Plasmodium knowlesi invasion following spread by infected mosquitoes, macaques and humans

Laith Yakob[1,*], Alun Lloyd[2], Rowland Kao[3], Heather M. Ferguson[3], Patrick Brock[3], Chris Drakeley[4], Michael Bonsall[5,6]

[1] Department of Disease Control, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

[2] Biomathematics Graduate Program, Center for Quantitative Sciences in Biomedicine and Department of Mathematics, North Carolina State University, Raleigh, NC 27695

[3] Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow G61 1QH

[4] Immunology and Infection Department, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT

[5] Mathematical Ecology Research Group, Department of Zoology, University of Oxford, Oxford OX1 3PS

[6] St Peter's College, Oxford, OX1 2DL

* laith.yakob@lshtm.ac.uk

Running Head: Plasmodium knowlesi invasion analysis

SUMMARY

Plasmodium knowlesi is increasingly recognised as a major cause of malaria in Southeast Asia. *Anopheles leucosphyrous* group mosquitoes transmit the parasite and natural hosts include long-tailed and pig-tailed macaques. Despite early laboratory experiments demonstrating successful passage of infection between humans, the true role that humans play in *P. knowlesi* epidemiology remains unclear. The threat posed by its introduction into immunologically naïve populations is unknown despite being a public health priority for this region. A two-host species mathematical model was constructed to analyse this threat. Global sensitivity analysis using Monte Carlo methods highlighted the biological processes of greatest influence to transmission. These included parameters known to be influential in classic mosquito-borne disease models (e.g., vector longevity); however, interesting ecological components that are specific to this system were also highlighted: while local vectors likely have intrinsic preferences for certain host species, how plastic these preferences are, and how this is shaped by local conditions, are key determinants of parasite transmission potential. Invasion analysis demonstrates that this behavioural plasticity can qualitatively impact the probability of an epidemic sparked by imported infection. Identifying key vector sub/species and studying their biting behaviours constitute important next steps before models can better assist in strategizing disease control.

Keywords: Invasion analysis; Plasmodium knowlesi; vector-borne disease; mathematical model; vector behaviour

INTRODUCTION

The major human malaria species *Plasmodium falciparum* and *P. vivax* infect approximately 200 million people every year, killing nearly 600,000 (WHO 2014). These parasites successfully established in human populations thousands of years ago following zoonotic emergence from ape hosts in Africa (Liu *et al.* 2010; Liu *et al.* 2014). In 2004, a surprisingly high prevalence of *P. knowlesi* was found in humans in Malaysian Borneo when diagnostic microscopy was replaced by the more discriminatory method of nested PCR (Singh *et al.* 2004). This ground-breaking study identified that all blood samples from 208 people reporting atypical malaria infection in Kapit division of Malaysian Borneo were *P. knowlesi*-positive but misidentified as the morphologically similar *P. malariae* – a result subsequently corroborated by a larger, follow-up study conducted by the same group (Cox-Singh *et al.* 2008). Although long- and pig-tailed macaques are the natural hosts for this species, *P. knowlesi* has now been described in humans across several SE Asian countries and is the leading cause of human malaria in Malaysian Borneo (Singh and Daneshvar 2013).

Mathematical models have been exploited in malaria research for a century and have produced considerable insight in both the epidemiology and control of infection (Smith *et al.* 2012). Model complexity has increased along with biological understanding and computational power; however, even the most complex ecological transmission models have fundamental elements that are identical, or analogous, to the original Ross–Macdonald formulations (Reiner *et al.* 2013). This family of models typically assume a single host species – an assumption that must be relaxed in the current context. Due to the relatively recent discovery of human infections with this species, and the correspondingly nascent understanding of infection processes, *P. knowlesi* models are relatively scarce and uncomplicated. The first published *knowlesi* malaria model expanded the Ross–Macdonald formula to account for heterogeneous biting of the vector (*Anopheles leucosphyrous* group) split between both macaque and human mammalian hosts

(Yakob *et al.* 2010). A game theoretic approach to evolutionary invasion analysis of this deterministic system of ordinary differential equations was used to calculate the conditions under which a parasite might switch natural hosts from macaques to humans (Yakob *et al.* 2010). Subsequent adaptations of this model were used to explore how vector control strategies could be optimised - both at larval and adult stages (Abdullahi *et al.* 2013); and, to explore how the basic reproduction number may be impacted by different ecological settings (Imai *et al.* 2014). Using a mathematical model, we build on this work to analyse the probability of successful parasite invasion into a host population following its introduction by an infected vector or host (either human or macaque).

Stochastic effects are known to be highly influential during the period immediately after the introduction of infection into a population (Bartlett 1956), and are accounted for in calculating the probabilities of successful invasion of *P. knowlesi* introduced into susceptible populations (ranging from exclusively macaque to exclusively human). We also incorporate a flexible formulation that allows for qualitatively distinct host-selection vector biting behaviours because this aspect remains largely unknown for local vector species while also being 1) critical to vector-borne disease epidemiology and control (Besansky *et al.* 2004); 2) likely to vary considerably (and not necessarily linearly) across differing proportionate representations of alternative mammalian hosts (Takken and Verhulst 2013); and 3) also likely to vary according to local vector sibling species (Gillies 1967). Insights gained into *P. knowlesi* epidemiology, including parasite invasion probabilities, are discussed along with proposed future research directions.

METHODS

Figure 1 depicts the different epidemiological compartments in the model and their connections. Being a severely neglected tropical disease, there is a general absence of longitudinal studies

detailing *P. knowlesi* malaria infection (Fornace *et al.* 2015). Consequently, a flexible and open-ended description of the transmission dynamics (Yakob 2016a; Yakob 2016b) is presented and used to calculate between-species parasite transmission numbers as well as invasion probabilities. Sensitivity analysis of the parameters underlying these thresholds will determine the aspects of unknown infection biology that might constitute priorities for future research.

Transmission dynamics

$$\frac{dS}{dt} = \mu H + \gamma I + \tau R - m p_H b_{VH} S Z - \mu S \quad \text{Eq 1}$$

$$\frac{dI}{dt} = m p_H b_{VH} S Z - (\gamma + \varepsilon + \pi + \mu) I \quad \text{Eq 2}$$

$$\frac{dR}{dt} = \varepsilon I + \kappa A - (\tau + m\theta p_H b_{VH} Z + \mu) R \quad \text{Eq 3}$$

$$\frac{dA}{dt} = \pi I + m\theta p_H b_{VH} Z R - \kappa A - \mu A \quad \text{Eq 4}$$

$$\frac{dX}{dt} = \mu_V V - \left(p_H b_{HV}(I + \sigma A) + (1 - p_H) b_{NV}(I_N + \sigma_N A_N)\right) X - \mu_V X \quad \text{Eq 5}$$

$$\frac{dY}{dt} = \left(p_H b_{HV}(I + \sigma A) + (1 - p_H) b_{NV}(I_N + \sigma_N A_N)\right) X - (\zeta + \mu_V) Y \quad \text{Eq 6}$$

$$\frac{dZ}{dt} = \zeta Y - \mu_V Z \quad \text{Eq 7}$$

$$\frac{dS_N}{dt} = \mu_N N + \gamma_N I_N + \tau_N R_N - m(1 - p_H) b_{VN} S_N Z - \mu_N S_N \quad \text{Eq 8}$$

$$\frac{dI_N}{dt} = m(1 - p_H) b_{VN} S_N Z - (\gamma_N + \varepsilon_N + \pi_N + \mu_N) I_N \quad \text{Eq 9}$$

$$\frac{dR_N}{dt} = \varepsilon_N I_N + \kappa_N A_N - (\tau_N + m\theta_N (1 - p_H) b_{VN} Z + \mu_N) R_N \quad \text{Eq 10}$$

$$\frac{dA_N}{dt} = \pi_N I_N + m\theta_N (1 - p_H) b_{VN} Z R_N - (\kappa_N + \mu_N) A_N \quad \text{Eq 11}$$

All variables depicting epidemiological categories are proportions. Susceptible humans (*S*) become infectious (*I*) following a bite from an infectious vector (*Z*). Infectious humans revert to susceptible at rate γ. Different parameterisation of the clearance rate of symptomatic infection (ε), the rate of reversion to full susceptibility (τ) and the susceptibility to asymptomatic infections (θ) affects the temporality of immunity. Human hosts can become asymptomatically infected (*A*)

directly progressing from symptomatic infection when the rate termed π is greater than 0, or following on from recovery (R) and subsequent reinfection ($\theta > 0$). Asymptomatic infection in macaques is assumed to be lifelong (by setting recovery from secondary infection, κ_N, to equal 0) whereas humans are assumed to be able to clear the parasites and recover at rate κ. Processes governing infection in the natural macaque hosts are denoted by subscript N. Susceptible vectors (X) become *infected* (Y) following a bite from an infectious host, and after the extrinsic incubation period ($1/\zeta$), become *infectious* (Z). The ratio of mosquitoes to total hosts is denoted m and the vector mortality rate is μ_V. Transmission coefficients are denoted by 'b' with associated subscripts (these are distinguished by the host species involved should species-specific estimates arise in the future e.g. b_{VH} is the transmission coefficient from vectors to human hosts and comprises the bite rate per vector multiplied by the probability of parasite transmission per bite). However, because there are two alternative host species, bites must be further partitioned according to which host species actually receives the bite from a vector. This required the following framework to apportion these bites among alternative host species as determined by both their relative abundances and intrinsic vector preferences for specific host species.

Functional responses in the human blood index

The proportion of bites on humans is determined by a flexible formula that allows for a wide range of different functional responses depicting distinct vector biting behaviours:

$$p_H = \frac{\dot{H}}{\dot{H}+\alpha(1-\dot{H})^{\beta}}\,. \qquad \text{Eq 12}$$

Here p_H is the 'human blood index' (Garret-Jones 1964); $\dot{H}$ is the availability of humans relative to all other potential hosts; α and β are parameters that shape the functional response of human bite proportion relative to all potential host species. Type I responses ($\alpha = \beta = 1$) assume bite

distribution among alternative host species that is directly proportionate to their relative availability; Type II human blood index responses ($\alpha < 1$ and $\beta \geq 1$) are convex-up with increasing human availability relative to alternative hosts and describe an anthropophilic vector; Type III responses ($\alpha \geq 1$ and $\beta > 1$) are s-shaped and depict a zoophagic vector that becomes increasingly anthropophilic with increased human encounters; Type IV responses ($\alpha > 1$ and $\beta \leq 1$) are convex-down and describe a zoophilic vector that only bites humans when there are few alternatives; and Type V responses ($\alpha \leq 1$ and $\beta < 1$) are s-shaped reflected in the y=x line and describe a negative prey-switching (Abrams *et al.* 1993) analogue, e.g., whereby anthropophilic vectors avoid a nuisance response. A fuller description of these functional responses can be found in (Yakob 2016b). Figure 2 illustrates the shape of association between the human blood index and human host availability relative to all potential blood hosts. A complete range of host availabilities is displayed – from entirely macaque populations (0 on the x-axis) to entirely human populations (1 on the x-axis), and everything in between e.g. at the half-way mark (0.5) of the x-axis, equal availability of humans and macaques is shown for a mixed population. This formula is used to assess the importance of different host availabilities (i.e. different environmental settings) and different host-feeding behaviours in the resulting between-species transmission rates and invasion analysis.

Calculation of the basic reproduction number: Entries of the next generation matrix

Standard theory states that the basic reproduction number, R_0, can be calculated as the largest eigenvalue (i.e. the spectral radius) of the next generation matrix, K (Diekmann and Heesterbeek 2000). In the present context, involving two types of hosts and one type of vector, K is a 3 by 3 matrix. Entries of K, which we write as K^{ij}, depict the expected number of infections of each type (human host, macaque host or vector) that are directly produced by an infectious individual of

each type (human, macaque or vector) when the system is at (or very near) the infection free equilibrium. Standard theory shows how the K^{ij} can be calculated by considering the linearized infected subsystems, decomposing each into two matrices (Diekmann *et al.* 2010): one depicting the infection transmission (T) and the other depicting all other transitions (Σ). Each K^{ij} is calculated as the spectral radius of the next generation matrix (NGM) for that component of the system calculated from $-T\Sigma^{-1}$ (Diekmann *et al.* 2010). For the present system, there are four non-zero entries of the next generation matrix (whose derivations are shown below): the average number of human cases arising from an infected vector (K^{VH}); the average number of macaque cases arising from an infected vector (K^{VN}); the average number of vector infections arising from an infected human (K^{HV}); and the average number of vector infections arising from an infected macaque (K^{NV}). These between-species transmission numbers and their sensitivities to the underlying model parameters are assessed in terms of the Spearman's rank correlation coefficient calculated from 5000 iterations of a Monte Carlo multivariate sensitivity analysis (whereby all parameters were assumed to have triangular probability distributions ±10% about the median values described in Table 1). Global sensitivity analysis was used to ascertain the processes that are most instrumental in *P. knowlesi* transmission rates.

Invasion probabilities

For deterministic model formulations, if the average number of secondary infections arising from a primary infection exceeds unity, the successful invasion of the pathogen into the host population is guaranteed. New epidemics driven by the imports of small numbers of infected hosts or vectors are less certain than implied by determinism: for instance, an initial infective could, with some probability, recover or die before causing any secondary infections. Calculation of invasion probabilities requires a stochastic model, a framework that can be obtained by

189 reinterpreting the rates of continuous movement between compartments in the deterministic
190 differential equation model as rates (probabilities per unit time) at which discrete transition
191 events occur in the stochastic model. Branching process theory has been used to calculate the
192 extinction probability of (potential) epidemics sparked by the introduction of infected individuals
193 (Athreya and Ney 1972) and this has recently been expanded to calculate invasion probabilities
194 for vector-borne disease systems allowing for two levels of host attractiveness (Lloyd *et al.* 2007).
195 In line with these previous developments, invasion probabilities among the different host types
196 are the same, in that an outbreak amongst one host type necessarily means ongoing infections
197 amongst other host types, even if this is just a spill-over effect. To the best of our knowledge the
198 current analysis constitutes the first to describe methods of invasion analysis for a real multi-host
199 vector-borne disease system. This theory requires the calculation of probability generating
200 functions, $G(s)$, that summarize the distributions of secondary infections of each type of species
201 that results from the introduction of an infected vector, macaque or human. In these functions,
202 secondary infections amongst vectors, macaques and humans are labelled using powers of s_v, s_n
203 and s_h respectively. As in the deterministic analysis, all quantities are calculated at the infection
204 free equilibrium. For the human host population, calculation of the probability generating
205 function needs to account for the fact that an infectious human host in the *I* compartment can
206 move to the asymptomatic (*A*) compartment and continue to cause infections. This is achieved by
207 calculating generating functions for infections produced while in the two compartments and
208 combining them, accounting for the probability of making the infected (*I*) to asymptomatic (*A*)
209 transition, to give the overall generating function for an infective human host. We remark that
210 the branching process analysis does not need to consider the transition from recovered (*R*) to
211 asymptomatic (*A*) (recovered individuals becoming re-infected) as the rate of this flow is negligible
212 near the infection free equilibrium. The generating function for the number of secondary
213 infections generated from the infected (*I*) class is

214 $G_I(s_v) = \frac{1}{1+R_1(1-s_v)}$ Eq 13

215 where R_1=$mb_{HV}/(\gamma + \varepsilon + \pi + \mu)$. The generating function for the asymptomatic (*A*) class is

216 $G_A(s_v) = \frac{1}{1+R_2(1-s_v)}$ Eq 14

217 where R_2= $\sigma mb_{HV}/(\kappa + \mu)$. With ϕ denoting the probability that an infected (*I*) individual will
218 become asymptomatic (*A*), i.e. ϕ=π/($\gamma + \varepsilon + \pi + \mu$), the generating function for the number of
219 secondary infections generated after departure from the infected (*I*) class is given by

220 $G_Z(s_v) = 1 - \phi + \phi G_A(s_v)$ Eq 15

221 Making use of the fact that the generating function for the sum of two independent
222 random variables is the product of their generating functions, we have that the generating
223 function for the secondary infections resulting from an infected human host is given by
224 $G_{HV}(s_v)=G_I(s_v).G_Z(s_v)$ and hence

225 $G(s_v) = \frac{1}{1+R_1(1-s_v)}\left\{1 - \phi + \phi \frac{1}{1+R_2(1-s_v)}\right\}$ Eq 16

226 The generating function, $G_{NV}(s_V)$, describing the distribution of the number of vectors
227 infected by an infectious macaque is obtained similarly. The generating function for the numbers
228 of humans and macaques infected by an infectious vector is $G_V(s_h,s_n)$, where

229 $G_V(s_h, s_n) = \frac{1}{1+K^{VH}(1-s_h)+K^{VN}(1-s_n)}$ Eq 17

230 As in Lloyd et al. (2007), extinction probabilities following an introduction of an infected
231 vector, human or macaque (s_v, s_h and s_n , respectively) are found by solving the set

232 $G_V(s_h, s_n) = s_v$

233 $G_{HV}(s_v) = s_h$ Eq 18

234 $G_{NV}(s_v) = s_n$.

235 This is most easily achieved by substituting the second and third of these equations into
236 the first, leaving an equation for s_v alone. This results in a fifth degree polynomial for which one
237 root is $s_v = 1$, and thus leaves a quartic polynomial to solve for s_v. This equation can be solved
238 numerically and s_h and s_n found by substitution. Standard theory shows that these invasion
239 probabilities are all zero when the basic reproduction number, R_0, of the system is less than one
240 and fall between 0 and 1 when R_0 is greater than one (i.e. invasion happens with some non-zero
241 probability, but is not guaranteed).

242 Previous explorations of multi-host systems have assumed that the proportion of bites on
243 alternative host species is directly proportional to their relative availability. Using the new
244 formulation that allows for qualitatively different functional responses in vector bite behaviours
245 (Eq 12), the sensitivity of invasion probabilities to this neglected aspect of disease vector ecology
246 was also assessed.

247

248 RESULTS

249 NGMs were used to calculate the expected number of infections of each type (human host,
250 macaque host or vector) that are directly produced by an infectious individual of each type:

251 $$K^{HV} = \frac{mb_{HV}p_H(\kappa+\mu+\pi\sigma)}{(\kappa+\mu)(\gamma+\pi+\varepsilon+\mu)}$$ Eq 19

252 $$K^{NV} = \frac{mb_{NV}(1-p_H)(\kappa_N+\mu_N+\pi_N\sigma_N)}{(\kappa_N+\mu_N)(\gamma_N+\pi_N+\varepsilon_N+\mu_N)}$$ Eq 20

253 $$K^{VH} = \frac{b_{VH}p_H\zeta}{\mu_V(\mu_V+\zeta)}$$ Eq 21

254 $$K^{VN} = \frac{b_{VN}(1-p_H)\zeta}{\mu_V(\mu_V+\zeta)}$$ Eq 22

255 The resulting basic reproduction number, R_0, is calculated as:

256 $$R_0 = \sqrt{(K^{HV}K^{VH} + K^{NV}K^{VN})}$$

257 Figure 3 describes the sensitivity of the parasite transmission numbers between species
258 to the parameter values in the form of tornado plots. Across the different functional response
259 Types, there is good qualitative consistency in the transmission numbers' sensitivity to underlying
260 parameters. Intuitively, both K^{VH} and K^{VN} are highly sensitive to the mosquito mortality rate – a
261 parameter that is well understood to be strongly influential in classic models of vector-borne
262 diseases (Macdonald 1956). Both K^{HV} and K^{NV} are similarly sensitive to the transmission
263 coefficients (b) and very insensitive to mammalian host longevity (inverse of their respective
264 mortality rates, μ and μ_N) as per traditional malaria models. Of note is the considerable variation
265 in transmission numbers in relation to the availability of humans relative to all alternative blood
266 hosts, $\dot{H}$, whereby $\dot{H}$ was the most influential parameter for all transmission numbers under a
267 Type III functional response (a zoophagic vector that becomes increasingly anthropophilic with
268 increased human encounters) and of markedly lower significance under a Type V response
269 (negative prey-switching). This result is apparent from Figure 2.

270 Sensitivity analysis was conducted at $\dot{H}$=0.5 (i.e. humans and macaques are equally
271 available) because this is where differences between the Types are most pronounced. The
272 gradient of the human blood index as a function of human availability relative to all blood meal
273 hosts is steepest for Type III and flattest for Type V at this cross-section. This ranking in sensitivity
274 will shift non-monotonically for the different functional types in vector biting behaviour across
275 the range of alternative host availabilities.

276 Figure 4 shows the invasion probabilities for *P. knowlesi* in relation to host availability and
277 vector host-selection behaviours. General trends arise when comparing these probabilities across
278 scenarios whereby the pathogen is introduced by vectors, humans and macaques: introduction

279 of the pathogen by an infected host is least likely to invade when the local host population is
280 dominated by heterologous species; and when *P*. *knowlesi* is introduced by an infected mosquito,
281 invasion potential is maximised in macaque-only populations. This can be explained by the
282 assumed superiority of macaques as parasite hosts (they are assumed to remain infectious for
283 life). However, an unanticipated result of the mosquito-driven invasion analysis is the fact that,
284 regardless of the assumed biting behaviour, minimal invasion probabilities corresponded with
285 non-trivial mixes of macaque and human hosts.

286 For the most part, the invasion probabilities behave distinctively across different
287 functional Types. Of note are the differences between scenarios whereby *P*. *knowlesi* can
288 successfully invade when introduced by a macaque: when humans constitute >30% of all blood-
289 hosts, invasion is precluded in a Type III (switched biting behaviour) entomological scenario but,
290 in a Type I (classic proportionate biting assumption) scenario, this complete exclusion is restricted
291 to settings in which there are no macaques.

292

293 DISCUSSION

294 Malaria caused by *Plasmodium knowlesi* can be a highly debilitating and potentially fatal disease.
295 To improve our understanding of this neglected tropical disease, we developed models to explore
296 the probability of *P*. *knowlesi* invasion into different populations.

297 Multivariate sensitivity analyses highlight aspects of vector and pathogen life history that
298 are most influential in disease transmission. Consistent with models of other malarias, disease
299 transmission is critically sensitive to vector longevity. Accurate age-grading for natural anopheline
300 mosquitoes remains a major hurdle and most estimates come from ovarian examination of the
301 number of gonotrophic cycles that females have undergone (Cook and Sinkins 2010). Not even

rough estimates produced through this indirect measuring method are yet available for members of *Anopheles leucosphyrous* group. Additionally, this group is made up of several species that are morphologically impossible to distinguish (Sallum *et al.* 2005) and whose life histories, bite behaviours and thus contribution to *P. knowlesi* transmission are only just beginning to be uncovered (Tan *et al.* 2008; Vythilingam *et al.* 2006; Wong *et al.* 2015). Future modelling efforts incorporating entomological parameters will require allowing for considerable uncertainty – as incorporated here – until empirical information becomes available.

The current study constitutes the first endeavour in determining the probability of successful invasion following a *P. knowlesi* introduction into a susceptible population. This is particularly relevant for newly emerging infectious diseases because of their vulnerability of fade-out through random effects when infection numbers are low. To conduct this invasion analysis, it was assumed that the human hosts were immunologically naïve. In terms of *P. knowlesi* transmission, over 70% of infections are in individuals over the age of 20 years (Grigg, William et al in prep). This is not the epidemiological profile that would be expected if acquired immunity were an important transmission determinant locally. There is good evidence that *P. knowlesi* exhibits unstable transmission in humans (with a strong seasonal effect). Indeed, unstable transmission would be expected for a spill-over parasite. Together, these factors suggest that human populations that suffer from *P. knowlesi* infection do so through the repeat invasion of the parasite into humans from the macaque reservoir; and, that sustained transmission within humans over prolonged periods is seldom (if ever) experienced. Therefore, the assumption of an immunologically naïve human population with which to simulate *P. knowlesi* invasion currently seems appropriate.

The current study highlights vector biting behaviours (anthropophilic, switching and zoophilic i.e. Types II, III and IV) which result in maximum human-elicited invasion probabilities

across broad host availabilities. Critical in ascertaining the true threat that humans pose in transporting infection between different populations will be identification of the functional response in vector biting behaviour to variations in the availability of alternative blood hosts.

An in-depth analysis was conducted into how vectors respond to differing availabilities of alternative blood sources in terms of their host selection and how this impacts transmission. When non-linear responses are accounted for, quantitative differences arise in the parasite transmission numbers between species but qualitative differences emerge in the invasion probabilities. For example, when humans constitute two-thirds or more of the available blood hosts, invasions sparked by infected macaques are completely precluded when spread by vectors exhibiting Type II, III or IV responses. Establishing how local vector biting behaviour responds to a changing environment as humans increasingly encroach upon and supplant macaque habitats will be key to addressing the likelihood of *P*. *knowlesi* spread by human (or macaque) importation. Semi-field experiments using varied availabilities of alternative hosts and testing blood-meals of fed mosquitoes could help improve understanding of this behaviour.

Following the precedents of the major human malaria species *P*. *falciparum* and *P*. *vivax*, *P*. *knowlesi* may be in the process of emerging as a substantive agent of malaria from primates into human populations – and recent field studies suggest that distinct parasite strains have invaded human populations (Ahmed *et al.* 2014; Divis *et al.* 2015; Pinheiro *et al.* 2015). This offers a unique opportunity to identify the environmental drivers behind the parasite's evolution. To this end, the current study in which methods are developed to calculate invasion probabilities for multi-host malaria infections advances our ability to explore these important questions.

The present study highlights areas requiring further investigation. Biological understanding for *P*. *knowlesi* is germinal (although burgeoning) and currently dictates the appropriate level of complexity for disease models. Numerous host, parasite and environmental

factors impact the epidemiology of all malarias and the coming years can be expected to better equip us in building upon this initial effort to simulate *P. knowlesi* invasion. For example, haemaglobinopathies are known to impact malaria epidemiology and (particularly beta thalassaemia) occur at high rates in *P. knowlesi*-endemic populations. Currently, it is unknown whether/how these haemaglobinopathies affect susceptibility to *P. knowlesi* infection and these were consequently omitted from the current analysis. Additionally, given the overlapping endemicity with other malaria species in some regions, a future direction of the current work would be the exploration of the effects of *P. knowlesi* invasion in regions with *P. falciparum* and/or *P. vivax* already. However, much of our parameterisation comes from studies in Sabah where levels of *P. falciparum* and *P. vivax* transmission are very low and unlikely to impact *P. knowlesi* invasion.

Another shortcoming arising from data paucity is the need to resort to parameter values gleaned from classic malaria entomological and epidemiological studies. Recent genetic analysis suggests a lack of clustering of parasite genotypes in humans or macaques, which may be suggestive of zoonotic rather than human-vector-human transmission (Divis *et al.* 2015; Lee *et al.* 2011). However, a similar result would be anticipated under the circumstance that human outbreaks were limited in size i.e., transmission chains were relatively short. A comprehensive multivariate sensitivity analysis allowed detection of the model parameters for which direct estimates were as yet unavailable and that were simultaneously highly influential in disease transmission. As described above, mosquito longevity is highly influential, but, so too is the vector biting behaviour. Additionally, seasonal effects on vector species' (or sibling species') abundance (absolute as well as relative to one another) have only recently been described for *A. balabacensis* (Wong *et al.* 2015), and the integration of these new data into seasonally-driven entomological models constitutes important future work.

Following a successful control campaign, malaria incidence in Malaysia has declined considerably in recent years and targets have been set for imminent elimination (Cotter *et al.* 2011). Unfortunately, the current endemicity of *P. knowlesi* threatens elimination in this region (William *et al.* 2013). While informing the epidemiology and control of a considerable public health threat, rapid knowledge development in the ecology of this newly emerging disease can also be expected to provide invaluable insight into the evolutionary processes underlying successful pathogen invasion into humans.

COMPETING INTERESTS

We have no competing interests.

AUTHOR CONTRIBUTIONS

LY and MBB conceived the study; LY produced the model; LY and ALL carried out model analysis. All authors interpreted model output; contributed important intellectual content; and gave their final approval of the version to be published.

REFERENCES

Abdullahi, M. B., Hasan, Y. A. and Abdullah, F. A. (2013). Optimal Control of *Plasmodium knowlesi* Malaria in Human and Macaques. *British Journal of Mathematics & Computer Science* **4,** 271-287.

Abrams, P. A., Matsuda, H. and Harada, Y. (1993). Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. *Evolutionary Ecology* **7,** 465-487.

Ahmed, A. M., Pinheiro, M. M., Divis, P. C., Siner, A., Zainudin, R., Wong, I. T., Lu, C. W., Singh-Khaira, S. K., Millar, S. B., Lynch, S., Willmann, M., Singh, B., Krishna, S. and Cox-Singh, J. (2014). Disease Progression in *Plasmodium knowlesi* Malaria Is Linked to Variation in Invasion Gene Family Members. *PLoS Negl Trop Dis* **8,** e3086. doi: 10.1371/journal.pntd.0003086

Anonymous (2010). Basic Population Characteristics by Administrative Districts. Malaysian Department of Statistics,

Athreya, K. B. and Ney, P. E. (1972). *Branching processes,* Springer, Berlin, Germany.

Bartlett, M. S. (1956). Deterministic and stochastic models for recurrent epidemics. In *Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 4 (ed. Neyman, J.), University of California Press, Berkeley.

Besansky, N. J., Hill, C. A. and Costantini, C. (2004). No accounting for taste: host preference in malaria vectors. *Trends in Parasitology* **20,** 249-251. doi: http://dx.doi.org/10.1016/j.pt.2004.03.007

Bonnet, S., Gouagna, L. C., Paul, R. E., Safeukui, I., Meunier, J. Y. and Boudin, C. (2003). Estimation of malaria transmission from humans to mosquitoes in two neighbouring villages in south Cameroon: evaluation and comparison of several indices. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **97,** 53-59.

Coatney, G. R., Collins, W. E., Warren, M. and Contacos, P. G. (2003). *The Primate Malarias,* Division of Parasitic Diseases, CDC, Atlanta, US.

Collins, W. E. (2012). *Plasmodium knowlesi*: a malaria parasite of monkeys and humans. *Annual Review in Entomology* **57,** 107-121.

Cook, P. E. and Sinkins, S. P. (2010). Transcriptional profiling of Anopheles gambiae mosquitoes for adult age estimation. *Insect Molecular Biology* **19,** 745-751. doi: 10.1111/j.1365-2583.2010.01034.x

Cotter, C., Gosling, R., Smith Gueye, C., Phillips, A. A., Feachem, R. G. A., Moyes, C. and Hay, S. I. (2011). Atlas of Asia Pacific Malaria Elimination Network (APMEN). University of California, San Francisco Global Health Group,

Cox-Singh, J., Davis, T. M. E., Lee, K.-S., Shamsul, S. S. G., Matusop, A., Ratnam, S., Rahman, H. A., Conway, D. J. and Singh, B. (2008). *Plasmodium knowlesi* Malaria in Humans Is Widely Distributed and Potentially Life Threatening. *Clinical Infectious Diseases* **46,** 165-171. doi: 10.1086/524888

Diekmann, O. and Heesterbeek, J. a. P. (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation,* Wiley, Chichester, UK.

Diekmann, O., Heesterbeek, J. a. P. and Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface* **7,** 873-885. doi: 10.1098/rsif.2009.0386

Divis, P. C. S., Singh, B., Anderios, F., Hisam, S., Matusop, A., Kocken, C. H., Assefa, S. A., Duffy, C. W. and Conway, D. J. (2015). Admixture in Humans of Two Divergent *Plasmodium knowlesi* Populations Associated with Different Macaque Host Species. *PLoS Pathog* **11,** e1004888. doi: 10.1371/journal.ppat.1004888

Fornace, K. M., Nuin, N. A., Betson, M., Grigg, M. J., William, T., Anstey, N. M., Yeo, T. W., Cox, J., Ying, L. T. and Drakeley, C. J. (2015). Asymptomatic and Submicroscopic Carriage of *Plasmodium knowlesi* Malaria in Household and Community Members of Clinical Cases in Sabah, Malaysia. *Journal of Infectious Diseases* **213,** 784-787. doi: 10.1093/infdis/jiv475

Franks, S., Koram, K. A., Wagner, G. E., Tetteh, K., Mcguinness, D., Wheeler, J. G., Nkrumah, F., Ranford-Cartwright, L. and Riley, E. M. (2001). Frequent and Persistent, Asymptomatic *Plasmodium falciparum* Infections in African Infants, Characterized by Multilocus Genotyping. *Journal of Infectious Diseases* **183,** 796-804. doi: 10.1086/318834

Garret-Jones, C. (1964). The human blood index of malaria vectors in relation to epidemiological assessment. *Bulletin of the World Health Organization* **30,** 241-261.

Gillies, M. T. (1967). Experiments on host selection in the *Anopheles gambiae* complex. *Annals of Tropical Medicine and Parasitology* **61,** 68-75.

Imai, N., White, M. T., Ghani, A. C. and Drakeley, C. J. (2014). Transmission and Control of *Plasmodium knowlesi*: A Mathematical Modelling Study. *PLoS Negl Trop Dis* **8,** e2978. doi: 10.1371/journal.pntd.0002978

Lee, K.-S., Divis, P. C. S., Zakaria, S. K., Matusop, A., Julin, R. A., Conway, D. J., Cox-Singh, J. and Singh, B. (2011). *Plasmodium knowlesi*: Reservoir Hosts and Tracking the Emergence in Humans and Macaques. *PLoS Pathog* **7,** e1002015. doi: 10.1371/journal.ppat.1002015

Liu, W., Li, Y., Learn, G. H., Rudicell, R. S., Robertson, J. D., Keele, B. F., Ndjango, J.-B. N., Sanz, C. M., Morgan, D. B., Locatelli, S., Gonder, M. K., Kranzusch, P. J., Walsh, P. D., Delaporte, E., Mpoudi-Ngole, E., Georgiev, A. V., Muller, M. N., Shaw, G. M., Peeters, M., Sharp, P. M., Rayner, J. C. and

Hahn, B. H. (2010). Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* **467,** 420-425. doi: http://www.nature.com/nature/journal/v467/n7314/abs/nature09442.html#supplementary-information

Liu, W., Li, Y., Shaw, K. S., Learn, G. H., Plenderleith, L. J., Malenke, J. A., Sundararaman, S. A., Ramirez, M. A., Crystal, P. A., Smith, A. G., Bibollet-Ruche, F., Ayouba, A., Locatelli, S., Esteban, A., Mouacha, F., Guichet, E., Butel, C., Ahuka-Mundeke, S., Inogwabini, B.-I., Ndjango, J.-B. N., Speede, S., Sanz, C. M., Morgan, D. B., Gonder, M. K., Kranzusch, P. J., Walsh, P. D., Georgiev, A. V., Muller, M. N., Piel, A. K., Stewart, F. A., Wilson, M. L., Pusey, A. E., Cui, L., Wang, Z., Färnert, A., Sutherland, C. J., Nolder, D., Hart, J. A., Hart, T. B., Bertolani, P., Gillis, A., Lebreton, M., Tafon, B., Kiyang, J., Djoko, C. F., Schneider, B. S., Wolfe, N. D., Mpoudi-Ngole, E., Delaporte, E., Carter, R., Culleton, R. L., Shaw, G. M., Rayner, J. C., Peeters, M., Hahn, B. H. and Sharp, P. M. (2014). African origin of the malaria parasite *Plasmodium vivax*. *Nat Commun* **5,** doi: 10.1038/ncomms4346

Lloyd, A. L., Zhang, J. and Root, A. M. (2007). Stochasticity and heterogeneity in host–vector models. *Journal of the Royal Society Interface* **4,** 851-863. doi: 10.1098/rsif.2007.1064

Macdonald, G. (1956). Epidemiological basis of malaria control. *Bulletin of the World Health Organization* **15,** 613-626.

Okell, L. C., Bousema, T., Griffin, J. T., Ouedraogo, A. L., Ghani, A. C. and Drakeley, C. (2012). Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. *Nature Communications* **3,** 1237.

Pinheiro, M. M., Ahmed, M. A., Millar, S. B., Sanderson, T., Otto, T. D., Lu, W. C., Krishna, S., Rayner, J. C. and Cox-Singh, J. (2015). *Plasmodium knowlesi* Genome Sequences from Clinical Isolates Reveal Extensive Genomic Dimorphism. *PLoS ONE* **10,** e0121303. doi: 10.1371/journal.pone.0121303

Reiner, R. C., Perkins, T. A., Barker, C. M., Niu, T., Chaves, L. F., Ellis, A. M., George, D. B., Le Menach, A., Pulliam, J. R. C., Bisanzio, D., Buckee, C., Chiyaka, C., Cummings, D. a. T., Garcia, A. J., Gatton, M. L., Gething, P. W., Hartley, D. M., Johnston, G., Klein, E. Y., Michael, E., Lindsay, S. W., Lloyd, A. L., Pigott, D. M., Reisen, W. K., Ruktanonchai, N., Singh, B. K., Tatem, A. J., Kitron, U., Hay, S. I., Scott, T. W. and Smith, D. L. (2013). A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *Journal of the Royal Society Interface* **10,** doi: 10.1098/rsif.2012.0921

Rickman, L., Jones, T. R., Long, G. W., Paparello, S., Schneider, I., Paul, C. F., Beaudoin, R. L. and Hoffman, S. L. (1990). *Plasmodium falciparum*-infected *Anopheles stephensi* inconsistently transmit malaria to humans. *American Journal of Tropical Medicine and Hygiene* **43,** 441-445.

Sallum, M. a. M., Peyton, E. L., Harrison, B. A. and Wilkerson, R. C. (2005). Revision of the Leucosphyrus group of Anopheles (Cellia) (Diptera, Culicidae). *Revista Brasileira de Entomologia* **49,** doi: 10.1590/s0085-56262005000500001

Singh, B. and Daneshvar, C. (2013). Human infections and detection of *Plasmodium knowlesi*. *Clinical Microbiology Reviews* **26,** 165-184.

Singh, B., Sung, L. K., Matusop, A., Radhakrishnan, A., Shamsul, S. S. G., Cox-Singh, J., Thomas, A. and Conway, D. J. (2004). A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *The Lancet* **363,** 1017-1024. doi: http://dx.doi.org/10.1016/S0140-6736(04)15836-4

Smith, D. L., Battle, K. E., Hay, S. I., Barker, C. M., Scott, T. W. and Mckenzie, F. E. (2012). Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. *PLoS Pathog* **8,** e1002588. doi: 10.1371/journal.ppat.1002588

Takken, W. and Verhulst, N. O. (2013). Host Preferences of Blood-Feeding Mosquitoes. *Annual Review in Entomology* **58,** 433-453.

Tan, C. H., Vythilingam, I., Matusop, A., Chan, S. T. and Singh, B. (2008). Bionomics of *Anopheles latens* in Kapit, Sarawak, Malaysian Borneo in relation to the transmission of zoonotic simian malaria parasite *Plasmodium knowlesi*. *Malaria Journal* **7,** 1-8. doi: 10.1186/1475-2875-7-52

Vythilingam, I., Tan, C. H., Asmad, M., Chan, S. T., Lee, K. S. and Singh, B. (2006). Natural transmission of *Plasmodium knowlesi* to humans by *Anopheles latens* in Sarawak, Malaysia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **100,** 1087-1088. doi: 10.1016/j.trstmh.2006.02.006

White, M. T., Griffin, J. T., Akpogheneta, O., Conway, D. J., Koram, K. A., Riley, E. M. and Ghani, A. C. (2014). Dynamics of the Antibody Response to *Plasmodium falciparum* Infection in African Children. *Journal of Infectious Diseases* **210,** 1115-1122. doi: 10.1093/infdis/jiu219

Who (2014). World Malaria Report 2014. World Health Organization, Geneva, Switzerland.

William, T., Rahman, H. A., Jelip, J., Ibrahim, M. Y., Menon, J., Grigg, M. J., Yeo, T. W., Anstey, N. M. and Barber, B. E. (2013). Increasing Incidence of *Plasmodium knowlesi* Malaria following Control of *P. falciparum* and *P. vivax* Malaria in Sabah, Malaysia. *PLoS Neglected Tropical Diseases* **7,** e2026. doi: 10.1371/journal.pntd.0002026

Wong, M. L., Chua, T. H., Leong, C. S., Khaw, L. T., Fornace, K., Wan-Sulaiman, W.-Y., William, T., Drakeley, C., Ferguson, H. M. and Vythilingam, I. (2015). Seasonal and Spatial Dynamics of the Primary Vector of *Plasmodium knowlesi* within a Major Transmission Focus in Sabah, Malaysia. *PLoS Negl Trop Dis* **9,** e0004135. doi: 10.1371/journal.pntd.0004135

Yakob, L. (2016a). Endectocide-treated cattle for malaria control: A coupled entomological-epidemiological model. *Parasite Epidemiology and Control* **1,** 2-9. doi: http://dx.doi.org/10.1016/j.parepi.2015.12.001

Yakob, L. (2016b). How do biting disease vectors behaviourally respond to host availability? *Parasites & Vectors* **9,** 468.

Yakob, L., Bonsall, M. B. and Yan, G. (2010). Modelling knowlesi malaria transmission in humans: vector preference and host competence. *Malaria Journal* **9,** 329.

Yanuar, A., Chivers, D. J., Sugardjito, J., Martyr, D. J. and Holden, J. T. (2009). The Population Distribution of Pig-tailed Macaque (*Macaca nemestrina*) and Long-tailed Macaque (*Macaca fascicularis*) in West Central Sumatra, Indonesia. *Asian Primates Journal* **1,** 2-11.

538 Table 1. *Plasmodium knowlesi* mathematical model parameters, descriptions, median values
539 and source.

	Definition	Median Values Humans (Macaques)	Source
b_{VH}	Transmission coefficient (to humans); bite rate x transmission probability	0.1; 1/3 x 0.3	(Rickman *et al.* 1990)
b_{VN}	Transmission coefficient (to non-humans); bite rate x transmission probability	0.1; 1/3 x 0.3	(Rickman *et al.* 1990)
b_{HV}	Transmission coefficient (humans→vectors); bite rate x transmission probability	0.007; 1/3 x .02	(Bonnet *et al.* 2003)
b_{NV}	Transmission coefficient (non-humans→vectors); bite rate x transmission probability	0.007	(Bonnet *et al.* 2003)
m	Ratio of mosquitoes to all hosts (macaques & humans)	10	Assumption
γ	Recovery rate	0.07 (0) day^{-1}	(Coatney *et al.* 2003)
ε	Clearance rate of symptomatic infection	0.07 (0) day^{-1}	(Coatney *et al.* 2003)
κ	Clearance rate of asymptomatic infection	0.01 (0) day^{-1}	(Franks *et al.* 2001)
π	Asymptomatic primary infection rate	0.14 (0.14) day^{-1}	Assumption
θ	Susceptibility to secondary asymptomatic infection	1 (0)	Assumption
τ	Full susceptibility reversion rate	0.0057 (0) day^{-1}; 1/(ln(2)x3 years)	(White *et al.* 2014)
σ	Adjustment factor for asymptomatic transmissibility to vector	0.25 (0.25)	(Okell *et al.* 2012)
μ	Birth and death rate of hosts (i.e. stable population)	$3.4x10^{-5}(2.7x10^{-4})$ day^{-1}	(Anonymous 2010; Yanuar *et al.* 2009)
μ_V	Birth (or maturation) and death rate of vectors (i.e. stable population)	0.1 day^{-1}	(Yakob *et al.* 2010)
ζ	Rate of parasite development within vector	0.1 day^{-1}	(Collins 2012)

540

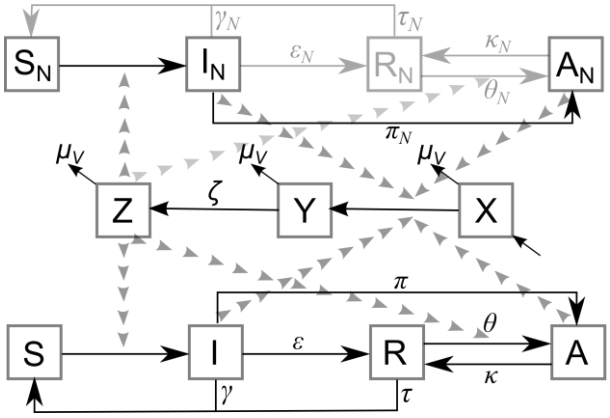
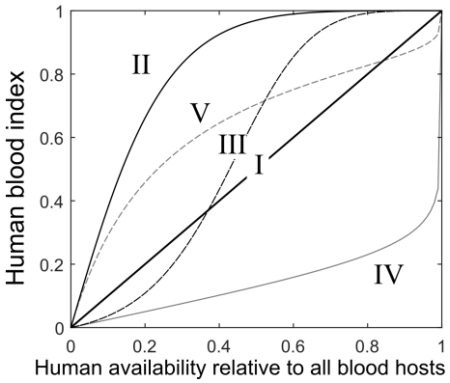


Figure 1. A general framework for multi-host vector-borne diseases. Top row: susceptible non-human hosts (S_N) become infectious (I_N) following an infectious bite from a vector, and then potentially recover (R_N) or become asymptomatically (and/or chronically) infected (A_N). Middle row: susceptible vectors (X) become infected (Y) and then infectious (Z), following successful pathogen transmission during a bloodmeal. Bottom row: susceptible human hosts (S) become infectious (I) following an infectious bite from a vector, and then potentially recover (R) or become asymptomatically (and/or chronically) infected (A). Current best understanding of this infection system is that macaques remain infected for many years (in the order of their lifetimes); but, should evidence arise that they clear infections (similar to the human system), the model allows for this development (shaded-out region of the transmission process).



559

560 Figure 2. The qualitatively distinct functional types in vector biting behaviour. Vector-borne
561 disease models ubiquitously assume that the human blood index is directly proportional to the
562 availability of humans relative to all blood hosts (Type I). In this study, alternative vector
563 behaviours are also modelled for comparative purposes. Parameterisation of Equation 12
564 needed to produce the curves for Types I-V were $\alpha=1, \beta=1$; $\alpha=0.25, \beta=4$; $\alpha=4, \beta=4$; $\alpha=4, \beta=0.25$;
565 $\alpha=0.25, \beta=0.25$.

566

567

568

569

570

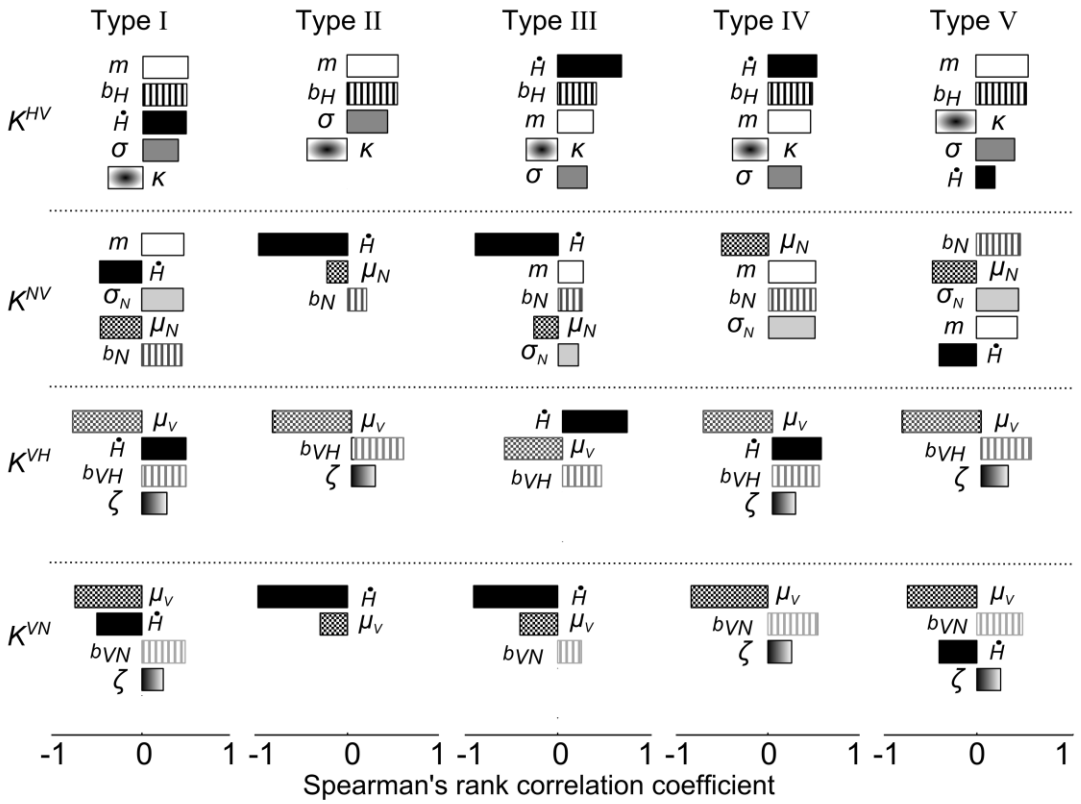
571

572

573

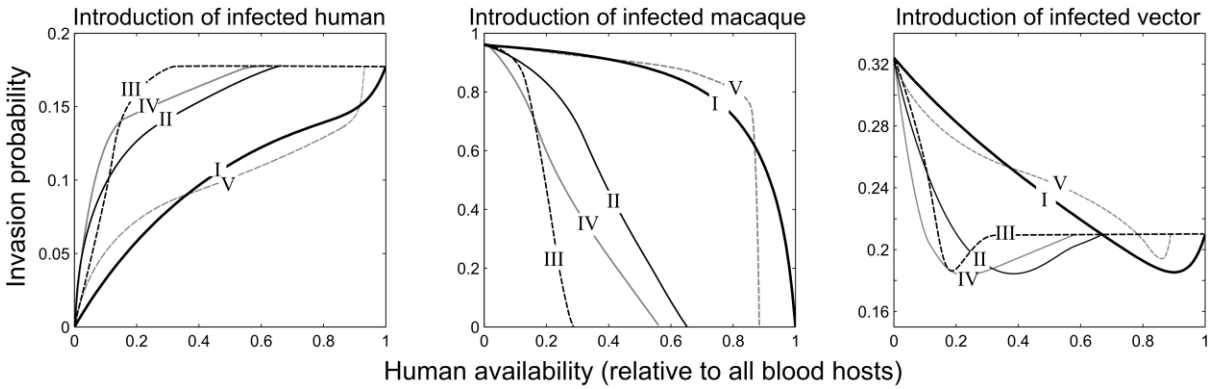
574

575



m: mosquito to host ratio; ***b***: transmission coefficient from host to vector (subscript ***H***:human, ***N***:nonhuman); $\dot{\boldsymbol{H}}$: human proportion of hosts; ***σ***: adjustment for asymptomatic transmissibility to vector (subscript ***N***:nonhuman); ***κ***: asymptomatic clearance; ***μ***: mortality (subscript ***N***:nonhuman host, ***V***:vector); ***ζ***: parasite development in vector

Figure 3. Multivariate sensitivity analysis for the different functional response Types. K^{VH}: average number of human infections arising from an infectious vector; K^{VN}: average number of macaque infections arising from an infectious vector; K^{HV}: average number of vector infections arising from an infectious human; K^{NV}: average number of vector infections arising from an infectious macaque. Results are shown for parameters that had Spearman's rank correlation coefficients of over 0.1 following 5000 iterations of a Monte Carlo simulation.



588

589 Figure 4. *Plasmodium knowlesi* invasion probabilities following introduction by infected human

590 (1-s_h), infected macaque (1-s_n) or infected vector (1-s_v). The lines are labelled with the different

591 functional Types in vector biting behaviour.