

# 1 **Updating estimates of *Plasmodium knowlesi* malaria risk in response to** 2 **changing land use patterns across Southeast Asia**

3 Ruarai J Tobin<sup>1, †</sup>, Lucinda E Harrison<sup>2</sup>, Meg K Tully<sup>1</sup>, Inke N D Lubis<sup>3</sup>, Rintis Noviyanti<sup>4</sup>, Nicholas M Anstey<sup>5</sup>,  
4 Giri S Rajahram<sup>6</sup>, Matthew J Grigg<sup>5</sup>, Jennifer A Flegg<sup>2</sup>, David J Price<sup>1,7</sup>, and Freya M Shearer<sup>1,8, †</sup>

5 <sup>1</sup>Infectious Disease Dynamics Unit, Melbourne School of Population and Global Health, The University of Melbourne,  
6 Melbourne, Australia

7 <sup>2</sup>School of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia

8 <sup>3</sup>Department of Paediatrics, Faculty of Medicine, Universitas Sumatera Utara, Medan, Indonesia

9 <sup>4</sup>Eijkman Research Center for Molecular Biology, BRIN, Jakarta, Indonesia

10 <sup>5</sup>Menzies School of Health Research and Charles Darwin University, Darwin, Australia

11 <sup>6</sup>Infectious Diseases Society Kota Kinabalu Sabah, Menzies School of Health Research Clinical Research Unit, Hospital  
12 Queen Elizabeth II, and Clinical Research Centre, Queen Elizabeth Hospital, Ministry of Health, Kota Kinabalu, Malaysia

13 <sup>7</sup>Doherty Institute for Infection and Immunity, The Royal Melbourne Hospital and The University of Melbourne,  
14 Melbourne, Australia

15 <sup>8</sup>Infectious Disease Ecology Modelling Group, Telethon Kids Institute, Perth, Australia

16 <sup>†</sup>Corresponding authors: Ruarai J Tobin ([ruarai.tobin@unimelb.edu.au](mailto:ruarai.tobin@unimelb.edu.au)) and Freya M Shearer

17 ([freya.shearer@unimelb.edu.au](mailto:freya.shearer@unimelb.edu.au))

## 19 **Abstract**

### 20 **Background**

21 *Plasmodium knowlesi* is a zoonotic parasite that causes malaria in humans. The pathogen has a natural host  
22 reservoir in certain macaque species and is transmitted to humans via mosquitoes of the *Anopheles* Leucos-  
23 phyrus Group. The risk of human *P. knowlesi* infection varies across Southeast Asia and is dependent upon  
24 environmental factors. Understanding this geographic variation in risk is important both for enabling ap-  
25 propriate diagnosis and treatment of the disease and for improving the planning and evaluation of malaria  
26 elimination. However, the data available on *P. knowlesi* occurrence are biased towards regions with greater  
27 surveillance and sampling effort. Predicting the spatial variation in risk of *P. knowlesi* malaria requires meth-  
28 ods that can both incorporate environmental risk factors and account for spatial bias in detection.

### 29 **Methods & Results**

30 We extend and apply an environmental niche modelling framework as implemented by a previous mapping  
31 study of *P. knowlesi* transmission risk which included data up to 2015. We reviewed the literature from Octo-  
32 ber 2015 through to March 2020 and identified 264 new records of *P. knowlesi*, with a total of 524 occurrences  
33 included in the current study following consolidation with the 2015 study. The modelling framework used  
34 in the 2015 study was extended, with changes including the addition of new covariates to capture the effect  
35 of deforestation and urbanisation on *P. knowlesi* transmission.

### 36 **Discussion**

37 Our map of *P. knowlesi* relative transmission suitability estimates that the risk posed by the pathogen is  
38 highest in Malaysia and Indonesia, with localised areas of high risk also predicted in the Greater Mekong  
39 Subregion, The Philippines and Northeast India. These results highlight areas of priority for *P. knowlesi*  
40 surveillance and prospective sampling to address the challenge the disease poses to malaria elimination  
41 planning.

## 42 **Author Summary**

43 *Plasmodium knowlesi* is a parasite that can cause malaria when it infects humans. Although most people do  
44 not experience severe illness from *Plasmodium knowlesi* infection, a small number will develop serious or  
45 even fatal disease. The parasite is found naturally in some monkeys throughout Southeast Asia, and spreads  
46 from these monkeys to humans through mosquitoes. Previous research predicted where the risk of being  
47 infected is highest according to what we know about the environment across Southeast Asia, such as if there  
48 are forests in an area or if the altitude is high. In this work, we extend this previous research with more up-to-  
49 date data on environmental conditions and infections to predict the risk of being infected with *Plasmodium*  
50 *knowlesi*. We show that the risk *Plasmodium knowlesi* poses to humans is high across much of Southeast  
51 Asia, and that the disease will continue to challenge national goals to eliminate malaria.

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## 52 Introduction

53 *Plasmodium knowlesi* is a zoonotic pathogen of growing public health concern in Southeast Asia. The pathogen  
54 has a reservoir in *Macaca fascicularis* and the closely related *Macaca nemestrina* and *Macaca leonina* macaques,  
55 and is transmitted between macaques and from macaques to humans via mosquito vectors of the *Anopheles*  
56 *Leucosphyrus* Group [1, 2]. Although demonstrated experimentally [3], evidence of direct human-to-human  
57 transmission of *P. knowlesi* occurring in nature is limited [4, 1, 5, 6, 7]. Infection by *P. knowlesi* most often  
58 causes mild to moderate illness in humans [8]. However, a range of outcomes are possible, with both asymp-  
59 tomatic infection [9, 10, 11] and severe disease being reported. Studies of patients presenting to health care  
60 facilities in Malaysia reported severe disease in around 6-9% of patients [12, 13].

61 For humans, the likelihood of contracting a *P. knowlesi* infection has been found to be dependent upon a  
62 range of risk factors, with case-control and seroprevalence studies demonstrating associations between envi-  
63 ronmental variables and the occurrence of infection. A seroprevalence survey performed in Malaysia and the  
64 Philippines found that prior infection with *P. knowlesi* was associated with the proximity of forested areas to  
65 an individual's home and the clearing of forest near their home [14]. A similar study performed in northern  
66 Sabah, Malaysia, found associations between prior infection and an individual reporting that they have had  
67 activity in forested areas or that they have had contact with macaques [15]. A population-based case-control  
68 study performed within Sabah, Malaysia, found an association between current *P. knowlesi* infection and an  
69 individual reporting either that they had recently cleared vegetation or that their home was in proximity to  
70 long grass [11].

71 The spatial epidemiology of *P. knowlesi* malaria has historically been poorly understood. This is partially due  
72 to widespread misdiagnosis. By clinical presentation, the symptoms of *P. knowlesi* infection can be easily mis-  
73 attributed to other major human species of malaria such as *P. vivax* or *P. falciparum* [16]. Under microscopic  
74 examination, the parasite appears almost identical to *P. malariae* [17] and the early ring stages of *P. faldi-*  
75 *parum* [18]. One review of historical microscopy diagnoses demonstrated that across 375 studies, 57% of *P.*  
76 *knowlesi* infections were misdiagnosed [17]. In addition to misdiagnosis, the understood spatial distribution  
77 of *P. knowlesi* malaria has been biased by differences in surveillance effort. Within peer-reviewed literature,  
78 reported *P. knowlesi* infections are most common in Malaysia, which is likely reflective of both high burden  
79 and a substantial surveillance effort in the country [19]. Indigenous cases of *P. knowlesi* malaria have also  
80 been detected in Brunei, Cambodia, Indonesia, Laos, Myanmar, the Philippines, Thailand, and Vietnam, but  
81 these have historically been the result of small-scale prospective sampling efforts and individual case reports.  
82 One study reported identifying *P. knowlesi* in India within the Andaman and Nicobar Islands [20].

83 The elimination of malaria in at least 20 countries by 2025 is listed as a key milestone of the World Health Or-  
84 ganisation's 2016–2030 *Global technical strategy for malaria* [21]. *P. knowlesi* presents a challenge to these ef-  
85 forts, since interventions that are effective against the human malaria species such as indoor residual spraying  
86 will be less effective against *P. knowlesi* due to the pathogen's persistence in wildlife reservoirs. Furthermore,  
87 cross-reactivity of antibodies between *P. knowlesi* and the closely genetically related *P. vivax* may provide pro-  
88 tection against *P. knowlesi* infection [22], implying that the elimination of *P. vivax* in a region could lead to  
89 reduced immunity and subsequently an increase in the number of *P. knowlesi* infections [19].

90 The incidence of *P. knowlesi* in humans appears to be increasing within Southeast Asia; in Malaysia, the num-  
91 ber of recorded human *P. knowlesi* infections doubled over the period from 2015 to 2018 [23]. A similar trend  
92 is visible in the rising number of case reports within Indonesia [24]. Though these trends may simply reflect  
93 improvements in surveillance [23], it has been suggested that deforestation in the region may be leading to a  
94 real increase in the number of human *P. knowlesi* infections [25, 24, 26, 27]. A primary driver of deforestation  
95 in the region is the development of oil palm or timber plantations, which produce an environment that is  
96 believed to be of enhanced risk for *P. knowlesi* infection, with plantation labourers being required to live and  
97 work in proximity to recently disturbed forests that may contain *P. knowlesi* reservoirs and vectors.

98 Sustained transmission of vector-borne zoonoses can only occur at the nidus where pathogen, host and vec-  
99 tor are present in sufficient abundance [28]. For each of these, certain constraints limit their distribution, for  
100 example: a pathogen may be unable to survive at certain temperatures; a host may be displaced by human  
101 activity; and a vector may be unable to reproduce without access to standing water. The field of geospatial  
102 information systems (GIS) provides a large amount of data on such environmental and anthropological fac-  
103 tors [29]. Environmental niche modelling utilises this geospatial data to identify relationships between the  
104 presence of a pathogen, host or vector and the environments in which they have been observed, allowing  
105 for prediction of the suitability for transmission of a vector-borne zoonosis such as *P. knowlesi* across a geo-  
106 graphic area of interest [30].

107 In 2015, Shearer and colleagues applied a niche modelling approach to produce the first predictive map of  
108 *P. knowlesi* malaria risk across Southeast Asia [31]. This map provided an initial evidence base for identifying  
109 areas where disease surveillance and epidemiological investigations would be most informative to improve  
110 understanding of *P. knowlesi* malaria risk. Since the publication of the 2015 occurrence database and risk  
111 map, the volume of *P. knowlesi* data has increased across Southeast Asia, with this including detections of  
112 the pathogen in new locations. As new data accrues, it is important to update risk predictions to ensure that  
113 the most up-to-date evidence is available to public health researchers, practitioners, and policymakers. Fur-  
114 thermore, since 2015, studies providing evidence of the importance of deforestation in the risk of *P. knowlesi*  
115 malaria have been published, and novel datasets characterising spatial and temporal variation in land use  
116 patterns have become available.

117 In this study, we present updates to the *P. knowlesi* infection database and risk map produced in 2015 [31].  
118 We perform a comprehensive review of the literature from October 2015 through to March 2020 to produce  
119 a consolidated database of *P. knowlesi* infection occurrences across Southeast Asia. By combining this oc-  
120 currence dataset with data on a range of environmental covariates using a niche modelling framework, we  
121 produce updated predictions of relative suitability for *P. knowlesi* transmission to humans at fine-scale across  
122 Southeast Asia. We compare the outputs of our model to those from the 2015 model.

## 123 **Methods**

### 124 **Infection Data**

125 The infection occurrence database is a listing of reported locations of *P. knowlesi* infections in either humans,  
126 macaques or mosquitoes. The infection occurrences used in the 2015 analysis were extracted from litera-  
127 ture published up to October 2015. In order to identify new occurrences, we searched the ‘Web of Science’  
128 database on March 2nd 2020, using the keywords “knowlesi” or “monkey malaria” and filtered for results pub-  
129 lished after October 2015 (Figure 1A). Following the exclusion of laboratory studies, we extracted infection oc-  
130 currence records from publications which utilised validated *P. knowlesi*-specific diagnostics (i.e. semi-nested  
131 PCR or a combination of microscopy and molecular techniques, as in the 2015 review [31]). The data col-  
132 lection protocol used was the same as in the 2015 analysis and further detail can be found therein [31]. We  
133 combined the collected infection occurrences produced by the current study ( $n = 264$ , Figure 1) with those  
134 identified in the 2015 analysis ( $n = 260$ ).

135 Each location in the infection occurrence database could either take the form of a point or a polygon record.  
136 We created point records where the likely exposure site was reported with enough precision that it could be  
137 assigned to a  $5 \times 5$  km grid cell. Where this level of precision was not available, we created *polygonal* records,  
138 assigning the likely exposure site to a region bounded by a polygon (Figure 1B). We created these polygons as  
139 either administrative level 1 (the first subdivision below national, e.g. state or province) or administrative level  
140 2 (the second subdivision below national, e.g. district or regency) then disaggregated these polygons onto to  
141 the  $5 \times 5$  km grid for model fitting, prediction and evaluation.

142 Prior to model fitting and evaluation, we excluded nine records which spanned an area greater than 1,000  
143 grid cells (approximately  $25,000 \text{ km}^2$ ). These records were unlikely to affect results given that each had sub-  
144 stantial overlap with other more precise spatial records.

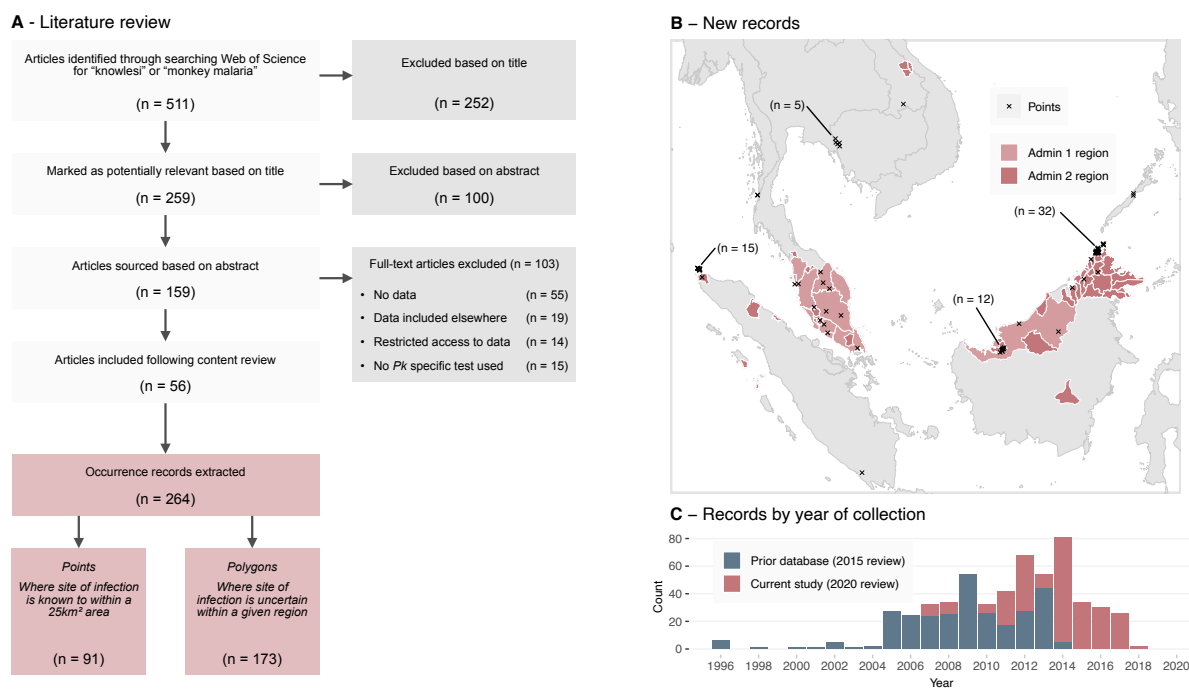


Figure 1: **A:** Study and sample selection process for the 2020 infection occurrence database. Records were produced via a literature review which was performed on March 2nd 2020, filtering for publications released after October 2015. **B:** Newly extracted point and polygon occurrence records across Southeast Asia by spatial type. Admin 1 regions are the first subdivision below national, e.g. state or province. Admin 2 regions are the second subdivision below national, e.g. district or regency. **C:** The number of occurrence samples in each occurrence database by the year the sample was collected.

## 145 Covariate Data

146 The infection risk model incorporated 20 environmental covariates (Table 1), each a 5 × 5 km gridded raster  
 147 covering Southeast Asia. Of these 20 covariates, we treated 14 as time-varying with an annual resolution,  
 148 allowing the model to associate each infection occurrence records with covariate values corresponding to the  
 149 year the infection was recorded, capturing the variation of risk factors over time. Data for these annually-  
 150 varying covariates were available for each year from 2001 to 2019, extending upon the coverage of the 2015  
 151 model (which covered 2001 to 2015). We assigned five samples which were collected before the year 2001  
 152 covariate values for the year 2001.

153 While tasseled-cap values (transformed Landsat imagery which can help differentiate areas of vegetation  
 154 and urbanisation) and human population density were included as synoptic (static) variables in the 2015  
 155 model, in this work we incorporated them as temporally-varying covariates. The 2015 model incorporated  
 156 an urban accessibility metric which defined the travel time to the nearest city of 50,000 people or more by  
 157 land- or water-based travel in the year 2000 [32]. Here, we instead used the healthcare accessibility surface  
 158 — a modelled measure of travel time to the nearest healthcare facility produced by the Malaria Atlas Project  
 159 which used data up to mid-2019 [33] — as a measure of urban accessibility.

160 We replaced the intact and disturbed forest coverage layers used in the 2015 model with covariates that bet-  
 161 ter captured the temporal and spatial dynamics of forest change in Southeast Asia. The forest coverage data  
 162 sets used in the 2015 model were derived from the Intact Forest Landscapes project, which utilised a strict,  
 163 manually assessed criteria for defining intact versus disturbed forest [34, 31, 35]. However, the temporal res-  
 164 olution of this dataset is low, with data only available for four distinct years (2000, 2013, 2016 and 2020). We  
 165 chose instead to utilise data provided through the Global Forest Change project, which provides annual data  
 166 on tree coverage over the last 20 years on forest presence at the resolution of 1 arc-second (roughly 30 m) [36].



167 We aggregated the Global Forest Change dataset up to the  $5 \times 5$  km grid over the Southeast Asia study region  
 168 through the calculation of both a tree coverage and a tree loss metric. We defined tree loss to be the proportion  
 169 of forest area lost within each  $5 \times 5$  km cell for each study year. Similarly, we defined tree coverage as the  
 170 proportion of land where forest coverage was present at the beginning of the Global Forest Change data period  
 171 and where no subsequent loss was recorded up until each study year. As the Global Forest Change project  
 172 has not calculated forest gain past the year 2012, we were not able to include any possible increase in forest  
 173 coverage.

Name	Description	Temporal resolution
<b>Host species distribution</b>		
<i>Macaca fascicularis</i> suitability	Modelled suitability of inhabitation by macaques of species <i>M. fascicularis</i> [35].	Synoptic
<i>Macaca nemestrina</i> suitability	Modelled suitability of inhabitation by macaques of species <i>M. nemestrina</i> [35].	Synoptic
<i>Anopheles</i> Leucosphyrus Group suitability	Modelled suitability of inhabitation by mosquitoes of the <i>Anopheles</i> Leucosphyrus Group [35].	Synoptic
<b>Environmental</b>		
SRTM elevation	Mean elevation [37]	Synoptic
Tasseled cap wetness s.d.	Tasseled-cap transformed MODIS data [38, 39]. <b>Now treated as temporally-varying.</b>	Annual
Tasseled cap wetness mean	“ “ “	Annual
Tasseled cap brightness s.d.	“ “ “	Annual
<i>Plasmodium falciparum</i> temperature suitability	Modelled temperature suitability index for <i>P. falciparum</i> transmission used as proxy for suitability of <i>P. knowlesi</i> [40].	Synoptic
Forest loss	Proportion of land where forest coverage has been lost in a given year [36]. <b>Replaced the disturbed forest dataset.</b>	Annual
Forest coverage	Proportion of land with forest coverage present in a given year [36]. <b>Replaced the intact forest dataset.</b>	Annual
<b>Sociodemographic</b>		
Healthcare accessibility	Modelled duration travel time to the nearest healthcare facility [33]. <b>Replaced the urban accessibility dataset.</b>	Synoptic
WorldPop human population	Mean human population density [41, 42]. <b>Now treated as temporally-varying.</b>	Annual
<b>MODIS/IGBP landcover</b>		
Open shrublands	Proportion of land with given land classification [43].	Annual
Woody savannas	“ “ “	Annual
Savannas	“ “ “	Annual
Grasslands	“ “ “	Annual
Permanent wetlands	“ “ “	Annual
Croplands	“ “ “	Annual
Cropland/natural vegetation mosaic	“ “ “	Annual
Urban and built up	“ “ “	Annual

Table 1: The set of raster covariate datasets used in model fitting and prediction. Differences in raster datasets between this work and those used in the 2015 *P. knowlesi* risk model appear in bold. STRM: Shuttle Radar Topography Mission, MODIS: Moderate Resolution Imaging Spectroradiometer, IGBP: International Geosphere-Biosphere Programme.

## 174 Model Fitting

175 We utilised a bootstrapped boosted regression tree modelling framework to characterise relationships be-  
176 tween a region's environment and the occurrence of *P. knowlesi* transmission. Regression trees produce an  
177 approximation of some latent function (e.g. the probability of a *P. knowlesi* infection occurring) by recursively  
178 splitting across potential predictor variables (e.g. environmental covariates). The points at which these splits  
179 occur and the value assigned across each split region are selected such that the error between the regres-  
180 sion tree and the observations is minimised [44]. Boosted regression trees extend upon the regression tree  
181 framework by producing a large number of trees and combining them in an ensemble (a process known as  
182 boosting) such that they better approximate the latent function [45]. Boosted regression trees are able to fit  
183 complex nonlinear responses including high-dimensional interactions between explanatory variables due to  
184 their hierarchical tree structure and have been shown to exhibit high predictive accuracy [46]. Finally, boot-  
185 strapping of the boosted regression tree process can be performed, allowing for uncertainty in the output to  
186 be estimated [47].

187 When applied to presence-absence data (such as from a systematic survey), niche models generally use a  
188 binomial likelihood to represent the probability of a species being present at a given location. Where most  
189 of the data available for modelling are presence-only, as is the case for *P. knowlesi* malaria, it is common  
190 practice in niche modelling to supplement occurrence records with “background” points to represent areas  
191 where the species or disease has not been reported [46]. A variety of approaches have been employed to select  
192 background points, including sampling to ensure that their spatial distribution emulates the sampling bias in  
193 the presence records [48].

194 Most *P. knowlesi* occurrences to date are recorded in Malaysia, Brunei and Singapore, with all three of these  
195 countries having eliminated the human malaria species (e.g., *P. vivax* and *P. falciparum*), such that that *P.*  
196 *knowlesi* is routinely considered a potential cause of malaria cases. Outside of these countries, surveillance for  
197 *P. knowlesi* is limited and infection records are sparse. As per the 2015 study, the goal of our niche modelling  
198 analysis is to predict broadly into the under-sampled regions outside of Malaysia, Brunei, and Singapore,  
199 using a model fit to data from within these three countries (i.e. the model training region, Figure 2A) where  
200 we can account for reporting bias through the selection of background points. Data from outside of these  
201 three countries formed the evaluation dataset (i.e. the model evaluation region, Figure 2B), which we used to  
202 assess the model's predictive ability outside of the training region.

203 To produce background points for the human and mosquito records, as in the 2015 model [31], we sampled  
204 points across the training region, with this sampling weighted by human population density [41] under the  
205 assumption that more populous areas would have a greater probability of reporting human cases and that the  
206 locations of mosquito infection studies were selected based on the presence of human *P. knowlesi* cases. To  
207 produce macaque background records, we sampled points from a survey of macaques and other mammals  
208 [35], as we expected this survey to have similar sampling bias to that of macaque *P. knowlesi* infection records.  
209 This approach is not biased by the under-ascertainment of *P. knowlesi* infections that arise due to asymp-  
210 tomatic/submicroscopic or spontaneously resolving disease [9, 10, 11], given such effects would be expected  
211 to be uniform geographically.

212 As in the 2015 model [31], the geographic distribution of *Macaca leonina* — a putative host species of *P.*  
213 *knowlesi* which was only classified as a species distinct from *Macaca nemestrina* in 2001 [49] — has not been  
214 included as an explanatory covariate in model fitting as the species is not found in the model training region.

215 To produce each bootstrap we performed sampling with replacement across each of the combined occur-  
216 rence polygons, occurrence point records and background points, using occurrence records present in the  
217 training region of Malaysia, Brunei and Singapore. We constrained this sampling so that at least 10 presence  
218 and 10 background points were present within each bootstrap. We then degraded the occurrence polygon  
219 records sampled to points via spatially uniform sampling of a singular point across the set of points bounded  
220 by each polygon (Figure 2A). For each bootstrap, we assigned weights to sampled points such that the sum of  
221 weights for presence points was equal to the sum of weights for the background points, and environmental  
222 values were assigned to each point from the set of covariate rasters corresponding to the spatial location and  
223 year the sample was recorded. We produced a covariate for host species, indicating if the sample was collected  
224 from a human, a mosquito or a macaque. We repeated this process to produce 500 bootstrapped datasets.

225 For each bootstrapped dataset, we fit boosted regression trees using the `gbm3` and `seegSDM` packages. Hy-  
226 perparameters for model fitting were unchanged from the defaults provided by `seegSDM` version 0.1-9 (initial  
227 trees = 10, learning rate/shrinkage = 0.005, tree complexity = 4, maximum trees = 10,000). We produced pre-  
228 dictions across each of the 500 bootstrapped models, with summary statistics including mean, variance, and  
229 interquartile range calculated for each  $5 \times 5$  km grid cell across Southeast Asia (Figure S1). As in the 2015  
230 model, we restricted predictions to areas within the range of macaque and mosquito species known to be re-  
231 quired for zoonotic transmission of *P. knowlesi* (i.e. the overlap in range maps of at least one reservoir and one  
232 vector species), using predicted species extent maps previously reported [35]. This includes areas where such  
233 populations may not yet be present, such as Sulawesi, where *M. fascicularis* and *M. nemestrina* macaques are  
234 currently kept as pets and there is the potential for a feral population to establish [35].

235 We produced a multivariate environment similarity surface (MESS) map (Figure 2C), indicating geographic  
236 areas where the value of at least one environmental covariate was outside the range of values present in the  
237 training data (i.e. the model is extrapolating) or vice-versa [50].

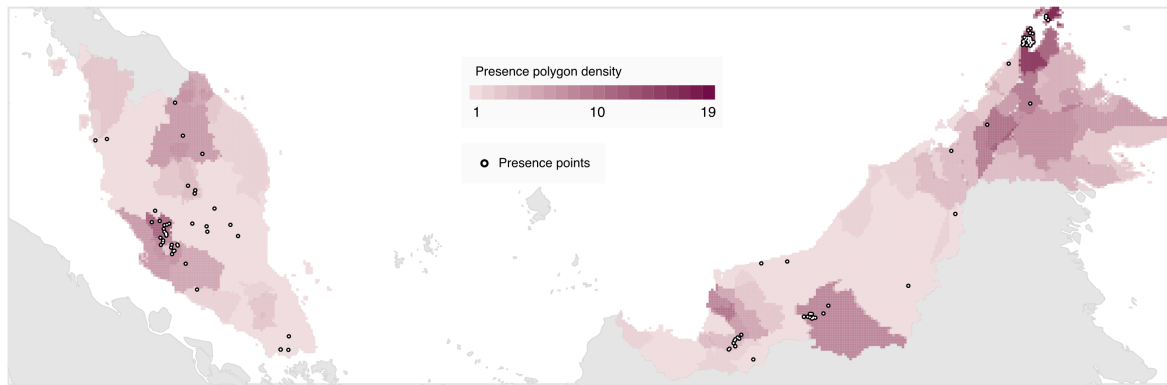
238 Prediction results for each bootstrapped model, rasters of summary statistics, the code used to produce  
239 results, and the updated occurrence database have been made available at [osf.io/k5bsa](https://osf.io/k5bsa) (DOI 10.17605/  
240 OSF.IO/K5BSA).

## 241 **Model Evaluation**

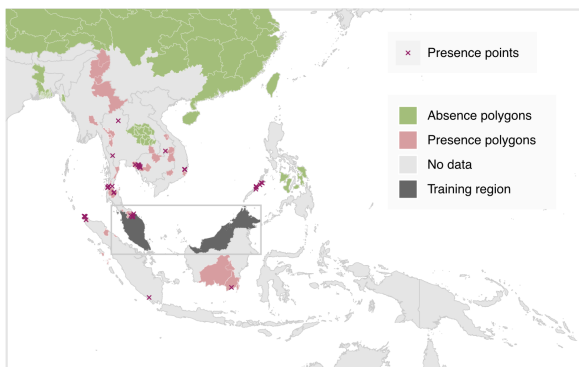
242 We evaluated the model's predictive performance by calculating the area under the curve (AUC) metric across  
243 both the training and evaluation datasets. For the training dataset, we estimated a 10-fold cross-validated  
244 AUC throughout the tree count optimisation process, and reported the training AUC for each bootstrap as that  
245 of the optimal model selected. Across the evaluation dataset, we calculated AUC across each bootstrapped  
246 model, with pairwise distance selection of samples performed to avoid spatial sorting bias [51].

247 We calculated covariate relative influence scores for each bootstrapped model, representing the number of  
248 times a variable is selected for regression tree splitting, weighted by the squared improvement to the model as  
249 a result of each split and averaged over all trees [52]. We summarised these scores across the models as means  
250 and 95% confidence intervals, with mean values also being used to rank the relative covariate importance.  
251 We further calculated accumulated local effect (ALE) scores to describe the average effect of a covariate on  
252 the prediction value across the range of each covariate. The ALE score achieves this by identifying how the  
253 model prediction changes in response to small changes in the covariate of interest while all other covariates  
254 are kept constant, allowing for the effects of covariates to be identified even when the covariates may be highly  
255 correlated [53].

A – Training region



B – Evaluation region



C – MESS

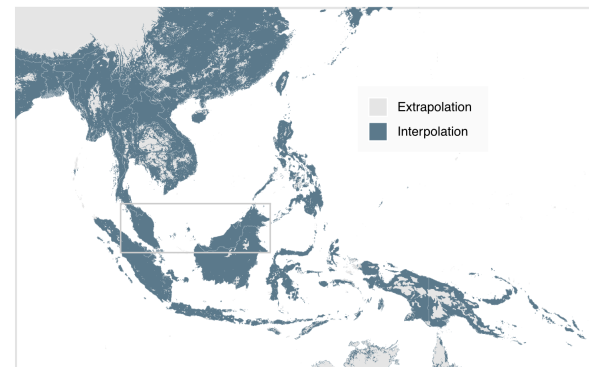


Figure 2: **A:** The data-set of occurrence points and polygons used for fitting the boosted regression tree model across the model training region of Malaysia, Brunei, and Singapore. Presence polygons are displayed as the number of polygons covering each given pixel, with this density being proportional to the probability distribution of points sampled from the polygons for each bootstrap. **B:** The presence and absence records used in the model evaluation process, across the evaluation region of Southeast Asia excluding Malaysia, Brunei and Singapore. **C:** Multivariate environmental similarity surface (MESS) for the model, where areas shaded in light grey indicate that at least one covariate value at that point is outside the range of values within the training data (extrapolation).

## 256 Results

### 257 Infection data

258 The literature review of articles including data on *P. knowlesi* infection occurrences published between Oc-  
259 tober 2015 and March 2020 returned 511 candidate articles. Following a review of titles and abstracts, 159  
260 articles were deemed likely to contain data for extraction, and 56 articles were identified as meeting the final  
261 criteria (Figure 1A). From these 56 articles, 264 occurrences of *P. knowlesi* were extracted, with 91 (34%) be-  
262 ing assigned a point record type and 173 (66%) being assigned a polygon record type. Of the 264 extracted  
263 records, 241 (91%) were infections identified in humans, with only 14 in macaques and nine in mosquitoes  
264 (Table S1). The number of records by year of sample collection was greatest in 2014 with 80 records across 14  
265 publications (Figure 1C).

266 A majority of records added to the 2015 database were collected in Malaysia ( $n = 201$ , 76%). Within Malaysia,  
267 the spatial distribution of records was highly heterogeneous (Figure 1, Table S1), with 127 polygon records as-  
268 signed to the region of Sabah in contrast to the three records identified in the capital region of Kuala Lumpur.  
269 Malaysia was also the location of eight of the nine observed infected mosquitoes in the dataset, consistent  
270 with the greater sampling effort within the country [19].

271 Our literature search reveals that more infection occurrences from Indonesia have been reported since 2015,  
272 comprising 17% ( $n = 45$ ) of the new presence records (where the prior 2015 literature search identified only  
273 five infections within the country). These records are the result of a small number of high-quality surveys and  
274 case reports from Aceh [54] and North Sumatra [55]. The literature review dataset contains three records from  
275 Laos, where the first confirmed human *P. knowlesi* infection was reported in 2016 [56].

276 In combination with the 260 occurrences used in the 2015 analysis, the total number of infection occurrences  
277 used in model fitting and evaluation was 524. Of these, 396 were within the training region of Malaysia, Brunei  
278 and Singapore, with the remaining 128 located elsewhere in Southeast Asia (Table S1).

### 279 Transmission Suitability Model Output

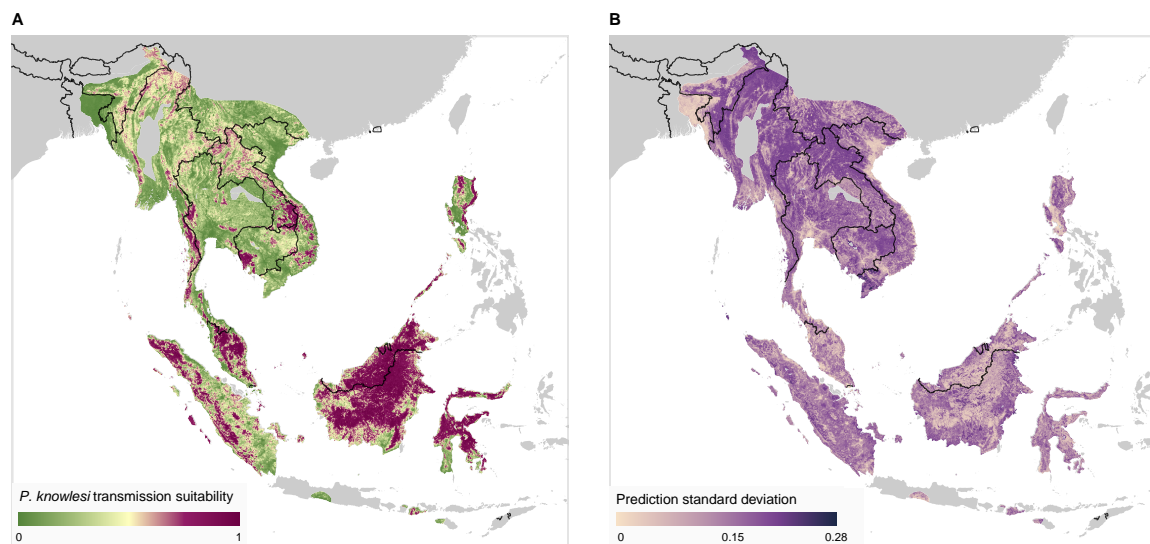


Figure 3: **A:** Modelled transmission suitability mean over Southeast Asia across the 500 bootstraps. Results are displayed only where an area is within the range of both a vector and reservoir species necessary for transmission (see Methods), regions outside of this range (displayed as grey) are considered to be very low risk for *P. knowlesi* transmission. Transmission suitability is a relative measure of the risk of *P. knowlesi* transmission from known reservoir species (via vector species) to humans. **B:** Standard deviation of the predicted transmission suitability across the 500 bootstraps.

280 The mean and standard deviation of predicted *P. knowlesi* transmission suitability across at-risk areas of  
281 Southeast Asia is presented in Figure 3. Further summary statistics of transmission suitability are presented  
282 in Figure S1.

283 The map of *P. knowlesi* transmission suitability (Figure 3A) shows highly heterogeneous levels of predicted  
284 risk across Southeast Asia. On the island of Borneo, all areas other than lower-lying coastal regions are ex-  
285 pected to have a relatively high risk of *P. knowlesi* transmission. Other more sparsely distributed areas of rela-  
286 tively high risk are predicted in Indonesia within the provinces of Sulawesi, Sumatra and West Nusa Tenggara.  
287 Peninsula Malaysia is predicted to have inland areas of high transmission risk. Thailand, Laos, Cambodia,  
288 Vietnam, Myanmar and the Philippine island of Luzon have smaller, localised areas of high predicted risk,  
289 with greater uncertainty in these predictions (Figure 3B) as a result of environmental differences to the model  
290 training region of Malaysia, Brunei and Singapore.

291 Within the training region, a mean area under the curve (AUC) of 0.81 was produced across the 500 boot-  
292 strapped models with a standard error of 0.001. For the evaluation region, the mean AUC was found to be 0.75  
293 with a standard error of 0.003. These values indicate a high degree of predictive performance.

294 Examining predictions within the evaluation region of the model (Southeast Asia excluding Malaysia, Brunei  
295 and Singapore), we may qualitatively assess the model's predictive performance. Regions with both a high  
296 modelled transmission suitability and previously identified occurrence samples of *P. knowlesi* — indicative of  
297 good model sensitivity — include the Aceh province of Sumatra island in Indonesia, the Koh Kong province  
298 in southern Cambodia and the Mimaropa region of the Philippines (Figure S2A). We also see that there are a  
299 substantial number of regions where the model predicts high transmission suitability where *P. knowlesi* oc-  
300 currence has not previously been identified as of the 2020 literature review (i.e. omission errors [57]). These  
301 areas include much of northern Sulawesi and the province of West Nusa Tenggara in Indonesia (Figure S2A).  
302 Such predictions may be suggestive of a lack of surveillance in these regions, or that an environment is con-  
303 ducive to transmission but currently lacking widespread occurrence of a necessary vector or host species (e.g.  
304 in Sulawesi, there are no native *M. nemestrina* or *M. fascicularis* macaques).



305 The covariate of human population density was found to have the highest ranked relative influence for the  
306 majority (496/500, 99.2%) of the bootstrapped models, closely followed by that of healthcare accessibility.  
307 Mean and 95% confidence intervals of relative influence scores across bootstraps are presented in Figure S5.  
308 The new covariates of tree coverage and forest loss were found to be highly influential; out of 21 covariates (20  
309 environmental covariates and the species covariate), the median rank for tree coverage was 5 (95% confidence  
310 interval: 2-11), and for forest loss was 10 (95% confidence interval: 6-15). Plots of the accumulated local  
311 effects (ALE) describing the influence of each continuous covariate across the covariate's range are presented  
312 in Figure S6.

## 313 Discussion

314 In this study, we utilised an environmental niche modelling approach to predict the relative suitability for *P.*  
315 *knowlesi* transmission to humans across Southeast Asia. We extended a previous analysis that incorporated  
316 data up to 2015 [31] by adding infection and environmental data up to 2020, and improving the utilisation  
317 of data on land use patterns. Through a review of literature published between October 2015 and March  
318 2020, we identified 264 published occurrences of *P. knowlesi*. This resulted in a total of 524 records being  
319 utilised in model fitting and evaluation for the current study. As changes in *P. knowlesi* transmission risk  
320 may be expected where substantial amounts of deforestation have occurred [25, 24, 26, 27], we now capture  
321 this in the model by deriving annual forest loss and coverage datasets. We predict that the distribution of *P.*  
322 *knowlesi* risk is highly heterogeneous across Southeast Asia, with the largest areas of predicted risk in Malaysia  
323 and Indonesia, and smaller, localised regions of high risk predicted in the Greater Mekong Subregion, The  
324 Philippines and Northeast India.

325 Our analysis can help to guide the prioritisation of locations for future sampling and surveillance for *P.*  
326 *knowlesi* malaria by highlighting areas of high predicted risk that may have been under-sampled. Since the  
327 publication of the 2015 analysis, there has been no change to the World Health Organization's malaria elim-  
328 ination status of any country believed to be at risk for indigenous *P. knowlesi* transmission [58]. However,  
329 within the Greater Mekong Subregion of Cambodia, Myanmar, Thailand, Laos and Vietnam, substantial de-  
330 clines have been observed in the total number of reported malaria cases as of 2021 [59]. The 2015 analysis  
331 noted that Laos, Myanmar, Thailand and Vietnam were likely high-value sites for future sampling efforts [31]  
332 and our literature search revealed only a small number of additional *P. knowlesi* occurrences in these coun-  
333 tries as of 2020 (Figure 1, Table S1). Our current analysis predicts localised areas of moderate-to-high relative  
334 transmission risk in this region (Figure 3), suggesting an ongoing need for surveillance of *P. knowlesi* malaria.

335 Indonesia has a stated goal to eliminate malaria by 2030 [60, 61], and may be on track given that a majority of  
336 administrative regions have declared elimination [62]. However, the presence of *P. knowlesi* across the country  
337 presents a serious challenge to these efforts. In March 2022, the WHO Malaria Policy Advisory Group (MPAG)  
338 concluded that certification of malaria elimination status should only occur where the risk of *P. knowlesi* was  
339 'negligible', i.e. below some low threshold of annual incidence [63, 58] – a requirement that has already pre-  
340 vented Malaysia from receiving elimination certification [64]. Given this requirement, continued surveillance  
341 and mitigation of *P. knowlesi* throughout at-risk regions of Indonesia will be important. Between the period of  
342 2015 and 2020, a small number of studies have identified substantial numbers of *P. knowlesi* infections within  
343 Indonesia [65, 66, 67], particularly within northern Sumatra [54, 68, 69], a region identified as a valuable target  
344 for surveillance effort in the 2015 model [31]. Despite this, the Indonesian region of Kalimantan on the island  
345 of Borneo still has a relative scarcity of occurrence data given its high predicted transmission suitability and  
346 the number of *P. knowlesi* cases reported in adjacent areas of Malaysia.

347 Our map of *P. knowlesi* transmission risk may also help to quantitatively guide site selection for public health  
348 surveillance or intervention. For example, surveillance sampling could be concentrated in regions where the  
349 model predicts a high transmission suitability but with a high variance, such that the understanding of the ge-  
350 ographical distribution of the disease is maximised for the least effort and that the uncertainty in these regions  
351 could be reduced in future risk mapping outputs. If value was instead placed on maximising the probability  
352 of identifying cases of *P. knowlesi*, sampling could be concentrated where a high transmission suitability is  
353 accompanied by lower variance. Efficient deployment of sampling resources could be achieved by combining  
354 the modelling outputs with constraints; for example, sites where access would require a prohibitive amount  
355 of travel time could be excluded [70].

356 As niche modelling frameworks are correlative, the secondary results described in this work should be in-  
357 terpreted with care. The relative influence scores (Figure S5) and the accumulated local effect plots (Figure  
358 S6) may provide insight into risk factors for *P. knowlesi* transmission. However, these results do not provide  
359 evidence for causal relationships, which would instead be more appropriately identified through studies util-  
360 ising a causal inference framework. For example, the covariate of healthcare accessibility, which ranks highly  
361 according to relative influence scores (Figure S5), could capture a direct causal effect on the risk of being diag-  
362 nosed with *P. knowlesi* (e.g. likelihood that someone is identified as having a *P. knowlesi* infection increasing  
363 with access to healthcare) or may simply be confounded by a common variable (e.g. likelihood of acquiring a  
364 *P. knowlesi* infection increasing for those who work at plantations, confounded by such plantations occurring  
365 in areas of lower healthcare accessibility).

366 Our model predicts the relative suitability for *P. knowlesi* transmission, not the prevalence of infection nor  
367 the incidence of cases (which would require different input data that are not widely available for *P. knowlesi*  
368 malaria). While transmission suitability is a useful metric for prioritising locations for future *P. knowlesi* sur-  
369 veys, the absolute values are specific to the input data and model parameterisation, and we therefore cannot  
370 directly compare absolute values produced by the model presented here and those from the model developed  
371 in 2015. Although we expect that the transmission suitability prediction produced by either of the models  
372 should be qualitatively related to the underlying ‘true’ risk of *P. knowlesi* infection, little can be said of this  
373 relationship other than that it is expected to be monotonic under the assumption that the background data  
374 points are biased in the same manner as the presence data [48]. This means, for instance, that any differences  
375 between the models that could arise as a result of dilation in this relationship (such as the upwards dilation  
376 observed in Figures S4A and S4B) cannot be taken alone as indicating a change in underlying transmission  
377 suitability.

378 Noting the limitations in these comparisons, we find that the predictions in our work and the 2015 model  
379 broadly align, though with clear differences in the local spatial variation of the prediction surface (Figures S3,  
380 S4A). As an example, on the island of Borneo our predictions form a smooth region of high predicted risk,  
381 whereas in the 2015 model predictions over the same area varied substantially at a small spatial scale; this  
382 pattern is repeated similarly elsewhere across Southeast Asia [31]. In countries such as Laos, Myanmar and  
383 Vietnam, we predict overall a lower transmission suitability than those presented in the 2015 model, though  
384 within these countries we continue to predict small areas of high transmission risk. Comparing the overall  
385 distributions of predicted transmission suitability between the 2015 and 2020 models shows that our new  
386 predictions produce a more highly contrasting bimodal distribution of risk compared to that produced by the  
387 2015 model [31] (Figure S4A).

388 The temperature suitability index covariate used in the model attempts to describe the effect of tempera-  
389 ture on the basic reproduction number for some combination of malaria parasites and mosquito vectors. As  
390 data on the incubation periods for *P. knowlesi* under differing temperatures and mosquito hosts is currently  
391 unavailable, no suitability index for the species can currently be produced. In this work, we instead utilise  
392 a proxy in the form of a suitability index for *P. falciparum* [40]. Even if this proxy does not itself accurately  
393 capture mechanistic limits on *P. knowlesi* reproduction, it is not immediately obvious what bias this would  
394 introduce into the results, if any, as the boosted regression tree model may still infer suitability under some  
395 transformation of the index. There is clear value in further laboratory research on the reproduction of *P.*  
396 *knowlesi* under different temperatures that could inform a species-specific suitability index.

397 It is believed that workers involved in the development and cultivation of oil palm plantations are at greater  
398 risk for developing *P. knowlesi* infection given their proximity to *P. knowlesi* vector and reservoir species [71].  
399 However, we were unable to include this as a covariate in our model as there is currently no published dataset  
400 of palm oil plantations with complete coverage across the Southeast Asia region.

401 Annual data is not available for some of the covariates used in the model where the underlying phenomena  
402 may be expected to change over time; the covariates of reservoir/vector species distribution and temperature  
403 suitability are dependent upon variables such as climate or land cover, and the covariate of healthcare acces-  
404 sibility is dependent upon changes in transportation infrastructure and locations of healthcare sites. In lieu  
405 of available data on change in these covariates over time they are instead assumed to be constant. In effect,  
406 this means that the modelled species distributions as of 2014, temperature suitability index for *P. falciparum*  
407 as of 2010 and healthcare accessibility as of 2019 are all assumed constant over the years 2001 to 2019. Future  
408 modelling efforts could be improved by considering the change in these covariates over time.

409 Our map of *P. knowlesi* transmission suitability predict high *P. knowlesi* disease risk across broad areas of  
410 Southeast Asia, with large regions of high predicted *P. knowlesi* risk that have not yet been sampled for the  
411 pathogen. Our work demonstrates the importance of continued surveillance and prospective sampling of the  
412 pathogen, especially in regions where malaria elimination is currently being pursued.

### 413 **Acknowledgements**

414 This study was supported through funding provided by the Australian Centre for International Agricultural  
415 Research (ACIAR), as part of the 'Evaluating zoonotic malaria transmission and agricultural and forestry land  
416 use in Indonesia' (ZOOMAL) project (LS/2019/116, [www.aciar.gov.au](http://www.aciar.gov.au)).

417 Further support for this project was provided by the National Health and Medical Research Council of Aus-  
418 tralia through its Centres of Research Excellence (ACREME, GNT1134989 [www.nhmrc.gov.au](http://www.nhmrc.gov.au)).

419 FMS was supported by the National Health and Medical Research Council of Australia Investigator Grant  
420 Scheme (Emerging Leader Fellowship, 2021/GNT2010051 [www.nhmrc.gov.au](http://www.nhmrc.gov.au)). JAF was supported by the  
421 Australian Research Council (ARC, FT210100034 and DP200100747 [www.arc.gov.au](http://www.arc.gov.au)). LEH was supported by  
422 a Melbourne Research Scholarship from the University of Melbourne ([www.unimelb.edu.au](http://www.unimelb.edu.au)). MJG was sup-  
423 ported by the National Health and Medical Research Council of Australia Investigator Grant Scheme (Emerg-  
424 ing Leader 2 Fellowship, 2023/GNT2017436 [www.nhmrc.gov.au](http://www.nhmrc.gov.au)). GSR and MJG were supported by the Na-  
425 tional Institutes of Health, USA (R01AI160457-01 [www.nih.gov](http://www.nih.gov)). GSR was also supported by the Malaysian  
426 Ministry of Health (Grant Number BP00500/117/1002 [www.moh.gov.my](http://www.moh.gov.my)).

427 This research was supported by The University of Melbourne's Research Computing Services and the Petas-  
428 cale Campus Initiative ([www.unimelb.edu.au](http://www.unimelb.edu.au)).

429 The funders had no role in study design, data collection and analysis, decision to publish, or preparation of  
430 the manuscript.

431 We thank Dr Timothy William for their support. We would like to also thank the Director General of Health  
432 Malaysia for the permission to publish this article.

### 433 **Data Availability**

434 Prediction results for each bootstrapped model, rasters of summary statistics, the code used to produce re-  
435 sults, and the updated occurrence database have been made available at [osf.io/k5bsa](https://osf.io/k5bsa) (DOI 10.17605/OSF.IO/K5BSA).

## 436 References

- 437 [1] I. Vythilingam, M. L. Wong, and W. S. Wan-Yussof. Current status of *Plasmodium knowlesi* vectors: a  
438 public health concern? *Parasitology*, 145(1):32–40, May 2016.
- 439 [2] William E. Collins and John W. Barnwell. *Plasmodium knowlesi*: Finally being recognized. *The Journal of*  
440 *Infectious Diseases*, 199(8):1107–1108, April 2009.
- 441 [3] William Chin, Edward Alpert, William E. Collins, Marvin H. Jeter, and Peter G. Contacos. Experimental  
442 Mosquito-Transmission of *Plasmodium Knowlesi* to Man and Monkey. *The American Journal of Tropical*  
443 *Medicine and Hygiene*, 17(3):355–358, May 1968.
- 444 [4] Pablo Ruiz Cuenca, Stephanie Key, Kim A. Lindblade, Indra Vythilingam, Chris Drakeley, and Kimberly  
445 Fornace. Is there evidence of sustained human-mosquito-human transmission of the zoonotic malaria  
446 *Plasmodium knowlesi*? a systematic literature review. *Malaria Journal*, 21(1), March 2022.
- 447 [5] Balbir Singh and Cyrus Daneshvar. Human Infections and Detection of *Plasmodium knowlesi*. *Clinical*  
448 *Microbiology Reviews*, 26(2):165–184, April 2013.
- 449 [6] George Robert Coatney. *The primate malarias*. US National Institute of Allergy and Infectious Diseases,  
450 1971.
- 451 [7] Paddy M. Brock, Kimberley M. Fornace, Minnie Parmiter, Jonathon Cox, Chris J. Drakeley, Heather M.  
452 Ferguson, and Rowland Raymond Kao. *Plasmodium knowlesi* transmission: integrating quantitative ap-  
453 proaches from epidemiology and ecology to understand malaria as a zoonosis. *Parasitology*, 143(4):389–  
454 400, January 2016.
- 455 [8] Bridget E. Barber, Matthew J. Grigg, Daniel J. Cooper, Donnelly A. van Schalkwyk, Timothy William, Giri S.  
456 Rajahram, and Nicholas M. Anstey. Clinical management of *Plasmodium knowlesi* malaria. In *Current*  
457 *research on naturally transmitted Plasmodium knowlesi*, pages 45–76. Elsevier, 2021.
- 458 [9] Mallika Imwong, Wanassanan Madmanee, Kanokon Suwannasin, Chanon Kunasol, Thomas J Peto, Ru-  
459 pam Tripura, Lorenz von Seidlein, Chea Nguon, Chan Davoeng, Nicholas P J Day, Arjen M Dondorp, and  
460 Nicholas J White. Asymptomatic natural human infections with the simian malaria parasites *Plasmod-*  
461 *ium cynomolgi* and *Plasmodium knowlesi*. *The Journal of Infectious Diseases*, 219(5):695–702, October  
462 2018.
- 463 [10] Kimberly M. Fornace, Nor Afizah Nuin, Martha Betson, Matthew J. Grigg, Timothy William, Nicholas M.  
464 Anstey, Tsin W. Yeo, Jonathan Cox, Lau Tiek Ying, and Chris J. Drakeley. Asymptomatic and submicro-  
465 scopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases  
466 in Sabah, Malaysia. *Journal of Infectious Diseases*, 213(5):784–787, October 2015.
- 467 [11] Matthew J Grigg, Jonathan Cox, Timothy William, Jenarun Jelip, Kimberly M Fornace, Patrick M Brock,  
468 Lorenz von Seidlein, Bridget E Barber, Nicholas M Anstey, Tsin W Yeo, and Christopher J Drakeley.  
469 Individual-level factors associated with the risk of acquiring human *Plasmodium knowlesi* malaria in  
470 Malaysia: a case-control study. *The Lancet Planetary Health*, 1(3):e97–e104, June 2017.
- 471 [12] Cyrus Daneshvar, Timothy M. E. Davis, Janet Cox-Singh, Mohammad Zakri Rafa’ee, Siti Khatijah Zakaria,  
472 Paul C. S. Divis, and Balbir Singh. Clinical and laboratory features of human *Plasmodium knowlesi* in-  
473 fection. *Clinical Infectious Diseases*, 49(6):852–860, September 2009.
- 474 [13] Matthew J Grigg, Timothy William, Bridget E Barber, Giri S Rajahram, Jayaram Menon, Emma Schimann,  
475 Kim Piera, Christopher S Wilkes, Kaajal Patel, Arjun Chandna, Christopher J Drakeley, Tsin W Yeo, and  
476 Nicholas M Anstey. Age-related clinical spectrum of *Plasmodium knowlesi* malaria and predictors of  
477 severity. *Clinical Infectious Diseases*, 67(3):350–359, March 2018.
- 478 [14] Kimberly M. Fornace, Lou S. Herman, Tommy R. Abidin, Tock Hing Chua, Sylvia Daim, Pauline J.  
479 Lorenzo, Lynn Grignard, Nor Afizah Nuin, Lau Tiek Ying, Matthew J. Grigg, Timothy William, Fe Espino,  
480 Jonathan Cox, Kevin K. A. Tetteh, and Chris J. Drakeley. Exposure and infection to *Plasmodium knowlesi*  
481 in case study communities in Northern Sabah, Malaysia and Palawan, The Philippines. *PLOS Neglected*  
482 *Tropical Diseases*, 12(6):e0006432, June 2018.

- 483 [15] Kimberly M Fornace, Paddy M Brock, Tommy R Abidin, Lynn Grignard, Lou S Herman, Tock H Chua,  
484 Sylvia Daim, Timothy William, Catriona L E B Patterson, Tom Hall, Matthew J Grigg, Nicholas M Anstey,  
485 Kevin K A Tetteh, Jonathan Cox, and Chris J Drakeley. Environmental risk factors and exposure to the  
486 zoonotic malaria parasite *Plasmodium knowlesi* across northern sabah, malaysia: a population-based  
487 cross-sectional survey. *The Lancet Planetary Health*, 3(4):e179–e186, April 2019.
- 488 [16] M Vadivelan, TK Dutta, et al. Recent advances in the management of *Plasmodium knowlesi* infection.  
489 *Trop Parasitol*, 4(1):31–4, 2014.
- 490 [17] Aongart Mahittikorn, Frederick Ramirez Masangkay, Kwuntida Uthaisar Kotepui, Giovanni De Jesus Mi-  
491 lanez, and Manas Kotepui. Quantification of the misidentification of *Plasmodium knowlesi* as *Plasmod-*  
492 *ium malariae* by microscopy: an analysis of 1569 P. knowlesi cases. *Malaria Journal*, 20(1), April 2021.
- 493 [18] Kim-Sung Lee, Janet Cox-Singh, and Balbir Singh. Morphological features and differential counts of  
494 *Plasmodium knowlesi* parasites in naturally acquired human infections. *Malaria Journal*, 8(1), April  
495 2009.
- 496 [19] Daniel J Cooper, Giri S Rajahram, Timothy William, Jenarun Jelip, Rashidah Mohammad, Joseph Bene-  
497 dict, Danshy A Alaza, Eva Malacova, Tsin W Yeo, Matthew J Grigg, Nicholas M Anstey, and Bridget E  
498 Barber. *Plasmodium knowlesi* malaria in Sabah, Malaysia, 2015–2017: Ongoing increase in incidence  
499 despite near-elimination of the human-only *Plasmodium* species. *Clinical Infectious Diseases*, 70(3):361–  
500 367, March 2019.
- 501 [20] Rupesh K Tyagi, Manoj K Das, Shiv S Singh, and Yagya D Sharma. Discordance in drug resistance-  
502 associated mutation patterns in marker genes of *Plasmodium falciparum* and *Plasmodium knowlesi*  
503 during coinfections. *J. Antimicrob. Chemother.*, 68(5):1081–1088, May 2013.
- 504 [21] World Health Organization. *Global technical strategy for malaria 2016-2030*. World Health Organization,  
505 2015.
- 506 [22] Fauzi Muh, Namhyeok Kim, Myat Htut Nyunt, Egy Rahman Firdaus, Jin-Hee Han, Mohammad Rafiul  
507 Hoque, Seong-Kyun Lee, Ji-Hoon Park, Robert W. Moon, Yee Ling Lau, Osamu Kaneko, and Eun-Taek  
508 Han. Cross-species reactivity of antibodies against *Plasmodium vivax* blood-stage antigens to *Plasmod-*  
509 *ium knowlesi*. *PLOS Neglected Tropical Diseases*, 14(6):e0008323, June 2020.
- 510 [23] Abraham Zefong Chin, Marilyn Charlene Montini Maluda, Jenarun Jelip, Muhammad Saffree Bin Jef-  
511 free, Richard Culleton, and Kamruddin Ahmed. Malaria elimination in Malaysia and the rising threat of  
512 *Plasmodium knowlesi*. *Journal of Physiological Anthropology*, 39(1), November 2020.
- 513 [24] Pablo Ruiz Cuenca, Stephanie Key, Amaziasizamoria Jumail, Henry Surendra, Heather M. Ferguson,  
514 Chris J. Drakeley, and Kimberly Fornace. Epidemiology of the zoonotic malaria *Plasmodium knowlesi*  
515 in changing landscapes. In *Current research on naturally transmitted Plasmodium knowlesi*, pages 225–  
516 286. Elsevier, 2021.
- 517 [25] Patrick M. Brock, Kimberly M. Fornace, Matthew J. Grigg, Nicholas M. Anstey, Timothy William, Jon  
518 Cox, Chris J. Drakeley, Heather M. Ferguson, and Rowland R. Kao. Predictive analysis across spatial  
519 scales links zoonotic malaria to deforestation. *Proceedings of the Royal Society B: Biological Sciences*,  
520 286(1894):20182351, January 2019.
- 521 [26] Gael Davidson, Tock H. Chua, Angus Cook, Peter Speldewinde, and Philip Weinstein. The role of ecolog-  
522 ical linkage mechanisms in *Plasmodium knowlesi* transmission and spread. *EcoHealth*, 16(4):594–610,  
523 January 2019.
- 524 [27] Danica J. Stark, Kimberly M. Fornace, Patrick M. Brock, Tommy Rowel Abidin, Lauren Gilhooly, Cyrlen  
525 Jalius, Benoit Goossens, Chris J. Drakeley, and Milena Salgado-Lynn. Long-tailed macaque response to  
526 deforestation in a *Plasmodium knowlesi*-endemic area. *EcoHealth*, 16(4):638–646, March 2019.
- 527 [28] William K. Reisen. Landscape epidemiology of vector-borne diseases. *Annual Review of Entomology*,  
528 55(1):461–483, January 2010.
- 529 [29] Huan Yu, Xiangmeng Liu, Bo Kong, Ruopu Li, and Guangxing Wang. Landscape ecology development  
530 supported by geospatial technologies: A review. *Ecological Informatics*, 51:185–192, 2019.



- 531 [30] Robert P. Anderson, Enrique Martínez-Meyer, Miguel Nakamura, Miguel B. Araújo, A. Townsend Peter-  
532 son, Jorge Soberón, and Richard G. Pearson. *Ecological Niches and Geographic Distributions (MPB-49)*.  
533 Princeton University Press, December 2011.
- 534 [31] Freya M Shearer, Zhi Huang, Daniel J Weiss, Antoinette Wiebe, Harry S Gibson, Katherine E Battle,  
535 David M Pigott, Oliver J Brady, Chaturong Putaporntip, Somchai Jongwutiwes, et al. Estimating ge-  
536 ographical variation in the risk of zoonotic *Plasmodium knowlesi* infection in countries eliminating  
537 malaria. *PLOS Neglected Tropical Diseases*, 10(8):e0004915, 2016.
- 538 [32] D. J. Weiss, A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, E. Poyart, S. Belchior,  
539 N. Fullman, B. Mappin, U. Dalrymple, J. Rozier, T. C. D. Lucas, R. E. Howes, L. S. Tusting, S. Y. Kang,  
540 E. Cameron, D. Bisanzio, K. E. Battle, S. Bhatt, and P. W. Gething. A global map of travel time to cities to  
541 assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336, January 2018.
- 542 [33] D. J. Weiss, A. Nelson, C. A. Vargas-Ruiz, K. Gligorić, S. Bavadekar, E. Gabrilovich, A. Bertozzi-Villa,  
543 J. Rozier, H. S. Gibson, T. Shekel, C. Kamath, A. Lieber, K. Schulman, Y. Shao, V. Qarkaxhija, A. K. Nandi,  
544 S. H. Keddie, S. Rumisha, P. Amratia, R. Arambepola, E. G. Chestnutt, J. J. Millar, T. L. Symons, E. Cameron,  
545 K. E. Battle, S. Bhatt, and P. W. Gething. Global maps of travel time to healthcare facilities. *Nature*  
546 *Medicine*, 26(12):1835–1838, September 2020.
- 547 [34] Peter Potapov, Aleksey Yaroshenko, Svetlana Turubanova, Maxim Dubinin, Lars Laestadius, Christoph  
548 Thies, Dmitry Aksenov, Aleksey Egorov, Yelena Yesipova, Igor Glushkov, et al. Mapping the World’s intact  
549 forest landscapes by remote sensing. *Ecology and Society*, 13(2), 2008.
- 550 [35] Catherine L Moyes, Freya M Shearer, Zhi Huang, Antoinette Wiebe, Harry S Gibson, Vincent Nijman,  
551 Jayasilan Mohd-Azlan, Jediah F Brodie, Suchinda Malaivijitnond, Matthew Linkie, et al. Predicting the  
552 geographical distributions of the macaque hosts and mosquito vectors of *Plasmodium knowlesi* malaria  
553 in forested and non-forested areas. *Parasites & vectors*, 9(1):1–12, 2016.
- 554 [36] Matthew C Hansen, Peter V Potapov, Rebecca Moore, Matt Hancher, Svetlana A Turubanova, Alexandra  
555 Tyukavina, David Thau, SV Stehman, Scott J Goetz, Thomas R Loveland, et al. High-resolution global  
556 maps of 21st-century forest cover change. *science*, 342(6160):850–853, 2013.
- 557 [37] Tom G Farr, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick,  
558 Mimi Paller, Ernesto Rodriguez, Ladislav Roth, et al. The shuttle radar topography mission. *Reviews of*  
559 *geophysics*, 45(2), 2007.
- 560 [38] S. E. Lobser and Warren B. Cohen. MODIS tasselled cap: land cover characteristics expressed through  
561 transformed MODIS data. *International Journal of Remote Sensing*, 28:5079 – 5101, 2007.
- 562 [39] Daniel J Weiss, Peter M Atkinson, Samir Bhatt, Bonnie Mappin, Simon I Hay, and Peter W Gething. An  
563 effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Pho-*  
564 *togrammetry and Remote Sensing*, 98:106–118, 2014.
- 565 [40] Peter W Gething, Thomas P Van Boeckel, David L Smith, Carlos A Guerra, Anand P Patil, Robert W Snow,  
566 and Simon I Hay. Modelling the global constraints of temperature on transmission of *Plasmodium faldi-*  
567 *parum* and *P. vivax*. *Parasites & vectors*, 4(1):1–11, 2011.
- 568 [41] Andrea E. Gaughan, Forrest R. Stevens, Catherine Linard, Peng Jia, and Andrew J. Tatem. High resolution  
569 population distribution maps for Southeast Asia in 2010 and 2015. *PLOS ONE*, 8(2):e55882, February  
570 2013.
- 571 [42] Andrea E Gaughan, Forrest R Stevens, Catherine Linard, Peng Jia, and Andrew J Tatem. High resolution  
572 population distribution maps for Southeast Asia in 2010 and 2015. *PLOS one*, 8(2):e55882, 2013.
- 573 [43] Mark A Friedl, Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley,  
574 and Xiaoman Huang. MODIS collection 5 global land cover: Algorithm refinements and characterization  
575 of new datasets. *Remote sensing of Environment*, 114(1):168–182, 2010.
- 576 [44] Glenn De'ath and Katharina E. Fabricius. Classification and regression trees: A powerful yet simple tech-  
577 nique for ecological data analysis. *Ecology*, 81(11):3178–3192, November 2000.

- 578 [45] J. Elith, J. R. Leathwick, and T. Hastie. A working guide to boosted regression trees. *Journal of Animal*  
579 *Ecology*, 77(4):802–813, July 2008.
- 580 [46] Jane Elith, Catherine H. Graham, Robert P. Anderson, Miroslav Dudík, Simon Ferrier, Antoine Guisan,  
581 Robert J. Hijmans, Falk Huettmann, John R. Leathwick, Anthony Lehmann, Jin Li, Lucia G. Lohmann,  
582 Bette A. Loiselle, Glenn Manion, Craig Moritz, Miguel Nakamura, Yoshinori Nakazawa, Jacob McC. M.  
583 Overton, A. Townsend Peterson, Steven J. Phillips, Karen Richardson, Ricardo Scachetti-Pereira, Robert E.  
584 Schapire, Jorge Soberón, Stephen Williams, Mary S. Wisz, and Niklaus E. Zimmermann. Novel methods  
585 improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–151, March  
586 2006.
- 587 [47] JR Leathwick, J Elith, MP Francis, T Hastie, and P Taylor. Variation in demersal fish species richness in the  
588 oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress*  
589 *Series*, 321:267–281, September 2006.
- 590 [48] Steven J. Phillips, Miroslav Dudík, Jane Elith, Catherine H. Graham, Anthony Lehmann, John Leathwick,  
591 and Simon Ferrier. Sample selection bias and presence-only distribution models: implications for back-  
592 ground and pseudo-absence data. *Ecological Applications*, 19(1):181–197, January 2009.
- 593 [49] Colin Groves. *Primate Taxonomy*. Smithsonian Series in Comparative Evolutionary Biology. Smithsonian  
594 Books, Washington, D.C., DC, May 2001.
- 595 [50] Jane Elith, Michael Kearney, and Steven Phillips. The art of modelling range-shifting species. *Methods in*  
596 *Ecology and Evolution*, 1(4):330–342, 2010.
- 597 [51] Robert J. Hijmans. Cross-validation of species distribution models: removing spatial sorting bias and  
598 calibration with a null model. *Ecology*, 93(3):679–688, March 2012.
- 599 [52] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*,  
600 pages 1189–1232, 2001.
- 601 [53] Christoph Molnar. *Interpretable Machine Learning: A Guide For Making Black Box Models Explainable*.  
602 Independently published, 2020.
- 603 [54] Herdiana Herdiana, Chris Cotter, Farah N Coutrier, Iska Zarlinda, Brittany W Zelman, Yusrifar Kharisma  
604 Tirta, Bryan Greenhouse, Roly D Gosling, Peter Baker, Maxine Whittaker, et al. Malaria risk factor  
605 assessment using active and passive surveillance data from Aceh Besar, Indonesia, a low endemic,  
606 malaria elimination setting with *Plasmodium knowlesi*, *Plasmodium vivax*, and *Plasmodium falciparum*. *Malaria journal*, 15(1):1–15, 2016.
- 608 [55] Inke N. D. Lubis, Hendri Wijaya, Munar Lubis, Chairuddin P. Lubis, Paul C. S. Divis, Khalid B. Beshir, and  
609 Colin J. Sutherland. Contribution of *Plasmodium knowlesi* to multispecies human malaria infections in  
610 North Sumatera, Indonesia. *The Journal of Infectious Diseases*, 215(7):1148–1155, February 2017.
- 611 [56] Moritoshi Iwagami, Masami Nakatsu, Phonepadith Khattignavong, Pheovaly Soundala, Lavy Lor-  
612 phachan, Sengdeuane Keomalaphet, Phonepadith Xangsayalath, Satoru Kawai, Bouasy Hongvanthong,  
613 Paul T Brey, et al. First case of human infection with *Plasmodium knowlesi* in Laos, 2018.
- 614 [57] A Townsend Peterson. *Mapping disease transmission risk: enriching models using biogeography and*  
615 *ecology*. JHU Press, 2014.
- 616 [58] World Health Organization. World malaria report 2022. 2022.
- 617 [59] World Health Organization. Accelerating malaria elimination in the greater mekong. 2022.
- 618 [60] Departemen Kesehatan. Keputusan Menteri Kesehatan Republik Indonesia Nomor  
619 293/MENKES/SK/IV/2009 28 April 2009 tentang Eliminasi Malaria di Indonesia, 2009.
- 620 [61] Ministry of Health Indonesia – Kementerian Kesehatan Republik Indonesia. Challenges Toward Malaria  
621 Elimination 2030. 2021.
- 622 [62] Vensya Sitohang, Elvieda Sariwati, Sri Budi Fajariyani, Dasom Hwang, Bayu Kurnia, Ratih Ketana Hap-  
623 sari, Ferdinand Johannis Laihad, Maria Endang Sumiwi, Paul Pronyk, and William A Hawley. Malaria  
624 elimination in Indonesia: halfway there. *The Lancet Global Health*, 6(6):e604–e606, June 2018.

- 625 [63] World Health Organisation et al. WHO Malaria Policy Advisory Group (MPAG) meeting: meeting report,  
626 March. 2022.
- 627 [64] Kimberly M Fornace, Gabriel Zorello Laporta, Indra Vythilingham, Tock Hing Chua, Kamruddin Ahmed,  
628 Nantha K Jeyaprakasam, Ana Maria Ribeiro de Castro Duarte, Amirah Amir, Wei Kit Phang, Chris Drake-  
629 ley, Maria Anice M Sallum, and Yee Ling Lau. Simian malaria: a narrative review on emergence, epidemi-  
630 ology and threat to global malaria elimination. *Lancet Infect. Dis.*, July 2023.
- 631 [65] I.N.D. Lubis, H. Wijaya, M. Lubis, C.P. Lubis, and C.J. Sutherland. Molecular identification of human *Plas-*  
632 *modium knowlesi* infections in North Sumatera, Indonesia. *International Journal of Infectious Diseases*,  
633 45:182, April 2016.
- 634 [66] Agus A. Wibowo, Sitti R. Umniyati, Jontari Hutagalung, and Tanti Rahayu. Confirmation of *Anopheles*  
635 *balabacensis* as natural vector of malaria caused by *Plasmodium knowlesi* inhabits forested areas in *Ke-*  
636 *camatan Balik Bukit*, pwestern lampung regency. 151:01028, 2020.
- 637 [67] Wuryantari Setiadi, Herawati Sudoyo, Hidayat Trimarsanto, Boy Adventus Sihite, Riahdo Juliarmen  
638 Saragih, Rita Juliawaty, Suradi Wangsamuda, Puji Budi Setia Asih, and Din Syafruddin. A zoonotic hu-  
639 man infection with simian malaria, *Plasmodium knowlesi*, in Central Kalimantan, Indonesia. *Malaria*  
640 *Journal*, 15(1), April 2016.
- 641 [68] Herdiana Herdiana, Irnawati Irnawati, Farah Novita Coutrier, Alfian Munthe, Mardiaty Mardiaty, Titik Yu-  
642 niarti, Elvieda Sariwati, Maria Endang Sumiwi, Rintis Noviyanti, Paul Pronyk, et al. Two clusters of *Plas-*  
643 *modium knowlesi* cases in a malaria elimination area, Sabang Municipality, Aceh, Indonesia. *Malaria*  
644 *journal*, 17(1):1–10, 2018.
- 645 [69] Farah Coutrier, Chris Cotter, Yusrifar K Tirta, Alanna Schwartz, Iska Zarlinda, Herdiana H Basri, Jutta  
646 Marfurt, Nicholas Anstey, William A Hawley, Asik Surya, et al. Serial molecular identification to con-  
647 firm the presence of *Plasmodium knowlesi* in Indonesia. In *American Journal of Tropical Medicine and*  
648 *Hygiene*, volume 93, pages 263–263, 2015.
- 649 [70] Joshua Longbottom, Ana Krause, Stephen J. Torr, and Michelle C. Stanton. Quantifying geographic  
650 accessibility to improve efficiency of entomological monitoring. *PLOS Neglected Tropical Diseases*,  
651 14(3):e0008096, March 2020.
- 652 [71] Nurul Athirah Naserrudin, Rozita Hod, Mohammad Saffree Jeffree, Kamruddin Ahmed, Richard Culleton,  
653 and Mohd Rohaizat Hassan. The role of human behavior in *Plasmodium knowlesi* malaria infection: A  
654 systematic review. *International Journal of Environmental Research and Public Health*, 19(6):3675, March  
655 2022.

656 **Supplementary Information**

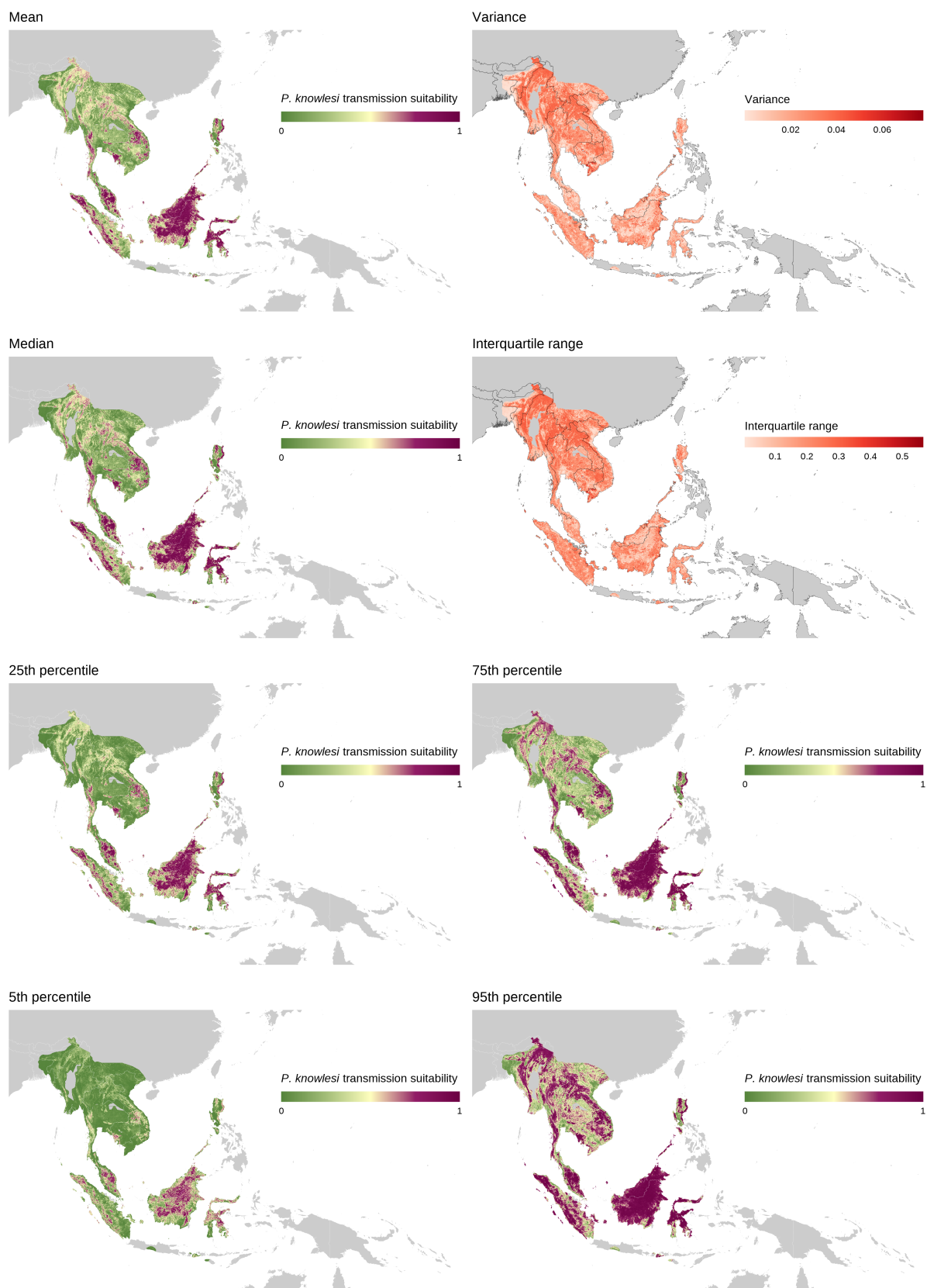


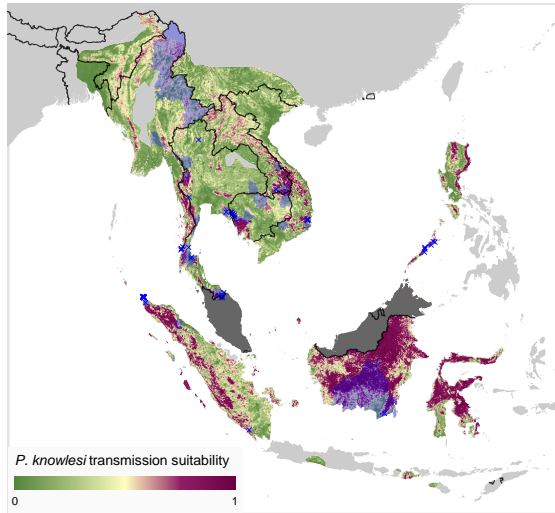
Figure S1: Summary statistics for modelled transmission suitability across Southeast Asia, calculated across the set of 500 bootstraps. Results are only displayed where an area is in the range of both a vector and reservoir species necessary for transmission (see Methods).

Country/Region	Human	Macaque	Mosquito	Totals	
	n	n	n	2020	(2015)
Brunei	0	0	0	0	(6)
Cambodia	5	0	0	5	(6)
Indonesia (Total)	44	1	0	45	(5)
Kalimantan Selatan	0	0	0	0	(3)
Kalimantan Tengah	1	0	0	1	(2)
Lampung	0	1	0	1	(0)
Nanggroe Aceh Darusalam	36	0	0	36	(0)
Sumatera Utara	7	0	0	7	(0)
Laos	3	0	0	3	(1)
Malaysia (Total)	181	8	12	201	(183)
Johor	3	0	0	3	(3)
Kedah	2	0	0	2	(1)
Kelantan	5	0	0	5	(17)
Melaka	2	0	0	2	(1)
Negeri	2	0	2	4	(3)
Pahang	3	0	1	4	(11)
Perak	3	0	1	4	(1)
Pulau Pinang	1	0	0	1	(1)
Sabah	127	7	0	134	(60)
Sarawak	22	1	5	28	(52)
Selangor	7	0	3	10	(30)
Terengganu	1	0	0	1	(2)
W.P. Kuala Lumpur	3	0	0	3	(0)
W.P. Labuan	0	0	0	0	(1)
Myanmar	2	0	0	2	(3)
Philippines	1	0	2	3	(7)
Singapore	0	0	0	0	(6)
Thailand	4	0	0	4	(32)
Vietnam	1	0	0	1	(11)

Table S1: The number of human, macaque and mosquito samples in the occurrence database produced by the 2015 literature review, with samples in Indonesia and Malaysia shown stratified by region (province, state or territory). Total counts are shown for records from both the 2020 literature review and 2015 literature review.



A – Evaluation region occurrence data



A – Evaluation region absence data

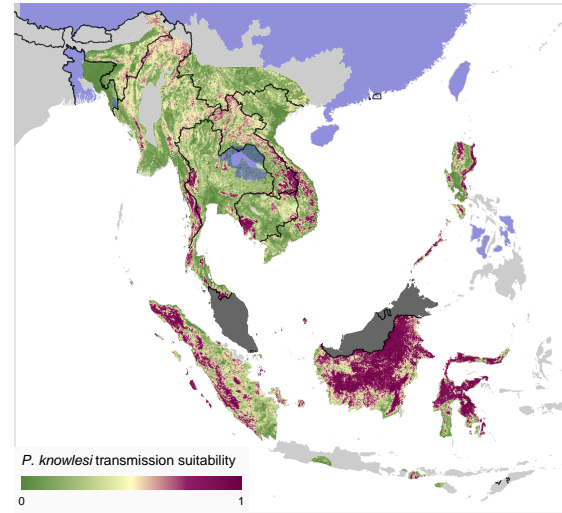
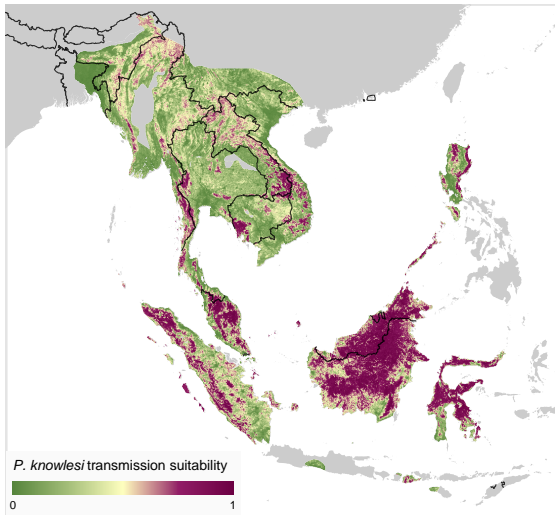


Figure S2: Predicted mean transmission suitability with overlay of infection occurrence data across the evaluation region. Results are only displayed where an area is in the range of both a vector and reservoir species necessary for transmission (see Methods). **A:** Predicted transmission suitability with infection occurrence polygons and points in blue. **B:** Predicted transmission suitability with infection absence polygons in blue.

A



B

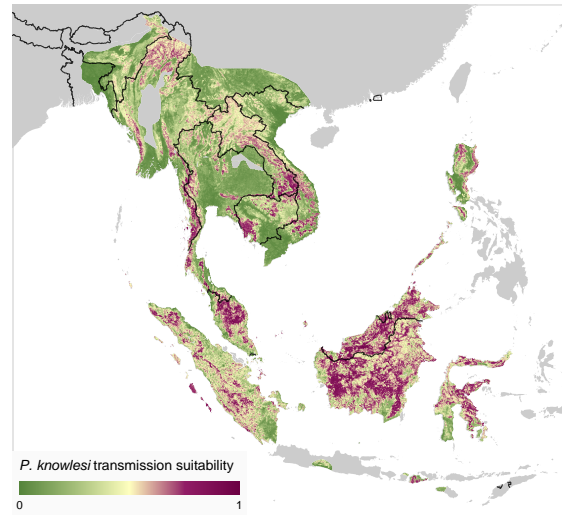


Figure S3: Comparison of the modelled mean transmission suitability value between the current work, with data up to 2020 (A), and the values presented in the 2015 model [31] (B). Note that the absolute value of predictions are not necessarily comparable given differences in model specification and training data. Results are only displayed where an area is in the range of both a vector and reservoir species necessary for transmission (see Methods).

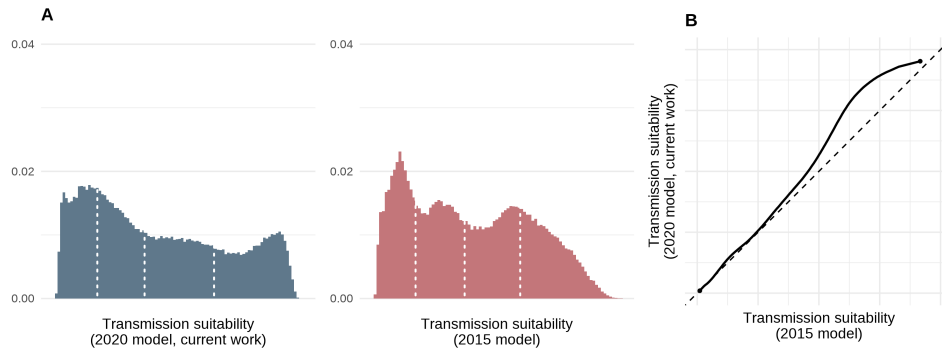


Figure S4: Histograms (A) and quantile-quantile plot (B) comparing the distributions of mean predicted transmission suitability for the 2015 and 2020 models of *P. knowlesi* transmission risk. Histograms are presented on a relative x-axis (ranging from minimal to maximal predicted mean risk), with quartiles of predicted risk displayed as dashed vertical lines.

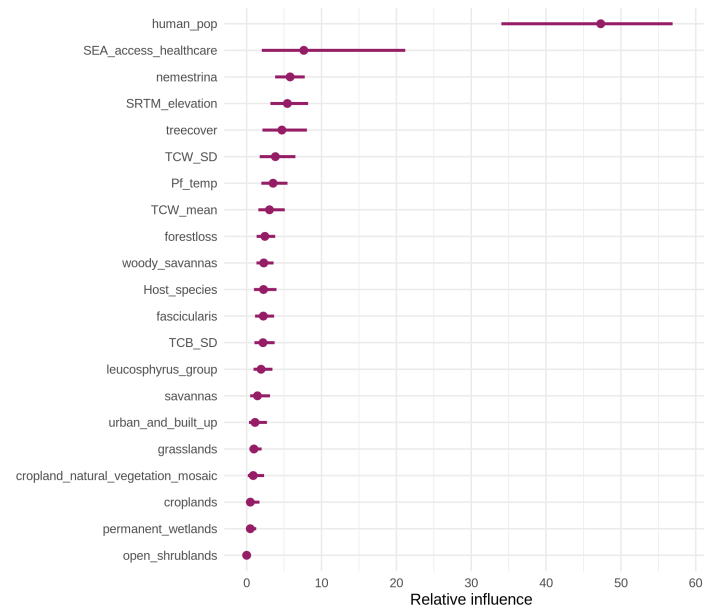


Figure S5: Relative influence scores for each covariate, calculated for each bootstrap, with points and lines representing median values and 95% confidence intervals respectively.

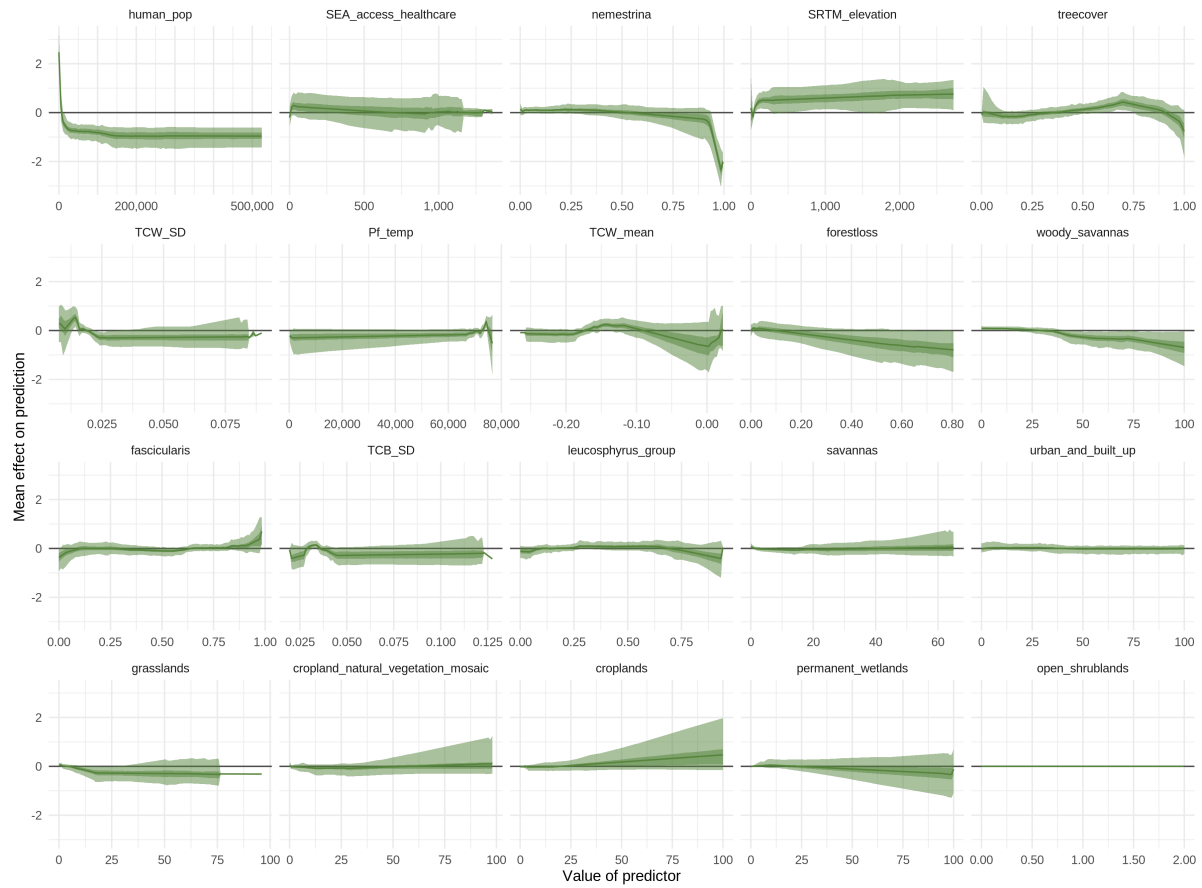


Figure S6: Accumulated local effect (ALE) plots for each covariate, indicating the mean effect of changing a covariate's value upon the prediction (on logistic scale) across the range of that covariate. The ALE values are calculated for each bootstrap, with the median value, 50% and 95% confidence intervals presented as lines, darker shaded regions and lighter shaded regions respectively.