

1 **Title:** Analysis of the malaria profiles in high-risk incidence municipalities in the Brazilian Amazon
2 using Principal Component Analysis in the period 2011-2013 and 2017-2019
3

4 **Authors**

5 Natália Martins Arruda* - National School of Public Health/Oswaldo Cruz Foundation (FIOCRUZ)

6 Vinícius de Souza Maia - Lund University

7 Bianca Cechetto Carlos - Population Studies Center 'Elza Berquó' (Nepo), University of Campinas
8 (Unicamp)

9 Carlos Eduardo Beluzo - University of Campinas (Unicamp) and Federal Institute of São Paulo

10 Luciana Correia Alves – Department of Demography, Institute of Philosophy and Human Sciences,
11 Population Studies Center 'Elza Berquó' (Nepo), University of Campinas (Unicamp)

12

13 * E-mail: natimarruda@gmail.com

14

15 **Author Summary**

16 Luciana C. Alves is an Associate Professor at the Institute of Philosophy and Human Sciences (IFCH)

17 at the University of Campinas (Unicamp) and research scientist at the Population Studies Center 'Elza

18 Berquó' (Nepo) and Chair of the Department of Demography at the University of Campinas. Natália

19 M. Arruda is a doctoral student in Epidemiology at the National School of Public Health/Fiocruz.

20 Vinícius S. Maia is a doctoral student at the Centre for Economic Demography at Lund University.

21 Bianca C. Carlos is an assistant researcher at the Population Studies Center 'Elza Berquó'

22 (Nepo)/Unicamp. Carlos E. Beluzo is a Professor at the Federal Institute of São Paulo and a doctoral

23 student at the Institute of Philosophy and Human Sciences (IFCH) at the University of Campinas

24 (Unicamp).

25

26

27 **Abstract** NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

28

29 **Background:** Malaria still is one of the most relevant infectious diseases in Brazil with 184,869 cases
30 in 2019 and 62.8% of these cases occurred in only 6.2% of municipalities that had high transmission.
31 The incidence of malaria is influenced by environmental, socioeconomic, demographic, and structural
32 factors (organization and effectiveness of health services, land use, and infrastructure).

33 **Methods:** We use data from the Malaria Epidemiological Surveillance System aggregated by period
34 and municipality, maintained by the Ministry of Health, from 2007 to 2019, and combine it with data
35 from the National Institute for Space Research's – *Satellite Monitoring of Brazilian Amazon Forest*
36 *Deforestation Project*, 2010 Demographic Census Microdata and Registry of Health Establishments
37 to analyze socioeconomic, demographic, environmental and health factors in two distinct periods.
38 We use the Principal Component Analysis algorithm to create different principal components
39 characteristics patterns in high-risk municipalities.

40 **Results:** The Principal Component Analysis allowed the creation of three profiles of high-risk
41 municipalities combining the contribution of different demographic, socioeconomic, environmental,
42 and health characteristics with the contribution of the municipalities in each principal component.
43 The first group with a vulnerability profile, that is, high-risk municipalities with cases more associated
44 with vulnerability characteristics, the second group is one with a profile of occupation and working
45 age, of high-risk municipalities with cases more related to socioeconomic and demographic variables
46 linked with specific occupations and a working-age population and the third group with municipalities
47 that has a mixed profile associated with different characteristics related to of *P. falciparum* in contrast
48 of *P. vivax* cases, such as the municipality of Rio Preto da Eva, which for the cases of *P. vivax*, may
49 be more related to children under 5 years and differently, *P. falciparum* in this municipality could be
50 more related to gold panning, deforestation and the presence of large bodies of water. **Conclusions:**
51 There are particularities in socioeconomic, demographic, environmental, and health characteristics
52 for infection by both types of *Plasmodium* and high-risk municipalities that become necessary to
53 understand the differences in the profile of the population affected by malaria together with the

54 environment in which they live, the weather, forest, hydrography and health conditions to assess the
55 structure in these regions, in addition to analyzing the cases of *P. falciparum* separately from the
56 cases of *P. vivax*, showing us that understand the socio-environmental determinants at the local level
57 is essential for the success of malaria prevention and control strategies.

58 **Keywords:** Malaria Epidemiology. Malaria Control. Risk Factors. Socioeconomic Factors.
59 Geographic Analysis.

60

61 **Introduction**

62 Around the world, there were 229 million malaria cases in 2019 in 87 countries. The malaria
63 incidence reduced from 80 cases per 1,000 population at risk in 2000 to 57 cases in 2019. In the
64 Americas, malaria cases have reduced by 40% (from 1.5 million to 0.9 million), in recent years. This
65 progress has been hampered by the major increase of cases in Venezuela, from 35,500 cases in 2000
66 to 467,000 in 2019. Brazil, Colombia, and Venezuela express over 86% of all cases in the region [1].

67 Worldwide, *Plasmodium falciparum* malaria is more prevalent than *Plasmodium vivax*. In
68 Brazil, the opposite is true and *P. vivax* cases correspond to 84% of total reported malaria [2]. Despite
69 *P. falciparum* malaria causes higher levels of morbidity and mortality than *P. vivax*, the last one is
70 also very important due to the peculiarities that make it harder to control, as the high number of
71 asymptomatic individuals, undetectable parasitemia, parasite latent stages (hypnozoites), early
72 production of gametocytes and suspicions of resistance to antimalarial drugs [3,4].

73 Between the period 2011 to 2019, Brazil went from 310,450 positive cases of malaria
74 registered per year to 184,869, a reduction of approximately 60%. As countries accelerate towards
75 malaria elimination, reduction in malaria transmission is highly variable leading to significant
76 heterogeneities in residual transmission across their territories [5,6,7,8].

77 The Legal Amazon (LA) is an endemic malaria region, consisting of 772 municipalities in the
78 states of Acre (AC), Amapá (AP), Amazonas (AM), Pará (PA), Rondônia (RO) and Roraima (RR),

79 Maranhão (MA), Mato Grosso (MT) and Tocantins (TO). Incidence of malaria varies among
80 municipalities and is measured by the Annual Parasite Incidence (API). Municipalities are classified
81 as having a low risk of incidence when the number of cases is less than 10 per thousand inhabitants,
82 medium risk when it is between 10 and 50 cases per thousand inhabitants, and high risk when it has
83 more than 50 cases. In 2019, 6.2% of municipalities had high transmission and included 62.8% of the
84 total cases.

85 Although environmental and social conditions are important for malaria-endemic levels,
86 factors such as access and quality of health services act on the dynamics of the disease [9,10]. The
87 interaction of environmental, health and socioeconomic factors favors variations in notifications,
88 many of which are associated with the flaws in epidemiological surveillance actions, responsible for
89 delays in diagnosis, treatment of the disease, and conditions of the vulnerability of the population
90 [11,12].

91 According to Confalonieri [13], the environmental and social characteristics of the Brazilian
92 Amazon are relevant to the determination of epidemiological patterns. The region's geographic and
93 ecological characteristics also substantially determine potential habitats for vector reproduction [14].
94 The incidence of malaria in municipalities is influenced by environmental, socioeconomic,
95 demographic, and structural factors (organization and effectiveness of health services, land use, and
96 infrastructure) [8].

97 Given the heterogeneity of malaria cases across regions, where there is an epidemiological
98 transition towards elimination in some, but in others, it has remained an endemic disease with
99 sustained transmission, it has become necessary to understand on the aggregate level, the main factors
100 related to the municipalities with the lowest and highest risk of malaria incidence [15,7,5].

101 Understanding all the factors associated with malaria among high-risk municipalities of Brazil
102 is necessary to promote better malaria control and more effective surveillance actions at a local level.
103 Our choice of period of analysis was the result of our understanding that from 2014 to 2016, the

104 reduction in the number of malaria cases was extraordinary. As a result, we have attempted to
105 compare periods where conditions were similar.

106 For the aforementioned reasons, the objective of this paper is to analyze the high incidence
107 risk profiles in the municipalities in the period 2011-2013 and 2017-2019, and for this analysis, the
108 Principal Component Analysis (PCA) algorithm was used. Principal Component Analysis (PCA)
109 algorithm is a method used to assist in understanding the most significant relationships between
110 variables, making it possible to analyze how each chosen variable is associated with each component
111 to identify these relationships and used as an exploratory tool [16,17].

112 113 **Methods**

114 Data were obtained from the Malaria Epidemiological Surveillance System (SIVEP-Malaria)
115 for the 2011 to 2019 period. This system records malaria cases in the Legal Amazon region, and each
116 malaria-diagnosed patient file is filled with over 40 demographic, socioeconomic, epidemiologic, and
117 health data including service, patient, and disease variables.

118 The environmental variables – Hydrography, Increment, Cloud, Deforested – were taken from
119 the National Institute for Space Research’s (INPE) *Satellite Monitoring of Brazilian Amazon Forest*
120 *Deforestation Project* (PRODES). The variable of crude migration was calculated from the 2010
121 Demographic Census Microdata conducted by the Brazilian Institute of Geography and Statistics
122 (IBGE).

123 The number of level general care establishments was taken from the Registry of Health
124 Establishments (CNES) which has information on healthcare infrastructure and services, type of care
125 offered, specialized services, number of available hospital beds, and existing capacity, to assist in
126 health planning.

127 Only positive cases were used, and grouped by the municipality of infection and year of
128 registration, counting the number of occurrences of each variable used in each category.

129 The results were divided between the periods 2011-2013 and 2017-2019, by high-risk
130 municipality through the API, by species of *Plasmodium* (*P. vivax* and *P. falciparum*), and by a group
131 of variables (Table 1): Environmental and Health; Demographic and Socioeconomic Variables.
132

Variables	Description	Unit
MUN_INFE	Infection municipality code	-
IPA_total_class	Malaria incidence rate by classified by low, medium and high risk categories	categories
Demographic and Socioeconomic Variables		
Illiterate Sch5to8grade_Incomplete CollegeComplete	Education level	%
Pregnant	1st, 2nd, 3rd trimester pregnant women and ignored gestational age	%
Non_Pregnant	Non-pregnant	%
Livestock Agriculture Tourism Domestic GoldPanning	Main activity performed by the patient in the last 15 days	%
Brown Indian Black White	Patient skin color	%
Male	Patient sex	%
Age_Under5 Age_15to49	Age of patient in years, by group	%
Environmental and Health Variables		
less_than_half_crosses	Less than half cross: number of crosses detected in the blood test	%
WithSymptoms	Has symptoms related to malaria	%
Hydrography	Area of rivers and lakes in the municipality	Km ²
Increment	Deforested Increment	-
crudemigrate_2010	Taxa de migração bruta daquele município em 2010	rate
propActiveDetection	Proportion of cases notified by active detection	%
Cloud	Quantity of clouds - proxy for rainfall in the municipality	absolute
esLevGeneralCare	number of health facilities by level of general care	absolute
Negative	Negative for other hemoparasites	%
Deforested	Deforested area in the period	Km ²

Table 1: Variables used on the PCAs.
Source: SIVEP – Malaria, 2011-2019.

133
134
135
136

137 Principal component analysis (PCA) is a method used to show strong patterns between
138 variables by reducing the number of characteristics when building the principal components,
139 sometimes referred to as dimensionality reduction. The principal components make it possible to
140 reduce a set of correlated variables into a smaller number of representative variables that collectively
141 explain most of the variability in the original set [16,17]. For the best performance of the PCA, the
142 data were scaled and centered.

143 In addition to generating principal components, the method is highly interpretable, allowing
144 the analyst to understand which variables were most strongly associated with each component as well
145 as which data points had a greater influence on the characterization of a given component. Each
146 component is a linear combination of the variables and the coefficient and each variable is called a
147 Loading (factor weights) and the contribution of the municipalities is called Score. The loading helps
148 us understand which variables are most important for each major component [16,17]. Where many
149 variables correlate with one another, they will all contribute strongly to the same principal component.

150 We have also included a descriptive map of municipalities by the level of risk for the two
151 periods analyzed, to familiarize the reader with the spatial distribution of this risk across the Brazilian
152 Amazon.

153 The PCA algorithm used was from the *prcomp* package and the models were executed in R
154 version 3.6.3.

155 **Ethics Statement**

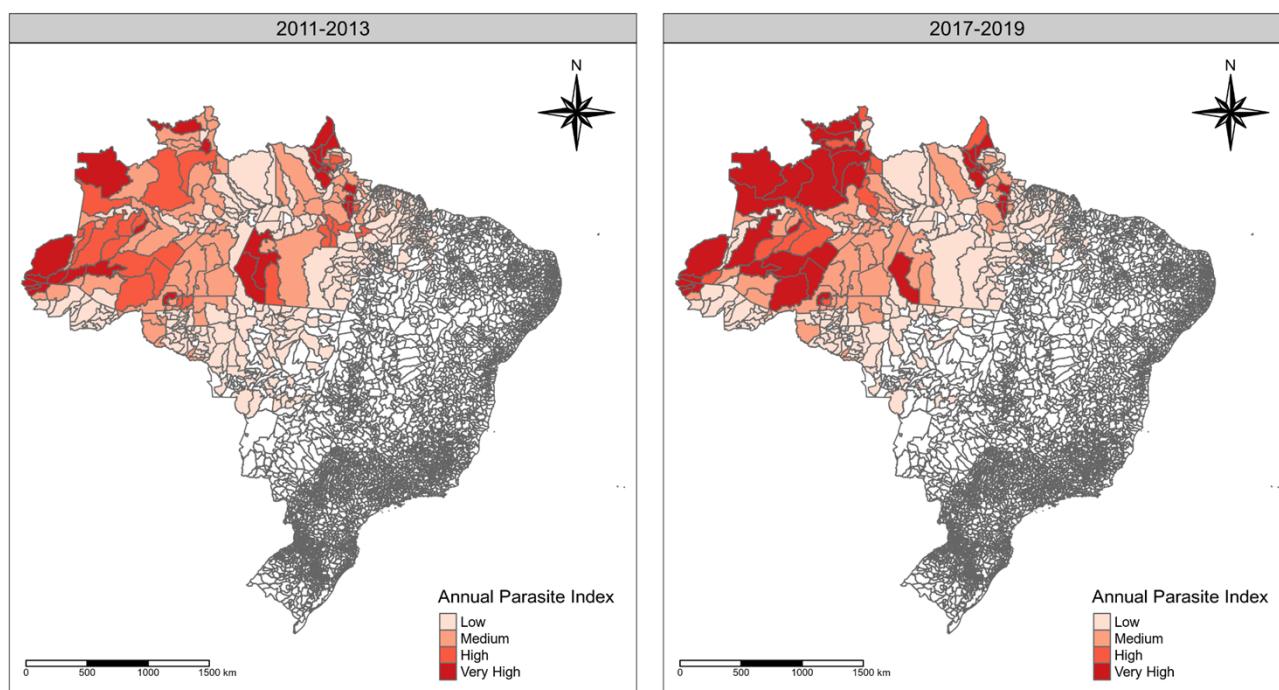
156 This work uses data in the public domain, made available upon request for access to information to
157 the Brazilian Ministry of Health, in accordance with Law No. 12,527, of November 18, 2011
158 (L12527). The data is de-identified, and the results present aggregated information, therefore, this
159 work is exempt from being evaluated by the Ethics and Research Committee, according to Resolution
160 No. 510, of April 7, 2016, of the National Health Council (Resolution CNS/MS 510/16).

161

162 **Results**

163 Figure 1 shows the risk of malaria transmission by a municipality in the two periods. The
164 municipalities that went from medium risk to very high risk in the two analyzed periods deserve
165 special attention: Santa Isabel do Rio Negro (state of Amazonas), Alto Alegre, Iracema, Pacaraima,
166 and Rorainópolis (all in the state of Roraima). When we look at the 46 municipalities with high
167 or very high risk in 2017-2019, 41% are municipalities in the state of Amazonas and 24% are in
168 Roraima. In comparison, of the 276 low-risk municipalities, 36% are in the state of Pará, only 7% are
169 in the state of Amazonas and 1% are in the state of Roraima. That is, there is a greater concentration

170 of high and very high-risk municipalities in the state of Amazonas and Roraima and a few



171 municipalities in a low-risk situation in these states.

172 **Figure 1. Risk of malaria transmission in Brazil, in the period 2011-2013 and 2017-2019.**

173 Source: SIVEP – Malaria, 2011-2013; 2017-2019.

174

175 Only 5 municipalities managed to change from high incidence in the period 2011-2013 to low
176 incidence in the period of 2017-2019: São Paulo de Olivença (AM), Tabatinga (AM), Anapu (PA),
177 Goianésia do Pará (PA) and Senador José Porfirio (PA).

178 There are 48 high-risk municipalities with positive cases for *P. vivax*. Regarding positive cases
179 for *P. falciparum*, we have 48 high-risk municipalities in the first period. Table 2 shows the first 4
180 main components that represent the highest percentage of variance from the PCA in the period 2011-
181 2013 and 2017-2019, and also facilitates the comparison between *P. vivax* and *P. falciparum* cases,
182 by groups of variables (demographic and socioeconomic; health and environmental).

183 There isn't an important difference in the 2011-2013 period between *P. vivax* and *P.*
184 *falciparum* concerning the explained variance on the PCA models. In the period 2017-2019, for cases
185 of *P. vivax*, the first 4 main components explain 70.10% in the group of demographic and

186 socioeconomic variables and 71.57% in the group of environmental and health variables. For *P.*
 187 *falciparum* cases, it explains 60.13% and 70.52%, respectively.

Period 2011 to 2013					
Principal Components	<i>P. vivax</i> (n=48)		Principal Components	<i>P. falciparum</i> (n=48)	
	Demographic and Socioeconomic (%)	Environmental and Health (%)		Demographic and Socioeconomic (%)	Environmental and Health (%)
PC1	30.79	31.81	PC1	30.52	31.13
PC2	12.92	15.14	PC2	13.97	15.86
PC3	11.34	15.02	PC3	10.14	13.58
PC4	10.12	10.82	PC4	8.81	10.66
Total Variance*	65.17	72.79	Total Variance*	63.44	71.23
Period 2017 to 2019					
Principal Components	<i>P. vivax</i> (n = 48)		Principal Components	<i>P. falciparum</i> (n = 45)	
	Demographic and Socioeconomic (%)	Environmental and Health (%)		Demographic and Socioeconomic (%)	Environmental and Health (%)
PC1	36.04	27.08	PC1	24.20	26.91
PC2	14.29	18.25	PC2	17.42	17.61
PC3	10.51	15.88	PC3	10.79	15.27
PC4	9.26	10.36	PC4	7.72	10.73
Total Variance*	70.1	71.57	Total Variance*	60.13	70.52

188 **Table 2: Percentage of total variance explained for the four main components selected, 2011 - 2013 and 2017-2019.**

189 Source: SIVEP – Malaria, 2011-2013; 2017-2019.

190 *sum of the percentage of the first four components of which one of the models.

191

192

193 **Principal Components Analysis in High-Risk Municipalities**

194 Figure 2 shows the loadings greater than 0.25 in absolute values for the first four principal
 195 components of all PCA models.

196 Looking at the positive cases of *P. vivax* for the period 2011-2013, the first principal
 197 component was positively correlated with being under 5 years old with complete higher education
 198 and negatively correlated with livestock, incomplete elementary school, male, white, and between 15
 199 to 49 years. The second has only two positively correlated variables – being indigenous and pregnant
 200 (PC2 – vulnerable subgroups), and negatively correlated with being illiterate and brown. From
 201 environmental and health characteristics (Figure 2b) we see a profile of negative from other parasites,
 202 hydrography, has less than half crosses on the exam and proportion of active detection (PC1- most of
 203 diagnoses characteristics – Figura 2b). It is worth mentioning that the third and fourth main
 204 components have only positively correlated variables, such as hydrography, deforested, increment,
 205 cloud, less than half crosses, and proportion of active detection.

206 From the 2017-2019 period (Figure 2c; 2d), the first principal component (Figure 2c) was
207 positively correlated with age under 5 and indigenous; and negatively correlated with male, black,
208 white, incomplete elementary school, and between 15 and 49 years old. The second principal
209 component has livestock occupation, illiteracy, tourism, male and indigenous as positively correlated,
210 and pregnant, not pregnant, and brown negatively correlated. In Figure 2d, from the first component
211 we can see positively correlated: proportion of active detection, negative to other parasites, less than
212 half crosses and hydrography and negatively correlated with increment on deforested area, deforested,
213 crude migration rate and the presence of symptoms. The second principal component has only
214 negatively correlated variables: number of health establishments, less than half crosses on the exam,
215 deforestation, and increment of deforestation in the period. The third component has a direct
216 association with environmental variables (cloud, hydrography, deforested, and increment).

217 Regarding infection by *P. falciparum*, we see a change in the profile in the two periods. In the
218 first period (Figure 2e), is 15 to 49 years old, male, incomplete elementary school, white and black
219 are positively correlated to the first principal component and negatively correlated with illiterate,
220 pregnant, indigenous, and under five years old. The second PC has an occupation in agriculture and
221 indigenous people as positively correlated, and non-pregnant women and brown negatively.
222 Hydrography and proportion of active detection were positively correlated with the first principal
223 component (Figure 2f).

224 In the period 2017-2019 instead (Figure 2g), under 5 years old, indigenous are positively
225 correlated, and black, brown, and 15 to 49 years old are negatively correlated. The second principal
226 component has male, tourism, and illiterate as positively correlated and pregnant and not pregnant
227 and elementary education incomplete negatively correlated. When we look at environmental and
228 health variables (Figure 2h), the presence of symptoms, crude migration rate, deforested and
229 Increment were positively correlated and negative for other parasites and the proportion of active
230 detection was negatively correlated.

231

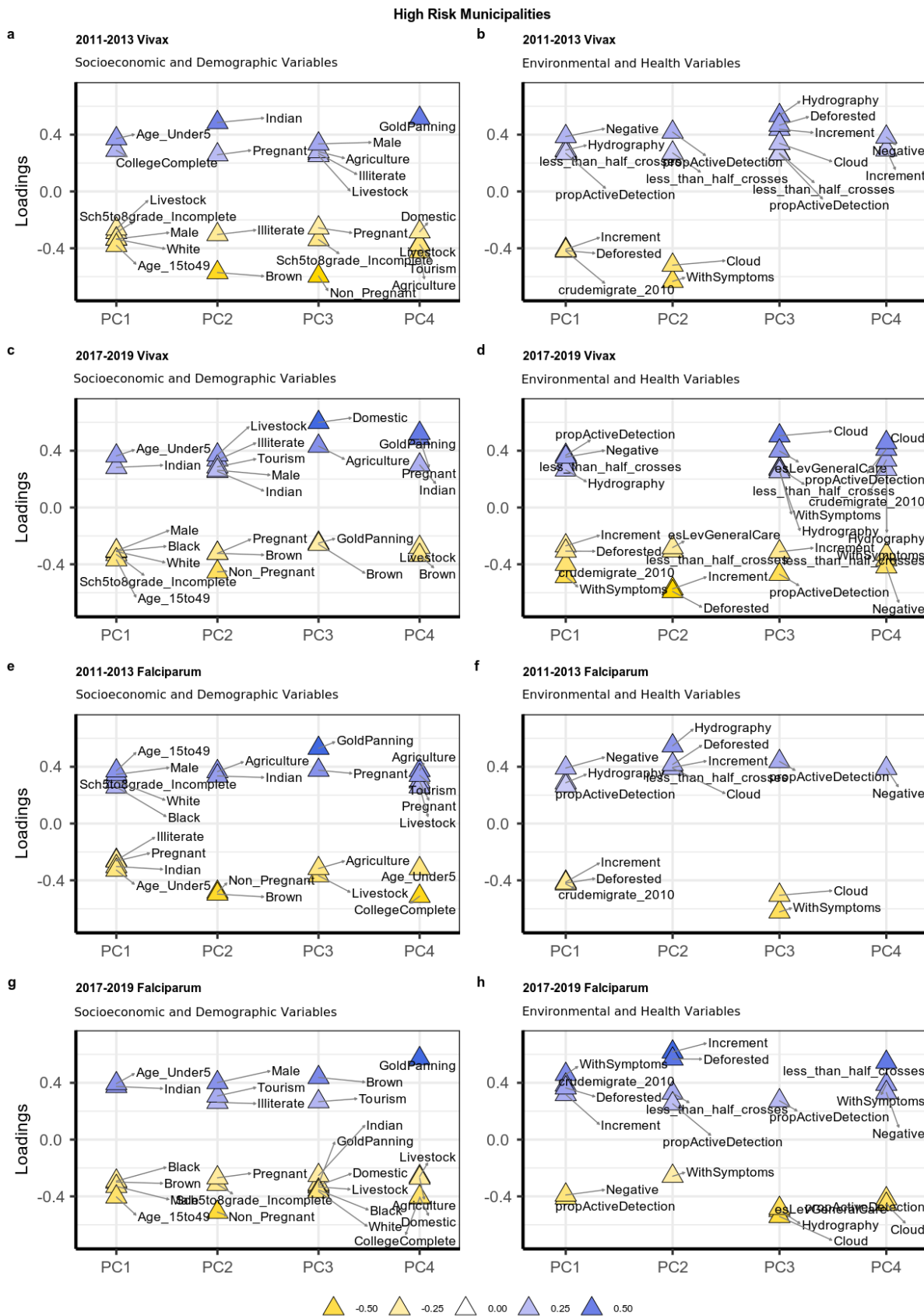


Figure 2: Loadings, High-risk municipalities, *P. vivax* and *P. falciparum*, 2011-2013 and 2017-2019 period.
Source: SIVEP – Malaria, 2011-2013; 2017-20.

232
233
234
235

236 Figure 3 shows all the high-risk municipalities with positive scores (contribution of the
237 observations – municipalities for every Principal Component) in the recent period of 2017-2019, so
238 from each component we can see the municipalities that most contribute to a specific principal
239 component and suggest that these municipalities have the profiles shown at Figure 2. As we can see,
240 it has some spatial patterns, fundamentally about environmental and health variables.

241 The municipalities represented by green circles have a profile of cases under 5 years old and
242 indigenous (Figure 3a) for the 2017-2019 period, the group of municipalities with a yellow circle has
243 a profile including livestock, illiterate, tourism, male and indigenous. The PC3 group of
244 municipalities has a profile of Domestic and also Agriculture, and the PC4 group Gold Panning,
245 Pregnant and indigenous.

246 Relative to Environmental and Health variables (Figure 3b) for *P. vivax*, the first group of
247 municipalities has higher positive loadings for the proportion of active detection cases, negative to
248 other parasites, with parasitemia less than one cross and Hydrography area. The second group (PC2
249 – yellow circle) has a profile of only negative correlation variables: number of health establishments,
250 deforested area, and increment of deforested area. The third group (PC3 – purple circle) has a high
251 correlation with the number of healthcare establishments, the area covered by water bodies, and
252 parasitemia with less than one cross, and the fourth group (PC4 – orange circle), high loadings of the
253 proportion of active detection and crude migration rate.

254 Regarding the *P. falciparum* cases and the socioeconomic and demographic profile (Figure
255 3c), the PC1 group of municipalities with high scores has an age under 5 and an indigenous profile.
256 The second group (PC2) of municipalities has a male, tourism, and illiterate profile. The third group
257 (PC3) of municipalities is associated with a brown and tourism profile and the fourth group (PC4) is
258 associated with a gold panning occupation profile.

259 Figure 3d shows the municipalities related to the *P. falciparum* cases and Environmental and
260 Health profiles and we can see that the first group of municipalities (PC1) characteristics is associated
261 with crude migration rate, symptoms, deforested area, and increment of deforested area. The second

262 group is also associated with deforested area and increment of deforested area and proportion of active
263 detection and exam with less than one cross (PC2). The third group (PC3) has a profile of proportion
264 of active detection and the fourth group (PC4) has a profile mostly of diagnostic variables: less than
265 one cross, the proportion of active detection, with symptoms and negative to other parasites.

266

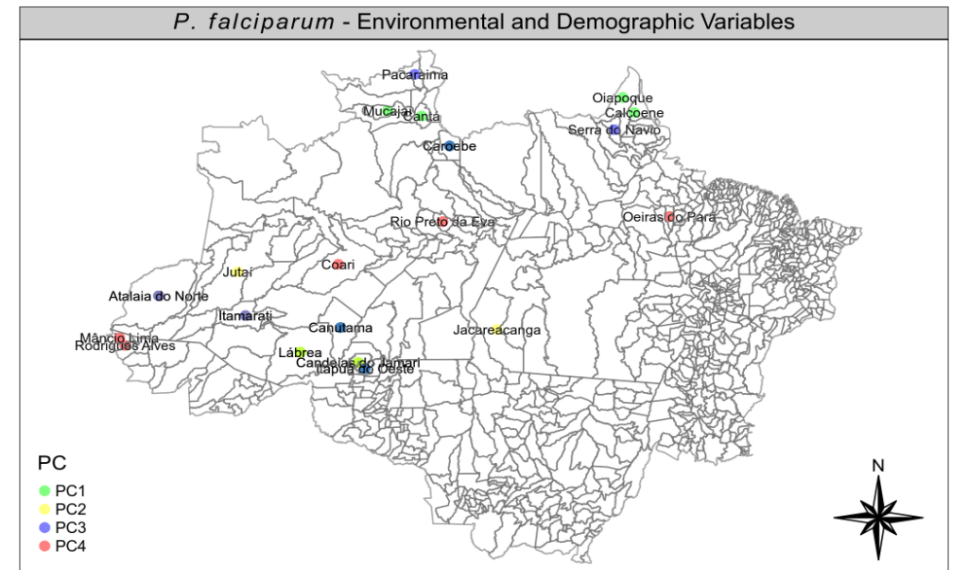
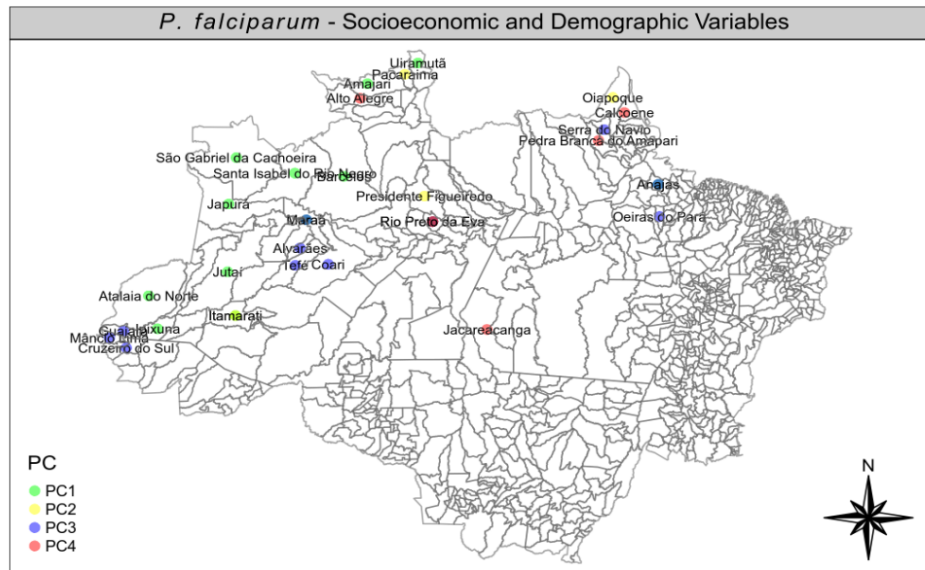
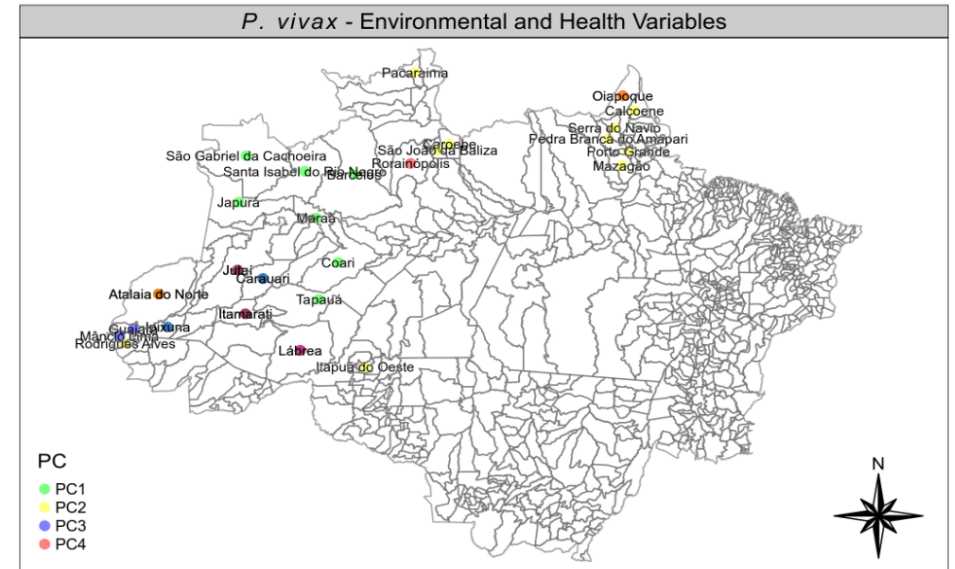
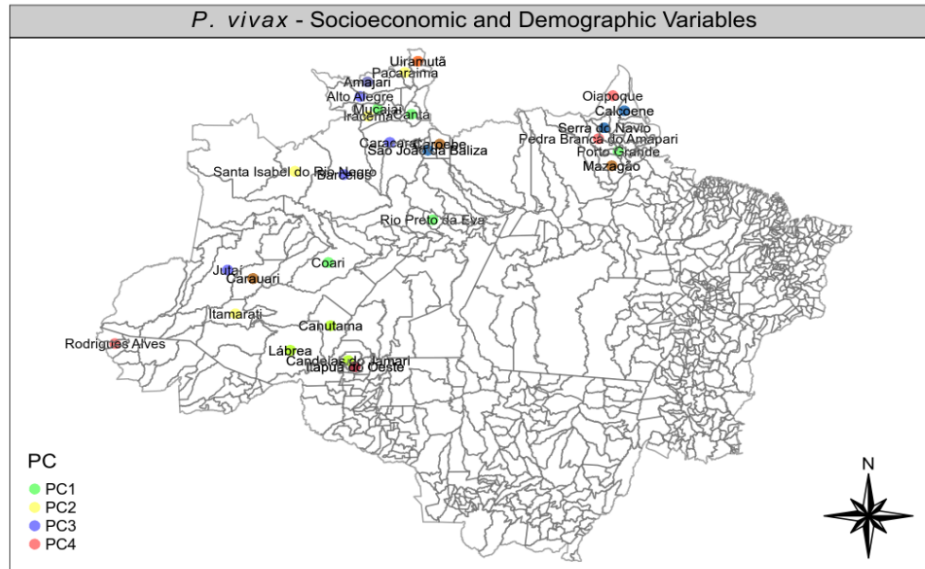


Figure 3. Scores High-risk municipalities, *P. vivax*, and *P. falciparum*, 2017-2019 period.

From this assessment, combining the loadings (Figure 2d;e;f;g) with the characteristics and scores (Figure 3) of municipalities we can draw the following profiles when we look at high-risk municipalities in the recent period (2017-2019):

Group 1: vulnerability profile, that is, high-risk municipalities with cases more associated with socioeconomic and demographic variables with vulnerability characteristics, such as children and the indigenous population, high proportions of illiteracy, incomplete secondary schooling, and little or no association with environmental or health variables, to the extent that they're captured in our data.

Group 2: profile of occupation and working age, of high-risk municipalities with cases more related to socioeconomic and demographic variables linked with specific occupations and a working-age population, such as men, people aged between 15 and 49 years, with low education and in occupations such as agriculture, livestock and gold panning in addition to environmental characteristics related to deforestation and hydrography.

Group 3: profile associated with different characteristics of *P. falciparum* and *P. vivax*, such as the municipality of Rio Preto da Eva, which for the cases of *P. vivax*, may be more related to children under 5 years and differently, *P. falciparum* in this municipality could be more related to gold panning, deforestation and the presence of large bodies of water (high loading in PC4). A few other municipalities appear to follow the same differentiated profile between the two types of *Plasmodium*, such as Coari, Canutama, and São Gabriel da Cachoeira.

It should be noted that the same municipality may have characteristics in more than one group above.

Discussion

This study analyzed the different high incidence risk profiles in the municipalities in the period in the Brazilian Amazon between 2011-2013 and 2017-2019. It was observed that the risk of having the disease was not uniform, which can be verified by the results of the annual parasite index (API).

The results show us that the main components described the majority of variation and explained the varied influence of the original characteristics. These influences, the loadings, show us different characteristics of the municipalities with high API.

There is a greater positive correlation for being under 5 years old, but also with complete higher education, men in the two periods of time, corroborating the finding of the study by Lana [7] which indicated that municipalities with higher API are highly characterized by peri-domestic transmission in which all segments of the population have reported cases and also indicates few changes in the infection profile between the two periods for these high-risk municipalities which is a concern because, unlike low-risk municipalities, they show no improvement trends, maintaining transmission or even increasing. [18], when studying the incidence of malaria and long-term epidemics, showed that municipalities with long-term epidemics were more frequent in the states with the highest incidence.

The study by Lana [7] demonstrates that there are distinct risk profiles between the municipalities and their transmission configurations, for example, notes that in a high-incidence municipality such as Anajás (Pará), a large proportion of cases occur in children. By contrast, in some low-risk municipalities, the majority of cases affect men aged 15 to 65 years old.

We found that municipalities with a high-risk transmission have heterogeneous profiles where, in some municipalities, there will be a greater relationship between the high number of cases with population vulnerability (such as age, indigenous population, education, etc.) and other groups of municipalities, the high-risk incidence could be more related to men in working age and be related to occupation.

The present study shows a predominance of high risk of Falciparum in border regions and mining regions, although other highly vulnerable occupational groups, such as tourism and agriculture, are also observed.

The gold panning occupations are generally located in areas with difficult access, with limited or non-existent health services and, in some cases, they operate illegally, factors that make it difficult

to cover timely diagnostic services. In addition, in some states that share an international border, for example, there is evidence of imported cases whose origin is from prospecting in other countries, but which are not registered as such. Another problem of malaria in mining is the high rate of *P. falciparum* (36%) and late and/or inadequate diagnosis and treatment [19].

Figures 2 and 3 show us municipalities with high scores on the groups of socioeconomic and demographic variables but don't appear on the maps of environmental and health groups. For example, the municipalities of Uiramutã (Roraima) and Rio Preto da Eva (Amazonas) have high scores on the demographic and socioeconomic variables and this could indicate that the high incidence cases of malaria in these municipalities are more strongly associated with population vulnerability than environmental and health variables when it comes to *P. vivax* infections.

Related to Group 1 presented in the previous section, studies show that the lower the years of education the higher the chances to have malaria and this could be related to a lack of opportunity, worse quality of life, and a general context of vulnerability [21,6,22]. Additionally, the high loading of the variable children under 5 years old could suggest that a significant proportion of malaria transmission occurs within households in these municipalities [12]. Children are population groups that have little immunity. Indoor contamination suggests the existence of precarious housing that does not offer protection, increasing contamination.

Group 2 shows us some municipalities with a high association with occupational and environmental characteristics. Corroborating with some studies [23,24,18] that indicate an association between the increase in cases of malaria and gold mining activity, as well as the association between deforestation for agriculture and livestock. Mining contributes to the spread of the disease, as it favors deforestation and the creation of pools of water, an ideal habitat for the reproduction of vectors that transmit malaria.

For instance, Oliveira [24] when analyzing the occurrence of malaria cases in municipalities in Amapá and its relationship with mining activity, found out that in the places where mining projects had been implemented, the number of cases was much higher compared to the places that did not

have mining activities. In addition, this profile is more associated with men, which can be explained by the fact that these occupations are generally associated with a male workforce, generally working in areas with poor infrastructure and favorable to the development of disease vectors. Thus, the predominance of cases in male populations in these municipalities in Group 2 is greater due to their higher exposure profile [25,26,27].

The associations between types of economic activity, social mobility, and regionalization of territorial occupation and the incidence of malaria confirmed that malaria infection could be associated with productive activity and labor flow linked to extractive activities and agricultural settlements as found in some principal component results like gold mining, agricultural and livestock occupation [8].

Hydrography and deforestation variables appeared with high loading and positive correlation in high-risk municipalities in Group 2. The high importance of the hydrography variable indicates a relationship between the cases of malaria and the high availability of natural breeding sites for mosquito vectors, since rainfall, river water levels, and climatic factors have been linked with malaria cases [14].

In a prior study, [28] also analyzed the influence of some variables such as temperature, precipitation, and water level on the peaks of malaria cases in four municipalities of Amazonas, and observed a statistically significant correlation between all analyzed variables and causes of disease. This result is explained by the fact that seasonality and the abundance of rainfall cause fluctuations in river water levels and consequent flooding create a favorable environment for mosquito breeding.

Relative to the high loading value in variables such as deforested area (total area deforested in the period) and increment in deforestation (difference in a deforested area in the period), there are some contradictory findings in the literature. While deforestation increased malaria incidence in some studies [29,30,31], forest conservation and high forest cover were associated with a higher incidence in other studies [32,33].

The relationship between deforestation and malaria is the object of some debate since there seems to be no one direction of causality and several concomitant factors may be interacting to determine it. For instance, in the case of "frontier malaria", the authors identify biological, ecological, and sociodemographic phenomena operating at distinct spatial scales to define it. Here, our inclusion of these variables is an attempt to include an ecological component that contributes to the axis of ecological receptivity of the area to disease vectors, rather than an attempt to summarize the effect of environmental variables on malaria incidence, such as the one attempted [34] for deforestation in several countries in the African continent [35].

Group 3 shows us municipalities with different characteristics between those who are infected by *P. vivax* and those who were infected by *P. falciparum*. [33] found some environmental differences between these two species, for example, *P. falciparum* malaria had a stronger relationship with deforestation than *P. vivax*. In addition, these municipalities show different patterns in socioeconomic and demographic characteristics, such as Rio Preto da Eva (Amazonas), in which transmission by *P. vivax* is related to children under 5 years and the indigenous population, suggesting a more frequent peridomestic infection, while in cases of *P. falciparum* is more related to gold panning.

According to the 2020 Epidemiological Bulletin, in the Amazon region, where 99.9% of malaria cases occur, around 80% of malaria was concentrated in 41 municipalities in 2019, with 16 in Amazonas (39.0%), 8 in Pará (19.5%), 7 in Roraima (17.1%), 4 in Amapá (9.8%), 3 in Acre (7.3%), 2 in Rondônia (4.9%) and 1 in Mato Grosso (2.4%). The National Program to Combat Malaria (PNCM) prioritizes the municipalities that together are responsible for 80% of the country's autochthonous cases of malaria. The present study uses machine learning algorithms for PCAs to find complex patterns and reinforces priority areas or municipalities previously established by the PNCM, as well as pointing to epidemiological indicators in Brazil.

The use of observational data and some contextual data from other sources is a limitation of this study because the authors didn't control the study design and quality of data reporting. The

inclusion of variables from distinct sources represents an attempt to control for confounding variables in a setting where primary data collection is not feasible.

The objective of this work was to draw profiles associated with the risk of incidence in the municipalities by combining data from the malaria surveillance and control data with environmental and health data from the municipalities.

It is necessary to understand the profile of the population affected by malaria together with the environment in which they live, the weather, forest, hydrography, and health conditions to assess the structure in these regions, in addition to analyzing the cases of *P. falciparum* separately from the cases of *P. vivax*, understanding there are particularities in socioeconomic, demographic, environmental and health characteristics for infection by both types of *Plasmodium*.

Thus, understanding the socio-environmental determinants at the local level is essential for the success of malaria prevention and control strategies.

Acknowledgments

We thank the Bill & Melinda Gates Foundation (Process ID INV 003970), the Brazilian Ministry of Health, the Department of Science and Technology (DECIT), and the National Council for Scientific and Technological Development (CNPq) (Process 443048/2019-3) by financial support.

References

1. WHO. World malaria report 2020. Geneva: World Health Organization; 2020. Available from: <https://www.who.int/publications/I/Item/9789240015791>.
2. Ministério da Saúde: Dados para cidadão – Malária – Brasil. [citado 10 Jun 2021] Disponível em: https://public.tableau.com/app/profile/mal.ria.brasil/viz/Dadosparacidado_201925_03_2020/Incio.
3. Ferreira MU, Castro MC. Challenges for malaria elimination in Brazil. *Malar J.* 2016;15(1).
4. Padilha MAO, Melo JO, Romano G, Lima MVM, Alonso WJ, Sallum MAM, Laporta GZ. Comparison of malaria incidence rates and socioeconomic-environmental factors between the states of Acre and Rondônia: a Spatio-temporal modeling study. *Malar J.* 2019;18.
5. Canelas T, Castillo-Salgado C, Ribeiro H. Analyzing the local epidemiological profile of malaria transmission in the Brazilian Amazon Between 2010 and 2015. *PLoS Curr.* 2018. pmid: 29623243.

6. Lana RM, Riback TIS, Lima TFM, Silva-Nunes M, Cruz OG, Oliveira FGS, et al. Socioeconomic and demographic characterization of an endemic malaria region in Brazil by multiple correspondence analysis. *Malar J.* 2017;16(1). pmid: 28969634.
7. Lana R, Nekkab N, Siqueira AM, Peterka C, Marchesini P, Lacerda M, et al. The top 1%: quantifying the unequal distribution of malaria in Brazil. *Malar J.* 2021;20(1). pmid: 33579298.
8. Souza PF, Xavier DR, Suárez Mutis MC, Mota JC, Peiter PC, Matos VP, et al. Spatial spread of malaria and economic frontier expansion in the Brazilian Amazon. *plos one.* 2019;14:e0217615. pmid: 31211772.
9. Ministério da Saúde. Secretaria de Vigilância em Saúde. Guia para gestão local do controle da Malária: diagnóstico e tratamento. Brasília: Ministério da Saúde; 2008. (Série B. Textos Básicos de Saúde).
10. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Guia prático de tratamento da Malária no Brasil. Brasília: Ministério da Saúde; 2010.
11. Ministério da Saúde. Boletim Epidemiológico. Brasília: Ministério da Saúde; 2020.
12. PAN-WHO. Report on the situation of Malaria in the Americas, 2000-2015. Geneva: PAN-WHO; 2016. (Document prepared by the Regional Malaria Program, Pan American Health Organization with data from Annual Country Reports-2016).
13. Confalonieri UEC. Saúde na Amazônia: um modelo conceitual para a análise de paisagens e doenças. *Dossiê Amazônia Brasileira I. Estud. av.* 2005;19(53):221-36.
14. Wolfarth-Couto B, Silva RA, Filizola N. Variability in malaria cases and the association with rainfall and rivers water levels in Amazonas State, Brazil. *Cad. Saude Publica.* 2019;35(2):e00020218.
15. Braz RM, Barcellos C. Analysis of the process of malaria transmission elimination with a spatial approach to incidence variation in the Brazilian Amazon, 2016. *Epidemiol Serv Saude.* 2018;27:e2017253.
16. Kuhn M, Johnson K. Applied predictive modeling. New York: Springer; 2013.
17. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning: with applications in R. eight ed. New York: Springer; 2017.
18. Braz RM, Duarte EC, Tauil PL. Characteristics of malaria epidemics in the municipalities of the Brazilian Amazon. *Cad Saude Publica.* 2013;29(5):935-44. pmid: 23702999.
19. Ministério da Saúde. Secretaria de Vigilância em Saúde. Plano de eliminação da Malária no Brasil. Brasília: Ministério da Saúde; 2015.
20. Lima ISF, Lapouble OMM, Duarte EC. Time trends and changes in the distribution of malaria cases in the Brazilian Amazon Region, 2004-2013. *Mem Inst Oswaldo Cruz.* 2017;112(1):8-18.
21. Lima JMT, Vittor A, Rifai S, Valle D. Does deforestation promote or inhibit malaria transmission in the Amazon? A systematic literature review and critical appraisal of current evidence. *Philos. Trans R Soc Lond B Biol Sci.* 2017;372(1722):20160125. pmid: 28438914.
22. Castillo-Salgado C. Mapeo geoepidemiológico en la nueva vigilancia de salud pública. El caso de la malaria en Chiapas, México, en 2002. *Gac Med Mex.* 2017;17;153Suppl 2:S5-S12.
23. Santos VR, Yokoo EM, Souza-Santos R, Atanaka-Santos M. Fatores socioambientais associados à distribuição espacial de malária no assentamento Vale do Amanhecer, Município de Juruena, Estado de Mato Grosso, 2005. *Rev Soc Bras Med Trop.* 2009;42(1):47-53.
24. Oliveira-Ferreira J, Lacerda MVG, Brasil P, Ladislau JLB, Tauil PL, Daniel-Ribeiro CT. Malaria in Brazil: an overview. *Malar J.* 2010;9. pmid: 20433744.

25. Lopes TMR, Ventura AMRS, Guimarães RJPS, Guimarães LHR. Situação epidemiológica da malária em uma região de Garimpo, na região da Amazônia brasileira, no período de 2011 a 2015. *Revista Eletrônica Acervo Saúde*. 2019;(25):e759.
26. Mesquita EM, Muniz TF, Sousa ALS, Brito CXL, Nunes, SCM, Grisotto MAG. Levantamento epidemiológico da malária no estado do Maranhão, Brasil nos anos de 2007 a 2012. *Rev Ciências Saúde*. 2013;15(1):11-18.
27. Amaral PM. Análise da influência de fatores socioambientais e climáticos sobre a incidência de dengue, malária e tuberculose. Vitória: Universidade Federal do Espírito Santo; 2015. (Monografia).
28. Wolfarth BR, Filizola N, Tadei W, Durieux L. Análise epidemiológica da malária e as relações com variáveis ambientais em quatro municípios do Amazonas, Brasil. *Anais do XV Symposium SELPER*; 2011.
29. Chaves LSM, Conn JE, López RVM, Sallum MAM. Abundance of impacted forest patches less than 5 km² is a key driver of the incidence of malaria in Amazonian Brazil. *Sci Rep*. 2018;8(1):7077. pmid: 29728637.
30. Hahn MB, Gangnon RE, Barcellos C, Asner GP, Patz JA. Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PLoS One*. 2014;9(1):e85725.
31. Santos AS, Almeida AN. The impact of deforestation on malaria infections in the Brazilian Amazon. *Ecol Econ*. 2018;154:247-56.
32. Valle D, Clark J. Conservation efforts may increase malaria burden in the Brazilian Amazon. *PLoS One*. 2013;8(3):e57519
33. MACDONALD AJ, MORDECAI RA. Amazon deforestation drives malaria transmission, and malaria burden reduces forest clearing. *PNAS*. 2019;116(44):22212-18.
34. Bauhoffa S, Busch J. Does deforestation increase malaria prevalence? Evidence from satellite data and health surveys. *World Development*. 2020;127:104734.
35. Castro MC, Monte-Mór RL, Sawyer DO, Singer BH. Malaria risk on the Amazon frontier. *PNAS*. 2017;103(7):2452-57.
36. Siteo GAA. Risco e vulnerabilidade social à malária na província de Zambézia, Moçambique. *Revista do Programa de Pós-Graduação em Geografia*. 2015;7(1):3-24.
37. Monteiro THA, Chaves TSS, Matos HJ, Lisboa NFS, Guimarães RJPS, Guimarães LHR, et al. Basic sanitation, socioeconomic conditions, and degree of risk for the presence and maintenance of malaria in a low-transmission area in the Brazilian Amazon. *Rev Soc Bras Med Trop*. 2015;48(5):573-79.