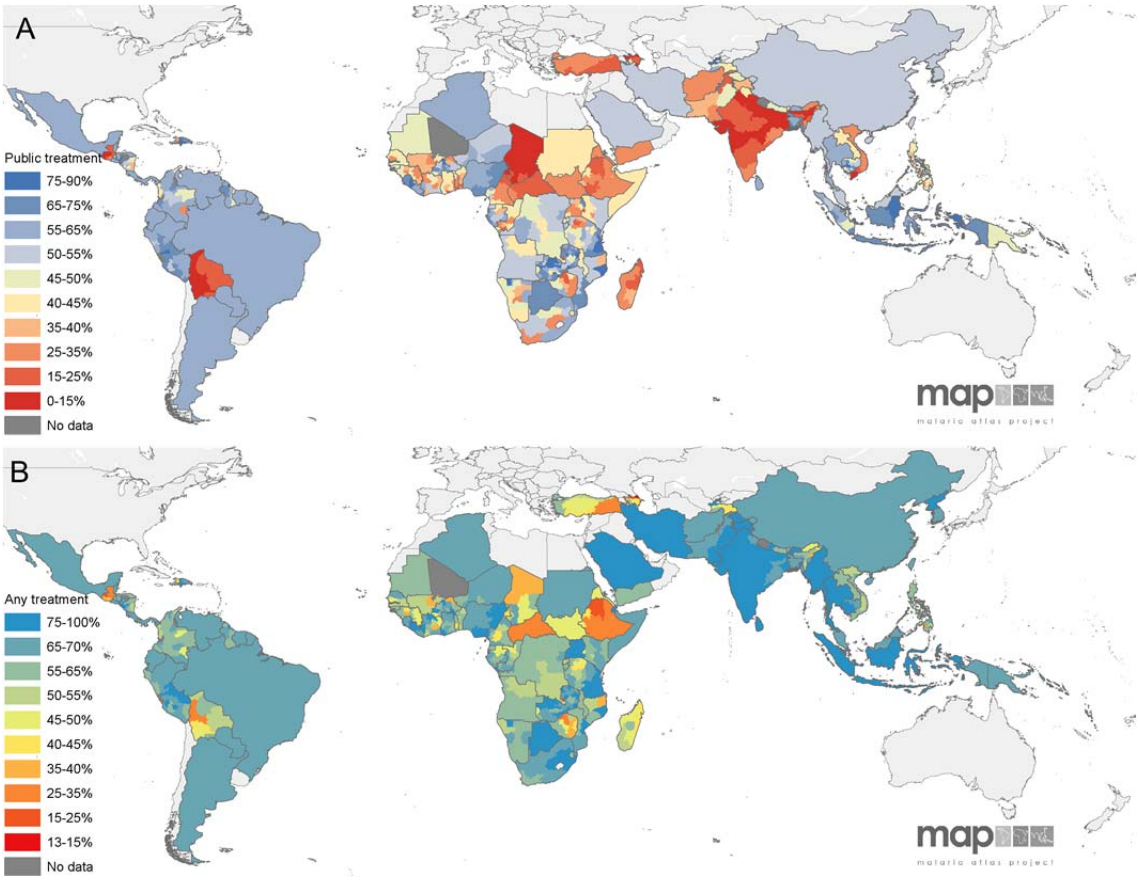


Figure 3. Observed and predicted treatment-seeking proportions. The observed treatment-seeking values in the regions shown in Figure 1 are mapped along with the national-level predicted values for the proportions seeking A) government/public treatment and B) any treatment.

5.5. Discussion

Information regarding the treatment-seeking behaviours of fever cases in MECs is essential to correctly interpret routine surveillance data from health facilities and obtain representative estimates of the number of cases due to *P. falciparum* and *P. vivax* malaria. Adjusted burden estimates rely primarily on data gathered from health facility reporting, including species-specific case data and slide positivity rates. Treatment-seeking data, which further adjust reported case numbers, however, cannot be collected from health facilities for the very reason that facility data only captures those that seek

treatment. Results are instead drawn from national level surveys of representative populations (i.e., not only people who are ill or already at the health facility). Survey data are increasingly available through DHS and MIS and are systematically curated to allow rapid processing through automation to find cluster, regional and national-level estimates of the proportion of children with fever that sought treatment at a facility that would be within the reporting system (public) as well as the proportion that sought treatment of any kind.

The upper and lower bounds of surveillance-based case estimates are differentiated by treatment-seeking variables [1, 17]. The higher case estimates are obtained by multiplying the adjusted cases by *1/p* (see formulae in Background). This correction scales up case estimates and implies an assumption that fever cases seeking treatment have the same slide positivity rate as those that did not. The lower estimate is obtained by multiplying the base case adjustment by *a/p* (*a* being the proportion of fever cases that sought any treatment). This scales the estimate down with the supposition that only those fevers that went for treatment were malaria cases. The upper estimates are likely to be more representative in areas where access to health facilities is poor and malaria cases are not able to seek treatment. Lower estimates are more characteristic of low endemicity regions where it can be assumed that nearly all cases seek and access treatment [1, 17].

Data obtained from national surveys (DHS and MIS) could be assembled for more than half of the MECs (60%, *n*=58). When national survey data are unavailable to estimate treatment-seeking behaviours for case estimates provided in the World Malaria Report, the WHO calculates the value from a regional average [3, 17]. There are likely to be substantial differences within regions, however, and the aim here was to refine that approach by instead modelling the treatment-seeking outcomes in countries without

national survey data to capture diversity within regions. Data that corresponded to the covariate variables identified in the review of the literature were readily available from the World Bank and produced models to show that government treatment-seeking and any treatment-seeking could be predicted at the national level from a limited set of covariate variables: year, WHO region, percent of pregnant women that receive pre-natal care, primary education completion rate, GDP growth, and national health expenditure (public and total). The percentage of women receiving pre-natal care was a strong indicator of fever treatment-seeking and likely resulted in the low model error values observed.

While the model fits were good, there were geographic patterns in both the outcomes and the certainty of the predictions (Figures 2 and 3). There were areas with low access or use of public treatment facilities in all regions of the malaria endemic world. Figure 3 highlights areas such as Central Africa and the Indian sub-continent. Accessing treatment of any kind was higher in all countries and these data revealed that treatment-seeking in some endemic areas, such as India, Pakistan and Afghanistan, was largely pursued outside the public sector. The confidence interval ranges (Figure 2) show that outcomes were well predicted in the Americas, but less so in Asia and least accurate in the Eastern Mediterranean countries. This implies that models and indicator variables were better suited to the Central and South American countries. Future predictions of this nature may be improved by additional covariates or, following further research into treatment-seeking indicators, different parameters for different regions.

Predictions and mapping treatment-seeking outcomes sub-nationally using cluster-level data to produce smooth surfaces like those produced by the Malaria Atlas Project for prevalence was also explored [5-7]. There were not sufficient covariate data available at smooth resolutions at the time of this analysis. The quantity and quality of higher

resolution sub-national covariate data that can be used in geostatistical analyses continue to improve with time and there may be greater potential for this type of analysis in the future [56].

5.6. Conclusion

Where data on treatment-seeking behaviours were not available from national surveys, relatively simple modelling techniques using freely available data were applied to fill data gaps**.** Both the results and methods presented have potential application beyond those described here and may inform the control and burden of other febrile diseases. In this context, data on treatment-seeking for fever are essential to malaria burden estimations. Gathering and visualizing these data for all MECs, sub-nationally when possible, is of use to estimate burden in areas of low endemicity where prevalence surveys are either not commonly performed or cannot detect sufficient levels of infection to generate modelled prevalence and burden estimates. This will enable a hybridized burden estimation approach that employs both surveillance-based and cartographic techniques in an effort to more accurately quantify the global burden of *P. falciparum* and *P. vivax* malarias.

5.7. Authors' contributions

KEB and PWG conceived the study and oversaw its implementation, with input from SIH and REH. KEB wrote the first draft of the manuscript and assembled treatment-seeking and indicator data with assistance from HG, DW, BM and UD. DB assisted the design of the modelling framework with input from SB and EC. All authors read and approved the final manuscript.

5.8. References

1. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C: **Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods**. *PLoS Med* 2011, **8**(12):e1001142.

2. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA *et al,*: **Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007**. *PLoS Med* 2010, **7**(6):e1000290.

3. WHO: **World Malaria Report 2014**. In. Geneva, Switzerland: World Health Organization; 2014: 226.

4. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA *et al,*: **Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013**. *Lancet* 2014, **384**(9947):1005-1070.

5. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U *et al,*: **The impact of malaria control on *Plasmodium falciparum* in Africa, 2000–2015**. *Nature* 2015, **Accepted**(MS ID: 2015-06-08556B).

6. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA *et al,*: **A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010**. *PLoS Negl Trop Dis* 2012, **6**(9):e1814.

7. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL *et al,*: **A new world malaria map: *Plasmodium falciparum* endemicity in 2010**. *Malar J* 2011, **10**:378.

8. Cameron E, Battle KE, Bhatt S, Weiss DJ, Bisanzio D, Mappin B *et al,*: **An emulator ensemble for modelling the relationship between *Plasmodium falciparum* parasite prevalence and clinical incidence**. *Nat Commun* 2015, **6**:8170.

9. Battle KE, Bhatt S, Cameron E, Guerra CA, Howes RE, Moyes CL *et al,*: **The global clinical burden of Plasmodium vivax**. *Nat Commun* 2015, **In prep.**

10. Hay SI, Smith DL, Snow RW: **Measuring malaria endemicity from intense to interrupted transmission**. *Lancet Infect Dis* 2008, **8**(6):369-378.

11. Ministry of Health Bhutan and World Health Organization and the University of California-San Francisco: **Eliminating Malaria: Case-study 9 | Climbing towards elimination in Sri Lanka**. In. Geneva: The World Health Organization; 2015.

12. Ministry of Health Sri Lanka and World Health Organization and the University of California-San Francisco: **Eliminating Malaria: Case-study 3 | Progress towards elimination in Sri Lanka**. In. Geneva: The World Health Organization; 2012.

13. Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, Smith DL *et al,*: **Operational strategies to achieve and maintain malaria elimination**. *Lancet* 2010, **376**(9752):1592-1603.

14. Gething PW, Noor AM, Gikandi PW, Ogara EA, Hay SI, Nixon MS *et al,*: **Improving imperfect data from health management information systems in Africa using space-time geostatistics**. *PLoS Med* 2006, **3**(6):e271.

15. Murray CJ, Lopez AD, Wibulpolprasert S: **Monitoring global health: time for new solutions**. *BMJ* 2004, **329**(7474):1096-1100.

16. Snow RW: **Sixty years trying to define the malaria burden in Africa: have we made any progress?** *BMC Med* 2014, **12**:227.

17. WHO: **World Malaria Report 2008**. In. Edited by WHO. Geneva, Switzerland: World Health Organization; 2008: 215.

18. Chaturvedi HK, Mahanta J, Pandey A: **Treatment-seeking for febrile illness in north-east India: an epidemiological study in the malaria endemic zone**. *Malar J* 2009, **8**:301.

19. Deressa W: **Treatment-seeking behaviour for febrile illness in an area of seasonal malaria transmission in rural Ethiopia**. *Malar J* 2007, **6**:49.

20. Deressa W, Ali A, Berhane Y: **Household and socioeconomic factors associated with childhood febrile illnesses and treatment seeking behaviour in an area of epidemic malaria in rural Ethiopia**. *Trans R Soc Trop Med Hyg* 2007, **101**(9):939-947.

21. Espino F, Manderson L: **Treatment seeking for malaria in Morong, Bataan, the Philippines**. *Soc Sci Med* 2000, **50**(9):1309-1316.

22. Fenny AP, Asante FA, Enemark U, Hansen KS: **Malaria care seeking behavior of individuals in Ghana under the NHIS: Are we back to the use of informal care?** *BMC Public Health* 2015, **15**:370.

23. Ilunga-Ilunga F, Leveque A, Ngongo LO, Laokri S, Dramaix M: **Treatment-seeking Paths in the Management of Severe Malaria in Children under 15 Years of Age Treated in Reference Hospitals of Kinshasa, Democratic Republic of Congo**. *Trop Med Health* 2015, **43**(1):11-19.

24. Thandar MM, Kyaw MP, Jimba M, Yasuoka J: **Caregivers' treatment-seeking behaviour for children under age five in malaria-endemic areas of rural Myanmar: a cross-sectional study**. *Malar J* 2015, **14**:1.

25. Xu JW, Xu QZ, Liu H, Zeng YR: **Malaria treatment-seeking behaviour and related factors of Wa ethnic minority in Myanmar: a cross-sectional study**. *Malar J* 2012, **11**:417.

26. Jensen TP, Bukirwa H, Njama-Meya D, Francis D, Kamya MR, Rosenthal PJ *et al,*: **Use of the slide positivity rate to estimate changes in malaria incidence in a cohort of Ugandan children**. *Malar J* 2009, **8**:213.

27. Petti CA, Polage CR, Quinn TC, Ronald AR, Sande MA: **Laboratory medicine in Africa: a barrier to effective health care**. *Clin Infect Dis* 2006, **42**(3):377-382.

28. Battle KE, Guerra CA, Golding N, Duda KA, Cameron E, Howes RE *et al,*: **Global database of *Plasmodium falciparum* and *P. vivax* incidence records, 1985-2013**. *Scientific Data* 2015, **2**(150012).

29. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM: **Malaria**. *Lancet* 2013.

30. **http://www.dhsprogram.com/** (Accessed 22 June 2015).

31. **http://www.malariasurveys.org/** (Accessed 22 June 2015).

32. Safe Software: **FME**. In.; 2015.

33. **www.ncbi.nlm.nih.gov/pubmed/** (Accessed 22 July 2015).

34. Hossain SM, Bhuiya A, Rasheed S: **Correlates of perceived malarial episodes and treatment-seeking behavior in a malaria-endemic rural area in Bangladesh**. *Southeast Asian J Trop Med Public Health* 2001, **32**(4):707-719.

35. Mustafa HS, Malik EM, Tuok HT, Mohamed AA, Julla AI, Bassili A: **Malaria preventive measures, health care seeking behaviour and malaria burden in different epidemiological settings in Sudan**. *Trop Med Int Health* 2009, **14**(12):1488-1495.

36. Nabyonga Orem J, Mugisha F, Okui AP, Musango L, Kirigia JM: **Health care seeking patterns and determinants of out-of-pocket expenditure for malaria for the children under-five in Uganda**. *Malar J* 2013, **12**:175.

37. Najnin N, Bennett CM, Luby SP: **Inequalities in care-seeking for febrile illness of under-five children in urban Dhaka, Bangladesh**. *J Health Popul Nutr* 2011, **29**(5):523-531.

38. Rutebemberwa E, Kallander K, Tomson G, Peterson S, Pariyo G: **Determinants of delay in care-seeking for febrile children in eastern Uganda**. *Trop Med Int Health* 2009, **14**(4):472-479.

39. Smith LA, Bruce J, Gueye L, Helou A, Diallo R, Gueye B *et al,*: **From fever to anti-malarial: the treatment-seeking process in rural Senegal**. *Malar J* 2010, **9**:333.

40. Sumba PO, Wong SL, Kanzaria HK, Johnson KA, John CC: **Malaria treatment-seeking behaviour and recovery from malaria in a highland area of Kenya**. *Malar J* 2008, **7**:245.

41. Klein EY, Lewis IA, Jung C, Llinas M, Levin SA: **Relationship between treatment-seeking behaviour and artemisinin drug quality in Ghana**. *Malar J* 2012, **11**:110.

42. Ewing VL, Lalloo DG, Phiri KS, Roca-Feltrer A, Mangham LJ, SanJoaquin MA: **Seasonal and geographic differences in treatment-seeking and household cost of febrile illness among children in Malawi**. *Malar J* 2011, **10**:32.

43. Holtz TH, Kachur SP, Marum LH, Mkandala C, Chizani N, Roberts JM *et al,*: **Care seeking behaviour and treatment of febrile illness in children aged less than five years: a household survey in Blantyre District, Malawi**. *Trans R Soc Trop Med Hyg* 2003, **97**(5):491-497.

44. Kassile T, Lokina R, Mujinja P, Mmbando BP: **Determinants of delay in care seeking among children under five with fever in Dodoma region, central Tanzania: a cross-sectional study**. *Malar J* 2014, **13**:348.

45. Matovu F, Nanyiti A, Rutebemberwa E: **Household health care-seeking costs: experiences from a randomized, controlled trial of community-based malaria and pneumonia treatment among under-fives in eastern Uganda**. *Malar J* 2014, **13**:222.

46. Muller O, Traore C, Becher H, Kouyate B: **Malaria morbidity, treatment-seeking behaviour, and mortality in a cohort of young children in rural Burkina Faso**. *Trop Med Int Health* 2003, **8**(4):290-296.

47. Tynan A, Atkinson JA, Toaliu H, Taleo G, Fitzgerald L, Whittaker M *et al,*: **Community participation for malaria elimination in Tafea Province, Vanuatu: part II. Social and cultural aspects of treatment-seeking behaviour**. *Malar J* 2011, **10**:204.

48. **http://data.worldbank.org/indicator/all** (Accessed 23 June 2015).

49. Wood SN: **Stable and efficient multiple smoothing parameter estimation for generalized additive models**. *J Am Stat Assoc* 2004, **99**(467):673-686.

50. Lin XH, Zhang DW: **Inference in generalized additive mixed models by using smoothing splines**. *J Roy Stat Soc B* 1999, **61**:381-400.

51. R Core Team: **R: A Language and Environment for Statistical Computing**. In. Vienna, Austria: R Foundation for Statistical Computing; 2013.

52. Bisanzio D, McMillan JR, Barreto JG, Blitvich BJ, Mead DG, O'Connor J *et al,*: **Evidence for West Nile virus spillover into the squirrel population in Atlanta, Georgia**. *Vector Borne Zoonot Dis* 2015, **15**(5):303-310.

53. Burnham KP, Anderson DR: **Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach**, Second edn. New York: Springer-Verlag; 2002.

54. Lund A, McMillan J, Kelly R, Jabbarzadeh S, Mead DG, Burkot TR *et al,*: **Long term impacts of combined sewer overflow remediation on water quality and population dynamics of Culex quinquefasciatus, the main urban West Nile virus vector in Atlanta, GA**. *Environ Res* 2014, **129**:20-26.

55. FAO: **The Global Administrative Unit Layers (GAUL): Technical Aspects**. In. Rome: Food and Agriculture Organization of the United Nations, EC-FAO Food Security Programme (ESTG); 2013.

56. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI *et al,*: **Re-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach**. *Malar J* 2015, **14**:68.

Chapter 6 – The global burden of *Plasmodium vivax* malaria in 2013

This culminating chapter of the thesis draws on the findings and outputs of the four preceding research chapters. Metrics described in Chapter 2, such as the population and risk and levels of endemicity, are revisited and updated here. The relapse zones and prevalence-incidence model of Chapters 3 and 4 are used to translate the predicted endemicity surface into a smooth map of the incidence of clinical disease. Treatment-seeking data from Chapter 5 are applied to reported case estimates which are used to calculate a surveillance-based clinical incidence measure. These two incidence rates are combined to produce the principal output of the thesis, a measure of the global burden of *Plasmodium vivax* malaria.

This chapter is in preparation for submission for publication. Three Supplementary Information Files accompany the manuscript and are shown in the Appendix.

6.1. Abstract

Plasmodium vivax occurs all across the endemic tropics and reaches into temperate climates. The infection causes acute attacks of malaria that may progress to severe and fatal illness. Clinical burden estimates, essential for treatment and control planning, have varied widely for *P. vivax* in the past. To develop an accurate measure of disease burden, two separate methods were employed which were adapted to the different types of data available from different regions. First, a surveillance-based estimate was derived from adjusted case reports. Second, a model-based estimate was generated from an updated map of *P. vivax* prevalence estimated by microscopic and rapid diagnostic mass blood surveys. Different countries were better suited to either method and a decision framework was designed to assign *P. vivax* endemic countries a surveillance- or model-based estimate. From this combined approach, between 8 and 19 million clinical cases were estimated to occur globally in 2013. These results highlight the regions with the greatest burdens as well as those where more complete surveillance and reporting would substantially improve the precision and confidence in such estimates.

6.2. Introduction

Over the past several years, *Plasmodium vivax* malaria has gained appreciation as a disease of great public health importance. It is the most geographically widespread of the human malarias [1, 2], affecting large, densely populated countries, and is associated with severe morbidity and mortality [3]. It can no longer be thought of as a 'benign' infection [4].

Assessment of the number of symptomatic cases that occur each year informs strategies for enhancing malaria control results and progress to elimination. Commonly referenced estimates of malaria disease burden measures [5] are not species-specific, and this hampers vivax-specific strategic planning [6]. There have been few efforts to estimate *P. vivax*-specific burden, and these have resulted in widely different estimates. Mendis and colleagues published an estimate of 72 to 80 million cases in 1999 based on case numbers and ratios of *P. vivax* to *P. falciparum* infections reported to World Health Organization (WHO) regional offices [7]. In 2007, Price, *et al.* estimated between 132 and 391 million cases based on non-falciparum malaria estimates generated by Hay, *et al.* [8, 9] using endemicity classes or levels of prevalence: (hypoendemic <10% prevalence of infection, mesoendemic 11–50%, hyperendemic 51–75% and holoendemic >75%)[10]. In 2013, the WHO began including *P. vivax* case estimates in the World Malaria Report (WMR). The WMR values are based on health-facility reports of passively-detected public-sector cases that are gathered by national agencies. These data are collated by the WHO and adjusted to account for under reporting, unconfirmed cases, and treatment seeking behaviours [11, 12]. The estimate published in 2014 was 11.9-22 million *P. vivax* clinical cases [13]. Determining where precisely the true burden of *P. vivax* malaria lies with respect to the wide range of these conflicting estimates (12 and 390 million) has been identified as a research priority [7, 9, 14].

Model-based, cartographic methods have been successfully applied to estimate *P. falciparum* malaria burden [15-17]. Applying the same methods across the entirety of the *P. vivax* world however is more problematic. Cartographic measures of disease burden are derived from prevalence estimates based on parasite rate (PR) surveys. *Plasmodium falciparum*, which is more widely studied than *P. vivax* [14, 18, 19], has

abundant PR data available from the published literature and more importantly, large-scale national surveys [20, 21]. Malaria parasite data from national surveys— typically Demographic Health Surveys (DHS) and Malaria Indicator Surveys (MIS)—are not as widely available from countries outside of Africa where *P. vivax* predominates. More important, *P. vivax* is poorly suited to detection in cross-sectional survey detection due to intrinsically low-density parasitaemia and hypnozoite forms in the liver. These result in relatively high frequencies of sub-patent and latent infections not detectable by standard field-based diagnostic techniques [22].

The current study incorporated both a surveillance- and cartographic-based estimation component. This dual approach exploited the strengths of both techniques and mitigated some of the weaknesses to deliver an optimized estimate of the global burden of illness attributable to infection by *P. vivax*.

6.3. Methods

6.3.1. Overview

The analysis presented is derived from a hybridized burden estimation framework. In countries where reporting appeared strong or where prevalence data were insufficient or non-representative, a surveillance-based estimate was applied. Where it was evident that reporting was non-representative or where coverage of PR data was high, a cartographic-based approach was applied. In the following sections we describe the application of this approach in a series of steps: (i) reported case data were used to generate a surveillance-based estimate of *P. vivax* cases; (ii) these estimates were applied to update the transmission limits of *P. vivax*; (iii) the predicted *P. vivax* infection prevalence values within the stable limits of transmission was updated; (iv) a

model relating infection prevalence and clinical incidence of *P. vivax* was used to calculate a cartographic-based estimate of *P. vivax* cases; and finally (v) a framework was developed to determine which countries were best suited to either method to create a hybridized map and global estimate of *P. vivax* incidence and burden of clinical disease.

6.3.2. Assembling API data

Annual parasite incidence (API) refers to routine surveillance data reporting the number of cases per year within a defined population under surveillance (for example within a sub-national administrative area). API data have been used previously by the Malaria Atlas Project (MAP) to define the geographical limits of transmission [1, 2, 23, 24]. The API data reported to MAP directly from national ministries of health in *P. vivax* malaria-endemic countries (*Pv*MECs) were averaged over the most recent five years of data to determine areas at no risk (no cases), unstable transmission (<0.1 cases per 1000 people per annum) or stable transmission (≥0.1 per 1000 p.a.). Here, *P. vivax*-specific incidence based on data reported to the WHO for the 2014 WMR was applied to define the limits of *P. vivax* transmission, as well as to calculate surveillance-based burden estimates.

The WHO makes adjustments to the number of cases reported from ministries or control programmes. Their formulae adjust the number of cases reported for under-reporting, for unconfirmed cases (using a *P. vivax*-specific slide positivity rate (SVR)), and for treatment-seeking behaviour. The upper and lower estimates are differentiated by the treatment-seeking behaviour data applied. To derive the upper estimate the number of cases is divided by the proportion estimated to seek treatment at a public facility to adjust for some cases not being captured by the government reporting system. The lower estimate is also multiplied by the proportion of fever cases that sought

treatment of any kind (public or private) to downscale cases on the assumption that fever cases that did not seek treatment were not caused by malaria (i.e., had an SVR of zero). The method for deriving the total number of cases and the adjustments applied here differed slightly from the WHO's methodologies [11, 12]. First, mixed infections were included here as *P. vivax* cases, and sub-national or modelled treatment-seeking data were used in place of national or regional average values. These adjustments are described in full in Supplementary File 1.

To calculate incidence rates, the adjusted surveillance-based case estimates were divided by the populations of the administrative units where the case data were available. Most population data were provided in the country case reports but, when not available, the population was summarized from a gridded population of the world (GPW) adjusted to the year 2013 using the 2010 and 2015 population grids [25].

6.3.3. Updating the Pv*PR database*

Endemicity or prevalence surfaces are predicted from local measures of the parasite rate (PR), obtained from cross-sectional surveys. The methodology used to develop and maintain the MAP database of such surveys is described in detail elsewhere [1, 8, 26] and summarized briefly in Supplementary File 2. Since the first iteration of the *P. vivax* endemicity map [1], the MAP database has grown substantially from the inclusion of national survey data. Although data from DHS and MIS have grown increasingly accessible [27] over recent years, these surveys rarely report *P. vivax* parasitaemia data. Outside of Africa, MIS are rarely undertaken and DHS do not gather data on malaria parasites. Surveys in African countries test for malaria, but often the malaria species is not specific and assumed to be *P. falciparum.* Only Uganda and the Democratic Republic of Congo (DRC) had DHS parasite data that reported test results for *P. vivax*. However, the DRC data was not included in this analysis because diagnosis was

performed using only polymerase chain reaction (PCR). PCR results were not included because the varying methods of PCR differ greatly in sensitivity [28, 29] and results could not be consistently reconciled with the microscopy and rapid diagnostic test (RDT) results already in the parasite rate database.

There were additional criteria for measures of *Pv*PR to be included in this analysis. Surveys must have been collected after 1 January 1985 and have been conducted in representative populations (i.e., not symptomatic patients or pregnant women only). The age and number of individuals sampled must have been reported, along with species-specific results, the diagnostic technique, and the start month and year of the survey. Finally, the survey location coordinates or site name must have been provided to allow reliable geopositioning [30-32]. In total, 12,782, spatially and temporally unique *Pv*PR observations were deemed valid for inclusion in prevalence mapping and modelling.

6.3.4. Modelling the P. vivax *parasite rate*

Building on the Bayesian model-based geostatistical (MBG) framework used previously [1], a latent Gaussian process model (LGM) was developed to represent the all-age *P. vivax* parasite rate ($PvPR_{1\text{-}99}$). Full details of the spatial model are given in Supplementary File 2. A minimum set of the covariates described by Weiss, *et al.* [33] were selected as candidate predictor variables for the mean *Pv*PR surface: day and night land surface temperature [33], *P. vivax* temperature suitability [34], enhanced vegetation index (EVI) [35], International Geosphere-Biosphere Programme (IGBP) land cover class 13 (urban and built-up areas) [36], and accessibility to cities with populations greater than 50,000 [37]. Given the discrete *Pv*PR observations and these environmental covariates, the LGM predicts a full Bayesian posterior distribution for the continuous surface of prevalence across the range of stable transmission (see below). For both model flexibility and computational efficiency the *P. vivax* endemic

world was divided into four geographic regions, which were fitted separately (see Figure S2.2 in Supplementary File 2): the Americas, Africa+ (Africa plus Yemen and Saudi Arabia), Asia (mainland Eurasia minus the Malaysian peninsula) and Asia-Pacific (the southern islands of Asia and the Malaysian peninsula). The mapped output represents the mean of the posterior distribution. The 95% credible intervals (CrIs) of the posterior distribution were also determined to quantify the uncertainty of the prediction of each mapped 5km x 5km pixel.

6.3.5. Defining the limits of transmission and population at risk

Using the reported case data described above, the limits of transmission were defined by classifying those areas with zero annual *P. vivax* incidence, or *Pv*API, as no risk, <0.1‰ cases as unstable transmission and ≥0.1‰ as stable transmission. In addition to those areas defined by reported data, areas that were unsuitable for malaria transmission based on temperature and aridity [24] and the *P. vivax* temperature suitability covariate [34] were further classed as unstable. Likewise, urban areas are less malarious than the surrounding rural environments due to the distinct ecological conditions presented by man-made environments [38-40]. Cities identified by the international travel and health guidelines [41, 42] and mapped using the Global Rural Urban Mapping Project (GRUMP) urban extents layer [43] were therefore also classed as malaria-free. The only exceptions were those cities within the range of *Anopheles stephensi* in Asia, which has been shown to breed efficiently in urban settings [44]. In areas where the predicted *Pv*PR_{1-99} was less than 1% *Pv*PR, it was not considered appropriate to describe these areas as stable transmission and these areas were reclassified as unstable transmission.

To estimate the global population at risk (PAR) of *P. vivax*, the proportion of the population predicted to be Duffy positive within the stable and unstable limits of

transmission was found. Duffy negative individuals, which are found in high frequencies in sub-Saharan Africa [45], are considered refractory to *P. vivax* infection because of the observed dependency of *P. vivax* on the Duffy antigen found on red blood cells for invasion [46]. The GPW 2013 surface, which provides the population in each 5km x 5km pixel of the world, was multiplied by the proportion of the population considered to be Duffy positive based on the analysis by Howes, *et al*. [45] shown in Figure 1. The resulting population surface was then summarized within the stable and unstable limits of the *Pv*MECs.

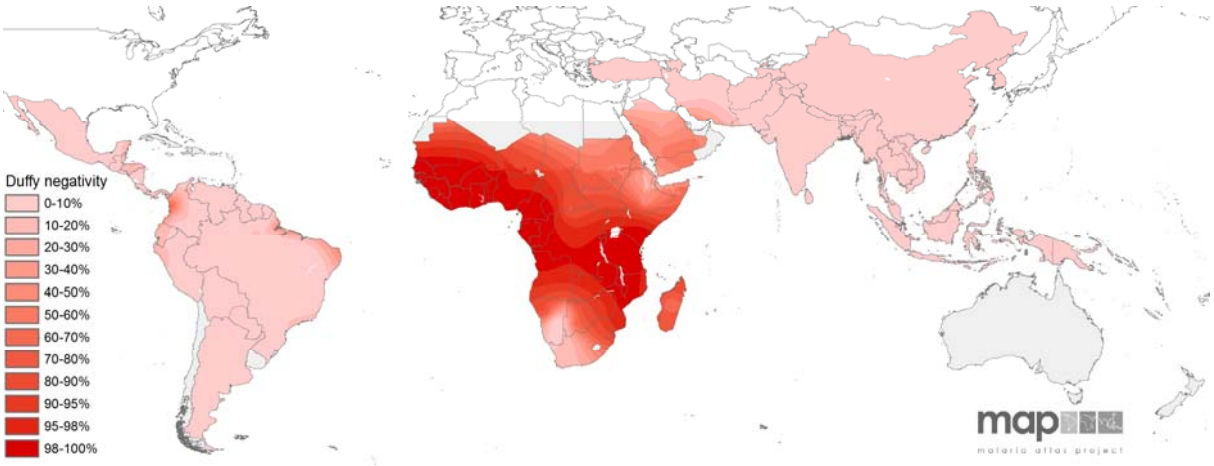


Figure 1. Global prevalence of the Duffy negativity phenotype. Reproduced from Howes, *et al*. [45].

6.3.6. Modelling P. vivax clinical incidence

Model-based estimates of clinical incidence were produced from the *Pv*PR surface using the functional relationship between *Pv*PR and clinical incidence developed by Battle, *et al.* [47]. Non-linear regression models were fitted via a hierarchical Bayesian model, and different models were fitted to the four global regions. These regions were stratified according to patterns of relapse rate (based on the previous analysis of average times from primary infection to first relapse [48]) because relapse patterns potentially

influence both incidence of disease and prevalence of blood-stage infection. The resulting models are illustrated in Figure S2.3 in Supplementary File 2.

Conditional simulations from both the *Pv*PR predictions and the prevalence-incidence model fits were used to generate robust estimates of uncertainty in the resulting incidence map and derived country and continental level case estimates. The GPW population surface (Figure S2.6) was used to calculate the number of cases from the mean and upper and lower CrIs of the predicted incidence rate. In areas of unstable transmission, the incidence was assumed to be between 0.01 and 0.1 cases per 1000. The mean of the predicted incidence in areas of stable transmission was used to generate national-level estimates of predicted burden.

6.3.7. Reconciling estimates of P. vivax *burden*

A classification system was developed based on the quality and quantity of both case reports and PR survey data available to determine which countries would be better represented by a burden estimate derived from surveillance reports or cartographic estimates derived from predicted PR. These methods are further explained in Supplementary File 3. In brief, a simple rubric was defined that allowed both the surveillance and PR data available from each country to be classified (in relative terms) as 'weak' or 'strong'. For PR data, this distinction was based on a threshold of available data above or below 90 data points per country (based on the minimum number of national survey records within a given country). For surveillance data, a 'strong' classification required all of the following: (i) case reports were available; (ii) case reports were *P. vivax* specific; (iii) *P. vivax* case reports were non-zero (since all countries included in this study were confirmed independently as being endemic for *P. vivax*).

For countries with strong PR data and weak surveillance data, only the cartographic-based approach was applied. For countries with strong surveillance data and weak PR data, only the surveillance-based approach was applied. For countries with both weak surveillance and PR data, the cartographic approach was applied with the rationale that this estimate, while based on a sub-optimal number of PR points, was at least able to provide a formal estimate of uncertainty around the prediction, and could draw strength from neighbouring country values and the influence of environmental covariates on *Pv*PR. For countries where both the PR and surveillance data were strong, the two estimates were combined to generate a joint incidence estimate based on the range of the upper and lower values of both the surveillance-based and cartographic-based incidence estimates. The median of the resulting distribution formed the point estimate and the 95% CrI the lower and upper bounds.

6.4. Results

6.4.1. Plasmodium vivax API

Plasmodium vivax annual parasite incidence (*Pv*API) was calculated from case reports submitted to the WHO for the 2014 WMR. Using this method, there were an estimated 10.5 (8.3 lower bound and 12.8 upper bound) million cases in 2013. Regional estimates, shown in Table 1, indicated that the largest burden was in Asia (6.66 million, 63%), followed by Africa+ (2.81 million, 27%), the Americas (0.54 million, 5%) and Asia-Pacific (0.50 million, 5%). Country-level estimates are provided in Supplementary File 1. The greatest clinical burden was found to be in India (4.47 million), Ethiopia (2.61 million), and Pakistan (1.36 million). These three countries comprised 80% of the estimated global burden alone. When the reported case estimates were divided by the population of the geographic unit they were reported by, estimates of the incidence rate

were produced. A map of *Pv*API is shown in Figure 2. Highest incidence values were observed in South America. It should be noted that this region also has some of the most complete reporting (refer to Figure S1.1 in Supplementary File 1).

Table 1. Regional clinical *Plasmodium vivax* burden estimates based surveillance-based, cartographic and combined methods.

	Surveillance (millions)			Cartographic (millions)			Combined (millions)		
Region	**Mean**	**Lower**	**Upper**	**Median**	**Lower**	**Upper**	**Point**	**Lower**	**Upper**
Africa+	2.81	1.66	3.96	2.14	1.31	3.30	2.33	1.62	3.26
Americas	0.54	0.43	0.66	3.06	0.93	10.75	0.70	0.44	1.58
Asia	6.66	5.75	7.58	15.69	8.10	38.67	6.75	5.34	9.61
Asia-Pacific	0.50	0.43	0.57	2.39	0.87	5.72	2.19	0.83	4.81
World	10.52	8.27	12.76	23.28	11.18	58.38	11.97	8.23	19.26

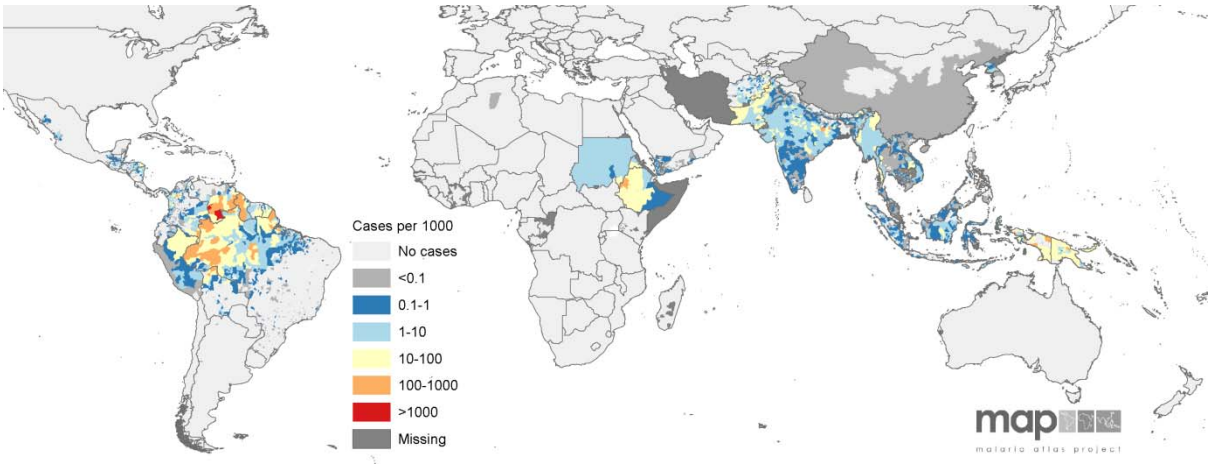


Figure 2. Annual *Plasmodium vivax* parasite incidence. Clinical incidence calculated from adjusted surveillance reports is shown on a spectrum of blue (low) to red (high). Those areas where data were missing are shown in dark grey, medium grey are areas considered to have unstable transmission (less than 1 case per 10,000), and lightest grey areas have no risk of *P. vivax*.

6.4.2. Populations at risk

API values were used to delineate areas at risk of *P. vivax* transmission. *Plasmodium vivax* was estimated to be endemic across some 31 million square kilometres globally.

The largest geographic area at risk of transmission was Africa+ (47%), followed by Asia (28%), the Americas (20%) and Asia-Pacific (6%). The largest area at risk of stable transmission was the Americas (5.1 million km^2, 51%), with Brazil contributing the largest country-level area of stable transmission (3.15 million km^2). Gridded population values (adjusted by the predicted Duffy positive prevalence) allowed the population at risk of each transmission class to be calculated. As shown in the regional summaries in Table 2, the largest populations at risk were in Asia, with 2.5 billion at any risk and nearly 700 million at risk of stable transmission. Globally, there were an estimated 2.9 billion people living at risk of *P. vivax* in 2013, with 2 billion living in areas where risk is low and case incidence is unlikely to exceed 1 case per 10,000 per year.

Table 2. Areas and populations at risk of *Plasmodium vivax* malaria in 2014.

	Area at risk (million km2)			Population at risk (millions)		
Region	**Unstable**	**Stable**	**Any risk**	**Unstable**	**Stable**	**Any risk**
Africa+	13.45	1.27	14.72	37.06	34.61	71.67
Americas	1.12	5.10	6.22	88.92	31.47	120.39
Asia	5.98	2.67	8.65	1,841.90	695.85	2,537.76
Asia-Pacific	0.82	0.94	1.76	110.49	29.98	140.847
World	21.37	9.97	31.34	2,078.37	791.91	2,870.29

6.4.3. Plasmodium vivax *endemicity*

Plasmodium vivax endemicity was predicted within areas of stable transmission. A smoothed map of the point estimates of the MBG outputs is shown in Figure 3. The $PvPR_{1\text{-}99}$ is shown on a scale of 1% to >7% (with <1% $PvPR_{1\text{-}99}$ being considered unstable transmission). The range of predicted values extended beyond 7% in some of the modelling regions. The maximum predicted $PvPR_{1\text{-}99}$ in Africa+ was 3.7% (in Ethiopia and Uganda) and 4.9% in the Americas (Brazil and Peru). The maximum estimate of $PvPR_{1\text{-}99}$ in Asia was 14.7% although there were only a few 5km x 5km

pixels in Myanmar and Cambodia where *Pv*PR$_{1\text{-}99}$ exceeded 10%. The largest range of predicted prevalence values was found in the Asia-Pacific, with a maximum prediction of 35.1% in Indonesia and Papua New Guinea (PNG). Again, only a small number of pixels were predicted to exceed 20% prevalence.

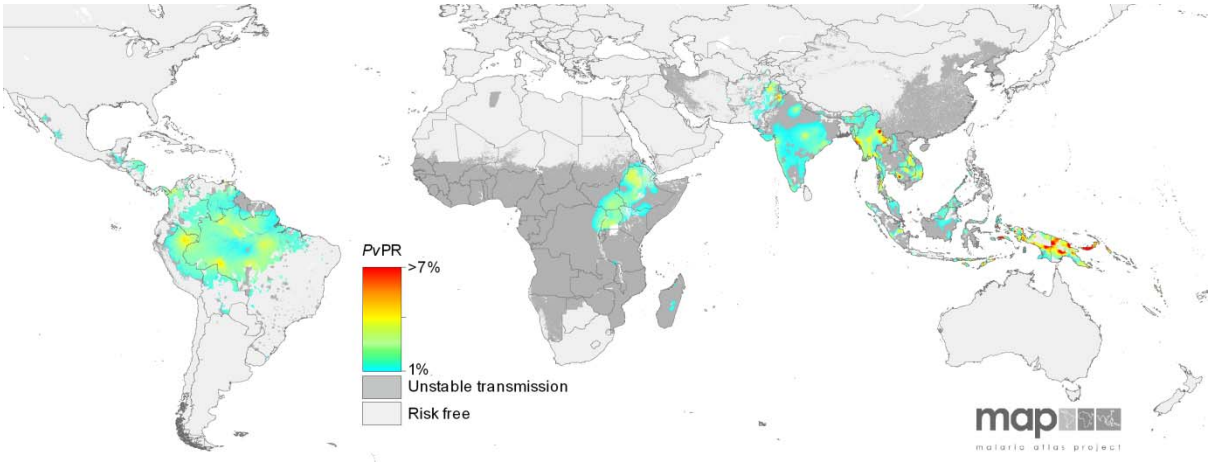


Figure 3. *Plasmodium vivax* in endemicity in 2013. Model-based geostatistical point estimates of the annual mean of *P. vivax* parasite rate in all ages (1-99) is show as a continuum of blue to red. Areas with a predicted PR <1% or API <0.1‰ are considered to have unstable transmission and are shaded dark grey. Light grey areas are free of *P. vivax* transmission.

Uncertainty was measured by the range of the 95% CrI of the *Pv*PR$_{1\text{-}99}$ posterior distribution obtained from conditional simulations (Figure 4). Predictions within Africa+ were the most certain, with greater uncertainty observed throughout the Americas, Asia and Asia-Pacific. The areas with the largest CrI range were in Asia-Pacific, particularly in parts of PNG where underlying PR observations in children ranged from 0 to 68%, resulting in a wide range of plausible *Pv*PR$_{1\text{-}99}$ values predicted by the model. Figure 4 also shows a population-weighted measure of uncertainty to highlight densely populated areas with greater uncertainty in the PR predictions. This map draws attention to India, Pakistan and Myanmar. Survey coverage in these countries was relatively sparse and while there were also few PR surveys available from

the Americas (see Figure S2.1 in Supplementary File 2), the endemic areas in that region are not heavily populated and there is therefore less operational impact resulting from uncertain estimates.

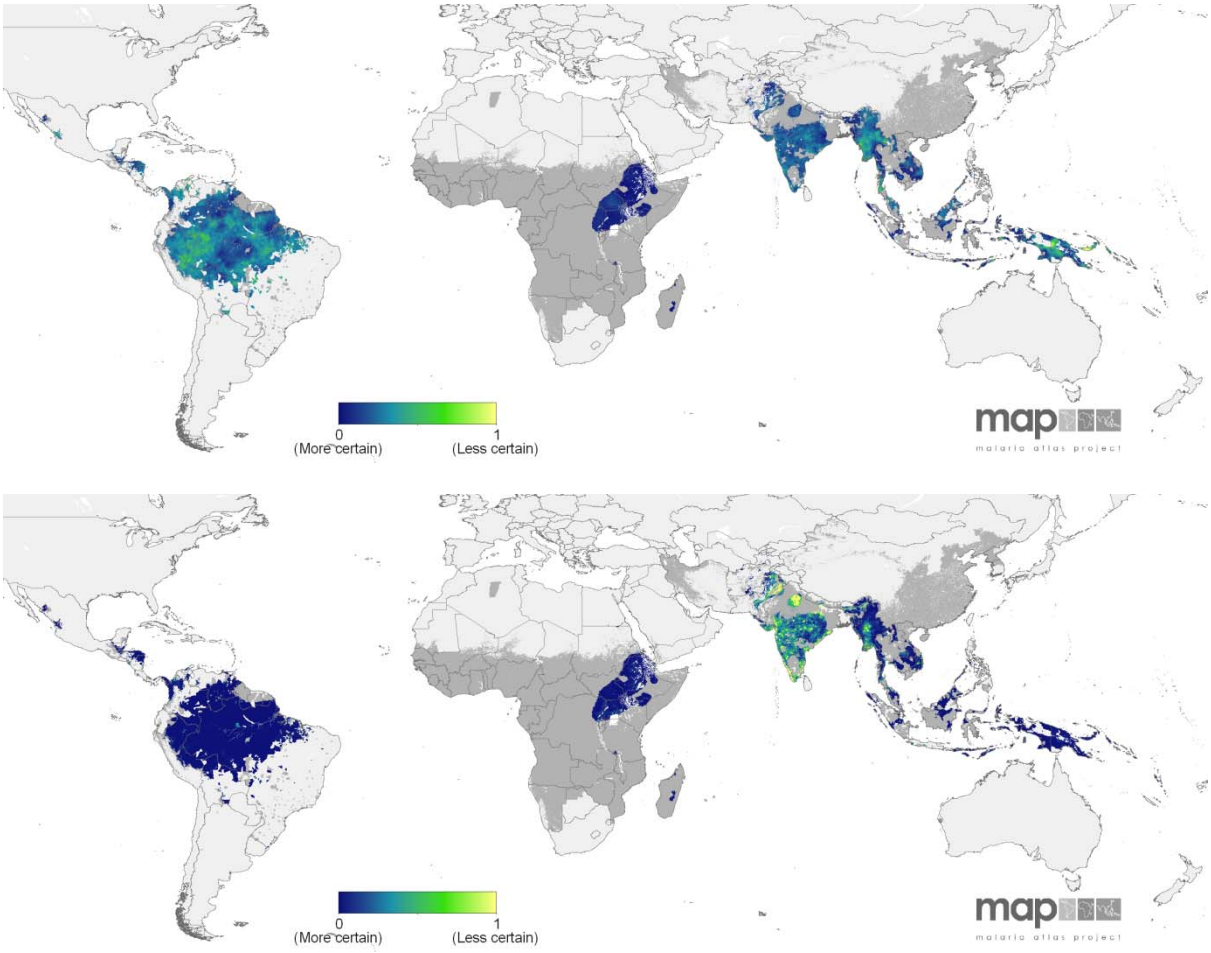


Figure 4. Uncertainty associated with *P. vivax* endemicity predictions. The top map shows uncertainty of the posterior 95% CrI range around the mean prediction at each pixel on a spectrum of blue (more certain) to yellow (less certain). Larger CrI ranges indicate a wider range of *Pv*$PR_{1\text{-}99}$ values as plausible for those pixels. The bottom maps shows the same index multiplied by the population of each pixel normalized to a scale of 0 to 1. Yellow areas show high uncertainty in large populations.

6.4.4. Modelled P. vivax *clinical incidence*

The predicted surface of *P. vivax* clinical incidence rate is shown in Figure 5. The associated uncertainty, represented again by the range of the 95% CrI of the incidence conditional simulation results, is shown in Figure S2.5 in Supplementary File 2. This

figure shows a high degree of uncertainty in the Americas and Asia-Pacific. As shown in Figure S2.3, the zone which includes India has the narrowest credible intervals and that region, as well as Africa, which was based on the pooled fit of the prevalence-incidence model (see Figure S2.3 in Supplementary File 2), had the narrowest 95% CrI ranges for the mean incidence fits.

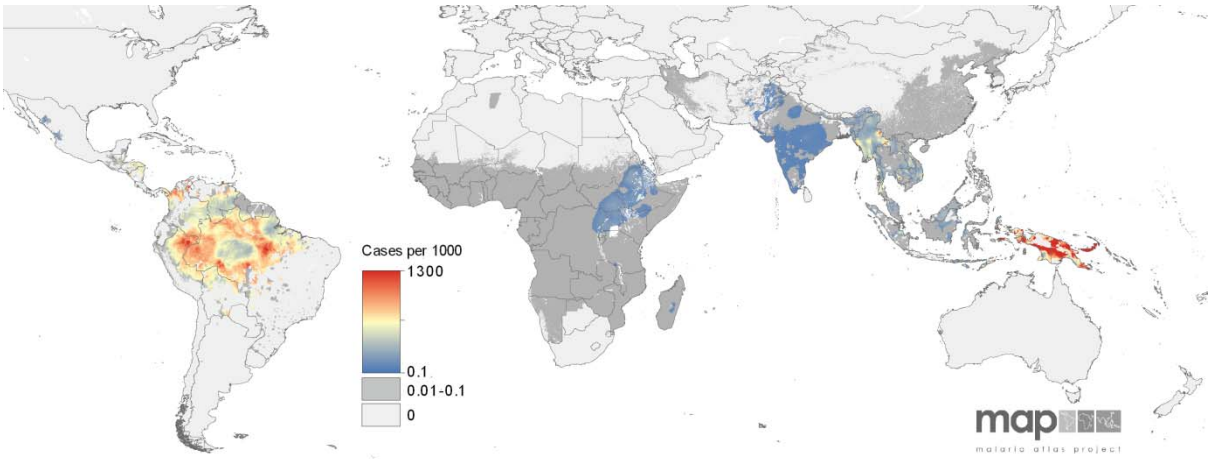


Figure 5. Modelled clinical incidence of *P. vivax*. The mean geostatistical estimates of incidence are shown from blue (0.1 cases per 1,000 per annum) to red (1,300 cases per 1,000 p.a.). Dark grey areas have low risk (0.01-0.1 cases per 1,000 p.a.) and light grey areas have no risk.

The modelled incidence output and gridded population surface were used to generate cartographic-based estimates of burden, with associated uncertainty shown by the 95% CrI bounds. Regional estimates are shown in Table 1 and country-level values are in Supplementary File 2. The same general pattern was observed here as with the cartographic estimates. The greatest burden was predicted in Asia (15.69 million, 67%). Based on the median estimates, the countries with the greatest burden were India (9.71 million, 42% of the global burden), Myanmar (2.49 million, 11%), Ethiopia (1.87 million, 8%) and Pakistan (1.65 million, 7%).

6.4.5. *Combining estimates of burden*

Following the categorization of countries based on data quality and availability, the modelled cartographic estimate was not considered representative for all countries. There were 44 countries that were considered to have both weak case reports and PR data, the majority of which were in sub-Saharan Africa (see Table S3.1 in Supplementary File 3). Nine countries had weak surveillance data, but had good PR data coverage through national surveys or large-scale studies. Twenty-nine countries had more representative reported case data. These were primarily countries in the Americas, with several Asian countries, and only Algeria from Africa. Finally, 11 countries had both detailed case reporting data and strong PR data coverage, and these were assigned the joint case estimate. A map that combines all three incidence measure categories is shown in Figure 6. The resulting case map derived from multiplying the combined incidence by the population grid (Figure S2.6) is shown in Figure 7.

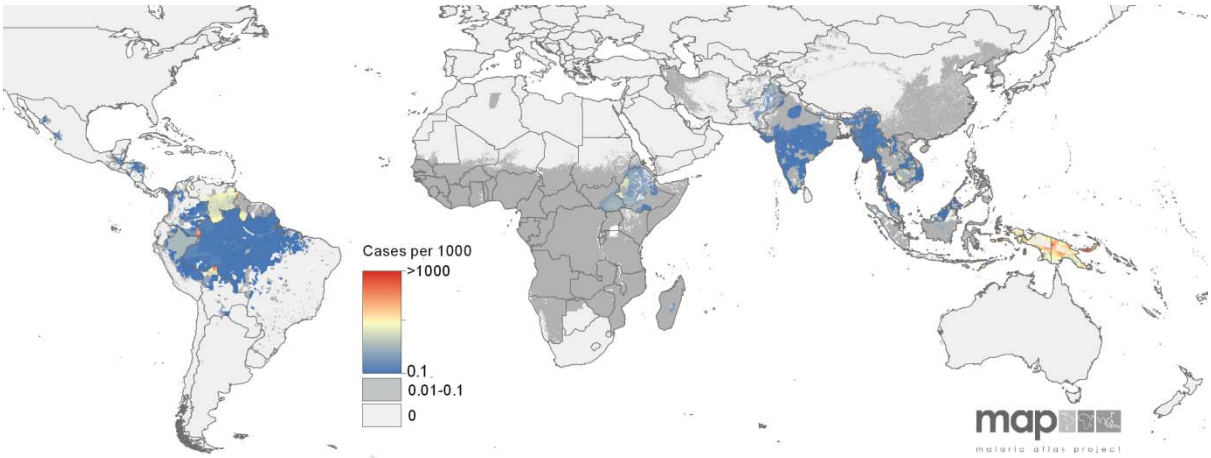


Figure 6. Combined surveillance-based and modelled incidence rates. Incidence rates estimated for countries with weak surveillance are shown as the modelled incidence value and those with strong reporting reflecting the adjusted reported annual *P. vivax* incidence. Those countries with a strong PR evidence base and surveillance are shown as a combined incidence rate.

6.4.6. *The global burden of* P. vivax *malaria*

After combining the model-based estimates of burden, the adjusted case reports, and joint estimates, the global burden of *P. vivax* was estimated to be 11.97 million cases per year, with a range of 8.23-19.26 million cases. Country-level estimates using the combined approach are shown in Supplementary File 3. The largest burdens were again observed in India, Ethiopia and Pakistan with 4.56, 2.06, and 1.65 million cases annually, respectively. According to this estimate, 69% of the global burden is found in these countries. However, Indonesia and PNG were also found to have large burdens: 1.26 and 0.833 million, respectively. Continentally, the largest burden was observed in Asia with 6.74 (5.34-9.61) million, followed by Africa+ with 2.33 (1.62-3.26) million. Asia-Pacific, though the smallest group of countries in this analysis, had 2.19 (0.83-4.81) million cases, while the Americas had a point estimate of less than 1 million: 0.70 (0.44-1.58) million.

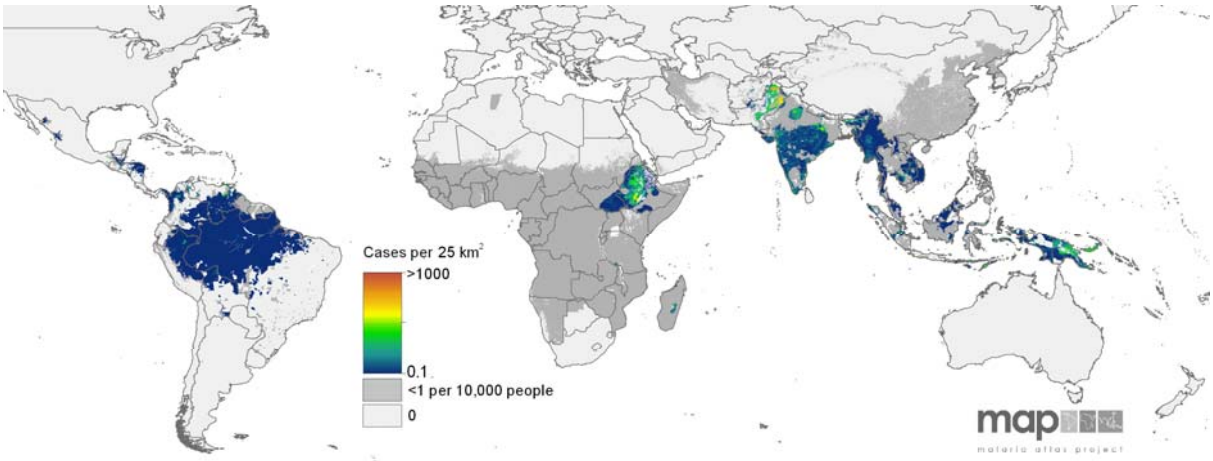


Figure 7. *Plasmodium vivax* cases based on a combined incidence rate. The incidence surface in Figure 5 and a grid of human population was used to produce a map of the number of cases per 5km x 5km pixel globally.

6.5. Discussion

Plasmodium vivax malaria is a disease of major public health importance. It is widespread, difficult to control, challenging to treat and manage, and associated with

significant morbidity and mortality. Updated transmission limits drawn from *P. vivax*-specific adjusted surveillance data show that 2.9 billion people, nearly one third the world's population, live at some risk of infection and almost 1 billion people live in higher-risk stable transmission areas. Within the geographical limits of stable transmission, prevalence of infection was predicted to be universally low with prevalence rarely exceeding 3% in Africa, 4% in the Americas and 7% in Asia and with a few small pockets of very high endemicity where prevalence exceeds 20%.

This study presents a novel approach to quantifying disease burden using both surveillance-based and cartographic-based techniques. The two methods yielded annual estimates ranging between 8.23 (lower limit combined method) and 58.38 million (upper limit cartographic method) *P. vivax* cases globally. Given the diversity in the epidemiology of the disease and structure of the health systems in *P. vivax* endemic countries, different countries were considered to be better suited to either method. This resulted in a hybridized approach that, based on various forms of uncertainty in the input data, estimated between 8.23 and 19.62 million cases globally in 2013.

6.5.1. Updated transmission limits and endemicity

The updated limits of transmission showed the geographic range of *P. vivax* transmission to be 31 million square kilometres compared to an estimated 44 million km^2 in 2010 [1]. This decrease is likely attributable to a genuine shrinking of the malaria map rather than the methodological differences in the derivation of the estimates. Significant progress has been made towards malaria elimination with Georgia, Kyrgyzstan and Uzbekistan are no longer considered endemic, and Azerbaijan, Sri Lanka and Turkey were observed and predicted to have no cases in 2013 (though they did in 2010). However, previous MAP limits were defined using API values reported directly from ministries of health and were not adjusted. The *Pv*API

applied in this analysis was calculated from case estimates reported to the WHO that were adjusted for reporting completeness, treatment seeking, and *P. vivax* slide positivity rate. Adjusting cases in this way would have increased the *Pv*API and therefore increased the areas of stable transmission, but would not have affected areas that were not at risk.

While the area at risk shrank, there was an increase in the population at risk between 2010 and 2013 (up to nearly 3 billion from 2.5 billion). *Plasmodium vivax* is endemic in the world's most populous countries: China, India, Indonesia, Brazil, and Pakistan. China was the only country to have only low risk unstable areas of transmission. Brazil had the largest estimated area of stable transmission areas, but it was found in the less-populated Amazonian region of the country. The regions of stable transmission in Asia and Asia-Pacific represent a significant challenge to the long-term ambitions for elimination due to large population movement and connectivity within these regions[49].

Patterns in endemicity have also changed since the 2010 iteration of the map, though these differences between the maps are confounded by changes in methodologies. Improvements were made to the modelling approach, including a more optimised suite of environmental covariates. Importantly, substantial additional *Pv*PR data (approximately 1,500 surveys) were used to generate the maps. The new map reveals universally low endemicity globally, as mentioned above. There are few places where prevalence exceeded 5%, none of which were outside of Asia. Hotspots of transmission were identified in Myanmar, Indonesia, and PNG.

These relatively low prevalence rates, compared to *P. falciparum* [2, 15], must be interpreted with caution. *Plasmodium vivax* circulates in the blood at lower parasite

densities than *P. falciparum*. Ideally, surveys using molecular techniques capable of detecting low and very low parasite densities, such as quantitative PCR (qPCR) and high-volume qPCR [28], would have been incorporated in the input dataset. However, none of these diagnostic methods are capable of detecting the hypnozoite reservoir. While sensitivity varies among the principle diagnostic techniques (microscopy, RDT or PCR), there is also variation observed within techniques due to training and quality of the microscopist [50] or the protocols applied to molecular methods [51]. Long-term, improved coverage with molecular diagnoses and standardization of methodologies to increase reproducibility, as has been done for microscopy [52], will help address these issues. An immediate goal should be to reconcile observed differences in diagnostic results for *P. vivax*, as previously done for *P. falciparum* [53, 54].

Microscopy and RDT diagnoses underestimate the true prevalence of blood-stage *P. vivax* across a range of transmission settings [55-59]. Studies comparing patent PR based on microscopy and more sensitive molecular techniques (PCR) show that microscopy fails to detect a large proportion of *P. vivax* infected individuals [60-62]. Work done on *P. falciparum* showed that underestimates in prevalence based on microscopy were greatest in regions of low endemicity, with a PCR to microscopy ratio of 5 to 1 where PCR prevalence was less than 10% (as compared to 2 to 1 when prevalence was between 25% and 75%) [54]. Low parasite densities indicate that these ratios would be even greater for *P. vivax*. Based on this logic, infection prevalence may be double or as much as five times greater than the patent microscopic PR values mapped here. Though, this is likely to be even greater because intuitively, the proportion of total infections that are sub-patent or latent is greatest in low endemicity, which is predominant in the *P. vivax* endemic world.

Plasmodium vivax prevalence may be further underestimated because *P. falciparum* is perceived as the more dangerous pathogen. In spite of lower blood-stage densities, *P. vivax* causes equivalent red blood cell reduction to *P. falciparum* [63], resulting in severe anaemia that is compounded by repeated bouts of infection due to relapses [4]. Regardless, there exists a strong reporting bias of mixed *P. vivax* and *P. falciparum* infections as *P. falciparum* mono-infections [64], potentially missing a large *P. vivax* infection reservoir.

The spatial variations in uncertainty of the model-based *Pv*PR$_{1\text{-}99}$ estimates reflect heterogeneities in both the coverage of available survey data and in the observed values themselves. *Plasmodium vivax* transmission can be highly focal [19], and varying PR observations (even within the same year and season) in similar environmental settings caused the model to produce a wide range of outputs in some areas (corresponding to high uncertainty values). Similarly, when the model was fitted to a limited amount of data, as in the Americas, the model was forced to make inferences across large areas with cross-sectional samples that are not necessarily representative of the area as whole. In contrast, countries with large national-level PR surveys available, such as Ethiopia and Uganda, were predicted with low uncertainty.

6.5.2. Plasmodium vivax *in Africa*

The prevalence and limits of stable *P. vivax* transmission in Africa are one of the most notable changes from the 2010 *Pv*PR$_{1\text{-}99}$. As seen in Figure 3, the 2013 iteration shows stable endemicity in Africa+ extended beyond where *P. vivax* is known to be present in the Horn of Africa (a population with higher proportions of Duffy positive populations, known to be susceptible to *P. vivax* infection [45, 65]). While *P. falciparum* is irrefutably the predominant parasite in Africa, a recent review of a variety of evidence including traveller infections, community prevalence surveys, local clinical case

reports, and entomological and serological studies showed an unambiguous presence of *P. vivax* in sub-Saharan Africa [65]. Analysis revealed that the low *Pv*PR observed in Africa was plausible with transmission being sustained exclusively by Duffy positive individuals. Reports of *P. vivax* infections in Duffy negative populations in Angola, Cameroon, Equatorial Guinea, Ethiopia, Madagascar, and Mauritania indicate that the protective efficacy of Duffy negativity may not be as definitive as previously thought [65]. However, there is clearly some limiting effect of the high prevalence of the Duffy negative phenotype in this part of the world because, given the climate and vector competence of the region [66], prevalence values similar to those in Asia-Pacific would otherwise be expected. Therefore, the population at risk was adjusted by a Duffy positive population to reflect the susceptible population globally.

The evidence of stable *P. vivax* transmission occurring in some sub-Saharan countries remains weak at this time [65], hence the *Pv*PR values (around 4%) obtained from the microscopy diagnoses from the national survey in Uganda [67] were questioned. The molecular diagnosis of *P. vivax* by PCR in the Democratic Republic of Congo [68] (DRC, not included in this analysis) support the possibility that there is, in fact, *P. vivax* in Uganda, though the rates reported are unlikely to be representative. As a result, for the cartographic and combined burden estimates, stable areas in the DRC, Kenya and Uganda were downgraded to unstable. The small, stable area of transmission for Madagascar was retained as there is conclusive evidence [65] of *P. vivax* transmission in this country where populations with Asian genetic decent result in lower prevalence of Duffy negativity [69]. Indeed transmission within Duffy-negative hosts has also been observed there [70].

6.5.3. Reconciling estimates of P. vivax *burden*

Determining which estimates of *P. vivax* burden were most representative for each of the 93 *P. vivax* endemic countries was an important aim of this analysis, as the type and quality of data available from each country differs (Supplementary File 3). Case estimates derived from adjusted *Pv*API reports were better suited for countries where *P. vivax* cases were both reported and confirmed. While the same formulae that the WHO uses to estimate surveillance-based burden were used, there were differences in some of the methods and data that the WHO employed for the case estimates reported in the WMR [13]. One of the most notable differences was in the use of data on treatment-seeking for fever. The WHO has previously applied national estimates obtained from household surveys and, where data are missing, a regional average is applied. The treatment-seeking data applied here were also derived from household surveys, but were inputted sub-nationally where the data were available, and missing data values were modelled based on social and economic indicator variables [71]. The WHO assumes near-complete treatment seeking for fever in countries nearing elimination because it is reasonable to presume that any malaria fever will seek care. In areas of Africa highly endemic for *P. falciparum*, fever is deemed an acceptable proxy for malaria because malaria is the primary cause of fever in many of these countries [72, 73], and thus the WHO estimates reflect a national value based on household surveys. In more developed countries, and as transmission decreases, the most common cause of fever will be other, less life-threatening causes. Nevertheless, the sub-national observed and national-level modelled values based on household survey data were not substituted [71]. This potentially over-estimated the reported case burden in countries nearing elimination. The use of modelled data and sub-national values likely resulted in an underestimated burden in more highly endemic areas, as the surveillance-based estimate

reported here is lower than that provided in the 2014 WMR. The use of a *P. vivax* specific SPR would have further contributed to the lower resulting estimate.

The application of modelled case estimates in countries where there was incomplete or missing case report data was straightforward. The challenge was in countries having both strong PR and case data. The decision to combine the estimates with a statistical procedure effectively weighting for their respective uncertainties was considered the best approach. Estimates may be refined in the future through developing a more detailed data scoring system.

It is also worth noting that the two types of burden estimates reported here essentially represent different types of burden. Surveillance-based estimates are made from case reports of people presenting at healthcare facilities with *P. vivax*. For the individual to seek care, it is likely to be an acute infection. The cartographic estimate is modelled on data obtained from active case detection (ACD) where households are visited and any individual with a fever is tested. Therefore, this estimate represents a broader-spectrum clinical definition. Both are relevant and important in generating estimates of both personal and societal costs.

6.5.4. The global burden of P. vivax

An accurate assessment of the global clinical burden of *P. vivax* malaria has been identified as a key knowledge gap. In order to benefit from the strengths of the two principal methods of burden estimation that use either adjusted reported case data or values modelled from predicted prevalence, a hybridized approach was applied. The resulting global burden of disease was between 8 and 19 million cases, with 0.4 to 1.6 million cases in the Americas, 2.0 to 3.0 in Africa+, 5.3 to 9.6 in Asia and 0.8 to 4.8 million in Asia-Pacific. While these estimates are much lower than those generated for

P. falciparum, it is important to consider both the epidemiology of the disease and methodology of burden estimation. Control measures are undoubtedly proving effective in curbing the burden of malaria in many regions[15]. The fact that our predicted estimates are in line with values presented in the 2014 WMR adds validity to surveillance-based estimates in the past. However, it is not possible to make comparisons between these values and previous estimates such as the 71-80 million cases reported by Mendis, *et al*. [7] from the 1990s. The methodologies have largely changed over the years and it is highly unlikely that a nearly 10-fold decline in *P. vivax* burden has occurred in the last 20 years, when little to no resources were directed at the species specifically [9, 14, 74], until very recently [75, 76]. Clinical incidence of *P. falciparum* was observed to have fallen over 40% over the last 15 years, but this was evidenced by the impact of large-scale targeted intervention efforts [15] for a species that is easier to control relative to *P. vivax* [77, 78]. Significant progress must still be made in order to quantify changes in *P. vivax* burden as a result of efforts made to control it. The methods and estimates presented here may serve as a benchmark for future work of this nature.

Despite shrinking transmission limits, it is important not to underestimate the potential impact of *P. vivax* malaria. It must be emphasized that the burden estimates presented here represent symptomatic clinical *P. vivax* and not parasite infection. Knowledge of the number of individuals who express clinical illness is of huge importance in developing treatment and control policies. Treatment adherence for *P. vivax* is particularly challenging due to the risks and prolonged regimen of primaquine therapy [79]. However, *P. vivax* is very often asymptomatic sub-patent and latent in the blood stage [22], which calls for a pause in the interpretation of these results.

While the prevalence mapped here shows only patent parasitaemia detected by microscopy or RDT (therefore underestimating the true prevalence), this would not translate to a significant under-estimate of clinical burden. The model of the relationship between prevalence of infection and clinical *P. vivax* malaria was also based on microscopy and RDT detected patent parasitaemia (PCR data were also excluded from that analysis) [47, 80]. This effectively removes any bias of estimating incidence from the patent $Pv\text{PR}_{1\text{-}99}$ map.

The clinical cases used to fit the prevalence-incidence were febrile with patent parasitaemia also by microscopy and RDT. What cannot be quantified is what proportion of these cases arose from a previously latent or sub-patent infection (likely many) and to what extent confining the case definition to patent parasitaemia underestimated the number of cases. A study done in Brazil detected at times >50% more patent parasitaemia in patients with symptoms when quantitative real-time PCR was used in place of conventional microscopy [81]. Therefore some fever cases found in ACD studies that were negative for *P. vivax* by microscopy may have indeed been parasite positive. The measures presented here utilize records of patent prevalence and clinical case incidence measured by diagnostic techniques suitable for large-scale field-based studies because these data are the most abundant and therefore the necessary tools for globally modelling analyses. As a result, the outputs show only the surface of the true endemicity and burden of *P. vivax*. This is of use when comparing endemicity and burden among regions, but to fully quantify the burden of *P. vivax* malaria more sensitive diagnostic techniques must be employed or standard methods must be adjusted for the extent to which they underestimate parasitaemia, as described above.

Finally, this study did not attempt to differentiate the severity of clinical episodes. Any such attempt would be complicated because severe morbidity results from the

cumulative effects of repeated infections [63, 82]. A primary infection does not have the same impact on a patient as a fourth or fifth relapse. However, to capture this effect in burden assessments will require a better understanding of the patterns of multiple relapses as well as the use of and adherence to radical cures. The burden of mixed *P. falciparum* and *P. vivax* infections also requires further exploration. As described above, these infections are often under-diagnosed and are associated with severe disease [83, 84].

The global clinical burden of *P. vivax* reported here represents the best approximation of the number of global cases annually based on data presently available. The estimates have the potential to be updated each year as new reports are submitted to the WHO for the WMR and more parasite rate data becomes available. Future generations of the maps presented here will help to quantify the impact of malaria control interventions and prioritise the allocation of resources. This knowledge will be crucial as *P. vivax* endemic countries progress towards the goal of elimination.

6.6. References

1. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA *et al,*: **A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010**. *PLoS Negl Trop Dis* 2012, **6**(9):e1814.

2. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL *et al,*: **A new world malaria map: *Plasmodium falciparum* endemicity in 2010**. *Malar J* 2011, **10**:378.

3. Baird JK: **Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria**. *Clin Microbiol Rev* 2013, **26**(1):36-57.

4. Anstey NM, Douglas NM, Poespoprodjo JR, Price RN: ***Plasmodium vivax*: clinical spectrum, risk factors and pathogenesis**. *Adv Parasitol* 2012, **80**:151-201.

5. Murray CJ, Ortblad KF, Guinovart C, Lim SS, Wolock TM, Roberts DA *et al,*: **Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013**. *Lancet* 2014, **384**(9947):1005-1070.

6. Feachem RG, Phillips AA, Targett GA, Snow RW: **Call to action: priorities for malaria elimination**. *Lancet* 2010, **376**(9752):1517-1521.

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

8. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW: **The global distribution and population at risk of malaria: past, present, and future**. *Lancet Infect Dis* 2004, **4**(6):327-336.

9. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM: **Vivax malaria: neglected and not benign**. *Am J Trop Med Hyg* 2007, **77**(6 Suppl):79-87.

10. Lysenko AJ, Semashko IN: **Geography of malaria. A medico-geographic profile of an ancient disease**. In: *Itogi Nauki: Medicinskaja Geografija.* Edited by Lebedew AW. Moscow: Academy of Sciences, USSR; 1968: 25-146.

11. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C: **Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods**. *PLoS Med* 2011, **8**(12):e1001142.

12. WHO: **World Malaria Report 2008**. In. Edited by WHO. Geneva, Switzerland: World Health Organization; 2008: 215.

13. WHO: **World Malaria Report 2014**. In. Geneva, Switzerland: World Health Organization; 2014: 226.

14. Mueller I, Galinski MR, Baird JK, Carlton JM, Kochar DK, Alonso PL *et al,*: **Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite**. *Lancet Infect Dis* 2009, **9**(9):555-566.

15. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U *et al,*: **The impact of malaria control on *Plasmodium falciparum* in Africa, 2000–2015**. *Nature* 2015, **Accepted**(MS ID: 2015-06-08556B).

16. Cameron E, Battle KE, Bhatt S, Weiss DJ, Bisanzio D, Mappin B *et al,*: **An emulator ensemble for modelling the relationship between *Plasmodium falciparum* parasite prevalence and clinical incidence**. *Nat Commun* 2015, **6**:8170.

17. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA *et al,*: **Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007**. *PLoS Med* 2010, **7**(6):e1000290.

18. Andersen F, Douglas NM, Bustos D, Galappaththy G, Qi G, Hsiang MS *et al,*: **Trends in malaria research in 11 Asian Pacific countries: an analysis of peer-reviewed publications over two decades**. *Malar J* 2011, **10**:131.

19. Carlton JM, Sina BJ, Adams JH: **Why is *Plasmodium vivax* a Neglected Tropical Disease?** *PLoS Negl Trop Dis* 2011, **5**(6):e1160.

20. **http://www.malariasurveys.org/** (Accessed 22 June 2015).

21. **http://www.dhsprogram.com/** (Accessed 22 June 2015).

22. Mueller I, Galinski MR, Tsuboi T, Arevalo-Herrera M, Collins WE, King CL: **Natural acquisition of immunity to *Plasmodium vivax*: epidemiological observations and potential targets**. *Adv Parasitol* 2013, **81**:77-131.

23. Guerra CA, Gikandi PW, Tatem AJ, Noor AM, Smith DL, Hay SI *et al,*: **The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide**. *PLoS Med* 2008, **5**(2):e38.

24. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperely WH *et al,*: **The international limits and population at risk of *Plasmodium vivax* transmission in 2009**. *PLoS Negl Trop Dis* 2010, **4**(8):e774.

25. **Gridded Population of the World, Version 3 (GPWv3): Population Density Grid**

26. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, Noor AM *et al,*: **Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project**. *Malar J* 2007, **6**:17.

27. Dalrymple U, Mappin B, Gething PW: **Malaria mapping: understanding the global endemicity of falciparum and vivax malaria**. *BMC Med* 2015, **13**:140.

28. Imwong M, Hanchana S, Malleret B, Renia L, Day NP, Dondorp A *et al,*: **High-throughput ultrasensitive molecular techniques for quantifying low-density malaria parasitemias**. *J Clin Microbiol* 2014, **52**(9):3303-3309.

29. Zimmerman PA, Howes RE: **Malaria diagnosis for malaria elimination**. *Curr Opin Infect Dis* 2015, **28**(5):446-454.

30. **Google Maps** [https://maps.google.com/]**,**

31. Microsoft: **Encarta Reference Library**. In. Seattle: Microsoft Corporation; 2007.

32. ESRI: **ArcGIS Desktop 10.3**. In. Redlands, CA: Environmental Systems Resource Institute; 2014.

33. Weiss DJ, Mappin B, Dalrymple U, Bhatt S, Cameron E, Hay SI *et al,*: **Re-examining environmental correlates of *Plasmodium falciparum* malaria endemicity: a data-intensive variable selection approach**. *Malar J* 2015, **14**:68.

34. Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW *et al,*: **Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax***. *Parasit Vectors* 2011, **4**(1):92.

35. Scharlemann JP, Benz D, Hay SI, Purse BV, Tatem AJ, Wint GR *et al,*: **Global data for ecology and epidemiology: a novel algorithm for temporal Fourier processing MODIS data**. *PLoS One* 2008, **3**(1):e1408.

36. Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A *et al,*: **MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets**. *Remote Sens Environ* 2010, **114**(1):168-182.

37. Nelson A: **A global map of accessibility**. In. Ispira, Italy: Environment Monitoring Unit - Joint Research Centre of the European Commission; 2008.

38. Hay SI, Guerra CA, Tatem AJ, Atkinson PM, Snow RW: **Urbanization, malaria transmission and disease burden in Africa**. *Nat Rev Microbiol* 2005, **3**(1):81-90.

39. Tatem AJ, Guerra CA, Kabaria CW, Noor AM, Hay SI: **Human population, urban settlement patterns and their impact on *Plasmodium falciparum* malaria endemicity**. *Malar J* 2008, **7**:218.

40. Qi Q, Guerra CA, Moyes CL, Elyazar IR, Gething PW, Hay SI *et al,*: **The effects of urbanization on global *Plasmodium vivax* malaria transmission**. *Malar J* 2012, **11**:403.

41. Centers for Disease Control and Prevention: **CDC Health Information for International Travel 2010**. In. Atlanta: U.S. Department of Health and Human Services, Public Health Service; 2009.

42. WHO: **International Travel and Health: Situation as on 1 January 2010**. In. Geneva: World Health Organization; 2010.

43. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A: **Determining global population distribution: methods, applications and data**. *Adv Parasitol* 2006, **62**:119-156.

44. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH *et al,*: **The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis**. *Parasit Vectors* 2011, **4**:89.

45. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, Gething PW *et al,*: **The global distribution of the Duffy blood group**. *Nat Commun* 2011, **2**:266.

46. Miller LH, Mason SJ, Clyde DF, McGinniss MH: **The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, FyFy**. *N Engl J Med* 1976, **295**(6):302-304.

47. Battle KE, Cameron E, Guerra CA, Golding N, Duda KA, Howes RE *et al,*: **Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease**. *Malar J* 2015, **14**:191.

48. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N *et al,*: **Geographical variation in *Plasmodium vivax* relapse**. *Malar J* 2014, **13**:144.

49. Cotter C, Sturrock HJ, Hsiang MS, Liu J, Phillips AA, Hwang J *et al,*: **The changing epidemiology of malaria elimination: new strategies for new challenges**. *Lancet* 2013, **382**(9895):900-911.

50. O'Meara WP, Barcus M, Wongsrichanalai C, Muth S, Maguire JD, Jordan RG *et al,*: **Reader technique as a source of variability in determining malaria parasite density by microscopy**. *Malar J* 2006, **5**.

51. Alemayehu S, Feghali KC, Cowden J, Komisar J, Ockenhouse CF, Kamau E: **Comparative evaluation of published real-time PCR assays for the detection of malaria following MIQE guidelines**. *Malar J* 2013, **12**:277.

52. WHO: **Microscopy for the detection, identification and quantification of malaria parasites on stained thick and thin blood films in research settings (version 1.0)**. In: *WHO Procedures Guidelines for the treatment of malaria.* vol. 1. Geneva: World Health Organization; 2015.

53. Mappin B, Cameron E, Dalrymple U, Weiss DJ, Bisanzio D, Bhatt S *et al,*: **Standardising *Plasmodium falciparum* infection prevalence measured via microscopy versus rapid diagnostic test**. *Malar J* 2015, **Under review**.

54. Okell LC, Bousema T, Griffin JT, Ouedraogo AL, Ghani AC, Drakeley CJ: **Factors determining the occurrence of submicroscopic malaria infections and their relevance for control**. *Nat Commun* 2012, **3**:1237.

55. da Silva NS, da Silva-Nunes M, Malafronte RS, Menezes MJ, D'Arcadia RR, Komatsu NT *et al,*: **Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia**. *Trans R Soc Trop Med Hyg* 2010, **104**(5):343-350.

56. Harris I, Sharrock WW, Bain LM, Gray KA, Bobogare A, Boaz L *et al,*: **A large proportion of asymptomatic *Plasmodium* infections with low and sub-microscopic parasite densities in the low transmission setting of Temotu Province, Solomon Islands: challenges for malaria diagnostics in an elimination setting**. *Malar J* 2010, **9**:254.

57. Katsuragawa TH, Gil LH, Tada MS, de Almeida e Silva A, Costa JD, Araujo Mda S *et al,*: **The dynamics of transmission and spatial distribution of malaria in riverside areas of Porto Velho, Rondonia, in the Amazon region of Brazil**. *PLoS One* 2010, **5**(2):e9245.

58. Mueller I, Widmer S, Michel D, Maraga S, McNamara DT, Kiniboro B *et al,*: **High sensitivity detection of *Plasmodium* species reveals positive correlations between infections of different species, shifts in age distribution and reduced local variation in Papua New Guinea**. *Malar J* 2009, **8**:41.

59. Steenkeste N, Rogers WO, Okell L, Jeanne I, Incardona S, Duval L *et al,*: **Sub-microscopic malaria cases and mixed malaria infection in a remote area of high malaria endemicity in Rattanakiri province, Cambodia: implication for malaria elimination**. *Malar J* 2010, **9**:108.

60. Cheng Q, Cunningham J, Gatton ML: **Systematic review of sub-microscopic *P. vivax* infections: prevalence and determining factors**. *PLoS Negl Trop Dis* 2015, **9**(1):e3413.

61. Imwong M, Nguyen TN, Tripura R, Peto TJ, Lee SJ, Lwin KM *et al,*: **The epidemiology of subclinical malaria infections in South-East Asia: findings from cross-sectional surveys in Thailand-Myanmar border areas, Cambodia, and Vietnam**. *Malar J* 2015, **14**(1):381.

62. Moreira CM, Abo-Shehada M, Price RN, Drakeley CJ: **A systematic review of sub-microscopic *Plasmodium vivax* infection**. *Malar J* 2015, **14**(1):360.

63. Douglas NM, Anstey NM, Buffet PA, Poespoprodjo JR, Yeo TW, White NJ *et al,*: **The anaemia of *Plasmodium vivax* malaria**. *Malar J* 2012, **11**:135.

64. Mayxay M, Pukrittayakamee S, Newton PN, White NJ: **Mixed-species malaria infections in humans**. *Trends Parasitol* 2004, **20**(5):233-240.

65. Howes RE, Reiner Jr RC, Battle KE, Longbottom J, Mappin B, Ordanovich D *et al,*: ***Plasmodium vivax* transmission in Africa**. *PLoS Negl Trop Dis* 2015, **Under review**.

66. Battle KE, Gething PW, Elyazar IR, Moyes CL, Sinka ME, Howes RE *et al,*: **The global public health significance of *Plasmodium vivax***. *Adv Parasitol* 2012, **80**:1-111.

67. Uganda Bureau of Statistics (UBOS) and ICF Macro: **Uganda Malaria Indicator Survey 2009**. In. Calverton, Maryland, USA: UBOS and ICF Macro; 2010.

68. Ministère du Plan et Macro International: **Enquête Démographique et de Santé, République Démocratique du Congo 2007**. In. Calverton, Maryland, U.S.A.: Ministère du Plan et Macro International; 2008.

69. Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, Malenke JA *et al,*: **African origin of the malaria parasite *Plasmodium vivax***. *Nat Commun* 2014, **5**:3346.

70. Ménard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, Ratsimbasoa A *et al,*: ***Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people**. *Proc Natl Acad Sci U S A* 2010, **107**(13):5967-5971.

71. Battle KE, Bisanzio D, Gibson HS, Bhatt S, Cameron E, Weiss DJ *et al,*: **Modelling national treatment-seeking rates to improve interpretation of malaria case reporting**. *Malar J* 2015, **submitted**(MS ID: MALJ-S-15-00880).

72. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM: **Malaria**. *Lancet* 2013.

73. Chandramohan D, Jaffar S, Greenwood B: **Use of clinical algorithms for diagnosing malaria**. *Trop Med Int Health* 2002, **7**(1):45-52.

74. Baird JK: **Neglect of *Plasmodium vivax* malaria**. *Trends Parasitol* 2007, **23**(11):533-539.

75. WHO: **World Malaria Report 2013**. In. Geneva, Switzerland: World Health Organization; 2013: 199.

76. WHO: **Control and elimination of *Plasmodium vivax* malaria: A technical brief**. In. Geneva, Switzerland: World Health Organization; 2015: 64.

77. Sattabongkot J, Tsuboi T, Zollner GE, Sirichaisinthop J, Cui L: ***Plasmodium vivax* transmission: chances for control?** *Trends Parasitol* 2004, **20**(4):192-198.

78. Wells TN, Burrows JN, Baird JK: **Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination**. *Trends Parasitol* 2010, **26**(3):145-151.

79. Baird JK: **Elimination therapy for the endemic malarias**. *Curr Infect Dis Rep* 2012, **14**(3):227-237.

80. Battle KE, Guerra CA, Golding N, Duda KA, Cameron E, Howes RE *et al,*: **Global database of *Plasmodium falciparum* and *P. vivax* incidence records, 1985-2013**. *Scientific Data* 2015, **2**(150012).

81. Barbosa S, Gozze AB, Lima NF, Batista CL, Bastos Mda S, Nicolete VC *et al,*: **Epidemiology of disappearing *Plasmodium vivax* malaria: a case study in rural Amazonia**. *PLoS Negl Trop Dis* 2014, **8**(8):e3109.

82. Baird JK: **Severe and fatal vivax malaria challenges 'benign tertian malaria' dogma**. *Ann Trop Paediatr* 2009, **29**(4):251-252.

83. Douglas NM, Lampah DA, Kenangalem E, Simpson JA, Poespoprodjo JR, Sugiarto P *et al,*: **Major burden of severe anemia from non-falciparum malaria species in Southern Papua: a hospital-based surveillance study**. *PLoS Med* 2013, **10**(12):e1001575; discussion e1001575.

84. Genton B, D'Acremont V, Rare L, Baea K, Reeder JC, Alpers MP *et al,*: ***Plasmodium vivax* and mixed infections are associated with severe malaria in children: a prospective cohort study from Papua New Guinea**. *PLoS Med* 2008, **5**(6):e127.

Chapter 7 – Discussion

The aim of this thesis was to generate a global estimate of *Plasmodium vivax* malaria burden. To do so, spatial patterns of relapse were identified and incorporated into a model of the relationship between prevalence of infection and incidence of clinical disease. The varying strength of health systems across *P. vivax* endemic countries result in different quantities and quality of available data globally. For that reason, to most accurately capture the burden of *P. vivax* malaria, it was necessary to employ more than one method to generate case estimates: (i) a surveillance-based approach that adjusts routine case reports and (ii) a cartographic approach that models clinical incidence from a map of *P. vivax* prevalence rate. Data on treatment-seeking behaviours were needed for the surveillance-based approach, and in areas without coverage of national household surveys it was necessary to model these values. Adjusted surveillance-based case estimates were used to define the geographic limits of *P. vivax* transmission, and prevalence of *P. vivax* infection was predicted within these bounds. Prevalence was translated into incidence and case number estimates. Case estimates from either the surveillance or cartographic method were applied to different endemic countries and used to produce a combined estimate that was considered to be the most representative of the true burden based on the data available.

This discussion will summarize the key findings from each chapter of the thesis, the strengths and limitations of the methodologies employed, and the implications of the results on the broader field of *P. vivax* epidemiology in the context of control and elimination goals. The research objectives of this thesis presented various challenges primarily rooted in the paucity of *P. vivax* epidemiological data. The modelling

approaches developed to meet these challenges and their resulting limitations are discussed. The uncertainty metrics produced alongside the model predictions are emphasized. These measures are used to suggest future surveillance priorities that will improve burden estimation, as well as areas where predictions made with high certainty can drive immediate policy intervention or change.

7.1. Chapter summary

The thesis began with an investigation into the global public health significance of *P. vivax*. The review in Chapter 2 served as a baseline assessment of the areas and populations at risk, the endemicity of *P. vivax* within those regions, and the vectors of *P. vivax*. The review also examined the methodologies used to generate such information. The large areas and populations at risk, but uniformly low prevalence, emphasized the distinctly different epidemiology of *P. vivax* compared to *P. falciparum*. Examination of the dominant vector species only existed for 'malaria' [1-4], and an investigation to find whether all *Anopheles* vectors that transmit *P. falciparum* were also competent *P. vivax* vectors revealed that there are in fact more species incriminated as vectors of *P. vivax* than of *P. falciparum*.

The ability of *P. vivax* to remain dormant in the human host and cause relapsing infections weeks or months following the primary infections is one of the most important challenges to controlling the parasite. While the biological trigger of relapse remains unknown [5-7], the results of this chapter confirmed a long standing hypothesis that there are distinct geographic patterns in relapse periodicity, such that tropical strains relapse quickly and temperate strains more slowly. The evidence base of >30,000 individual records of time from primary infection to first relapse revealed that

the geographic pattern of relapse frequency is more complex than a simple binary tropical and temperate division [8]. The meta-analysis in Chapter 3 showed that observed relapse periods were best characterized by nine geographic zones modified from those first identified by George Macdonald in 1957 [9]. The zoo-ecological zones were delineated based on mosquito fauna and climatic conditions. This demonstrated that while relapse may indeed by triggered by other systemic infections [5], there are evolutionary adaptations driven by environmental influences to optimise the survival of the parasite in areas where climatic conditions prevent perennial transmission.

Chapter 4 developed a functional relationship between the prevalence of infection and incidence of clinical disease. This was the necessary key for translating easily obtainable measures of infection prevalence into programmatically useful clinical case estimates. Fitting separate relationships for each of the relapse zones identified in Chapter 3 showed that there were indeed different patterns of infection and disease risk in regions characterized by different relapse phenotypes. An average pooled relationship was also generated so that data-poor relapse zones could nevertheless be estimated with suitable statistical confidence.

Like the preceding chapters, Chapter 5 was an integral step for downstream analyses, while also yielding operationally insightful outputs in its own right. The thesis aimed to estimate *P. vivax* burden using both surveillance- and cartographic-based approaches. In order to estimate cases from routine surveillance reports accurately, they must be adjusted for (i) underreporting, (ii) slide positivity rates (for unconfirmed cases) and (iii) treatment-seeking behaviours. The two first datasets were available from the WHO, while a novel methodological approach was developed here to address treatment-seeking behaviour. Two types of treatment-seeking data were used for these adjustments – those who sought treatment at public facilities and those who sought any

formal treatment. The proportion of febrile individuals that sought treatment at a public facility provided information about the proportion of potential malaria cases that are captured by the surveillance system. The proportion of cases seeking any treatment provided a proxy to estimate the number of fever cases that were likely malaria. Multiplying by the proportion of fever cases that sought any treatment supposes that only fevers that sought treatment were in fact malaria and down-scales the case estimate. Cases that did not seek care were assumed to be non-malarial fevers with an effective slide positive rate (SPR) of zero. Leaving this proportion out of the case estimate formula assumes that the rate of malaria (or SPR) was the same in fever cases whether they sought treatment or not and thus provides the upper estimate of cases. The upper and lower estimates, determined by treatment-seeking data, were then averaged to report an overall estimate. The WHO uses national-level data and a regional-average where treatment-seeking data are missing. In an effort to refine these methods, household survey data were used to generate sub-national estimates, and values for missing countries were modelled using social and economic indicator variables, such that every malaria endemic country (MEC) had a unique estimate for public and any treatment-seeking proportions.

Chapter 6 was the culminating chapter of the thesis, integrating products of all the previous chapters to generate an output of operational significance. First, country case reports were assembled, adjusted as described above, and mapped to produce a global estimate of surveillance-based burden and a map of annual parasite incidence (API) of *P. vivax*. API and ecological masks were used to define the limits of transmission in 2013, which allowed updated estimates of areas and populations at risk to be calculated. Using an updated global parasite rate (PR) database, the prevalence of *P. vivax* within the stable limits of transmission was estimated along with associated uncertainty. The

prevalence-incidence model developed in Chapter 4 was applied to find the predicted incidence of *P. vivax* clinical disease (again bounded by uncertainty metrics). A framework was developed to categorize the data available from each country to determine which countries would be best described by a surveillance-based, cartographic or combined approach. The result was a global estimate of 8-19 million cases of *P. vivax* in 2013.

7.2. Methodological discussion

This section will critique the statistical methods used in each chapter, and highlight the strengths and limitations of each. Epidemiological data are relatively limited for *P. vivax*, and the analytical approaches used here were designed to fully utilise the data available for each research question as well as produce rigorous associated measures of uncertainty. A summary of the principal research outputs of this thesis and the data and analytical framework used to the generate them is summarized in Table 7.1.

7.2.1. Methodological strengths

Although *P. vivax* is less studied that *P. falciparum* and as a result has less data available [10], the work presented here produced some of the largest open-access assemblies of *P. vivax* epidemiological data available, all of which have been published either as stand-alone databases or as supplements to the analyses they enabled. The analyses were therefore supported by the strongest evidence-base possible given the current state of *P. vivax* research prioritization. The first database made available was of the individual records of time to first relapse from over 30,000 individuals that were gathered from both formal literature searches and informal requests to colleagues for unpublished reports and documents.

Table 7.1. Summary of data types and analyses used to generate key outputs of thesis chapters.

	Inputs	Data quantity	Analysis/modelling framework	Outputs	Uncertainty
Relapse	Records of time to relapse	30,049 individual records	Mixed-effects meta-analysis to determine geographic system that best describes relapse variation	Observed and predicted incidence rate and mean time to relapse in each zone	95% CI based on raw and modelled data
Prevalence-incidence model	Space time-matched records of incidence from ACD and PR from XSS	176 matched records	Non-linear regression models were fitted via a hierarchical Bayesian according to relapse zones	Zone-specific prevalence-incidence relationships	Point-wise 95% CrI of pooled and zone-specific fitted functions
Treatment-seeking	Proportion of fevers in children that sought public or any treatment and social and economic covariates	58 country-level estimates from DHS and MIS	Generalized additive mixed models	Predicted treatment-seeking proportions for 38 MECs	95% CI from error of model averages
Surveillance-based clinical incidence	Cases, reporting rates and SPR from WHO and treatment-seeking from DHS, MIS, and above models	10,939 admin-level records for 96 MECs	Algebraic formulas from WMR that estimate upper and lower range of cases [11,12]	Average, lower and upper case estimates by admin region	No formal uncertainty; range of estimates only
Endemicity (*Pv*PR$_{1\text{-}99}$)	Geopositioned records of *Pv*PR and a suite of environmental covariates	12,782 *Pv*PR records	Latent Gaussian process model with Bayesian inference	Smooth map of predicted *Pv*PR in all ages (*Pv*PR$_{1\text{-}99}$)	95% CrI from conditional simulations
Cartographic-based clinical incidence	*Pv*PR$_{1\text{-}99}$ surface within the stable limits of transmission	Predicted *Pv*PR in 5km^2 pixels	Bayesian non-linear regression model (above)	Smooth map of *P. vivax* clinical incidence	95% CrI from conditional simulations
Global clinical burden	Country-level cartographic- and surveillance-based incidence multiplied by global population surface	Dual estimates for 93 *Pv*MECs	Decision framework to assign one or a combination of the two estimates	A combined estimate of clinical cases for *Pv*MECs	Formal uncertainty for modelled and combined estimates (95% CrI); upper and lower ranges for surveillance estimates

ACD = active case detection; CI = confidence interval; CrI = credible interval; DHS = Demographic and Health Surveys; MECs = malaria endemic countries; MIS = Malaria Indicator Surveys; PR = parasite rate; SPR = slide positivity rate; WMR = World Malaria Report; XSS = cross-section survey

The model in Chapter 4 was fitted to data from a large scale literature search which assembled as many records of clinical incidence measured by active case detection (ACD) as possible. The analysis required records of the ACD to be spatially and temporally matched to measures of PR. When PR data were not available from the ACD publication, matching was supplemented by data from the Malaria Atlas Project (MAP) PR database, which has been curated since 2005 and contains records from the last 30 years [13,14]. Finally, though the *P. vivax* parasite rate (*Pv*PR) database is dwarfed by that of *P. falciparum*, which benefits from an abundance of data from national household surveys [15], the *Pv*PR evidence-base was made as comprehensive as possible through rigorous searches and data requests.

In light of these sets of data, a suite of modelling approaches was developed to address each research question in the thesis. All modelled outputs were produced with rigorous associated measures of uncertainty (Table 7.1) and allowed for (non-parametric) flexibility in the input data. The relapse and treatment-seeking data were characterised by mixed effects models to allow for additive non-linear functional relationships between the response data and predictor variables.

The PR and prevalence-incidence models were built in a Bayesian framework. Bayesian geostatistics were well suited to the limited dataset available here because it: (i) allowed for highly flexible non-parametric models, (ii) leveraged correlated random effects to improve model prediction, and (iii) allowed for a realisation-based approach where a large number of PR and clinical incidence 'candidate' maps could be produced. These realisations not only allowed for rigorous estimation of point-wise uncertainty, but made possible aggregated uncertainty measures and full propagation of errors between models [16]. The mean value of all the *Pv*PR simulations was presented as the endemicity map in Chapter 6 and the *Pv*PR uncertainty was the 95% credible interval

(CrI) range of the candidate maps. Two hundred realisations of the *Pv*PR map were then translated through the functional prevalence-incidence model from Chapter 4, where uncertainty from the *Pv*PR model was propagated and combined with parameter uncertainty from the incidence model. The resulting clinical incidence map shown in Chapter 6 thereby incorporates the uncertainty of both predictions. The number of conditional simulations was made possible by using an integrated nested Laplace approximation (INLA) [17], which greatly improved upon the computational processing times used in similar work conducted previously [18,19].

7.2.2. Methodological limitations

Though the modelling approaches used in this thesis were chosen for their ability to cope with relatively sparse data, the principal challenges met in the analyses in each chapter surrounded data quality and quantity. The data on relapse had to be derived from patients not treated with a radical cure, which was difficult to find in contemporary data sources. Thus, the data used was biased towards historical records from drug trials around the time of the Second World War, records from U.S. and British soldiers returned home with *P. vivax* malaria after serving in the Pacific theatre and malaria therapy studies from when *P. vivax* was used to treat patients with neurosyphilis.

The model developed of the relationship between prevalence of infection and incidence of clinical disease was also restricted by the data available. As Figure 7 in Chapter 4 illustrates, there were only a few points available from some of the relapse zones, and some areas, such as the African zone, lacked data entirely. In addition, a consistent problem with the data used in this thesis was the potential for bias stemming from those areas studied being selected because they have more malaria or represent a unique

transmission setting relative to surrounding areas. However, at each analysis stage, minimal exclusions were set in order to include as many records as possible.

Plasmodium vivax PR data was similarly disparate relative to *P. falciparum* and the paucity of national household surveys meant that there was likelihood that the cross-sectional surveys used were not all population-representative samples. The PR data included in this thesis were only based on microscopy and rapid diagnostic test (RDT) diagnoses. Expanding the inclusion criteria to include molecular diagnosis would have increased data volume and reliability of the observed PR estimates. However, incorporation of this data first requires reconciling the difference in sensitivities among the different diagnostic techniques and varying methodologies used in molecular diagnosis. There were only a few sites where paired diagnostic results were available, but this will likely to change with time and the increasing use of the molecular diagnosis for *P. vivax,* which can be difficult to detect and is often misdiagnosed [20,21].

As described above, the modelling approaches used were designed to accommodate the characteristics of the available data and to produce measures of uncertainty to highlight regions where the model was not able to make reliable predictions. The only datasets that could not be brought into a strict probabilistic format and therefore lacked formal uncertainty metrics were the surveillance-based case measures. The values of the parameters used to adjust reported case numbers (i.e. reporting completeness and SPR) were provided as estimates by the WHO without background data. Without precise quantification of these input values, determining the uncertainty of the output was not possible. The certainty of the adjustment parameters remains one of the primary barriers to fully unravelling the problem of global burden estimation. However, the quality of these parameters increases in areas where malaria prevalence is low and surveillance

systems are strong. Therefore, the countries with surveillance-based estimates are likely to have the most accurate input data.

7.3. Interpretation, implications and future priorities

In light of the strengths and limitations outlined above, this final section discusses how the key findings of the thesis may provide insight into the epidemiology and future control efforts for *P. vivax* malaria.

It should be noted that the models and output data presented in this thesis are the first of their kind for *P. vivax*. Until now, methodologies specifically adapted to the intricacies of the biology of *P. vivax* have not existed, and these therefore pave the way for evidence-based decision-making for the control of this pernicious and resilient malaria species. Evidence of the value of the products developed here can be seen by their rapid uptake by the WHO, notably in their recently published Technical Brief on *P. vivax* [22], which informed the 2016-2030 Global Technical Strategy for Malaria adopted by the World Health Assembly in May 2015 [23].

The hypnozoite, with its treatment difficulties and the unknown mechanism through which it causes relapsing infections, is the greatest challenge facing *P. vivax* control, and perhaps malaria in general [24]. Understanding how rates of relapse differ geographically, and mapping estimates of the time and incidence of relapse, can help elucidate the relative impact of relapse on the malaria burden of an area. Studies have found that half of blood-stage infections and more than half of clinical episodes were attributable to relapse [25] in areas with short relapse frequency, such as Papua New Guinea. In the absence of studies of parallel cohorts that do and do not receive hypnozoitocidal treatment across a range of endemic settings, the patterns of relapse

described here can help predict the likely impact of effective radical cure on transmission.

Building upon the work done in Chapter 3, the role of relapse in *P. vivax* transmission has been modeled [26] in an effort develop a complete transmission model for *P. vivax,* something conspicuously absent from the field of malaria epidemiology. To understand the full impact of relapses on transmission, it will also be necessary to understand the proportion of primary infections that relapse. This too was observed to vary based on the geographic origin of the strain of *P. vivax* [6,27]. Finally, the number of multiple relapses will also be an essential component to understanding the potential force of infection from hypnozoite parasites.

Burden estimates for any disease are gathered to measure the impact of the disease relative to other diseases and to provide benchmarks to assess progress of control programmes. The case estimates produced here will therefore help countries assess the magnitude of the *P. vivax* problem within their borders, which is of particular value for countries without the infrastructure to accurately capture this information through their routine health surveillance systems. Knowledge of the *P. vivax*-specific malaria burden is of use to public health planning because it helps to budget for *P. vivax* treatment and perhaps prioritize testing for glucose-6-phosphate dehydrogenase (G6PD) deficiency, a genetic condition which leaves patients vulnerable to haemolysis if treated with primaquine, the only licensed drug to treat the hypnozoite stage. Plans are already underway by collaborators at the Mahidol Oxford Tropical Medicine Research Unit in Bangkok to incorporate the clinical burden estimates produced here into assessment of the economic burden of *P. vivax*, not only in terms of financial costs, but also the disability adjusted life years (DALYs) attributable to *P. vivax.*

To fully understand the contribution of *P. vivax* to malaria morbidity, the problem of multiple relapses must be unravelled. Relapses have a cumulative effect on the health of the patient, taxing the system with recurrent bouts of haemolysis [28-30]. The burden attributable to *P. vivax* is further complicated by comorbidities with other diseases (i.e. sepsis) and undiagnosed *P. falciparum* co-infection that are often associated with severe adverse effects or death [28,29]. Estimates of *P. vivax* mortality were not calculated alongside the clinical burden estimates in this thesis. *Plasmodium vivax* has only been recognized as a potentially fatal disease in recent years [29]. In order to truly understand the extent of *P. vivax* severe disease and death, formal criteria for severe *P. vivax* malaria must be defined [31] to allow comparison among studies and to enable the appropriate classification of cause of death [29].

The uncertainty estimates generated in this thesis are a key output in their own right. Areas with low uncertainty, where estimates are deemed to be reliable, can be targeted for immediate control or intervention. For both the endemicity and clinical burden estimates, areas with wide confidence intervals highlight areas where improved surveillance would greatly improve the evidence bases available for control planning. To strengthen their reporting systems, country-level control programs must also be made aware of the type of data required to accurately generate burden outputs: species-specific case data, the number of slides or tests examined relative to the number positive, and the proportion of individuals with malaria or fever that sought treatment. The national survey coverage and resulting high certainty in the predictions made in the East African countries highlight the utility of representative sampling. The results presented here and in recent associated work [32] will hopefully call attention to the need for *P. vivax* surveillance. While *P. vivax* is unlikely to be considered a significant public health problem in Africa, its ability to sustain transmission in very low density

Duffy positive populations may increase in importance in coming years as countries begin to lower *P. falciparum* case burdens.

Plasmodium vivax transmission is not likely to be sustained by the relatively low numbers of symptomatic clinical cases quantified here. The true burden of *P. vivax* on control and elimination goals is the large biomass of asymptomatic and undetectable infections. Even in low transmission settings, most *P. vivax* infections appear to be microscopically sub-patent and asymptomatic [33] and there remains no diagnostic capacity for detecting the hypnozoite stage. Strategies to find and safely eliminate these hidden parasites will be the final hurdle to any elimination goals.

7.4. Conclusions

The burden of *P. vivax* malaria has been identified as a key knowledge gap and was previously entirely unknown in areas where case data is not widely shared or entirely missing. The maps and estimates presented here are not definitive outputs of *P. vivax* malaria endemicity or clinical burden, but rather serve as the best representation of our knowledge given the data available. The maps serve as a mechanism to highlight where the relative impact of *P. vivax* is greatest and where uncertainty indicates that surveillance is needed. The results benefit from an analytical framework that is able to be continually updated as improvements to input data are made available. Continued collection and sharing of these data will require interdisciplinary efforts and political support, but would bolster all of the principle outputs of this thesis: the role of relapse, prevalence of infection, treatment coverage and burden of clinical disease.

This thesis opened by discussing the relative neglect of *P. vivax* malaria relative to *P. falciparum* despite the epidemiological intricacies of *P. vivax*. It is hoped that this thesis

demonstrates that the unique challenges of *P. vivax* present a substantial public health problem that requires greater research attention. Integration of a variety of epidemiological data into flexible modelling frameworks can produce quantitative and visual outputs that can help to inform public health decision making so that control gains made, such as those observed since the start of this doctoral research, can continue to be made. The significant advances in technical capacity and downstream programmatic decision-making enabled by the work presented here are made evident by the contrasts between the perspectives of *P. vivax* epidemiology dated to the start of this doctoral research, which were summarised in Chapter 2, relative to the outputs of the subsequent chapters and their culmination in Chapter 6 as estimates of *P. vivax* clinical case burden. It is the hope that this thesis provides a strong platform for further advances in the years to come as elimination efforts are increased worldwide.

7.5. References

1. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, *et al*: **The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis.** *Parasit Vectors* 2010, **3:**117.

2. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, *et al*: **The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic precis.** *Parasit Vectors* 2010, **3:**72.

3. Sinka ME, Bangs MJ, Manguin S, Chareonviriyaphap T, Patil AP, Temperley WH, *et al*: **The dominant *Anopheles* vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic precis.** *Parasit Vectors* 2011, **4:**89.

4. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, *et al*: **A global map of dominant malaria vectors.** *Parasit Vectors* 2012, **5:**69.

5. Shanks GD, White NJ: **The activation of vivax malaria hypnozoites by infectious diseases.** *Lancet Infect Dis* 2013, **13:**900-906.

6. White NJ: **Determinants of relapse periodicity in *Plasmodium vivax* malaria.** *Malar J* 2011, **10:**297.

7. White NJ, Imwong M: **Relapse.** *Adv Parasitol* 2012, **80:**113-150.

8. Lover AA, Coker RJ: **Quantifying effect of geographic location on epidemiology of *Plasmodium vivax* malaria.** *Emerg Infect Dis* 2013, **19:**1058-1065.

9. Macdonald G: **Local features of malaria.** In *The Epidemiology and Control of Malaria.* London: Oxford University Press; 1957: 63-99

10. PATH: **Staying the Course? Malaria Research and Development in a Time of Economic Uncertainty.** Seattle, U.S.A.: PATH; 2011.

11. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C: **Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods.** *PLoS Med* 2011, **8:**e1001142.

12. WHO: **World Malaria Report 2008.** (WHO ed. Geneva, Switzerland: World Health Organization; 2008:215.

13. Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, Noor AM, *et al*: **Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project.** *Malar J* 2007, **6:**17.

14. *Malaria Atlas Project.* www.map.ox.ac.uk.

15. Dalrymple U, Mappin B, Gething PW: **Malaria mapping: understanding the global endemicity of falciparum and vivax malaria.** *BMC Med* 2015, **13:**140.

16. Patil AP, Gething PW, Piel FB, Hay SI: **Bayesian geostatistics in health cartography: the perspective of malaria.** *Trends Parasitol* 2011, **27:**246-253.

17. Rue H, Martino S, Chopin N: **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.** *Journal of the Royal Statistical Society Series B-Statistical Methodology* 2009, **71:**319-392.

18. Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA, *et al*: **A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010.** *PLoS Negl Trop Dis* 2012, **6:**e1814.

19. Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, *et al*: **Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007.** *PLoS Med* 2010, **7:**e1000290.

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

21. Rosenberg R: ***Plasmodium vivax* in Africa: hidden in plain sight?** *Trends Parasitol* 2007, **23:**193-196.

22. WHO: **Control and elimination of *Plasmodium vivax* malaria: A technical brief.** Geneva, Switzerland: World Health Organization; 2015:64.

23. WHO: **Global Technical Strategy for Malaria 2016–2030.** Geneva: World Health Organization; 2015:32.

24. Wells TN, Burrows JN, Baird JK: **Targeting the hypnozoite reservoir of *Plasmodium vivax*: the hidden obstacle to malaria elimination.** *Trends Parasitol* 2010, **26:**145-151.

25. Betuela I, Rosanas-Urgell A, Kiniboro B, Stanisic DI, Samol L, de Lazzari E, *et al*: **Relapses contribute significantly to the risk of *Plasmodium vivax* infection and disease in Papua New Guinean children 1–5 years of age.** *J Infect Dis* 2012, **206:**1771-1780.

26. White MT, Karl S, Battle KE, Hay SI, Mueller I, Ghani AC: **Modelling the contribution of the hypnozoite reservoir to *Plasmodium vivax* transmission.** *Elife* 2014, **3**.

27. Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, *et al*: **Geographical variation in *Plasmodium vivax* relapse.** *Malar J* 2014, **13:**144.

28. Anstey NM, Russell B, Yeo TW, Price RN: **The pathophysiology of vivax malaria.** *Trends Parasitol* 2009, **25:**220-227.

29. Baird JK: **Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria.** *Clin Microbiol Rev* 2013, **26:**36-57.

30. Mendis K, Sina BJ, Marchesini P, Carter R: **The neglected burden of *Plasmodium vivax* malaria.** *Am J Trop Med Hyg* 2001, **64:**97-106.

31. Rahimi BA, Thakkinstian A, White NJ, Sirivichayakul C, Dondorp AM, Chokejindachai W: **Severe vivax malaria: a systematic review and meta-analysis of clinical studies since 1900.** *Malar J* 2014, **13:**481.

32. Howes RE, Reiner Jr RC, Battle KE, Longbottom J, Mappin B, Ordanovich D, *et al*: ***Plasmodium vivax* transmission in Africa.** *PLoS Negl Trop Dis* 2015, **Under review**.

33. Mueller I, Galinski MR, Tsuboi T, Arevalo-Herrera M, Collins WE, King CL: **Natural acquisition of immunity to *Plasmodium vivax*: epidemiological observations and potential targets.** *Adv Parasitol* 2013, **81:**77-131.

APPENDIX

Appendix Table of Contents

Appendix to Chapter 3 – Geographical variation in *Plasmodium vivax* relapse

---

This Appendix includes:

1. Additional File 1: Additional results and statistical analyses.

2. Additional File 4: Individual-level data reference list.

Additional File 1

Geographical variation in *Plasmodium vivax* relapse

Authors: Katherine E Battle[1§], Markku S Karhunen[1], Samir Bhatt[1], Peter W Gething[1], Rosalind E Howes[1], Nick Golding[1], Thomas P Van Boeckel[2], Jane P Messina[1], G Dennis Shanks[3], David L Smith[4,5], J Kevin Baird[6,7], Simon I Hay[1,5§]

[§]Authors for correspondence: katherine.battle@zoo.ox.ac.uk and simon.hay@zoo.ox.ac.uk

This file includes:

Additional results and statistical analyses

Initial data collection

Following the literature search, data were first recorded as aggregated records of relapse. In this initial dataset, there were 163 records of relapse from 121 references on 56,649 patients, of which 11,152 experienced at least one relapse. Relapse was measured from a cohort of patients and reported as minimum, mean, median and maximum time to first relapse of the patients that relapsed. However, few of the studies provided exact measures of the total follow-up time. Thus, it was necessary to collect the individual-level records and calculate the person time from these for analysing the incidence rate. Ninety-two of the original references in the aggregated dataset contained individual level data. Following application of exclusion criteria, eliminating records that aggregated time >1 month and instance of relapse earlier than 14 days, the 44,624 patients records (8,395 relapse) was reduced to 30,049 (5,731 relapse) from 87 references as shown in Figure 3 in the main text.

The publication year of studies included in the analysis ranged from 1920 to 2013, with a mean and median value of 1949 and 1962, respectively. However, given that several later studies included a large number of patients, based on individual-level data the mean and median publication years were 1982 and 1990. Figure A1 illustrates the temporal distribution and size of studies included by country.

The number of patients observed and those that relapsed are shown by country, treatment type, patient type and Macdonald zone in Table A1. As shown in Figure A2, the majority of the patients were from India, but a low proportion experienced a relapse. Of the 23 known strains, the most commonly studied was the Chesson strain (39%), followed by the Madagascar (17%) and St. Elizabeth (16%) strains. The wild infections originated from 19 different countries and regions. The most common drug therapy used was chloroquine given in combination with primaquine at dosages deemed insufficient as an effective radical cure (3 or 5 day primaquine courses) [1]. Outpatients and military personnel were the most frequently observed patient type. Malaria therapy patients and prison "volunteers" comprised only 2% and 3% of the dataset, respectively.

Summary statistics of the observed time to relapse as well as details of follow up times for the individual level data are reported by ecological zone in Table A2.

Table A1 Summary of important aspects of individual data
The values represent the number of total individuals observed and the number of individuals who experienced at least one relapse.

Country/Region	**total**	**relapse**
Brazil	60	38
Cameroon	6	1
China	3	3
Comoros	3	1
El Salvador	314	306
Ethiopia	292	16
French Guiana	13	3
Greece	94	72
Guyana	22	2
India	23537	1931
Indonesia	106	33
Iran	3	3
Korea (North and South)	101	46
Macedonia	20	14
Madagascar	328	174
Mediterranean	437	126
Mexico	70	44
Myanmar	235	80
New Guinea	1081	752
Nicaragua	6	4
Pacific	1019	753
Pakistan	10	10
Panama	750	129
Peru	51	29
Russia	242	151
Solomon Islands	863	695
Thailand	39	8
USA	335	299
Vietnam	9	8
TOTAL	30049	5731

Treatment	**total**	**relapse**
8-aminoquinolines	18129	1281
Chloroquine	8029	1926
Mepacrine	1837	1330
Proguanil	42	36
Quinine	680	467
SN	165	109
None	307	228
Other	76	62
Unknown	784	292

Subjects*	**total**	**relapse**
ACD	1617	177
ACD and PCD	348	326
Malaria therapy	596	355
Military	3723	2137
Outpatients	22719	1953
Prison volunteers	816	584
Travellers	3	1
Unknown	227	198

Macdonald zone	**total**	**relapse**
1. North America	335	299
2. Central America	1140	483
3. South America	146	72
4. N. Europe and Asia	245	154
5. Mediterranean	551	212
6. Sahara-Sahel	0	0
7. Sub-Saharan Africa	629	192
8. Monsoon Asia	23550	1944
9. Himalaya-Mekong	0	0
10. South East Asia	360	124
11. China-Korean Pen.	101	46
12. PNG + Solomon Is.	2992	2205

*****ACD = active case detection, PCD = passive case detection

Table A2 Summary statistics of observed time to first relapse by zone
The total number of patients and relapses observed by zone are shown along with the minimum, 1st quartile (1st Qu), median, mean, 3rd quartile (3rd Qu) and maximum time to relapse in days. The 95% confidence interval (CI) of time to relapse is also shown. The minimum, median, mean and maximum follow up (FU) observed in each zone is also shown.

Zone	Total patients	Total relapses	Min	1st Qu	Median	Mean	3rd Qu	Max	95% CI	Min FU	Med FU	Mean FU	Max FU
1	335	299	14	240	270	239.4	300	364	228.5, 250.2	14	360	395.6	1778
2	1140	483	30	120	168	180.6	224	1582	171.3, 190.0	35	180	246.6	2884
3	146	72	14	51.75	68.5	88.56	92.5	360	72.7, 104.4	28	180	188.6	360
5 + 6	551	212	14	90	119	122	162.8	300	114.0, 130.0	30	240	213.1	365
7	629	192	14	40.25	90.5	100	140	260	90.3, 109.7	28	245	193.5	407
8	23550	1944	14	60	120	153	240	450	148.4, 157.6	30	360	340.1	609
9+10	360	124	14	21	28	30.81	28	360	24.1, 37.5	14	28	44.45	730
11+4	346	200	30	216	300	289	360	630	271.5, 306.5	120	730	637.7	990
12	2992	2205	14	33	46	52.58	62	426	51.3, 53.9	14	112	116.9	546

Additional statistical analyses

Alternative modelling strategies

We attempted to construct individual-based survival models for relapse time. This seemed like a natural choice, given the fact that we had collected individual-based data. Initially, we tried to pool the data across geographic areas, but the log-rank test soon revealed that the inter-study variation could not be ignored in these data (see also Table 4 in the main text). Figure 5 in the main text is based on this type on pooling; while we regard this as an informative figure, the confidence intervals are not justified in the strictest statistical sense. Subsequently, we attempted to model the data by using mixed-effects Cox regression implemented in the R package coxme [2]. However, this led to numerical convergence problems for some model variants (geographic classification systems). Finally, we attempted to replicate the results of R/metafor by using mixed-effects Poisson regression (R/lme4), which operates on average incidence rate within each study, just as metafor [3]. We were unable to make the optimisation algorithm converge using lme4 in some of the model variants. Based on these results, we judged that the only reliable results were obtained from metafor, and consequently, we report these as the main finding of this study.

Throughout most of the various analyses performed, we found that the Macdonald system was the most justified description of the data, judging by statistical criteria. This effect persisted in initial survival analyses done on data pooled by geographic zone (see above), and also in the final meta-analysis. Furthermore, this same relationship seemed to carry over to mixed-effects survival analysis (R/coxme), but as noted above, we do not present the results from those analyses here.

Sensitivity of the meta-analysis

We tried to improve the predictive power of the Macdonald system by combining some of the geographic zones, but only in one case did this yield a slight improvement. The various zone combinations and the resulting effects on the predictive power of the model are shown in Table A3. Trial F was the system that was applied in the analysis because it combined zones with little or no data with geographically contiguous zones with data. This slightly improved the predictive power of the system.

In addition to analysing discrete geographic zones, we also tried to model the data by using a measure of the longest transmission suitability period as a continuous moderator. This variable was generated from combining monthly measures of temperature suitability [4] and enhanced vegetation index (EVI; a measure of vegetation and proxy for moisture). The resulting global index indicated the number of months per year each 5 km x 5 km pixel is suitable for malaria transmission (Figure A3). However, the model with the continuous covariate provided statistics that were only modest compared to those

of the geographic systems (compare Tables A3 and A4). Thus we conclude that there is an association between the transmission suitability and the relapse rate, but the geographic zones capture this association better, than any of the continuous transformations tried here. In addition, they may capture other relevant features of the *P. vivax* life cycle.

Table A3 Testing for improved Macdonald systems
This table presents statistics on meta-analysis models where different modifications of the Macdonald system are used as moderators. The statistics are: pseudo-R^2, the amount of heterogeneity accounted for; AIC, Akaike information criterion; BIC, Bayesian information criterion; and AICc, corrected AIC. The values of AIC, BIC and AICc are based on the restricted maximum likelihood estimator.

Trial	Zone combinations	Pseudo-R^2	AIC	BIC	AICc
A	**Original (no combinations)**	59.9 %	612.7	649.2	614.1
B	**2+3**	56.8 %	627.2	660.4	628.3
C	**5+6**	60.1 %	612.9	646.1	614.0
D	**10+12**	60.0 %	613.6	646.9	614.8
E	**4+5+6+11**	60.0 %	613.9	643.8	614.8
F	**5+6, 9+10, 4+11**	59.9 %	612.7	649.2	614.1
G	**2+3, F**	56.8 %	627.2	660.4	628.3
H	**10+12, F**	56.8 %	627.2	660.4	628.3
I	**All changes**	57.0 %	628.4	655.0	629.1

Table A4 Using suitability as a continuous covariate
This table presents statistics on meta-analysis models where different transformations of the climatic suitability of transmission (Figure A3) are used as a continuous moderating variable, in place of the geographic zones (Macdonald system). The statistics are: pseudo-R^2, the amount of heterogeneity accounted for; AIC, Akaike information criterion; BIC, Bayesian information criterion; and AICc, corrected AIC. The values of AIC, BIC and AICc are based on the restricted maximum likelihood estimator.

	Pseudo-R^2	**AIC**	**BIC**	**AICc**
Macdonald	60 %	613	649	614
Empty model	0 %	778	785	778
X	0 %	777	787	777
log (X + 1/30)	2 %	772	782	772
$\sqrt{X}$	1 %	774	784	774
X^2	0 %	778	788	778
$X + X^2$	0 %	778	788	778

We reran the mixed-effects meta-analysis by using different transformations of data. For square-root and Freeman-Tukey transformed data, White's 5-class system (3 phenotype zones, differentiated by Old and New World) yielded the best results; whereas White's original 3-class system was the best description of the untransformed data. However, we argue that there is no particular reason to trust these results more than the log-transformed data presented in the main matter, as the Shapiro-Wilk test showed that the log-transformation produced the least deviations from normality (Figure A4). We also tried to include patient medication (if known) or the type of study population as additional moderators in the meta-analysis framework, and calculated the model choice criteria again. The Macdonald system also performed best in these analyses (Table A5).

Table A5 Results of the addition of subject and medication type to meta-analysis
The AICc values reflect the addition of adding the type of subject, medication used or both to the various geographic system models.

	Only zones	**Zones+ Subjects**	**Zones+ Medication**	**Zones+ Subjects+Medication**
White-3	708	674	678	656
Lover	729	631	659	597
White-5	624	585	573	546
Macdonald	614	579	573	548

Finally, we investigated the data for outliers. Figure A5 presents externally standardised residuals from the modified Macdonald system. From this figure we observe that most studies seem to be in good alignment with the incidence rates predicted for them, except in zone 8 (Monsoon Asia), there are three studies that could be interpreted potentially as outliers. These included one very large study (n=6393) with few patients who relapsed, and two small treatment studies with short follow up. We tried removing these data points, but this did not affect the model choice and there was no biological or methodological justification for doing so. Also, the model-based estimates of incidence rate were very similar to those shown in Table 2 in the main matter. In analogy to Figure A5, a forest plot (Figure A6) showed that the incidence rates in most studies were in a good alignment with the mean incidences in their respective geographic zones, but there were some cases where the incidences did not match. Thus, we concluded that the use of mixed-effects meta-analysis (which contains random effects attributable to individual studies), was warranted in this case. This is also observed in most studies of this type.

Figures

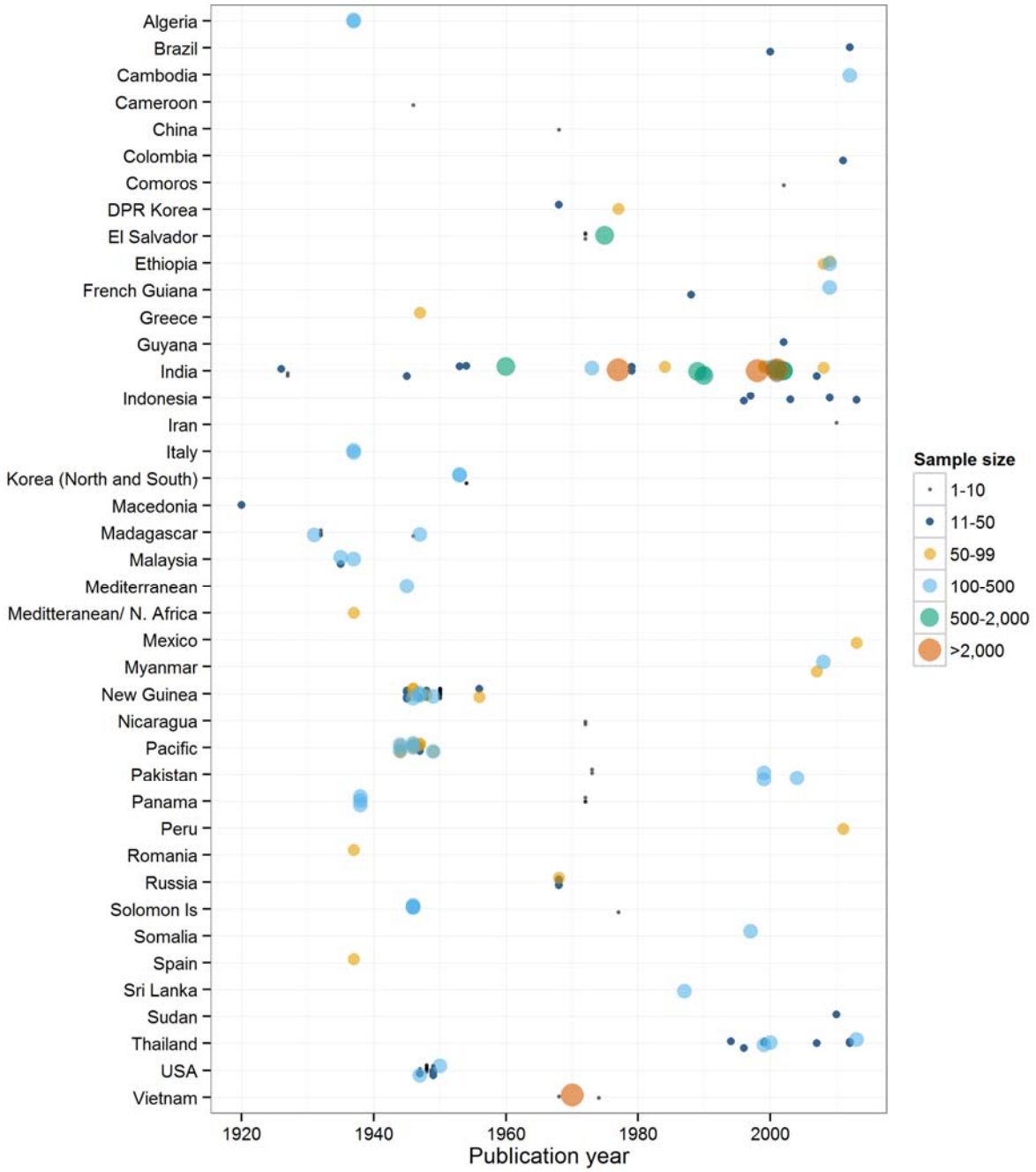


Figure A1 Temporal distribution and size of studies per country or region

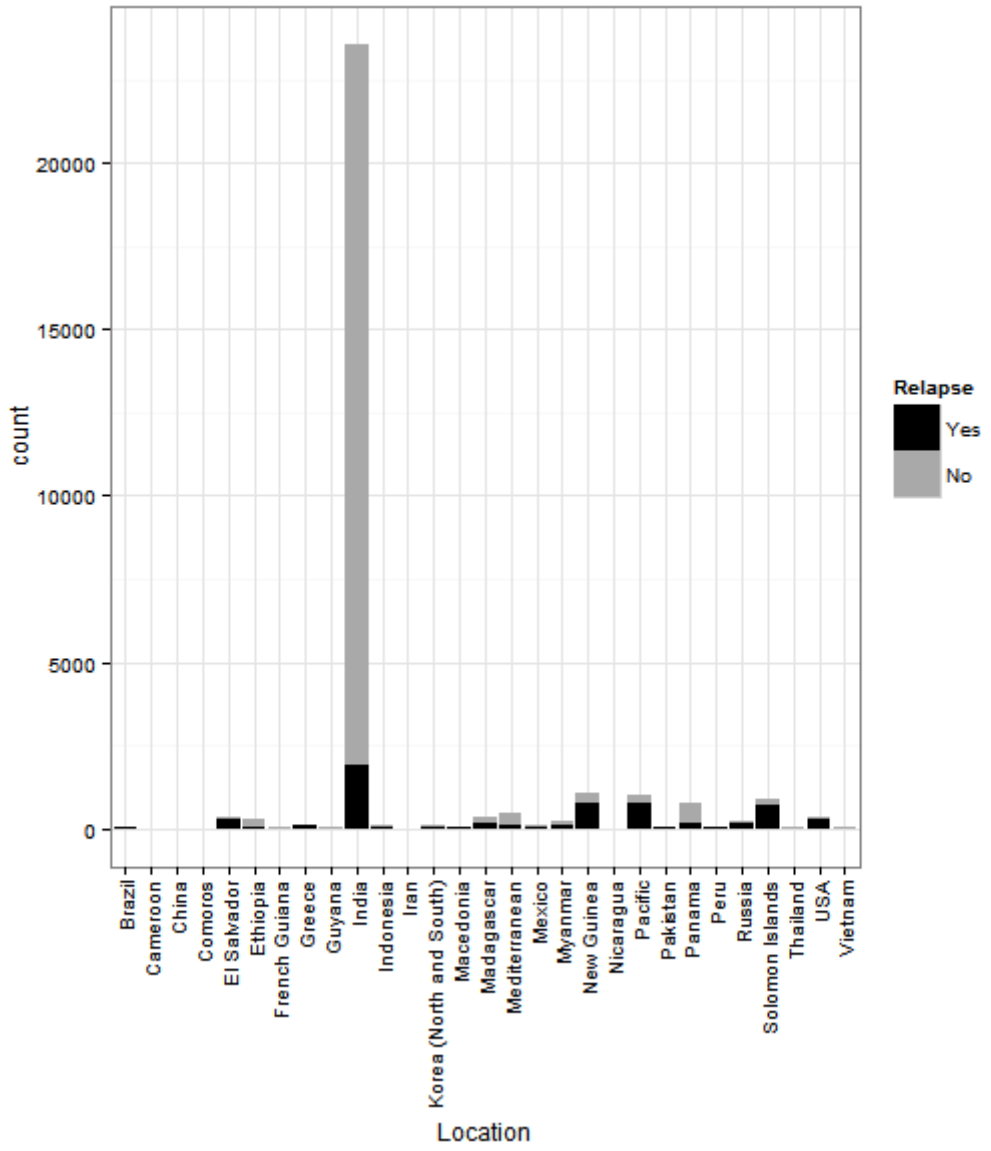


Figure A2 Number of total cases and number of relapses observed per country or region

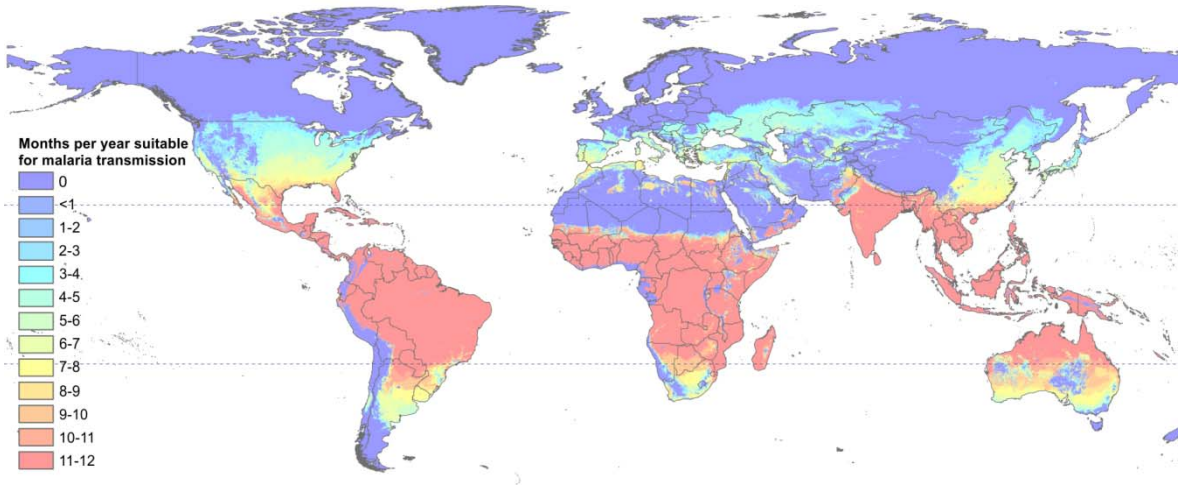


Figure A3 Global variation in *Plasmodium vivax* transmission suitability
The longest suitability period index classifies regions based on the number of months conditions are suitable for malaria transmission, ranging from year round (pink) to never suitable (blue). Values were calculated using monthly measures of *P. vivax* temperature suitability [4] and EVI (see main text).

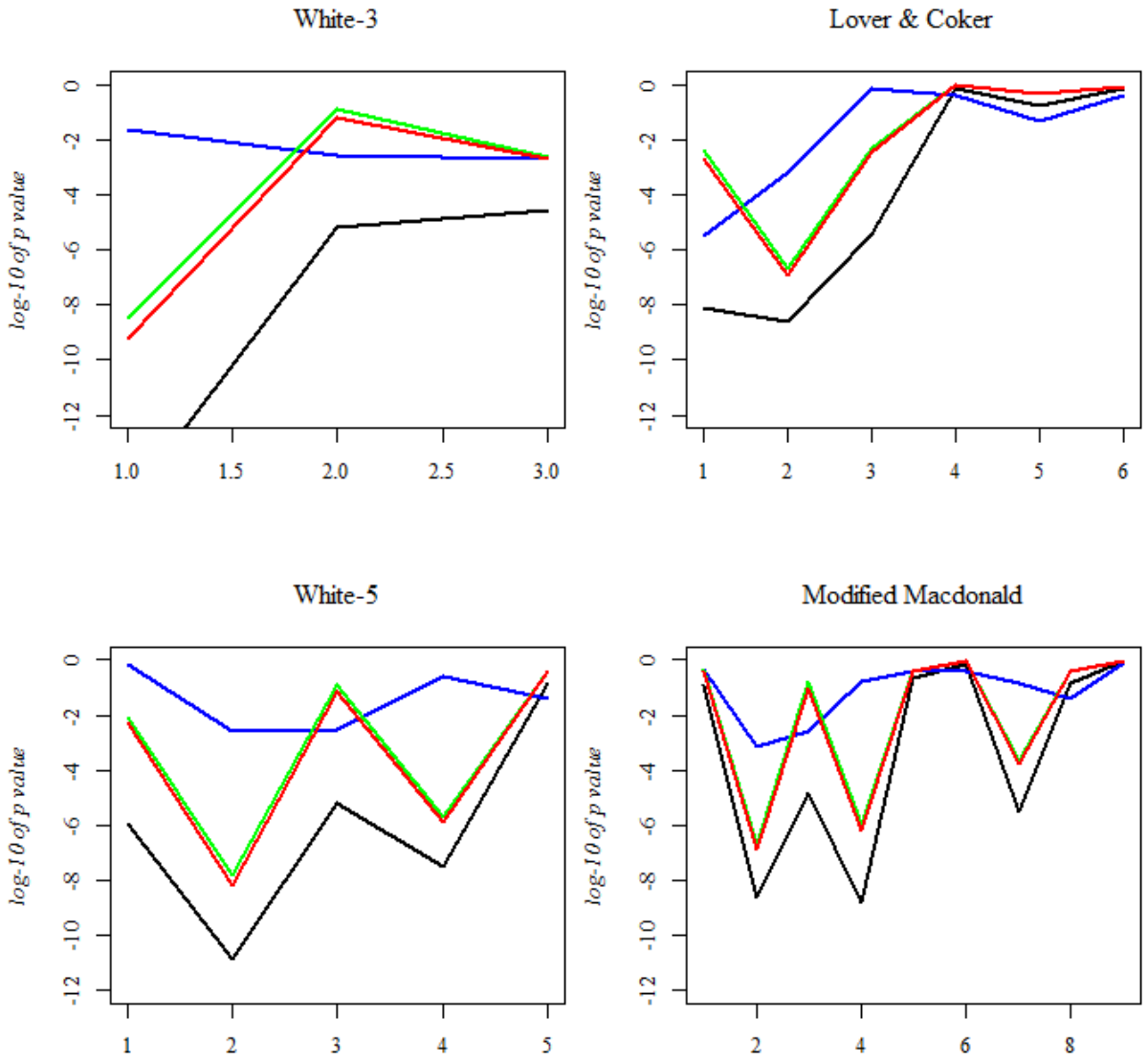


Figure A4 Results of Shapiro-Wilk test for normality

The Shapiro-Wilk test is a commonly used method for testing normality, which in turn is an underlying assumption of the meta-analysis models (i.e. R/metafor). Here we present the p values from this test. Points on the x-axes represent different geographic zones, while the lines themselves represent different transformations, such that logarithmic is blue, square root is green, Freeman-Tukey is red, and identity (no transformation) is black.

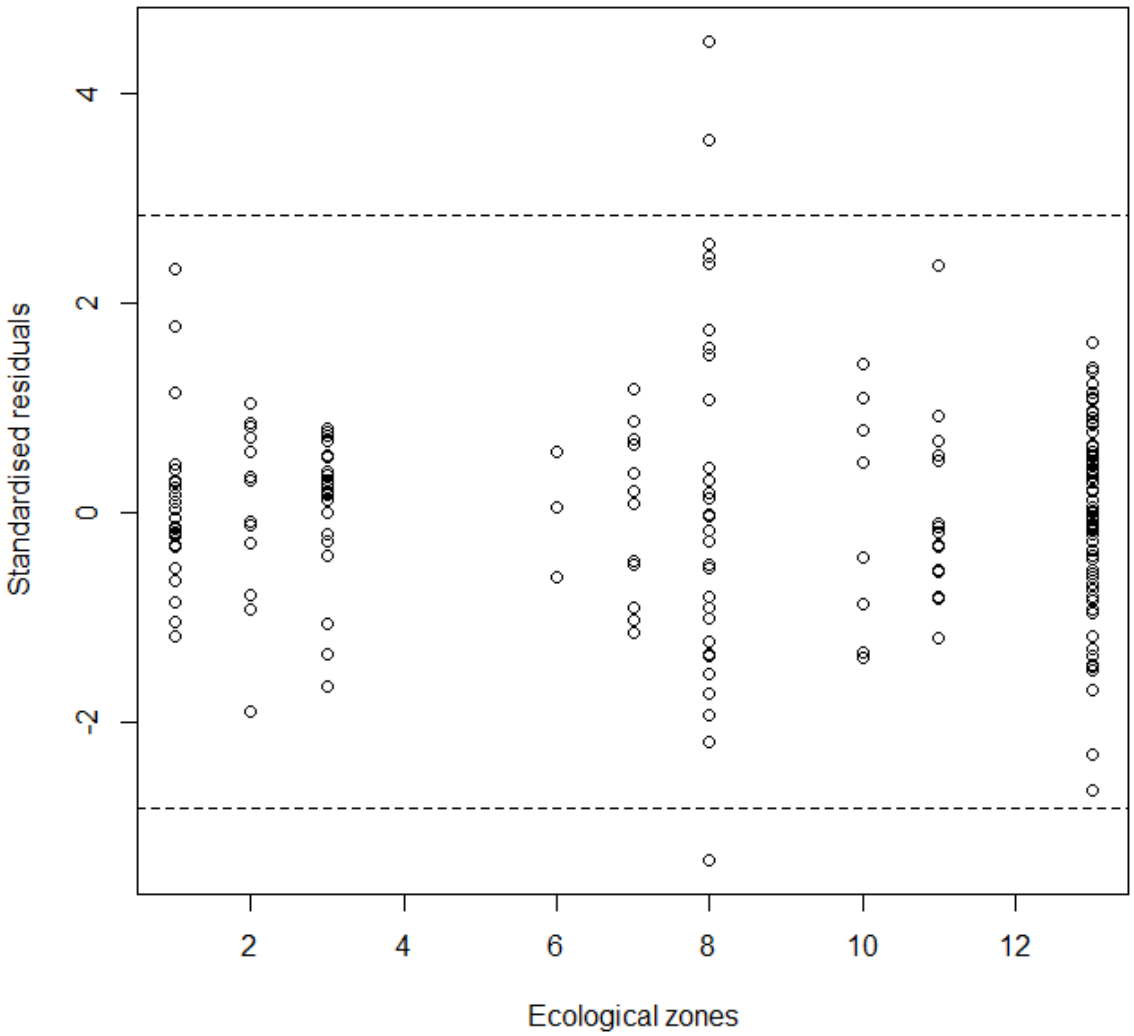


Figure A5 Detecting outliers in the modified Macdonald system
Each data point represents one study. The standardised residuals of the study effects are used here as a measure of model misfit. The horizontal lines are chosen such that there is *a priori* a 50 % chance that one of the residuals in a meta-analysis of this size falls outside these limits. Consequently, three of the studies could be interpreted as outliers.

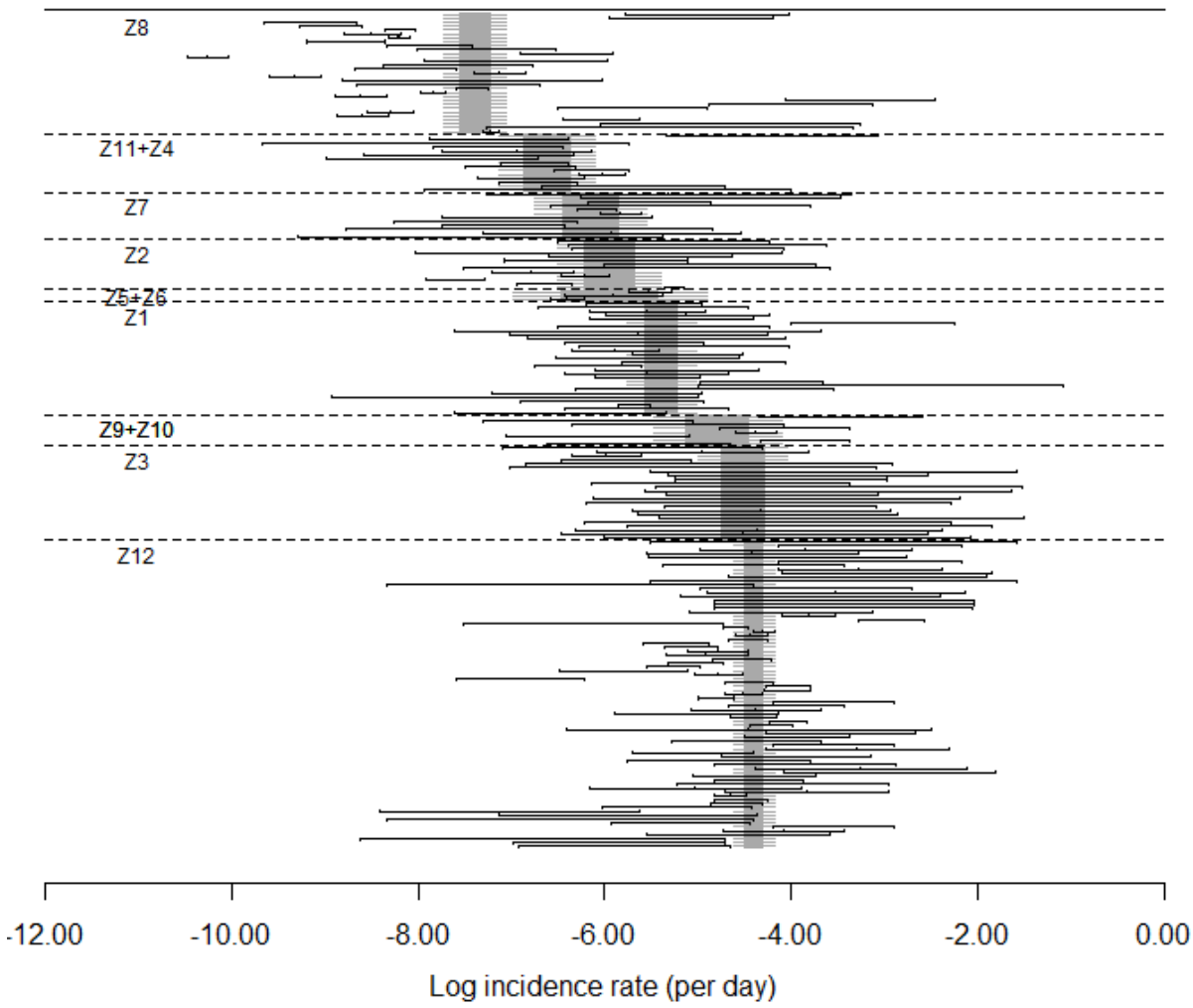


Figure A6 Forest plot for the modified Macdonald system
This figure has been produced by R/metafor [5]. The x axis represents log-incidence rate, running from -12 to 0 with even spaces. The grey diamonds represent the confidence intervals of fixed effects, whereas the horizontal lines represent CIs of the incidence rate within each study. The two do not always overlap due to study-specific random effects. Note that the scale of the x axis is natural logarithm.

References

1. Gogtay NJ, Desai S, Kamtekar KD, Kadam VS, Dalvi SS, Kshirsagar NA: **Efficacies of 5- and 14-day primaquine regimens in the prevention of relapses in *Plasmodium vivax* infections.** *Ann Trop Med Parasitol* 1999, **93:**809-812.
2. Therneau T: **coxme: Mixed Effects Cox Models. R package version 2.2-3.** 2012.
3. Bates D, Maechler M, Bolker B, Walker S: **lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5.** 2013.
4. Gething PW, Van Boeckel TP, Smith DL, Guerra CA, Patil AP, Snow RW, Hay SI: **Modelling the global constraints of temperature on transmission of *Plasmodium falciparum* and *P. vivax*.** *Parasit Vectors* 2011, **4:**92.
5. Viechtbauer W: **Conducting meta-analyses in R with the metafor package.** *J Stat Softw* 2010, **36:**1-48.

Additional File 4

Geographical variation in *Plasmodium vivax* relapse

Authors: Katherine E Battle[1§], Markku S Karhunen[1], Samir Bhatt[1], Peter W Gething[1], Rosalind E Howes[1], Nick Golding[1], Thomas P Van Boeckel[2], Jane P Messina[1], G Dennis Shanks[3], David L Smith[4,5], J Kevin Baird[6,7], Simon I Hay[1,5§]

[§]Authors for correspondence: katherine.battle@zoo.ox.ac.uk and simon.hay@zoo.ox.ac.uk

This file includes:

Individual-level relapse data literature references

References listed in Additional File 3

1. Adak T, Sharma VP, Orlov VS: **Studies on the *Plasmodium vivax* relapse pattern in Delhi, India**. *Am J Trop Med Hyg* 1998, **59**(1):175-179.
2. Alving AS, Pullman TN, Craige B, Jones R, Whorton M, Eichelberger L: **The clinical trial of eighteen analogues of Pamaquin (Plasmochin) in vivax malaria, Chesson strain**. *J Clin Invest* 1948, **27**(3):34-45.
3. Appavoo NC, Roy RG, Kapali V: **Results of 3-day radical treatment of *Plasmodium vivax* in North Arcot and South Arcot Districts of Tamil Nadu**. *Indian J Malariol* 1984, **21**(1):21-24.
4. Arnold J, Alving AS, Hockwald RS, Clayman CB, Dern RJ, Beutler E: **Natural history of Korean vivax malaria after deliberate inoculation of human volunteers**. *J Lab Clin Med* 1954, **44**(5):723-726.
5. Baird JK, Leksana B, Masbar S, Fryauff DJ, Sutanihardja MA, Suradi, Wignall FS, Hoffman SL: **Diagnosis of resistance to chloroquine by *Plasmodium vivax*: timing of recurrence and whole blood chloroquine levels**. *Am J Trop Med Hyg* 1997, **56**(6):621-626.
6. Baird JK, Sustriayu Nalim MF, Basri H, Masbar S, Leksana B, Tjitra E, Dewi RM, Khairani M, Wignall FS: **Survey of resistance to chloroquine by *Plasmodium vivax* in Indonesia**. *Trans R Soc Trop Med Hyg* 1996, **90**(4):409-411.
7. Baird JK, Tiwari T, Martin GJ, Tamminga CL, Prout TM, Tjaden J, Bravet PP, Rawlins S, Ferrel M, Carucci D *et al*: **Chloroquine for the treatment of uncomplicated malaria in Guyana**. *Annals of Tropical Medicine & Parasitology* 2002, **96**(4):339-348.
8. Basavaraj HR: **Observations on the treatment of 678 malaria cases with primaquine in an area free from malaria transmission in Mysore State, India**. *Indian J Malariol* 1960, **14**(3):269-281.
9. Berliner RW, Earle DP, Taggart JV, Welch WJ, Zubrod CG, Knowlton P, Atchley JA, Shannon JA: **Studies on the chemotherapy of the human malarias. VII. The antimalarial activity of pamaquine**. *J Clin Invest* 1948, **27**(3 Pt 2):108-113.
10. Coatney GR, Cooper WC, et al.: **Studies in human malaria. VII. The protective and therapeutic action of quinine sulfate against St. Elizabeth strain vivax malaria**. *Am J Hyg* 1948, **47**(1):120-134.
11. Coatney GR, Cooper WC, Eyles DE, Culwell WB, White WC, Lints HA: **Studies in human malaria. XXVII. Observations on the use of pentaquine in the prevention and treatment of Chesson strain vivax malaria**. *J Natl Malar Soc* 1950, **9**(3):222-233.
12. Coatney GR, Cooper WC, Young MD: **Studies in human malaria. XXX. A summary of 204 sporozoite-induced infections with the Chesson strain of *Plasmodium vivax***. *J Natl Mal Soc* 1950, **9**(381-396).
13. Coatney GR, Cooper WC, Young MD, Burgess RW: **Studies in human malaria. IV. The suppressive action of a phenanthrene amino alcohol, NIH-204 (SN-1796) against sporozoite-induced vivax malaria (St. Elizabeth strain)**. *Am J Hyg* 1947, **46**(1):132-140.
14. Coatney GR, Cooper WC, Young MD, Burgess RW, Smarr RG: **Studies in human malaria. II. The suppressive action of sulfadiazine and sulfapyrazine against sporozoite-induced vivax malaria (St. Elizabeth strain)**. *Am J Hyg* 1947, **46**(1):105-118.

15. Coatney GR, Ruhe DS, et al.: **Studies in human malaria. X. The protective and therapeutic action of chloroquine (SN 7618) against St. Elizabeth strain vivax malaria**. *Am J Hyg* 1949, **49**(1):49-59.
16. Contacos PG, Collins WE, Jeffery GM, Krotoski WA, Howard WA: **Studies on the characterization of *Plasmodium vivax* strains from Central America**. *Am J Trop Med Hyg* 1972, **21**(5):707-712.
17. Contacos PG, Coatney GR, Collins WE, Briesch PE, Jeter MH: **Five day primaquine therapy--an evaluation of radical curative activity against vivax malaria infection**. *Am J Trop Med Hyg* 1973, **22**(6):693-695.
18. Contacos PG, Collins WE, Chin W, Jeter MH, Briesch PE: **Combined chloroquine-primaquine therapy against vivax malaria**. *Am J Trop Med Hyg* 1974, **23**(2):310-312.
19. Cooper WC, Coatney GR, Culwell WB, Eyles DE, Young MD: **Studies in human malaria. XXVI. Simultaneous infection with the Chesson and the St. Elizabeth strains of *Plasmodium vivax***. *J Natl Malar Soc* 1950, **9**(2):187-190.
20. Cooper WC, Coatney GR, Jeffery GM, Imboden CA, Jr.: **Studies in human malaria. XXVIII. Observations on the use of chloroguanide against the Chesson strain of *Plasmodium vivax***. *J Natl Malar Soc* 1950, **9**(4):366-376.
21. Cooper WC, Ruhe DS, Coatney CR, Josephson ES, Young MD: **Studies in human malaria. VIII. The protective and therapeutic action of quinacrine against St. Elizabeth strain vivax malaria**. *Am J Hyg* 1949, **49**(1):25-40.
22. de Araujo FC, de Rezende AM, Fontes CJ, Carvalho LH, Alves de Brito CF: **Multiple-clone activation of hypnozoites is the leading cause of relapse in *Plasmodium vivax* infection**. *PLoS One* 2012, **7**(11):e49871.
23. Downs WG: **Results in an infantry regiment of several plans of treatment for vivax malaria**. *Am J Trop Med Hyg* 1946, **26**:67-86.
24. Dua VK, Sharma VP: ***Plasmodium vivax* relapses after 5 days of primaquine treatment, in some industrial complexes of India**. *Annals of Tropical Medicine & Parasitology* 2001, **95**(7):655-659.
25. Gentzkow CJ, Callender GR: **Malaria in the the Panama Canal Department, United States Army II. results of treatment with quinine, atabrine and plasmochin**. *Am J Epidemiol* 1938, **28**(2):174-189.
26. Gogtay NJ, Desai S, Kamtekar KD, Kadam VS, Dalvi SS, Kshirsagar NA: **Efficacies of 5- and 14-day primaquine regimens in the prevention of relapses in *Plasmodium vivax* infections**. *Ann Trop Med Parasitol* 1999, **93**(8):809-812.
27. Gogtay NJ, Desai S, Kadam VS, Kamtekar KD, Dalvi SS, Kshirsagar NA: **Relapse pattern of *Plasmodium vivax* in Mumbai: a study of 283 cases of vivax malaria**. *J Assoc Physicians India* 2000, **48**(11):1085-1086.
28. Gonzalez-Ceron L, Mu J, Santillan F, Joy D, Sandoval MA, Camas G, Su X, Choy EV, Torreblanca R: **Molecular and epidemiological characterization of *Plasmodium vivax* recurrent infections in southern Mexico**. *Parasit Vectors* 2013, **6**(1):109.
29. Gordon HH, Dieuaide FR, Marble A, Christianson HB, Dahl LK: **Treatment of *Plasmodium vivax* malaria of foreign origin. A comparison of various drugs**. *Arch Intern Med* 1947, **79**(4):365-380.
30. Guthmann JP, Pittet A, Lesage A, Imwong M, Lindegardh N, Min Lwin M, Zaw T, Annerberg A, de Radigues X, Nosten F: ***Plasmodium vivax* resistance to chloroquine in Dawei, southern Myanmar**. *Trop Med Int Health* 2008, **13**(1):91-98.
31. Horing RO: **Induced and war malaria**. *J Trop Med Hyg* 1947, **50**(8):150-159.

32. Imboden CA, Jr., Cooper WC, Coatney GR, Jeffery GM: **Studies in human malaria. XXIX. Trials of aureomycin, chloramphenicol, penicillin, and dihydrostreptomycin against the Chesson strain of *Plasmodium vivax***. *J Natl Mal Soc* 1950, **9**(4):377-380.
33. James SP: **A malaria survey and some epidemiological observations in England**. In: *Malaria at Home and Abroad.* London: John Bale, Sons & Danielsson, LTD.; 1920: 80-93.
34. James SP: **Some general results of a study of induced malaria in England**. *Trans R Soc Trop Med Hyg* 1931, **24**(5):478-525.
35. Jeffery GM: **Relapses with Chesson strain *Plasmodium vivax* following treatment with chloroquine**. *Am J Trop Med Hyg* 1956, **5**(1):1-13.
36. Jiang JB, Huang JC, Liang DS, Liu JX, Zhang SW, Cheng FC: **Long incubation of *Plasmodium vivax multinucleatum* as demonstrated in three experimental human cases**. *Trans R Soc Trop Med Hyg* 1982, **76**(6):845-847.
37. Jones R, Jr., Pullman TN, et al.: **The therapeutic effectiveness of large doses of paludrine in acute attacks of sporozoite-induced vivax malaria, Chesson strain**. *J Clin Invest* 1948, **27**(3 Pt1):51-55.
38. Ketema T, Getahun K, Bacha K: **Therapeutic efficacy of chloroquine for treatment of *Plasmodium vivax* malaria cases in Halaba district, South Ethiopia**. *Parasit Vectors* 2011, **4**:46.
39. Lints HA, Coatney CR, Cooper WC, Culwell WB, White WC, Eyles DE: **Studies in human malaria. XXII. Prolonged suppression of Chesson strain vivax malaria by the weekly administration of chlorguanide or chloroquine**. *J Natl Mal Soc* 1950, **9**(1):50-58.
40. London IM, Lavietes PH, et al.: **The effects of quinacrine (atabrine) suppression on the course of Pacific vivax malaria**. *Am J Med* 1946, **1**(6):615-620.
41. Maisonneuve H, Joly F, John M, Carles G, Rossignol JF: **[Efficacy of halofantrine in *Plasmodium falciparum* or *Plasmodium vivax* malaria in a resistance area (French Guiana)]**. *Presse Med* 1988, **17**(3):99-102.
42. USA, MRU: **Malaria Report No. 306**. In: *Board for the Coordination of Malarial Studies: Relapse rates of benign tertian malaria following therapy.* Washington, D.C. and Oxford; 1944.
43. Alving AS: **Malaria Report No. 445**. In: *Board for the Coordination of Malarial Studies: Summary of clinical tests at stateville on 29 June 1945.* Washington, D.C.; 1945.
44. Alving AS: **Malaria Report No. 563**. In: *Board for the Coordination of Malarial Studies: Summar of clinical tests at Stateville on 13 December 1945* Washington, D.C.; 1945.
45. Alving AS: **Malaria Report No. 645**. In: *Board for the Coordination of Malarial Studies: Summary of clinical tests at Stateville.* Washington, D.C.; 1946.
46. Alving AS: **Malaria Report No. 672**. In: *Board for the Coordination of Malarial Studies: Summary of clinical tests at Stateville on 1 April 1946.* Washington, D.C.; 1946.
47. Alving AS: **Malaria Report No. 680**. In: *Board for the Coordination of Malarial Studies: Summary of clinical tests at Stateville on 1 May 1946.* Washington, D.C.; 1946.
48. Alving AS: **Malaria Report No. 702**. In: *Board for the Coordination of Malarial Studies: Summary of clinical tests at Stateville on 1 June 1946.* Washington, D.C.; 1946.

49. Mason J: **Patterns of *Plasmodium vivax* recurrence in a high-incidence coastal area of El Salvador, C. A**. *Am J Trop Med Hyg* 1975, **24**(4):581-585.
50. McLester JB: **Studies of relapse in malaria**. In: *Proceedings of the Conference of Army Physicians, Central Mediterranean Forces: 1945; Rome*: Institute Superiore di Sanita Viale Regina Marguerita; 1945: 4-7.
51. Most H, Kane CA, et al.: **Combined quinine-plasmochin treatment of vivax malaria; effect of relapse rate**. *Am J Med Sci* 1946, **212**(5):550-560.
52. Most H, London IM, Kane CA, Lavietes PH, Schroeder EF, Hayman JM: **Chroloquine for treatment of acute attacks fo vivax-malaria**. *JAMA* 1946, **131**(12):963-967.
53. Nateghpour M, Mavi SA, Keshavarz H, Rezaei S, Abedi F, Edrissian G, Raeisi A: **Molecular monitoring of *Plasmodium vivax* infection after radical treatment in southeastern Iran**. *Iran J Arthropod Borne Dis* 2010, **4**(1):24-30.
54. Pukrittayakamee S, Vanijanonta S, Chantra A, Clemens R, White NJ: **Blood stage antimalarial efficacy of primaquine in *Plasmodium vivax* malaria**. *J Infect Dis* 1994, **169**(4):932-935.
55. Pullman TN, Craige B, Jr., et al.: **Comparison of chloroquine, quinacrine (atabrine) and quinine in the treatment of acute attacks of sporozoite-induced vivax malaria, Chesson strain**. *J Clin Invest* 1948, **27**(3 Pt1):46-50.
56. Ranque S, Badiaga S, Delmont J, Brouqui P: **Triangular test applied to the clinical trial of azithromycin against relapses in *Plasmodium vivax* infections**. *Malar J* 2002, **1**:13.
57. Roy RG, Chakrapani KP, Dhinagaran D, Sitaraman NL, Ghosh RB: **Efficacy of 5-day radical treatment of *P. vivax* infection in Tamil Nadu**. *Indian J Med Res* 1977, **65**(5):652-656.
58. Roy RG, Shanmugham CA, Chakrapani KP, Ganesan AV: **Results of 5-day course of radical treatment of *Plasmodium vivax* in six districts of Tamil Nadu**. *Indian J Med Res* 1979, **69**:939-943.
59. Ruhe DS, Cooper WC, Coatney GR, Josephson ES, Young MD: **Studies in human malaria: IX. The protective and therapeutic action of SN 6911 (Sontochin) against St. Elizabeth stain vivax malaria**. *Am J Hyg* 1949, **49**:41-48.
60. Ruhe DS, Cooper WC, Coatney CR, Josephson ES: **Studies in human malaria. XII. The protective and therapeutic action of SN 5241 against St. Elizabeth strain vivax malaria**. *Am J Hyg* 1949, **49**(3):346-354.
61. Ruhe DS, Cooper WC, Coatney GR, Josephson ES: **Studies in human malaria: XIV. The ineffectiveness of Colchicine (SN 12,080), SN 7266 and SN 8557 as curative agents against St. Elizabeth strain vivax malaria**. *Am J Hyg* 1949, **49**:361-366.
62. Saint-Yves IF: **Comparison of treatment schedules for *Plasmodium vivax* infections in the Solomon Islands**. *P N G Med J* 1977, **20**(2):62-65.
63. Shannon JA: **Rationale underlying the clinical evaluation of antimalarial drugs**. In: *A Survey of Antimalarial Drugs: 1941–1945*. Edited by Wiselogle FY. Ann Arbor, Michigan: J.W. Edwards; 1946: 177-220.
64. Shannon JA, Earle DP, Berliner RW, Taggart JV: **Studies on the chemotherapy of the human malarias. I. Method for the quantitative assay of suppressive antimalarial action in vivax malaria**. *J Clin Invest* 1948, **27**(3):66-74.
65. Sharma MID, Sehgal PN, Vaid BK, Dubey RC, Nagendra S, Paithne PK, Joshi ML: **Effectiveness of drug schedule being followed under the National Malaria Eradication Programme, India, for radical cure of vivax malaria cases**. *J Commun Dis* 1973, **5**(4):167-174.

66. Sharma RC, Gautam AS, Orlov V, Sharma VP: **Relapse pattern of *Plasmodium vivax* in Kheda district, Gujarat**. *Indian J Malariol* 1990, **27**(2):95-99.
67. Shute PG: **Latency and long-term relapses in benign tertian malaria**. *Trans R Soc Trop Med Hyg* 1946, **40**(2):189-200.
68. Singh J, Ray AP, Basu PC, Nair CP: **Preliminary studies on 8-aminoquinolines**. *Indian J Malariol* 1953, **7**(3):289-294.
69. Singh J, Ray AP, Misra BG, Nair CP: **Antirelapse treatment with primaquine and pyrimethamine**. *Indian J Malariol* 1954, **8**(2):127-136.
70. Singh N, Mishra AK, Sharma VP: **Radical treatment of vivax malaria in Madhya Pradesh, India**. *Indian J Malariol* 1990, **27**(1):55-56.
71. Sinha S, Dua VK, Sharma VP: **Efficacy of 5 day radical treatment of primaquine in *Plasmodium vivax* cases at the BHEL industrial complex, Hardwar (U.P.)**. *Indian J Malariol* 1989, **26**(2):83-86.
72. Sinton JA: **Studies in malaria with special reference to treatment. VII. The intravenous injection of sodium stovarsol in the treatment of benign tertian malaria**. *Indian J Med Res* 1927, **15**:287-299.
73. Srivastava HC, Yadav RS, Joshi H, Valecha N, Mallick PK, Prajapati SK, Dash AP: **Therapeutic responses of *Plasmodium vivax* and *P. falciparum* to chloroquine, in an area of western India where *P. vivax* predominates**. *Ann Trop Med Parasitol* 2008, **102**(6):471-480.
74. Sumawinata IW, Bernadeta, Leksana B, Sutamihardja A, Purnomo, Subianto B, Sekartuti, Fryauff DJ, Baird JK: **Very high risk of therapeutic failure with chloroquine for uncomplicated *Plasmodium falciparum* and *P. vivax* malaria in Indonesian Papua**. *Am J Trop Med Hyg* 2003, **68**(4):416-420.
75. Sutanto I, Suprijanto S, Nurhayati, Manoempil P, Baird JK: **Resistance to chloroquine by *Plasmodium vivax* at Alor in the Lesser Sundas Archipelago in eastern Indonesia**. *Am J Trop Med Hyg* 2009, **81**(2):338-342.
76. Swellengrebel NH, De Buck A: **Plasmoquine prophylaxis in benign tertian malaria**. *Proc Koninklijke Nederlandse Akademie Wetenschappen* 1932, **35**:912-914.
77. Teka H, Petros B, Yamuah L, Tesfaye G, Elhassan I, Muchohi S, Kokwaro G, Aseffa A, Engers H: **Chloroquine-resistant *Plasmodium vivax* malaria in Debre Zeit, Ethiopia**. *Malar J* 2008, **7**:220.
78. Tiburskaja NA, Vrublevskaja OS: **The course of infection caused by the North Korean strain of *Plasmodium vivax***. *WHO/MAL/77* 1977, **895**:1-19.
79. Tiburskaja NA, Sergiev PG, Vrublevskaja OS: **Dates of onset of relapses and the duration of infection in induced tertian malaria with short and long incubation periods**. *Bull World Health Organ* 1968, **38**(3):447-457.
80. Trager W, Bang FB, Hairston NG: **The effect of four different therapies on the relapse rate of vivax malaria**. *Am J Hyg* 1947, **45**(1):43-57.
81. Van den Eede P, Soto-Calle VE, Delgado C, Gamboa D, Grande T, Rodriguez H, Llanos-Cuentas A, Anne J, D'Alessandro U, Erhart A: ***Plasmodium vivax* sub-patent infections after radical treatment are common in Peruvian patients: results of a 1-year prospective cohort study**. *PLoS One* 2011, **6**(1):e16257.
82. Villalobos-Salcedo JM, Tada MS, Kimura E, Menezes MJ, Pereira da Silva LH: **In-vivo sensitivity of *Plasmodium vivax* isolates from Rondonia (western Amazon region, Brazil) to regimens including chloroquine and primaquine**. *Ann Trop Med Parasitol* 2000, **94**(8):749-758.
83. Walsh DS, Looareesuwan S, Wilairatana P, Heppner DG, Jr., Tang DB, Brewer TG, Chokejindachai W, Viriyavejakul P, Kyle DE, Milhous WK *et al*: **Randomized dose-ranging study of the safety and efficacy of WR 238605 (Tafenoquine) in the**

prevention of relapse of *Plasmodium vivax* malaria in Thailand. *J Infect Dis* 1999, **180**(4):1282-1287.

84. White WC, Cooper WC, et al.: **Studies in human malaria. XXI. The cure of St. Elizabeth strain vivax malaria by pentaquine-quinine, administered during acute attacks or during latency**. *J Natl Malar Soc* 1948, **7**(4):316-321.
85. Yadav RS, Ghosh SK: **Radical curative efficacy of five-day regimen of primaquine for treatment of *Plasmodium vivax* malaria in India**. *J Parasitol* 2002, **88**(5):1042-1044.
86. Yeshiwondim AK, Tekle AH, Dengela DO, Yohannes AM, Teklehaimanot A: **Therapeutic efficacy of chloroquine and chloroquine plus primaquine for the treatment of *Plasmodium vivax* in Ethiopia**. *Acta Trop* 2010, **113**(2):105-113.
87. Yorke W: **Further observations on malaria made during treatment of general paralysis**. *Trans R Soc Trop Med Hyg* 1926, **19**(3):108-131.

Appendix to Chapter 4 – Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease

---

This Appendix includes:

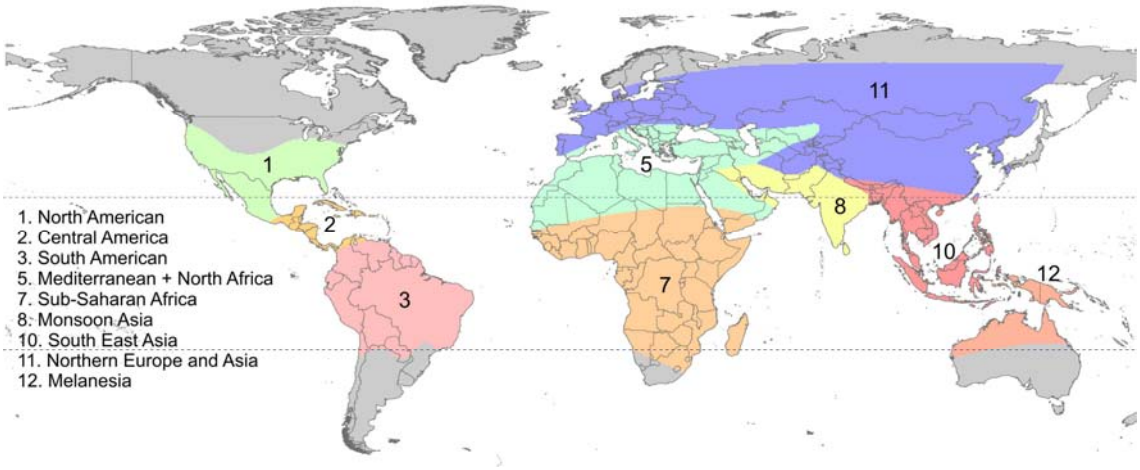
1. Additional File 1: Geographic zones of relapse phenotype. Relapse patterns of strains of *P. vivax* are proposed to differ among the nine ecological zones shown.
2. Additional File 2: Incidence records plotted versus the predicted MAP-based *Pv*PR values and observed concurrent *Pv*PR values. Incidence points versus MAP *Pv*PR values are shown in black and those points using concurrently measured *Pv*PR values are shown in blue.
3. Additional File 3: Approximate and exact person-time shown in plots of incidence per 1,000 person-years versus parasite rate. The incidence records with concurrent *Pv*PR estimates are plotted below on linear (A) and log scales (B) below. The blue points are those with approximated person time and those in grey had exact person-time reported.
4. Additional File 4: Case parasite density threshold shown in scatter plots of incidence per 1,000 person-years versus parasite rate. The incidence records are plotted below on linear (A) and log scales (B) below. The grey points are studies that used any parasitaemia in the case definition. Blue points are studies that defined a case as ≥500 parasites/μl of blood and red points, 1000 parasites/μl.
5. Additional File 5: The posterior for the non-parametric fitted function giving the impact of ACD frequency on the rate of detected clinical incidence cases. A

non-parametric statistical distribution the frequency of ACD was fit under a monotonicity restriction, which forces the posterior to preserve a strict ordering of the observed incidence scaling with respect to ACD frequency.

6. Associated publication: Battle, K.E., Guerra, C.A., Golding, N., Duda, K.A., Cameron, E., Howes, R.E., Elyazar, I.R.F., Baird, J.K., Reiner Jr, R.C., Smith, D.L., Gething, P.W. and Hay, S.I. (2015). Global database of *Plasmodium falciparum* and *P. vivax* incidence records, 1985-2013. *Scientific Data* 2, 150012.

Additional Files: Supplementary Figures 1-5.

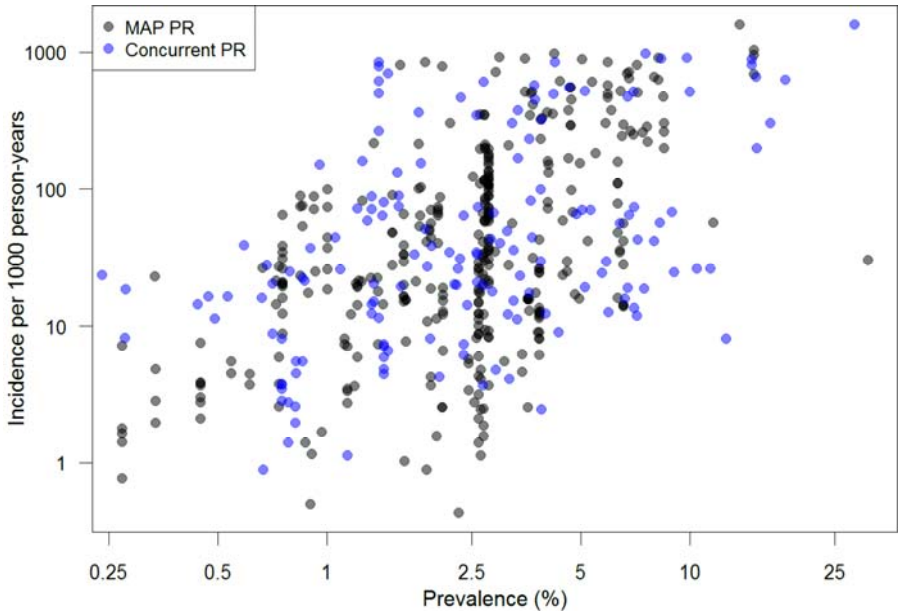
Additional File 1



Title: Geographic zones of relapse phenotype.

Description: Relapse patterns of strains of *P. vivax* are proposed to differ among the nine ecological zones shown above [28].

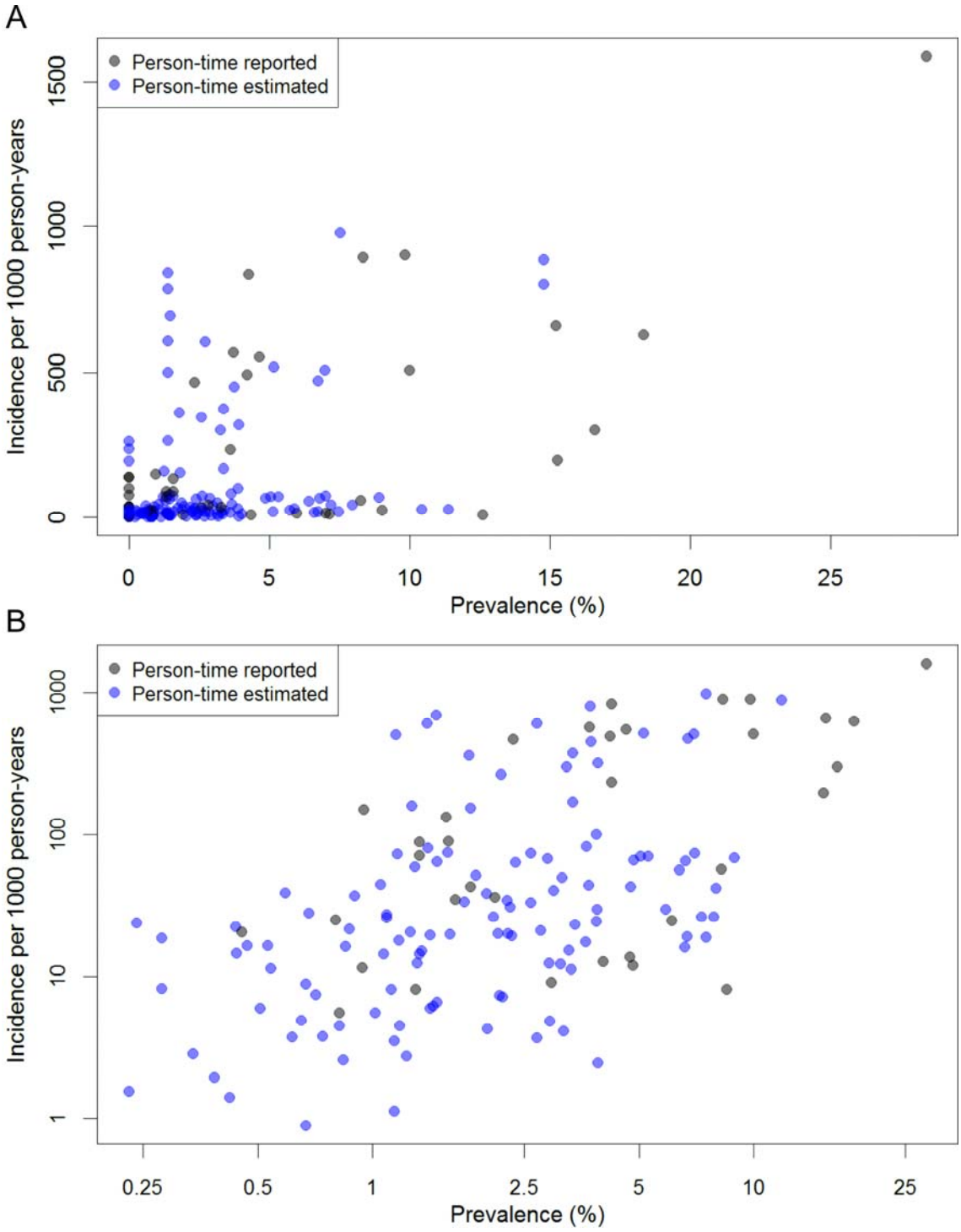
Additional File 2



Title: Incidence records plotted versus the predicted MAP-based *Pv*PR values and observed concurrent *Pv*PR values.

Description: Incidence points versus MAP *Pv*PR values are shown in black and those points using concurrently measured *Pv*PR values are shown in blue.

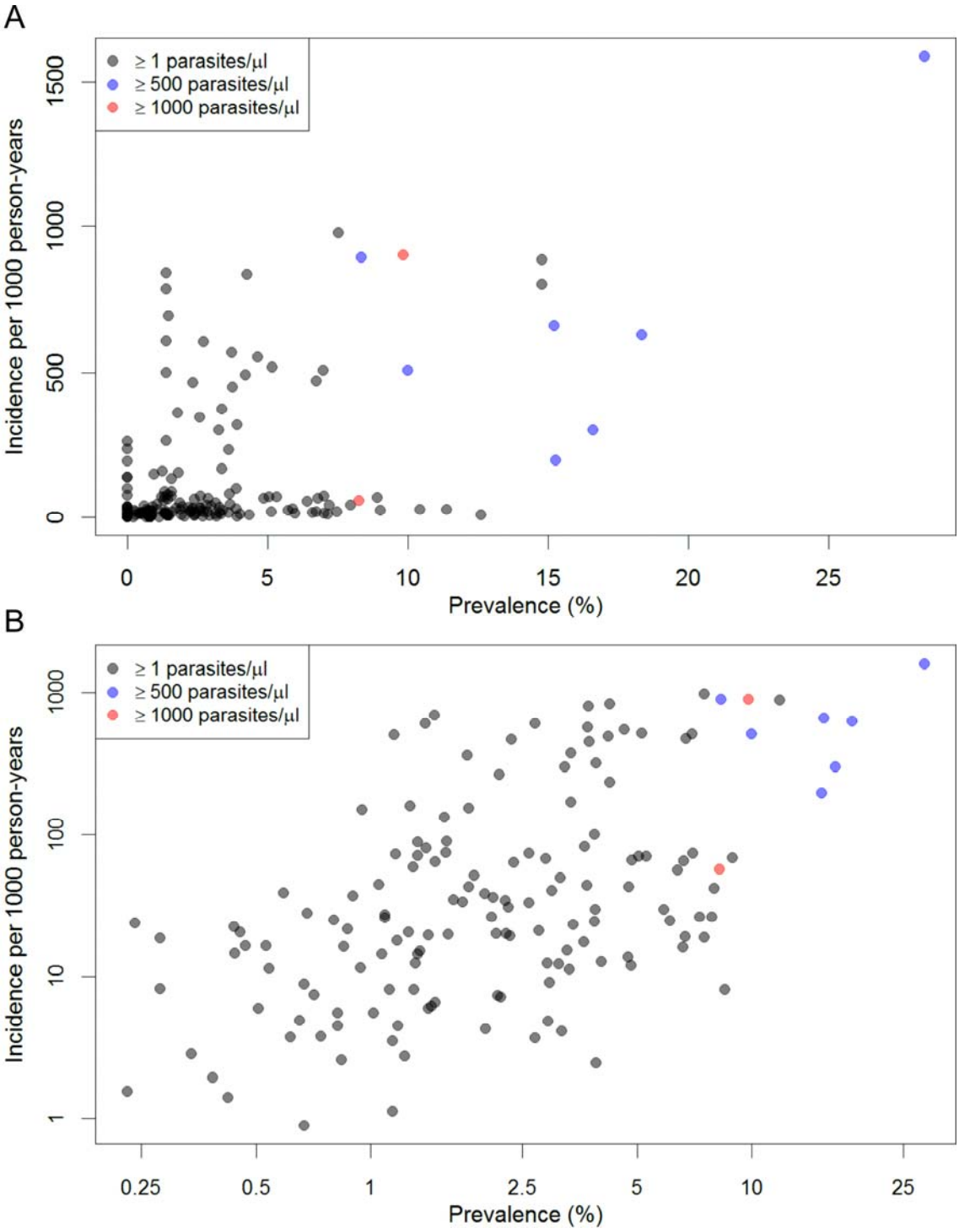
Additional File 3



Title: Approximate and exact person-time shown in plots of incidence per 1,000 person-years versus parasite rate.

Description: The incidence records with concurrent *Pv*PR estimates are plotted below on linear (A) and log scales (B) below. The blue points are those with approximated person time and those in grey had exact person-time reported.

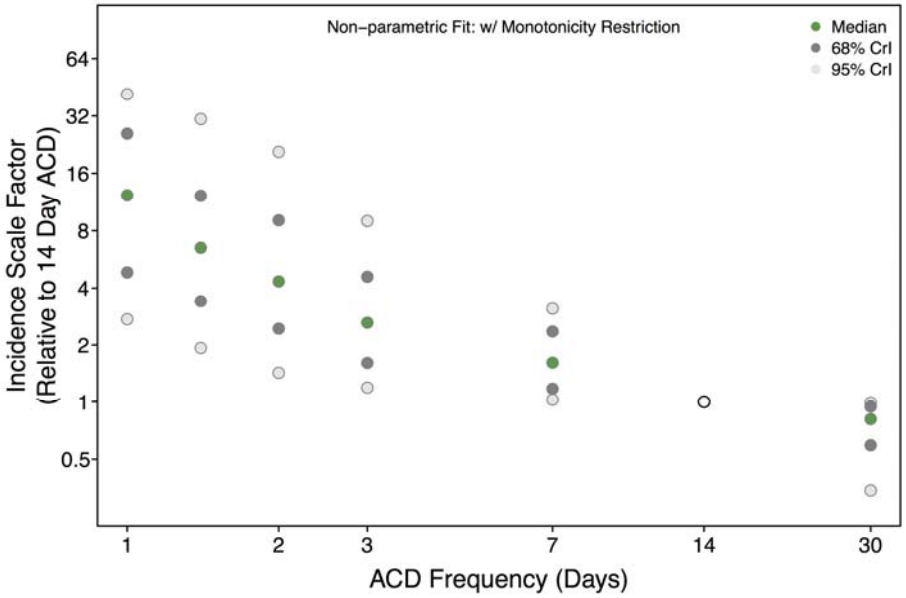
Additional File 4



Title: Case parasite density threshold shown in scatter plots of incidence per 1,000 person-years versus parasite rate.

Description: The incidence records are plotted below on linear (A) and log scales (B) below. The grey points are studies that used any parasitaemia in the case definition. Blue points are studies that defined a case as ≥500 parasites/µl of blood and red points, 1000 parasites/µl.

Additional File 5



Title: The posterior for the non-parametric fitted function giving the impact of ACD frequency on the rate of detected clinical incidence cases.

Description: A non-parametric statistical distribution the frequency of ACD was fit under a monotonicity restriction, which forces the posterior to preserve a strict ordering of the observed incidence scaling with respect to ACD frequency.

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Epidemiology

» Malaria

» Literature mining

» Data mining

Global database of matched *Plasmodium falciparum* and *P. vivax* incidence and prevalence records from 1985–2013

Katherine E. Battle[1], Carlos A. Guerra[2], Nick Golding[1], Kirsten A. Duda[1], Ewan Cameron[1], Rosalind E. Howes[1], Iqbal R.F. Elyazar[3], J. Kevin Baird[3,4], Robert C. Reiner Jr.[5], Peter W. Gething[1], David L. Smith[1,6] & Simon I. Hay[1,6]

Received: 22 December 2014

Accepted: 02 March 2015

Published: 18 August 2015

Measures of clinical incidence are necessary to help estimate the burden of a disease. Incidence is a metric not commonly measured in malariology because the longitudinal surveys required are costly and labour intensive. This database is an effort to collate published incidence records obtained using active case detection for *Plasmodium falciparum* and *Plasmodium vivax* malaria. The literature search methods, data abstraction procedures and data processing procedures are described here. A total of 1,680 spatio-temporally unique incidence records were collected for the database: 1,187 for *P. falciparum* and 493 for *P. vivax*. These data were gathered to model the relationship between clinical incidence and prevalence of infection and can be used for a variety of modelling exercises including the assessment of change in disease burden in relation to age and control interventions. The subset of data that have been used for such modelling exercises are described and identified.

Design Type(s)	observation design • epidemiological study • data integration objective
Measurement Type(s)	infectious disease incidence
Technology Type(s)	Epidemiology
Factor Type(s)	
Sample Characteristic(s)	*Plasmodium falciparum* • anthropogenic habitat • *Plasmodium vivax*

[1]Spatial Ecology and Epidemiology Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. [2]Sanaria Institute for Global Health and Tropical Medicine, Rockville, Maryland 20850, USA. [3]Eijkman-Oxford Clinical Research Unit, Jalan Diponegoro No 69, Jakarta 10430, Indonesia. [4]Nuffield Department of Medicine, Centre for Tropical Medicine, University of Oxford, Oxford OX3 7FZ, UK. [5]Indiana University School of Public Health, Bloomington, Indiana 47405, USA. [6]Fogarty International Center, National Institutes of Health, Bethesda, Maryland 20892, USA. Correspondence and requests for materials should be addressed to K.E.B. (email: katherine.battle@zoo.ox.ac.uk)

Background & Summary

The global clinical burden of malaria has proven difficult to enumerate. Previous efforts have estimated the clinical incidence of *Plasmodium falciparum* malaria using adjusted case reports[1,2], modelling based on study-level incidence and mortality rates[3] and cartographic modelling techniques[4–6]. All of these methods require malaria incidence measures in some part of their estimation or validation procedures. The surveillance-based approach relies on routinely reported case numbers, which are adjusted by country to account for incomplete reporting, the proportion of cases confirmed with routine diagnostics, and health facility use and access[1]. The Global Burden of Disease (GBD) study uses an amalgam of three methods to estimate case incidence. First, reported cases are used from a small subset of countries deemed to have reliable reporting. Second, corrected reporting (similar to the surveillance-based approach methods) is applied to a larger set of countries. And third, study-level incidence, such as the records reported here, along with mortality rates and a variety of parameters such as age and detection methods are used to model country-level incidence[3]. The cartographic approach, employed by the Malaria Atlas Project (MAP), also uses a tiered approach. Case estimates from countries with reliable estimate are used directly. Regions designated as unstable transmission regions[6,7] are assigned an incidence of 0.1 cases per 1,000 per year. For regions without accurate reporting in areas stable malaria transmission, a modelled relationship between study-level measures of incidence matched in space and time to prevalence surveys is applied to a smooth endemicity (prevalence) surface[7] and multiplied by a global population grid[4]. Such matched incidence and prevalence data are presented here.

Whilst approximations of 'non-*P. falciparum*' malaria exist, the burden of *Plasmodium vivax* malaria is considered to be largely unknown[8–10]. The primary reason for these knowledge gaps is that measures of malaria incidence are rarely undertaken as they are logistically demanding and thus expensive. To accurately measure clinical incidence of malaria, longitudinal studies must be conducted that include regular visits made to communities to check for symptomatic individuals through active case detection (ACD)[11]. The database described here aimed to compile as many ACD studies for *P. falciparum* and *P. vivax* as possible from 1985 to 2013 and represents a significant expansion upon previously published assemblies of incidence data[5].

A more commonly measured malaria metric is prevalence, also known as parasite rate (PR)[12]. As demonstrated by Patil *et al.*[5], Cameron *et al.*[13] and Battle *et al.*[14], ACD incidence records can be matched to PR measures to model the relationship between prevalence of infection and incidence of clinical disease. These models then have the potential to transform existing endemicity maps[7,15] into global burden estimates with known precision[5]. For this purpose, each ACD observation in this database has been matched to a concurrently measured PR value or an extracted spatially and temporally matched modelled PR[7,15].

All data curation and abstraction procedures to obtain the 1,680 incidence records, including geo-positioning and prevalence matching, are described here. The structure of the final database and technical validation efforts are also described along with notes to facilitate the replication of the analyses in Cameron *et al.*[13] and Battle *et al.*[14] Such validation yields a powerful mathematical tool supporting efforts to reliably estimate global burdens of disease imposed by the parasites causing human malaria.

Methods

Data collection

Here we provide additional detail on methodology to that provided in Cameron *et al.*[13] and Battle *et al.*[14], which utilize only a subset of the data presented here. PubMed (http://www.ncbi.nlm.nih.gov/pubmed) was searched on 27 November 2013 using the following search string: ((malaria[MeSH Terms]) AND ('Incidence' [Mesh] OR 'Epidemiology' [Mesh] OR 'epidemiology' [Subheading])) AND ('1985/01/01'[Date—Publication]: '3000'[Date—Publication]). This selected references published after 1 January 1985 and the Medical Subject Headings (MeSH; http://www.ncbi.nlm.nih.gov/mesh) ensured that all pseudonyms were included in the search. The cut off of 1985 was used to match the year range of PR records included in the MAP database[16]. The literature search returned 11,272 citations and was augmented with a further 25 references from previously published analyses[5,17]. A total of 15 search strings were tested varying the terms used and application of MeSH terms. The number of references returned ranged from 58 (((malaria[MeSH Terms]) AND 'active case detection') AND ('1985/01/01'[Date—Publication]: '3000'[Date—Publication])) to 1,291,787 (((((malaria[MeSH Terms]) AND incidence[MeSH Terms]) OR epidemiology[MeSH Terms]) OR epidemiology[MeSH Subheading]) AND ('1985/01/01'[Date—Publication]: '3000'[Date—Publication])). The search string used was chosen because it was a reasonable number of titles for a small team to sort through, while still capturing the majority (87%, 55/63) of the references used in a previously published collection[5].

Abstracts were reviewed to determine if the reference might contain *P. falciparum* or *P. vivax* incidence data. References excluded at this stage included reviews, case studies, vector-only, reports on animal or non-*P. falciparum* or *P. vivax* malaria, reports of imported malaria and technical articles including mathematical modelling and genetic analyses. The full list of 11,297 references was narrowed down to 898 references for full text review. Seventy-eight references known to contain incidence data from the work by Patil *et al.*[5] and Griffin *et al.*[17] were also set aside.

Full texts were obtained for the 976 references identified for review. The criteria for inclusion were: (i) longitudinal studies using ACD, (ii) symptomatic/clinical cases were the subjects of detection,

(iii) studies were conducted in the general community (not patient sub-groups or hospital-based studies), and (iv) diagnosis using microscopy or rapid diagnostic test (RDT). Studies done only on pregnant women were excluded due to their increased susceptibility to malaria[18]. Conversely, studies using only infants aged less than three months were excluded due to their potential temporary immunity to *P. falciparum* from maternal antibodies[19]. *Plasmodium vivax* has been shown to cause significant morbidity in young infants[20,21], however none of the studies conducted on infants measured *P. vivax* incidence. In-house language skills only allowed for the inclusion of articles written in English, French, Portuguese and Spanish. Twelve publications out of the 976 identified for review were in other languages (one Turkish and 11 Chinese) and therefore excluded. Articles that did not have enough information to determine the number of cases and the person-time observed (length of follow up for each cohort member) were excluded. Initially, there were no restrictions placed on the length or frequency of follow up, as long as it was explicitly reported. Based on the above criteria, data were abstracted from a total of 230 references, the majority of which measured incidence of *P. falciparum* malaria. Data on *P. falciparum* and *P. vivax* incidence were identified in 226 and 99 of these references, respectively. Literature review procedures are outlined in Fig. 1.

Geo-positioning

All available location information was extracted from the published sources. Incidence records were first positioned to a MAP region: Africa+ (Africa plus Saudi Arabia and Yemen), the Americas, and Central and Southeast Asia (CSE Asia). The number of records by species and region are shown in Table 1. Next, they were assigned to a country (based on 2013 boundaries) and place. The place was considered the location of the study and latitude and longitude coordinates for each site were found using values given in the paper (converting to decimal degrees where necessary) or where this was unavailable, manually digitized using Google Maps[22] or Microsoft Encarta[23]. Contextual information from the paper was used to differentiate when two places with the same name were within one country or to narrow down a region to be scanned for names that could be different spellings or translations of the site name (for example, Sissé rather than Cisse). If the site was a village, town or city, the latitude and longitude were taken from centre, unless a specific part of the town or city were specified as the study site. If the only location information given was a larger area such as a district administrative unit, as defined by the Food and Agriculture Organization (FAO) Global Administrative Unit Layers (GAUL) coding[24], the centroid of the region was found using geographic information systems (GIS) software[25]. If the location could not be determined by any of these means, the authors were contacted for further information. The locations of each record are shown in Fig. 2 and the distribution of the data records over time are shown by country in Fig. 3. Records were also matched to zoo-epidemiological zones originally defined by Macdonald[26] and modified by Battle *et al.*[27] to describe the geographic variation observed in *P. vivax* relapse phenotypes (Fig. 4). These classifications enable the relationship between prevalence and incidence to be modelled separately by region.

Calculating incidence

Incidence describes the number of events that occur within a specified time period. In this context, the events are symptomatic cases (fever cases confirmed by RDT or microscopy) of *P. falciparum* or *P. vivax* identified within a study population. Cases of *P. vivax* may arise from new mosquito-borne infections, recrudescence from treatment failure or relapses from hypnozoites (the dormant liver stage). Because there is no reliable way to differentiate relapses from new infections or recrudescence, the incidence of *P. vivax* reported here includes the cases from all origins.

Values for the number of cases and time period the study population was observed (recorded in person-years) were needed for modelling purposes and every effort was therefore made to extract or derive those values. If incidence was given, then the number of cases or person-time was derived from other information provided in the publication. In the few studies that provided age-specific data without age-specific population data, a general population structure was applied[28] to the whole population to determine the size of the composite age groups. Where person-time was not explicitly reported, this was calculated by multiplying the population number by the length of the study period. This was necessary for the majority of studies for both *P. falciparum* (61%, 722/1187) and *P. vivax* (77%, 378/493).

Matching prevalence to incidence

So that the data may be used to model the relationship between prevalence (PR) and incidence of clinical disease, an estimate of prevalence was spatially and temporally matched to each incidence record in the database. If the incidence publication source contained empirical prevalence data, this was abstracted to provide a space-time match between PR and incidence. Averages were taken for those studies that reported more than one cross-sectional survey (XSS) for the same community. The number and timing of each survey was recorded where available. Half of the records had PR data available from the same reference (840/1680). Some additional space-time prevalence matches were added using the MAP PR database[16] if a separate publication measured PR in the same community and year as the incidence data (6%, 97/1680). For the remaining 44% of the data (743/1680) without concurrent prevalence data, PR values were extracted for all incidence records using 2010 point estimates of the annual mean modelled *P. falciparum*[7] and *P. vivax*[15] endemicity values (shown in Fig. 1) using GIS software[25]. For *P. vivax*, areas

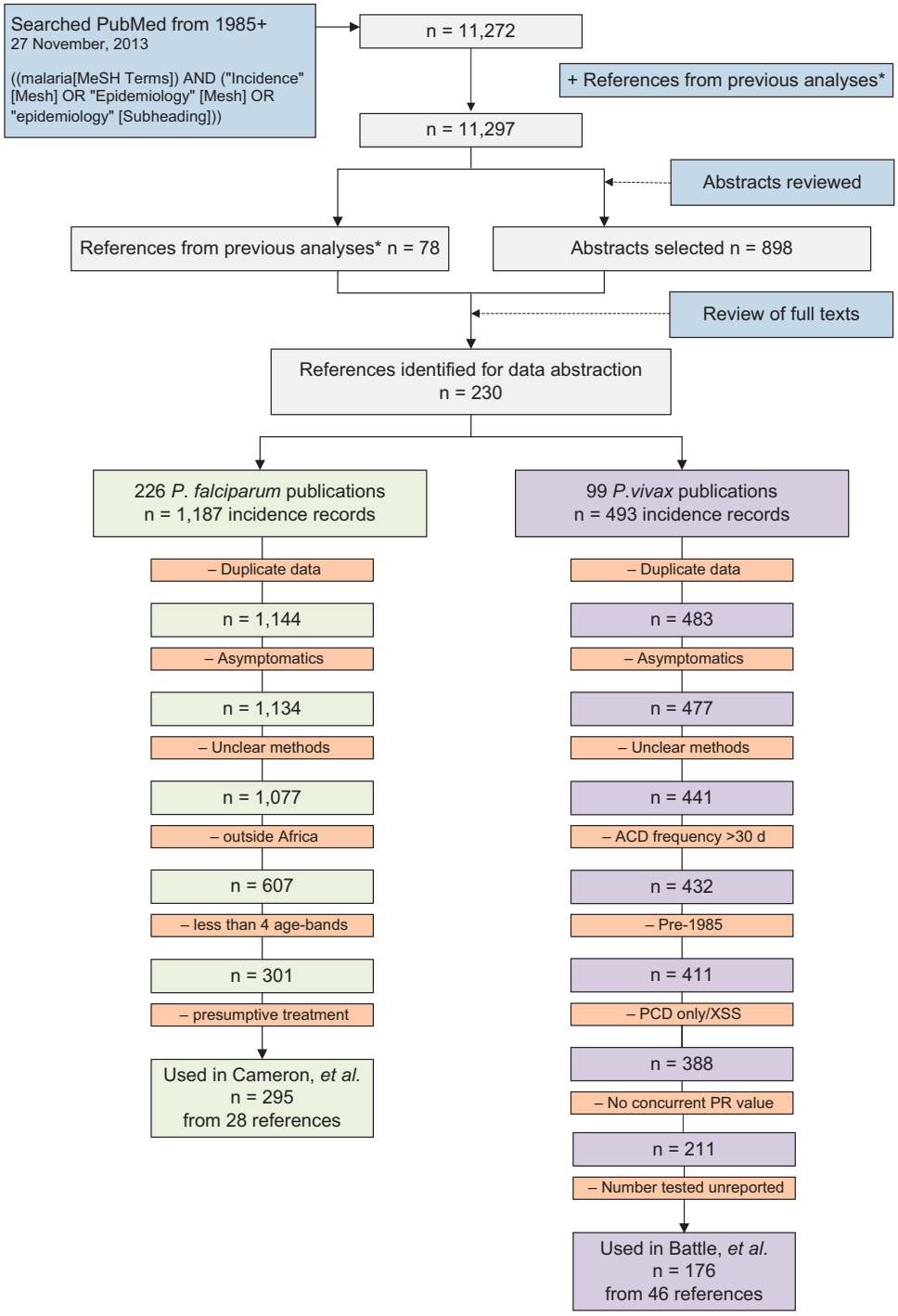


Figure 1. Schematic overview of the literature search procedure and results. The data exclusions to obtain clinical incidence records of use for model implementation for the *P. falciparum* (Cameron *et al.*[13]) and *P. vivax* (Battle *et al.*[14]) models are also shown. References from previous analyses* include those used by Patil *et al.*[5] and Griffin *et al.*[17]

where Duffy negative allele frequency was predicted to exceed 90% (ref. 29) is shown in hatching and *P. vivax* PR was predicted at $< 1\%$ in those areas. The PR values for *P. falciparum* were predicted for two to ten year-olds, whereas for *P. vivax*, the predictions were made for all ages (one to 99 years). Because these age ranges did not always correspond to age groups in the incidence data, these data were age standardized using a bespoke model parameterized for *P. falciparum*[30] and *P. vivax*[15]. The age standardization model was also applied as needed to PR values obtained from publications, as not all prevalence records had matching age ranges to the incidence data. The number of parasite positive

Region	*P. falciparum*	*P. vivax*	Total
Africa+	661	11	672
America	117	106	223
CSE Asia	409	376	785
Total	1,187	493	1,680

Table 1. Data records for *P. falciparum* and *P. vivax* by MAP region.

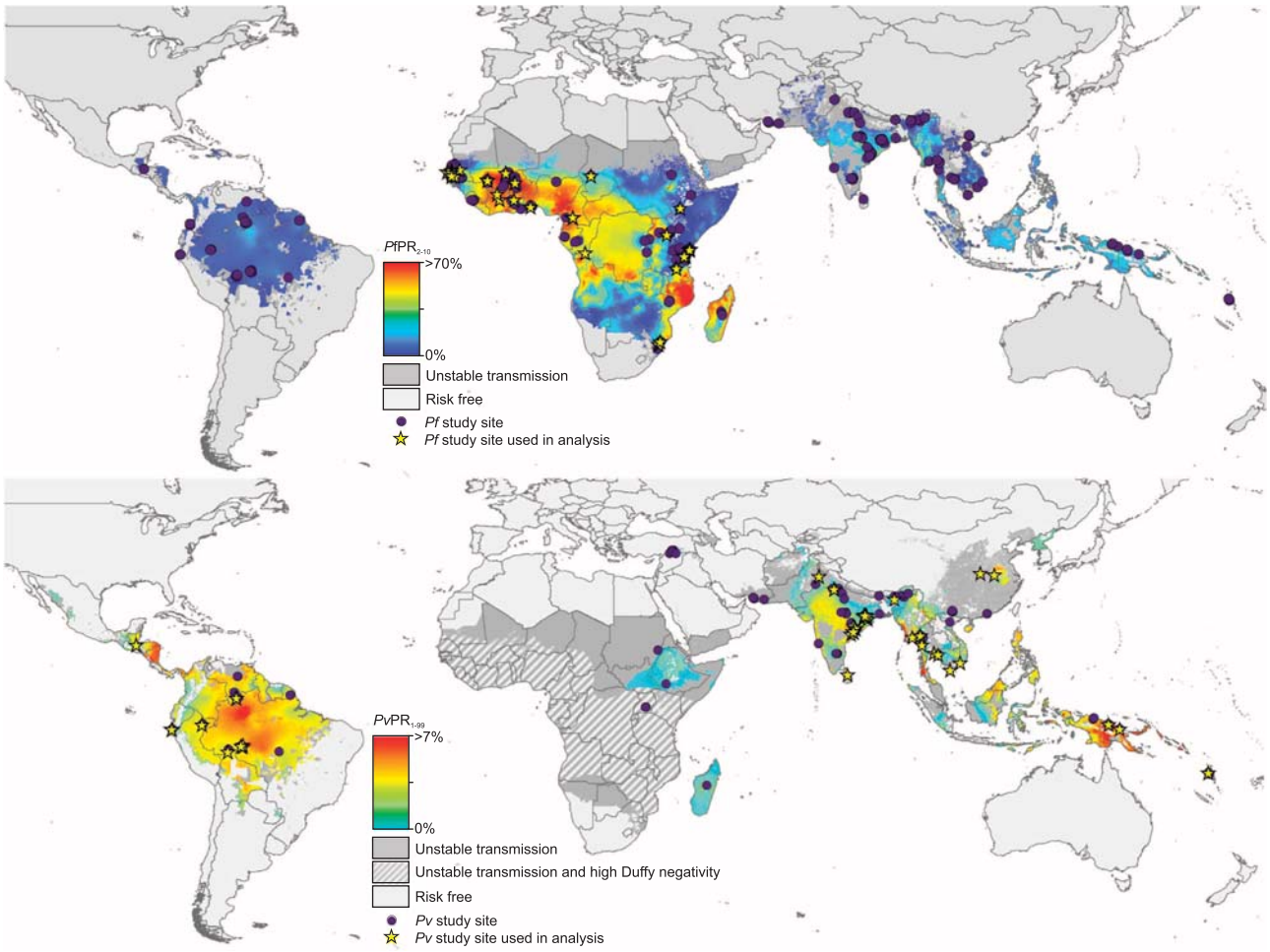


Figure 2. Distribution of incidence records of 2010 prevalence surfaces for *P. falciparum* and *P. vivax*. The geographic locations of the incidence records for *P. falciparum* (top panel) and *P. vivax* (bottom panel) are shown over the model-based geostatistics (MBG) point estimates of the annual mean *Pf*PR_{2-10} and *Pv*PR_{1-99} for 2010 within the spatial limits of stable limits of transmission (annual parasite index (API) ≥ 0.1 per 1,000 *per annum* (p.a.)), displayed on a continuum from blue (0% PR) to red (70% PR for *P. falciparum* and >7% PR for *P. vivax*). Dark grey areas were predicted to be unstable (API ≤ 0.1 per 1,000 p.a.) and light grey areas were classified as risk free. Areas in which Duffy negative allele frequency is predicted to exceed 90% (ref. 29) are shown in hatching for additional context in the *P. vivax* map. Study sites used in the *P. falciparum* (Cameron *et al.*[13]) and *P. vivax* (Battle *et al.*[14]) models are shown as yellow stars and other sites included in this dataset not used in the cited analyses are shown as purple points.

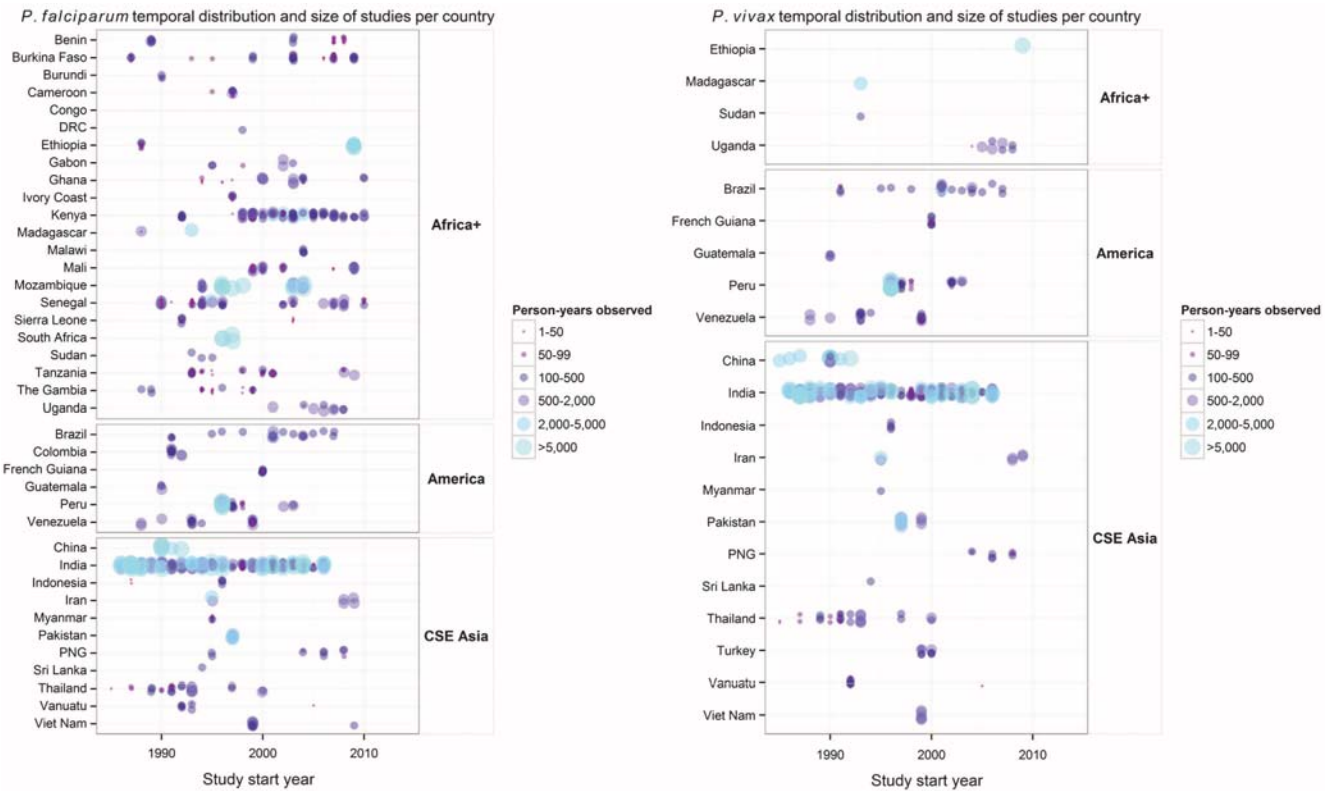


Figure 3. Temporal distribution and person-years observed for all incidence records by region and country. *Plasmodium falciparum* (left panel) and *P. vivax* (right panel) records are shown as points along an axis of the study start year. The points vary in colour and size based on the number of person-years observed in each study, such that studies with smaller sample sizes are small dark purple points and larger studies are large light blue points. The points are jittered so that overlapping points can be seen.

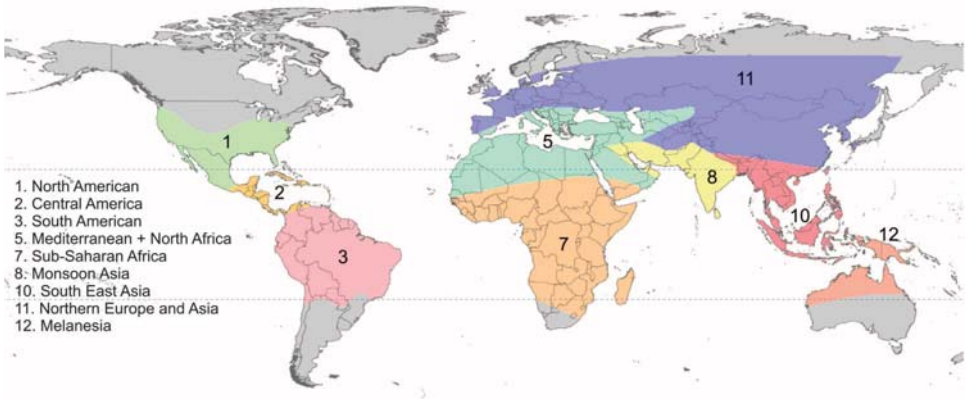


Figure 4. Geographic zones of varying malaria epidemiology and *P. vivax* relapse phenotypes. The zones were used by Battle *et al.*[27] to illustrate large-scale patterns in of relapse behaviour. Zones 4, 6 and 9 are not shown because they were joined with other zones as described in Battle *et al.*[27]

individuals was then adjusted to match the age-standardized prevalence value. See results in Data Citation 1. A schematic illustration the process of matching prevalence and incidence data is shown in Fig. 5.

Code availability

All age standardisation routines were implemented in an open-source software package ageStand[31], implemented in the R statistical programming environment[32]. The package contains one function,

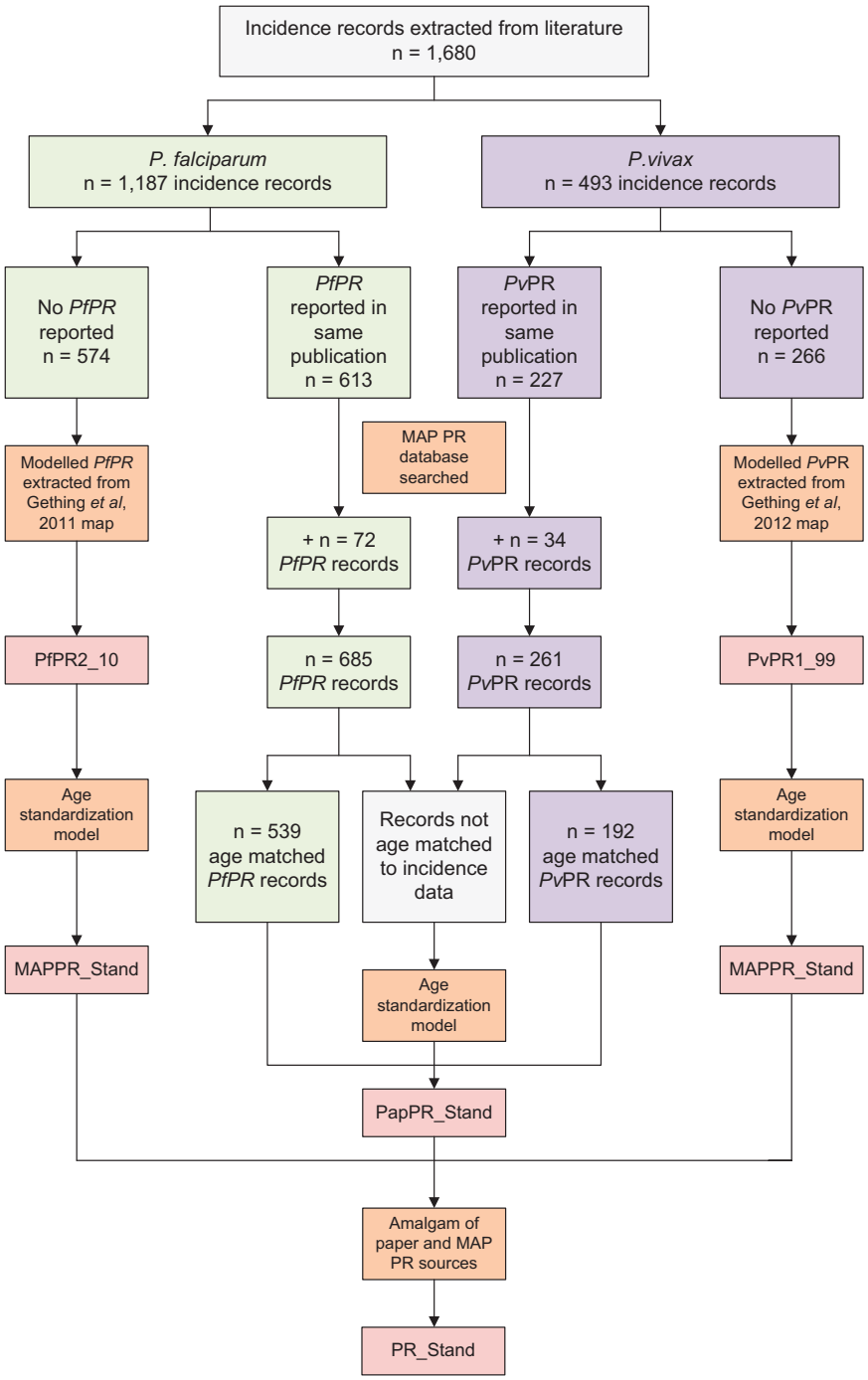


Figure 5. Schematic overview of the procedure of matching prevalence to incidence. Rectangles referring to all data are shown in grey, *P. falciparum* data in green, and *P. vivax* data in purple. Orange rectangles indicate data processing procedures and red rectangles symbolize fields in the final database.

`convertPrevalence`, which simplifies the conversion of prevalence estimates between age bounds. Five arguments were specified for the function. The first, `prevalence`, is a vector specifying the prevalence, or which PR field in this case, to convert from. Next, `age_min_in` and `age_max_in`, are vectors that specify the minimum and maximum ages associated with the estimates given in `prevalence`. `age_min_out` and `age_max_out` are vectors that provide the lower and upper bounds of the age range that the prevalence is to be converted to. Finally, the `parameters` argument specifies a set of parameters to be used in the model and was set to 'Pf_Smith2007' for all *P. falciparum* conversions and 'Pv_Gething2012' for all *P. vivax* conversions, referring the papers where the models were originally published for each species[30,33].

Data Records

Data on each species was abstracted separately and data were disaggregated by age groups where possible. Values were input into a spreadsheet containing 63 fields:

1. Identification

ENL_ID. Data source identification number.
PI_ID. Unique identifier for each record.
INC_AUTHOR. First author surname of incidence data publication.
INC_PUBYEAR. Publication year of incidence data publication.
GRIFFIN. Identifies a record used in the analysis by Griffin *et al.*[17]
CAMERON. Identifies a record used in the analysis by Cameron *et al.*[13]
BATTLE. Identifies a record used in the analysis by Battle *et al.*[14]
EXCLUSION. Potential exclusion criteria as described in Table 2.
SPECIES. *P. falciparum* (Pf) or *P. vivax* (Pv).

2. Geo-positioning

REGION. MAP regions are America, Africa+ and CSE Asia.
COUNTRY. Country where ACD was conducted.
ACD_LOCATION. The town/village/district where the ACD was conducted.
LAT. Latitude in decimal degrees (WGS1984 datum).
LONG. Longitude in decimal degrees (WGS1984 datum).
LATLONG_SOURCE. Source of the coordinates: Paper (from the publication), Google[22], Encarta[23], Other (other online sources or databases), Pers. Comm. (personal communication, usually reporting GPS-read coordinate), GIS (centroid of administrative unit found using ArcGIS)[25], or Combination (a combination of the aforementioned methods).
GEOPOS_NOTES. Further information about how geo-positioning was carried out.
EPIZONE. Numerical code for geographic epidemiological zones as defined in Battle *et al.*[27] and as shown in Fig. 4.
BATTLEZONE. The full name of the EPIZONE described above.

3. Incidence time

START_MONTH. Survey starting month.
START_YEAR. Survey starting year.
END_MONTH. Survey ending month.
END_YEAR. Survey ending year.
TIME. Number of years of observation, using fractions to represent <1 year (1 for 12 months, 0.67 for 8 months, 1.33 for 16 months, etc.)
TIME_CAT. Categorizes the TIME column into six categories for the purpose of further exclusion if needed (<6 months, 6–11 months, 12 months, 13–23 months, 24 months, >24 months).
FREQ_ACD. Frequency of ACD written as text (every 2nd day, weekly, fortnightly, etc.). Weekly* indicates a record where the frequency of ACD was not explicitly reported in the study and assumed to be one week.

Potential exclusion criteria*	Description	Pf	Pv	Total
Duplicate	Data from different studies reporting the same data or data from the total population where age-specific data were also reported	15	10	25
Asymptomatic	Papers that did not diagnose based on clinical symptoms, but on infection alone, and therefore asymptomatic cases would be included in the incidence estimates	11	6	17
Unclear	Methods regarding the ACD were vague; often the frequency of ACD was not reported	50	36	86
Infrequent ACD	Studies that carried out ACD at intervals greater than 30 days	22	9	31
Pre-1985	Studies that were published after 1985, but contain incidence data gathered before 1985	38	21	59
XSS	Studies that appear to be cross-sectional surveys rather than longitudinal ACD studies	10	10	20
PCD	Studies that appeared to use only passive case detection	54	13	67
<4 age bands	Studies that either did not stratify incidence by age or did so with less than four age groups	859	371	1,230
Approximate person-time	Person-years observed (PYO) was approximated by multiplying the study population by study time period	722	300	1,022
Rx	Population was given presumptive treatment prior to ACD observation period	35	2	37

Table 2. Exclusion criteria applied to the initial dataset for *P. falciparum* and *P. vivax*. *Figure 1 illustrates how these criteria were applied to the species-specific data records.

FREQ_ACD_NUM. This will express the frequency of ACD numerically should scaling be applied downstream. Records the number of days between each visit (every 2nd day = 2, every fortnight = 14, every month = 30).

PCD. Passive Case Detection. *Yes/No* for if Passive Case Detection was conducted alongside the ACD.

POP. Number of people observed for TIME; the study population.

d. The number of positive species-specific clinical cases. Asymptomatic and mixed infections were not included. Mixed infections were often a negligible proportion of the total infections and present a challenge because it is not possible to determine the parasite that caused the symptomatic episode.

PYO. If person-time is specifically reported in the paper, that value was used after converting to person-years. If person-time was not explicit, the length of the study period (TIME) was multiplied by the study population (POP) such that PYO = TIME * POP.

PYO_APPR. PYO approximated. This is a binary entry to indicate if the PYO was approximated or if an exact PYO was provided in the paper. If PYO = POP*TIME then it was approximated (value 1), and if PYO was reported in the paper (even if it is converted to years from days/weeks/months), then it was exact (value 0).

INC. Incidence = (d/PYO)*1,000. Incidence may have been explicitly reported in the paper. However, the likelihoods are derived from *d* and PYO and therefore if incidence was provided, the PYO and *d* were calculated from INC.

INC_NOTES. Description for how incidence values were obtained or derived from information given in the publication.

DIAGNOSTIC. Diagnostic technique used (*Microscopy* or *RDT*). Data based on serology or PCR were not included.

CLINICAL_DEF. Clinical Definition. The definition of a clinical case as given in the paper: fever+any parasitaemia or fever+parasitaemia within a fixed or age-dependent threshold (fixed was used if the study reported both).

CASE_DENS_THRESH. Case definition parasite threshold. Some publications specified a minimum parasite load for a patient to be considered a positive case. If any parasite density was permitted, a threshold of 1 was entered, otherwise the value specified in the paper was entered.

4. Incidence population

INC_LAR. Incidence lower age range. If there were multiple age groups studied in one paper, they were entered as separate rows. If no lower age was given, it was assumed to be zero.

INC_UAR. Incidence upper age range. If no upper age was given, it was assumed to be 85.

EIR. Entomological inoculation rate. This was recorded if given in the reference to provide a measure of transmission intensity.

INTERVENTION. Any interventions taking place in the study population. Control and intervention arms should be entered in different rows. For control groups or if there was no intervention, *None* was entered.

5. Prevalence data

PR_AUTHOR. The author of the source of the parasite rate (PR) data. If the data was found in the same paper as the incidence data, it was entered as *Same*.

PR_PUBYEAR. The year that the reference that cites the parasite rate was published. If it is the same paper as the incidence paper, it was entered as *Same*.

PR_MONTH. The month prevalence that the survey was conducted. If not provided, *NA* was recorded.

PR_YEAR. The year that the prevalence survey was conducted.

N_SURVEYS. This refers to the number of prevalence surveys (XSS) included in the PR estimate reported. Several studies reported PR values that are averaged from more than one survey.

PR_LAR. Lower age range of individuals tested in XSS. If no lower age was given, it was assumed to be zero.

PR_UAR. Upper age range of individuals tested in XSS. If no upper age was given, it was assumed to be 85.

AGE_MATCH. If the PR age range was the same as the incidence age range, the value was 1. If they did not match, the value was 0. If there was no PR value from the paper, the value was 99.

N. Number of individuals examined in the prevalence survey, or if slide positivity rate was reported, the number of slides examined was used.

N_POS. Number of individuals positive for parasite in question.

N_POS_ADJ. Adjusted number positive based on age-standardized prevalence value (PapPR_Stand or PR_Stand below).

PR. Calculated parasite rate = (N_POS/N).

PR_NOTES. Description of how a concurrent PR estimate was obtained.

PR_DIAGNOSTIC. Diagnostic technique used to identify *P. falciparum* or *P. vivax* prevalence within the population (microscopy or RDT). Data based serology or PCR were not included.

6. Matched prevalence data

PfPR2_10. Predicted *Pf*PR values from the *P. falciparum* MAP endemicity surface[7].

PvPR1_99. Predicted *Pv*PR values from the *P. vivax* MAP endemicity surface[15].

MAPPR_Stand. Estimate from the *P. falciparum* or *P. vivax* MAP surface age-standardized to the age-range used in the incidence data.

PapPR_Stand. Concurrent PR estimate age-standardized to the incidence age-range.

PR_Stand. If a concurrent PR estimate was available, the age-standardized one is used here, if not, the age-standardized MAP estimate is used.

7. Citations

REF_ACD. Reference for the incidence data.

PMID. PubMed identification number for ACD reference. Unpublished sources were left blank, but the type of source (e.g. thesis or conference proceedings) was noted in the full reference given in REF_ACD.

REF_PR. Reference for the PR data (if available).

PMID_PR. PubMed identification for PR reference. Unpublished sources were left blank, but the type of source (e.g. thesis or conference proceedings) was noted in the full reference given in REF_PR. If no concurrently measured PR was found, this was field also left blank.

Technical Validation

There were 1,680 rows of incidence data following initial data extraction (Data Citation 1). All records were entered by one team member and checked by a second. Cells where there was disagreement were highlighted and checked by a third person where possible. Checking was done to ensure that entries were accurate and that the inclusion criteria outlined above were met. Some exceptions to inclusion criteria described were made to allow for studies used in previous analyses[5,17] to be added to the database.

To record any exceptions to the inclusion criteria used in the Cameron *et al.*[13] and Battle *et al.*[14] studies and to flag other records for potential exclusion in future analyses, an additional field was added to the database (see EXCLUSION field above). The first exclusion, which applied to both *P. falciparum* and *P. vivax* data, was to remove records from different studies that reported the same data (same population at the same time). The records prioritized for inclusion were those that had been included in previous analyses[5,17]. Next, studies that measured both symptomatic and asymptomatic cases that passed the first inclusion stage, but found during validation, were marked for potential exclusion as the incidence measure would not be specific to clinical cases. Studies with unclear methods, such as un-specified frequency of detection or the number of cases or person-time could not be derived, were also marked for potential exclusion. For *P. falciparum*, the remaining potential exclusion criteria based on the analysis by Cameron *et al.*[13] was to not include studies that (i) had fewer than four age-specific estimates from the same population during the same time to remove studies under-powered for inference of the *P. falciparum* age-incidence relationship and or (ii) where the population were treated presumptively at the start of the transmission season. There were only six *P. falciparum* records excluded from the Cameron *et al.* analysis for presumptive treatment, but it was noted in INTERVENTION field for other records that had been flagged for potential exclusion for other reasons. For the *P. vivax* analysis by Battle *et al.*[14], studies that made ACD visits more than one month apart were excluded, as were studies conducted prior to 1985. Incidence reports from retrospective analyses or passive case detection (PCD)-only were marked in the EXCLUSION field, as were XSS because they are not longitudinal and measure both symptomatic and asymptomatic cases. The records flagged as PCD and XSS had originally been abstracted because they were used in previous analyses[5,17].

A summary of the exclusion criteria described above is shown in Table 2 and a schematic of the exclusion procedures is shown in Fig. 1. Table 3 shows the regional distribution of the 328 *P. falciparum* and the 152 *P. vivax* records remaining after the species-specific exclusions described above were applied. The study by Cameron *et al.*[13] was restricted to Africa, but the lack of data in Africa in the *P. vivax* analysis conducted by Battle *et al.*[14] represents a genuine absence of *P. vivax* data in the region. Note that all 1,680 records originally abstracted remain in the database so that customized exclusions can be applied for any future analyses using this data.

Region	*P. falciparum*—Cameron*	*P. vivax*—Battle†
Africa+	295	—
America	—	43
CSE Asia	—	133
Total	**295**	**176**

Table 3. Data records for used in analysis by Cameron *et al.*[13] and Battle *et al.*[14] by MAP region. *The study by Cameron *et al.* was restricted to Africa. †The study by Battle *et al.* was global and therefore lack of records in Africa+ represents a genuine absence of data that matched the inclusion criteria from that region.

Usage Notes

This dataset was generated for the purpose of modelling the relationship between incidence of clinical malaria and the more commonly measured PR. This database has been directly applied to the models described in Cameron *et al.*[13] and Battle *et al.*[14], and is similar to the smaller dataset published by Patil *et al.*[5], with the intention of developing species-specific global burden maps for *P. falciparum* and *P. vivax* malaria. Those in turn directly inform global estimates of the burden of clinical disease attributable to each species. This information is critical to efficiently allocate resources and direct efforts to combat these illnesses.

As described above, this database reports incidence of all infections, not only new infections, and therefore include relapses (*P. vivax* only), recrudescences and reinfections in both the prevalence and incidence measures[27,34,35]. For *P. vivax* in particular, these data could be used in conjunction with data on patients who have received radical cure treatment (with either primaquine or tafenoquine) or treatment without primaquine to determine the incidence of new infections or relapses, respectively. This would be done by taking the overall incidence in a location, as reported here, and subtracting the incidence of new infections from patients treated with a radical cure or without a radical cure. This would be essential data for determining sporozoite- and hypnozoite-specific attack rates, the relative proportions of which may directly inform the character of interventions against endemic malaria.

It has been hypothesised that a key driver of relapse in *P. vivax* is infection with *P. falciparum*. To allow for investigation of the potential interactions between the endemicity of one species on the incidence of another, a prevalence measure for both *P. vivax* and *P. falciparum* is provided for each entry.

This database may also be of use for other analyses of clinical burden. Where possible, data has been disaggregated by age. This allows for studies of how burden of disease changes with age, as was done by Griffin *et al.* using a smaller dataset of *P. falciparum* in children in Africa[12]. The database also contains incidence data from intervention studies with data from both intervention and control arms entered. This would offer insight into the impact of control on incidence of disease as compared to prevalence of infection. Finally, the collection of data from 1985 until the present may improve our understanding in the change of malaria burden over time.

References

1. Cibulskis, R. E., Aregawi, M., Williams, R., Otten, M. & Dye, C. Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. *PLoS Med.* **8,** e1001142 (2011).
2. WHO. *World Malaria Report* 199 (World Health Organization, 2013).
3. Murray, C. J. *et al.* Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384,** 1005–1070 (2014).
4. Hay, S. I. *et al.* Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. *PLoS Med.* **7,** e1000290 (2010).
5. Patil, A. P. *et al.* Defining the relationship between *Plasmodium falciparum* parasite rate and clinical disease: statistical models for disease burden estimation. *Malar. J.* **8,** 186 (2009).
6. Hay, S. I. *et al.* A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* **6,** e1000048 (2009).
7. Gething, P. W. *et al.* A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10,** 378 (2011).
8. Mendis, K., Sina, B. J., Marchesini, P. & Carter, R. The neglected burden of *Plasmodium vivax* malaria. *Am. J. Trop. Med. Hyg.* **64,** 97–106 (2001).
9. Mueller, I. *et al.* Key gaps in the knowledge of *Plasmodium vivax*, a neglected human malaria parasite. *Lancet Infect. Dis.* **9,** 555–566 (2009).
10. Price, R. N. *et al.* Vivax malaria: neglected and not benign. *Am. J. Trop. Med. Hyg.* **77,** 79–87 (2007).
11. Snow, R. W., Menon, A. & Greenwood, B. M. Measuring morbidity from malaria. *Ann. Trop. Med. Parasitol.* **83,** 321–323 (1989).
12. Hay, S. I., Smith, D. L. & Snow, R. W. Measuring malaria endemicity from intense to interrupted transmission. *Lancet Infect. Dis.* **8,** 369–378 (2008).
13. Cameron, E. *et al.* Defining the relationship between infection prevalence and clinical incidence of *Plasmodium falciparum* malaria. *Nat. Commun.* 6:8170 doi: 10.1038/ncomms9170 (in the press).
14. Battle, K. E. *et al.* Defining the relationship between *Plasmodium vivax* parasite rate and clinical disease. *Malar. J.* **14,** 191 (2015).
15. Gething, P. W. *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl. Trop. Dis.* **6,** e1814 (2012).
16. Guerra, C. A. *et al.* Assembling a global database of malaria parasite prevalence for the Malaria Atlas Project. *Malar. J.* **6,** 17 (2007).
17. Griffin, J. T., Ferguson, N. M. & Ghani, A. C. Estimates of the changing age-burden of *Plasmodium falciparum* malaria disease in sub-Saharan Africa. *Nat Commun* **5,** 3136 (2014).
18. Rogerson, S. J., Hviid, L., Duffy, P. E., Leke, R. F. & Taylor, D. W. Malaria in pregnancy: pathogenesis and immunity. *Lancet Infectious Diseases* **7,** 105–117 (2007).
19. Hviid, L. & Staalsoe, T. Malaria immunity in infants: a special case of a general phenomenon? *Trends Parasitol.* **20,** 66–72 (2004).
20. Baird, J. K. Evidence and implications of mortality associated with acute *Plasmodium vivax* malaria. *Clin Microbiol. Rev.* **26,** 36–57 (2013).
21. Poespoprodjo, J. R. *et al.* Vivax malaria: a major cause of morbidity in early infancy. *Clinical Infectious Diseases* **48,** 1704–1712 (2009).
22. Google Maps www.google.co.uk/maps/ (Google, Mountain View, CA, 2013).
23. Microsoft. *Encarta Reference Library*. (Microsoft Corporation, 2007).
24. FAO. *The Global Administrative Unit Layers (GAUL): Technical Aspects*. (Food and Agriculture Organization of the United Nations, EC-FAO Food Security Programme (ESTG, 2008).
25. ESRI. *ArcGIS Desktop 10.0*. (Environmental Systems Resource Institute, 2010).
26. Macdonald, G.. *The Epidemiology and Control of Malaria*. (Oxford University Press, 1957).
27. Battle, K. E. *et al.* Geographical variation in *Plasmodium vivax* relapse. *Malar. J.* **13,** 144 (2014).
28. OpenMalaria. ModelDemography: Human population demographics 2014).
29. Howes, R. E. *et al.* The global distribution of the Duffy blood group. *Nature Commun.* **2,** 266 (2011).

30. Smith, D. L., Guerra, C. A., Snow, R. W. & Hay, S. I. Standardizing estimates of the *Plasmodium falciparum* parasite rate. *Malar. J.* **6,** 131 (2007).
31. Golding, N. ageStand R package. GitHub https://github.com/SEEG-Oxford (2014).
32. Core Team., R. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
33. Gething, P. W. *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl. Trop. Dis.* **6,** e1814 (2012).
34. Imwong, M. *et al.* The first Plasmodium vivax relapses of life are usually genetically homologous. *J. Infect. Dis.* **205,** 680–683 (2012).
35. Imwong, M. *et al.* Relapses of *Plasmodium vivax* infection usually result from activation of heterologous hypnozoites. *J. Infect. Dis.* **195,** 927–933 (2007).

Data Citation

1. Battle, K. E. *et al. Dryad* http://dx.doi.org/10.5061/dryad.j0kv0 (2015).

Acknowledgements

S.I.H. is funded by a Senior Research Fellowship from the Wellcome Trust (095066), which also supports K.E.B., C.A.G., K.A.D. and R.E.H. N.G. is funded by a grant from the Bill & Melinda Gates Foundation (OPP1053338). J.K.B. is supported by Wellcome Trust grant (B9RJIXO). I.R.F.E. is funded by a Public Health and Tropical Medicine Fellowship of the Wellcome Trust (B9RZGS0). P.W.G. is a Medical Research Council Career Development Fellow (K00669X) and receives support from the Bill and Melinda Gates Foundation (OPP1068048), which also supports E.C. D.L.S. is funded by a grant from the Bill & Melinda Gates Foundation (OPP1110495), which also supports R.C.R. R.C.R., D.L.S. and S.I.H. also acknowledge funding support from the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, and the Fogarty International Center, National Institutes of Health. The authors also thank Joshua Longbottom his for proof-reading and comments.

Author Contributions

K.E.B. drafted the manuscript with editing and approval of all authors. K.E.B. and R.E.H. developed data search and abstraction protocols. K.E.B., C.A.G. and K.A.D. compiled the data records. K.E.B. and C.A.G. performed the technical validation. N.G. wrote statistical software to process and standardise data. E.C. lead the model development. I.R.F.E. and J.K.B. provided data. R.C.R. and J.K.B. provided feedback on data implementation and D.L.S. provided the age-standardization model and feedback on modelling procedures. S.I.H. and P.W.G. conceived the database design and advised on data abstraction and validation procedures.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Battle, K. E. *et al.* Global database of matched *Plasmodium falciparum* and *P. vivax* incidence and prevalence records from 1985–2013. *Sci. Data* 2:150012 doi: 10.1038/sdata.2015.12 (2015).