





## RESEARCH ARTICLE

# REVISED Modelling sociodemographic factors that affect malaria prevalence in Sussundenga, Mozambique: a cross-sectional study. [version 2; peer review: 2 approved]

Joao Ferrao <sup>1</sup>, Dominique Earland<sup>2</sup>, Anisio Novela<sup>3</sup>, Roberto Mendes<sup>4</sup>, Marcos Ballat <sup>5</sup>, Alberto Tungadza<sup>6</sup>, Kelly Searle<sup>7</sup>

<sup>1</sup>Engineering & Agriculture, 1Instituto Superior de Ciências e Educação a Distância, Beira, Sofala, Mozambique

<sup>2</sup>School of Public Health, University of Minnesota, Twin City, Minnesota, USA

<sup>3</sup>Hospital Distrital de Sussundenga, Direccao Distrital de Saude, Sussundenga, Manica, Mozambique

<sup>4</sup>GIS - Faculdade de Economia e Gestao, Universidade Catolica de Mocambique, Beira, Sofala, Mozambique

<sup>5</sup>Faculdade de Engenharia, Universidade Catolica de Mocambique, Chimoio, Manica, Mozambique

<sup>6</sup>Faculdade de Ciências de Saúde, Universidade Católica de Moçambique, Chimoio, Manica, Mozambique

<sup>7</sup>School of Public Health, University of Minnesota, Twin City, Minnesota, USA

**V2** First published: 14 Feb 2022, 11:185  
<https://doi.org/10.12688/f1000research.75199.1>

Latest published: 05 May 2022, 11:185  
<https://doi.org/10.12688/f1000research.75199.2>

## Abstract

**Background:** Malaria is still one of the leading causes of mortality and morbidity in Mozambique with little progress in malaria control over the past 20 years. Sussundenga is one of most affected areas. Malaria transmission has a strong association with environmental and sociodemographic factors. The knowledge of sociodemographic factors that affects malaria, may be used to improve the strategic planning for its control. Currently such studies have not been performed in Sussundenga. Thus, the objective of this study is to model the relationship between malaria and sociodemographic factors in Sussundenga, Mozambique.

**Methods:** Houses in the study area were digitalized and enumerated using Google Earth Pro version 7.3. In this study 100 houses were randomly selected to conduct a community survey of *Plasmodium falciparum* parasite prevalence using rapid diagnostic test (RDT). During the survey, a questionnaire was conducted to assess the sociodemographic factors of the participants. Descriptive statistics were analyzed and backward stepwise logistic regression was performed establishing a relationship between positive cases and the factors. The analysis was carried out using SPSS version 20 package.

**Results:** The overall *P. falciparum* prevalence was 31.6%. Half of the malaria positive cases occurred in age group 5 to 14 years. Previous malaria treatment, population density and age group were significant predictors for the model. The model explained 13.5% of the variance in malaria positive cases and sensitivity of the final model was 73.3%.


**Conclusion:** In this area the highest burden of *P. falciparum* infection

## Open Peer Review

Approval Status  

	1	2
<b>version 2</b>		
(revision)		
05 May 2022		
<b>version 1</b>		
14 Feb 2022		

1. **Ewan Cameron**, Curtin University, Perth, Australia

2. **Gabriel Zorello Laporta** , Centro Universitario FMABC, Santo André, Brazil

Any reports and responses or comments on the article can be found at the end of the article.

was among those aged 5–14 years old. Malaria infection was related to sociodemographic factors. Targeting malaria control at community level can combat the disease more effectively than waiting for cases at health centers. These findings can be used to guide more effective interventions in this region.

### Keywords

sociodemographic, social determinants of health, malaria, prevalence, Sussundenga,



This article is included in the **Pathogens** gateway.



This article is included in the **Emerging Diseases and Outbreaks** gateway.



This article is included in the **Sociology of Health** gateway.

**Corresponding author:** Joao Ferrao ([jferrao@isced.ac.mz](mailto:jferrao@isced.ac.mz))

**Author roles:** **Ferrao J:** Conceptualization, Data Curation, Methodology, Writing – Original Draft Preparation; **Earland D:** Writing – Review & Editing; **Novela A:** Investigation, Supervision; **Mendes R:** Formal Analysis; **Ballat M:** Data Curation; **Tungadza A:** Data Curation; **Searle K:** Funding Acquisition, Methodology, Validation

**Competing interests:** No competing interests were disclosed.

**Grant information:** This study was funded through a SEED grant from the University of Minnesota Center for Global Health and Social Responsibility (CGHSR). They provided funds to carry out the study.

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

**Copyright:** © 2022 Ferrao J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Ferrao J, Earland D, Novela A *et al.* **Modelling sociodemographic factors that affect malaria prevalence in Sussundenga, Mozambique: a cross-sectional study. [version 2; peer review: 2 approved]** F1000Research 2022, 11:185 <https://doi.org/10.12688/f1000research.75199.2>

**First published:** 14 Feb 2022, 11:185 <https://doi.org/10.12688/f1000research.75199.1>

**REVISED Amendments from Version 1**

This is a new version of the manuscript. The model equation was changed and more text was added in the introduction and discussion as a result of the valuable contribution of the reviewers.

**Any further responses from the reviewers can be found at the end of the article**

**Background**

Malaria is a serious and sometimes fatal disease caused by a *Plasmodium* spp. parasite that commonly infects *Anopheles* spp. mosquitos which feed on humans. Although malaria can be a deadly disease, infection and death can be prevented.<sup>1</sup> Almost half of the world's population lives in areas at risk of malaria transmission. Six countries account for more than half of all malaria cases worldwide and Mozambique is among them.<sup>2</sup>

In Mozambique, a country in Sub-Saharan Africa, with a population of over 30 million, malaria is one of the leading causes of mortality and morbidity. In 2018, Mozambique recorded the third largest number of malaria cases in the world, accounting for 5% of all cases.<sup>3</sup>

The country has made little progress in malaria control. Indoor residual spraying (IRS), insecticide treated bed nets (ITNs), and parasitological diagnosis in health facilities using rapid diagnostic test (RDTs) with effective artemisinin combination therapy (ACT) are the forms of malaria intervention currently being used. The entire country uses RDTs with ACT as the standard of care in public health facilities and ITNs are only available at antenatal clinics, indicated for pregnant women and children under five.<sup>4</sup>

Manica Province in central Mozambique has the second highest number of malaria incidences in the country. In the first quarter of 2020, there were 1,039,283 recorded cases with an incidence of 371 per 1000 inhabitants.<sup>5</sup> Sussundenga village, in Manica Province is one of most affected areas, with 31,397 malaria cases reported in 2019.

Malaria risk, disease severity, and clinical outcome depend on environmental, sociodemographic, economic, and behavioral factors.<sup>6-12</sup> A study in Chimoio, the provincial capital of Manica, close to Sussundenga Village, modelled the influence of climate on malaria occurrence. The study indicated that selected environmental characteristics accounted for 72.5% of malaria incidences, implying that non-environmental factors such as sociodemographic, economic, cultural and behavioral traits would account for the rest.<sup>13</sup>

While Mozambique is a country with one of the highest incidences and prevalence of malaria in the region and, it accounts for nearly half of childhood deaths, little is known about the epidemiology to inform appropriate and effective interventions. This is one of two major barriers to expanding control measures in the country with the other being limited funding.

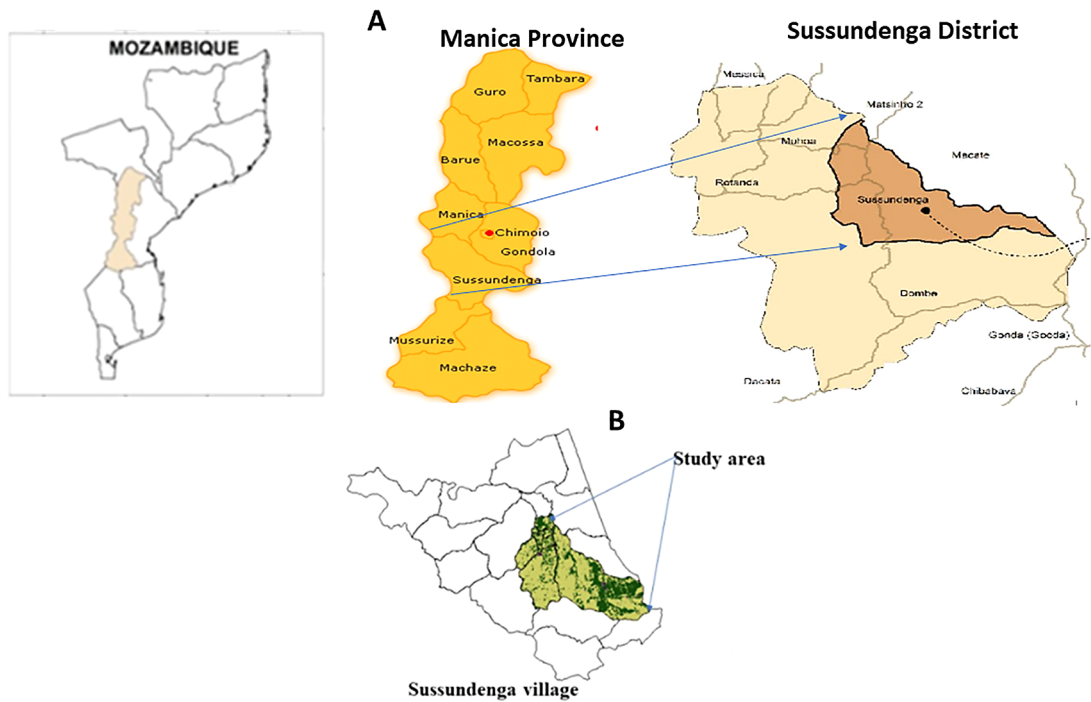
In the country, malaria transmission occurs all year round and, the knowledge of sociodemographic factors that affect malaria is crucial for informing the implementation of the most appropriate and effective malaria interventions to achieve control. In Sussundenga no studies are known in this field. Therefore, the objective of this study was to model the relationship between malaria and sociodemographic factors in Sussundenga's rural municipality.

**Methods****Study area**

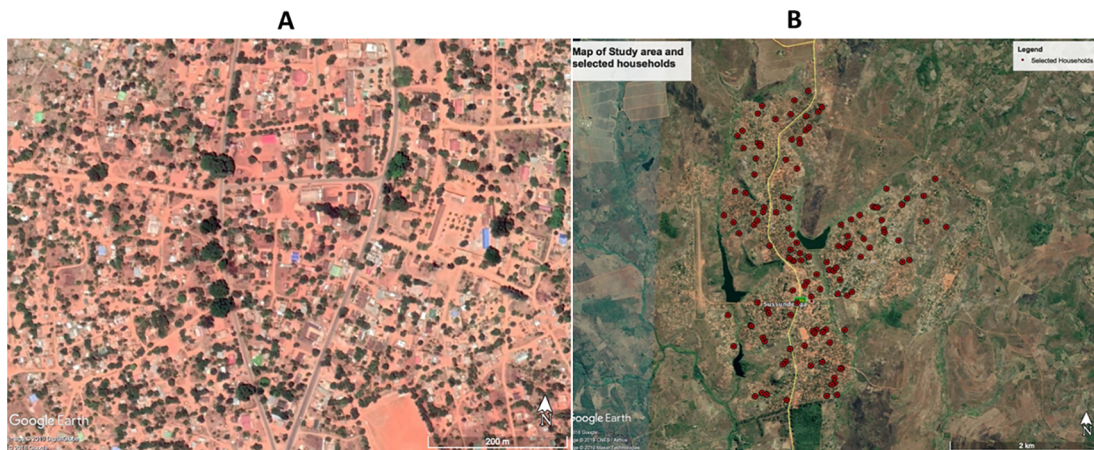
The village of Sussundenga is a rural, agrarian community 40 km from the Zimbabwe border, and is 40 km from the provincial capital of Chimoio (Figure 1).

Sussundenga is within an area of 156.9 km<sup>2</sup>, has an estimated population of 31,429 inhabitants, 47% males and 53% females. The age distribution is: 19.5% from 0 to 4 years old, 29.9% from 5 to 14-years-old, 20.5% from 15 to 24 years old and 30.1% with over 24 years old. The village is divided administratively in 17 residential areas (neighborhoods).<sup>15</sup>

The climate is tropical with an average annual precipitation of 1,200 mm. The rainy season occurs from November to April. The average minimum temperature is 6.3°C in the month of July and the average maximum temperature is 38.9°C in the month of January and the annual average is 21.2°C.<sup>16</sup>



**Figure 1. Study area.** A. Map of Mozambique, Manica province and Sussundenga district: adapted from National Cartography and Remote Sensing Centre (CENACARTA).<sup>14</sup> B. Sampled site in Sussundenga village: adapted from CENACARTA.



**Figure 2. A.** High-resolution imagery of Sussundenga village from Google Earth Pro™ (Google Earth, 2019 Google image, 2019 CNES/Airbus, image 2019 Maxar Technology). **B.** Selected households from Google Earth Pro™<sup>17</sup> (Google Earths Image 2019 Terra metrics, 2019, Google).

### Data collection

GoogleEarth Pro™<sup>17</sup> Google Earth Pro version 7.3 (Google, Amphitheatre Pkwy, Mountain View, CA, USA). satellite imagery was used to digitize and enumerate all household structures in the village of Sussundenga (Figure 2). This was a pilot study to determine malaria prevalence, risk factors, and health seeking behaviors. The sample size was determined by feasibility for the study team and study design of the community based cross-sectional survey. All households in the study area were digitized and enumerated using Google Earth Pro. With the aim of enrolling 100, a random sample of 125 households was taken, as backup for refusals and errors in the digitizing process (misclassified non-household structures).

Coordinates of the households were extracted using a GPS device and maps of the selected households to conduct study visits. The study involved two visits to the selected households. The first was a notification visit where the study team introduced themselves to the head of the household and explained the objectives and procedures of the study. It is customary for the head of household to provide permission to the study team before any activities take place at the household involving other household members. Once the head of household gave permission, the study team conducted a household census with the head of household and begin the process of individual written informed consent with the household residents, for all adult (18+ years) residents and parental permission and consent from minors.

After obtaining consent from the household residents, the study team informed participants when they would return the following day to conduct the study activities. The only eligibility requirement was that the residents live in the household full time. Data collectors verbally administered a questionnaire to collect the basic demographics. The field study was carried out from December 2019 to January 2020.

The study nurse collected current malaria specific symptoms by self-report and took participant's temperature using a digital thermometer (GP-300, RoHS:ISO 9000). They then collected a finger prick blood sample to administer a Rapid Diagnostic Test (RDT), RightSign Biotest<sup>R</sup> (Biotest, Hangzhou Biotest Biotech Co, China, Ref.No:IMPF – C51S). According to the manufacture, this test captures the HRP2 antigen on the strip and has a sensitivity is >99.0%. The results were recorded and, in the event, that a participant was positive for malaria, the study nurse referred them to the Sussundenga rural health center (RHC) for diagnosis confirmation and treatment. The questionnaire was conducted using tablet computers with the REDCap a secure, web-based data capture tool. Study data were collected and managed using REDCap electronic data capture tools hosted at University of Minnesota, downloaded to an Excel sheet for analysis.<sup>18</sup> REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing: 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources.

### Data analysis

This study was a cross-sectional community-based survey. The analyses were conducted on datasets downloaded from REDCap to an Excel spread sheet. A binary variable was used to represent the dependent variable, malaria infection, to show whether malaria was present (positive) to RDT or absent (negative) was used.

The explanatory variables analyzed were the following sociodemographic factors: age, if the person was an adult or child, age category (0 to 4, 5 to 14, 14 to 24 and <24), sex (male and female), history of malaria treatment, if the person had paid employment, cell phone ownership, education level, population density of the neighborhood, location (neighborhood), household category or type (hut or conventional) and household size.

The malaria prevalence, was calculated by dividing positive cases of malaria by the study population tested at the time multiplied by 100.<sup>19</sup>

$$\text{Prevalence (\%)} = \frac{\text{Persons having malaria}}{\text{Tested during the period}} \times 100 \quad (1)$$

Chi-square for proportion of age group and sex was tested. To establish the relationship between malaria prevalence and sociodemographic factors, logistic backward stepwise logistic regression was used with the following model:

$$X_i : G(P_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \quad (2)$$

Where:  $G(P_i)$  = link function

$P_i$  = likelihood of response for the -i-th factor

$\beta_0$  = intercept

$\beta_i$  = coefficient

$x_i$  = independent variable.

This method starts with a full (saturated) model and each step gradually eliminates variables that do not contribute. Allowing for a reduced model that best explains the data. This method is useful since, it reduces the number of predictors, reducing multicollinearity and resolves overfitting.<sup>20</sup>

To test the goodness of fit for the model, the Hosmer–Lemeshow (1989) test was performed.<sup>21</sup> To build the final model, the independent variables  $p < 0.05$  were included. Outcomes such as scores statistic's, regression coefficient's, significance levels of variable coefficients and, overall classification accuracy were performed.

The sensitivity (conditional probability of a positive test given that the patient has malaria) of the final model measures the proportion of positive that were correctly identified and, was calculated using<sup>22</sup>:

$$\text{Sensitivity}(\%) = \frac{\text{Number of true malaria positive}}{(\text{Number of true malaria positive} + \text{Number of false malaria negative})} \times 100 \quad (3.1)$$

To measure performance of the binary model, sensitivity and specificity tests carried out. The specificity (conditional probability of a negative test given that the patient is well) of the final model measures the proportion of negative case correctly identified and was calculated using<sup>21</sup>:

$$\text{Specificity}(\%) = \frac{\text{Number of true negatives}}{(\text{Number of true malaria negatives} + \text{Number of false malaria positives})} \times 100 \quad (3.2)$$

Positive predictive value (PPV) that is, the conditional probability, whether the screened people who tested positive do or do not actually have malaria was calculated using<sup>22</sup>:

$$\text{PPV}(\%) = \frac{\text{Number of true malaria positive}}{(\text{Number of true malaria positive} + \text{Number of false malaria positive})} \times 100 \quad (3.3)$$

Negative predicted value (NPV) that is, the conditional probability that an individual with a test indicative of no malaria infection is actually disease free, was calculated using<sup>22</sup>:

$$\text{NPV}(\%) = \frac{\text{Number of true malaria negatives}}{(\text{Number of true malaria negatives} + \text{Number of false malaria negative})} \times 100 \quad (3.4)$$

All tests were carried out using SPSS IBM version 20 (IBM Corporation, Armonk, New York, USA) (RRID: SCR\_002865).<sup>23</sup>

## Results

### Malaria prevalence, sex, age and, age group and education level of participants

From 125 selected households 100 were visited [Figure 3](#) presents the positive and negative cases per visited site. Of the 358 participants tested and, interviewed 108 (31.6%) tested positive for malaria. There was an equal distribution of the enrolled participants among sex, 55% were female and 45% males, Chi-squared = 1.28,  $P = 0.2578$ , Degree of freedom (DF) = 1.

The age of participants varied from 1 to 80 years old, with a median of 17 years and an average of 21 standard deviation (SD), 16.2 years old. The participants' education level varied, where 35.1% had no education or less than primary (5 grades), 47.4% had primary or basic school (grades 5 to 10) and 17.5% had secondary and higher education.

### Malaria prevalence by age category

[Figure 4](#) presents the malaria positivity results for age categories. Half of the malaria positive cases occurred among those 5 to 14 years age category. This category comprises has 32.7% of the Sussundenga population according to the National Institute of Statistics (INE). The age category of over 24 years presented 17.6% of the malaria cases, this age category comprises 30.4% of the Sussundenga population according to the INE. There was a statistically significant difference in positive malaria cases among groups, Chi-squared = 25.857,  $P = 0.0022$ , DF = 9.

### Association between malaria infection and sociodemographic factors

The backward stepwise regression selection of predictors into the binary logistic model produced a series of models and, in this study, we only present the relevant, initial models and other outputs can be found in appendix 1.

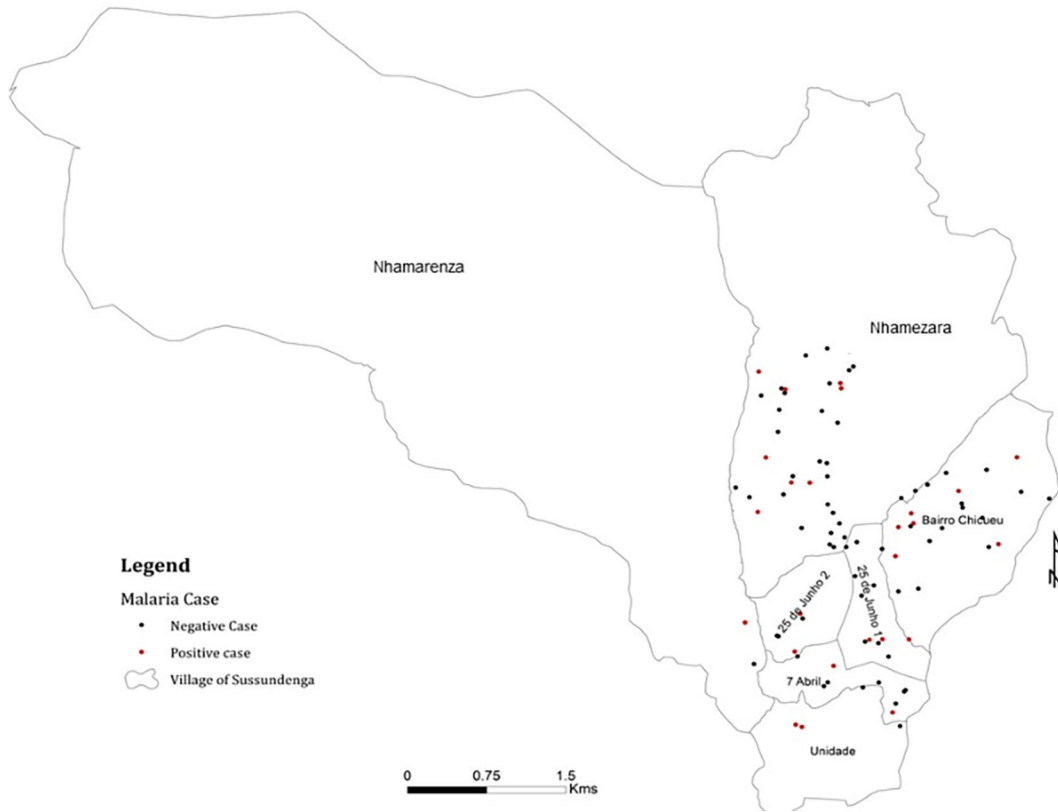


Figure 3. Malaria positive and negative cases in Sussundenga village.

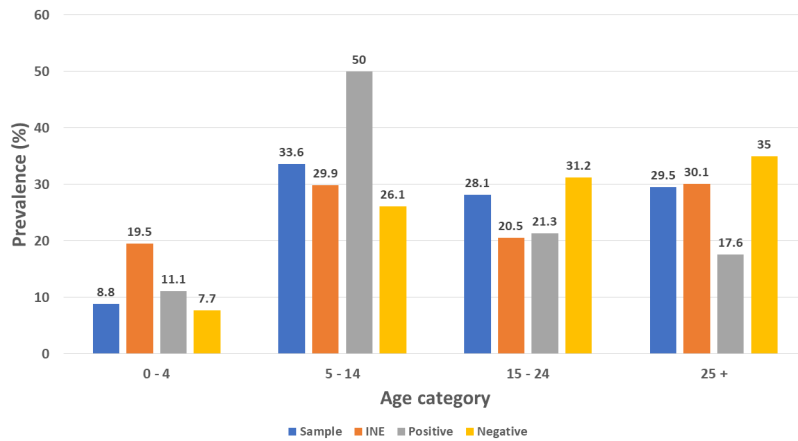


Figure 4. Malaria prevalence by age group in Sussundenga Village, INE = National Institute of Statistics.

Table 1. Backward stepwise model summary.

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	408.482 <sup>a</sup>	.109	.151
9	413.304 <sup>b</sup>	.096	.135

<sup>a</sup>Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

<sup>b</sup>Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

**Table 1** presents the backward stepwise (Wald) model summary and the Nagelkerke’s R<sup>2</sup> in final step is 0.135. This suggests that presence of malaria variation shown in the dependent variable of this model is approximately 13.5%.

**Table 2** presents the Hosmer and Lemeshow test, indicating that this model fit the data.

**Table 3** presents the classification table of the final model, that is, the models capability to predict malaria positive cases, indicating a model accuracy of 71.6%. The sensitivity of the final model in classifying malaria positive cases is 73.3% and specificity of the final model to classify malaria negative cases is 93.3%. The positive predictive value is 66% and, the negative predictive value is 72.5% meaning that, the final model is able to predict 66% of malaria positive tests and, 72.5% negative malaria tests.

**Table 4** presents the Wald’s test of significance and the odds ratio predictors variables in the final model. From the results, pervious malaria treatment ( $p=0.15$ ), population density ( $p=0.05$ ), and age group ( $p=0.00$ ) were significant predictors

**Table 2. Hosmer-Lemeshow test.**

Step	Chi-squared	DF	Sig.
1	8.558	8	.381
9	5.990	8	.648

Df = degrees of freedom, Sig. = Wald’s.

**Table 3. Final backward stepwise (Wald) model classification table.**

Observed			Predicted		
			Malaria results		Percentage correct
			Negative	Positive	
Step 1	Malaria result	Negative	218	22	90.8
		Positive	77	39	33.6
	Overall percentage				72.2
Step 9	Malaria result	Negative	224	16	93.3
		Positive	85	31	26.7
	Overall percentage				71.6

Wald = Wald test.

<sup>a</sup>The cut-off value is .500.

**Table 4. Final model Wald’s of significance and odds ratio of predictor variables.**

		Constant (B)	S.E.	Wald	DF	Sig.	Exp (B)	95% CI for EXP(B)	
								Lower	Upper
Step 9 <sup>a</sup>	Previous malaria treatment	-.607	.249	5.941	1	.015	.545	.335	.888
	Population density	-.0001	.000	3.830	1	.050	1.000	1.000	1.000
	Household category	-.601	.315	3.651	1	.056	.548	.296	1.016
	Age category			18.890	3	.000			
	Age category (0 to 4 years)	1.040	.458	5.155	1	.023	2.829	1.153	6.944
	Age category (5 to 14 years)	1.289	.317	16.573	1	.000	3.631	1.952	6.755
	Age category (> 14 years)	.472	.339	1.934	1	.164	1.603	.824	3.117
	Constant	-.821	.305	7.232	1	.007	.440		

B = regression coefficients, S. E = standard errors, Wald = Wald test, Df = degrees of freedom, Sig. = Wald’s significance, Exp(B) (OR = odds ratio, 95% CI = confidence interval of the odds ratio).

<sup>a</sup>Variable(s) entered on step 1: Adult or child, Sex, Previous malaria treatment, Employment, Age, Cell phone, Education, Population density, Household size, HH category, Age category, Location.



while, household category did not add significantly to the model. The table indicates that the age category 0 to 4 years old as almost three times more likely to test positive for malaria (OR 2.829, 95% CI 1.153–6.944), 3.6 times (OR 3.61, 95% CI 1.952 – 6.755) for age group 5 to 14 years and, 1.6 times for the age group of 15 or older (OR 1.603, 95% CI 0.824 – 3.117).

The built model is:

$$G(Pi) = - .821 - .607 \text{ Previous malaria treatment} - 0.0001 \text{ Population density} \\ + 1.040 \text{ Age category (0 to 4 years)} + 1.289 + (\text{Age category (5 to 14 years)}).$$

## Discussion

In this study, malaria prevalence was 31.6% for Sussundenga Village, much higher than the prevalence recorded in Chimoio city (20.1%).<sup>24</sup> In the neighboring Zimbabwe, malaria prevalence was 19.5% in Mutare and 50.9% in Mutasa districts in 2016.<sup>25</sup> In southern Zambia a study in 2020, reported parasite prevalence between 0.7 and 1.8%<sup>26</sup> and, 34% in Malawi in 2016.<sup>27</sup>

No significant difference was found between different sexes in this study. Similar results were reported in Chimoio, Mozambique in 2018,<sup>24</sup> in Malawi in 2020<sup>28</sup> and in Zimbabwe<sup>29</sup> in 2021.

This study recorded half of the malaria prevalence in the 5 to 14 years age category and, an odds ratio of 3.61. In Ghana this age groups accounted for 43.3% and, in Rwanda the odds of infection by malaria were reported to be 1.817 times for this age category.<sup>30,31</sup> Studies in Kenya indicated that highest malaria prevalence occurs in children between ages of 11 to 14 and, children from 5 to 18 years as the most at-risk age category.<sup>32,33</sup> Contrarily, in Chimoio, Mozambique it was reported 52% of malaria cases are found in children under five,<sup>24</sup> this discrepancy may due to the fact that the present study was carried out at community level while, the Chimoio study was carried out from health center data.

This study suggests that recent diagnosis and treatment for malaria infection reduces the odds of subsequent infection approximately by 54.5%. Similar results were reported in Mozambique, Ghana, Comoros, Kenya, Indonesia and India.<sup>34–39</sup> This reduction in odds is likely due to prophylactic effect of ACT. It provides protection from 2 weeks to 1 month after completion. After repeated infections, the individual develops a certain degree of immunity. Also, when re-infected, patients tend to present a mild form of the diseases without symptoms and, natural active immunity is established after ten or more *P. falciparum* infections, which can be sufficient to suppress symptoms and clinical signs.<sup>40</sup>

Different results were reported in Angola where women who had a previous malaria infection during pregnancy also had a higher risk to contract malaria.<sup>41</sup> This is likely because pregnant women may take sulfadoxine-pyrimethamine rather than ACT.

In this study population density was found as a significant predictor for an individual to test positive for malaria. Similar results were reported in Chimoio<sup>24</sup> in 2016, in a study in 14 endemic African countries<sup>42</sup> in 2017 and in Ethiopia<sup>43</sup> in 2015.

The variables age, if the person was an adult or child, sex, paid employment, cell phone ownership, education level, location (Bairro) and household size were removed from the model due to redundancy and for not adding significance to the model.

The age category is a good proxy for age group and, household size for household category. Paid employment and cell phone ownership variables were included in this study, as rural wealth indicators. These were not found significant predictors contrary to a study in Mozambique that indicated that, children from higher income families (58%) tend to be at lower risk for malaria compared to children from lower income families (43%).<sup>44</sup> Another study in sub-Saharan Africa<sup>45</sup> showed that, malaria prevalence increases with a decrease in income in 2018.

Education level was not finding significant predictor in this study. Similar results were reported in Malawi in 2018,<sup>46</sup> Indonesia and India.<sup>38,39</sup> There were conflicting results reported in Mozambique<sup>47</sup> in 2011, in Ghana in 2014<sup>30</sup> and in Sub-Saharan Africa<sup>45</sup> in 2018.

In this study it is suggested that approximately 13.5% of the variation in malaria infection can be attributed to sociodemographic and economic traits. A previous study modelled the influence of climate on malaria occurrence in

Chimoio and indicated that environmental traits accounted for 72.5% of malaria occurrences.<sup>13</sup> This implies that non-environmental factors such as sociodemographic, economic, cultural and behavioral traits could partially account for the remaining percentage, consistent with the present study. Environmental factors related to malaria cases were reported in Burundi (82%), in Nigeria (66%) and by Global Fund (90%).<sup>11-13</sup>

The capability model using social, economic, and demographic variables to predict malaria positive cases (model accuracy), was 72.3% in this study. A logistic regression model analyzing hematological parameter and age in Ghana reported 77.4%.<sup>30</sup> The sensitivity of the final model in classifying malaria positive cases was 73.3% and the final model was able to predict 66% (PPV) meaning that the model is very effective in predicting malaria infection using socio-demographic characteristics. In Iran a model predicting malaria re-introduction reported 81.8% positive predictive value<sup>39</sup> and 52.72% in Ghana in a model analyzing hematological parameter and age.<sup>30</sup>

### Limitations of the study

Data collection for this study was conducted in December and January during the rainy and wet season which is also the peak malaria transmission season. Because of this, it is likely that we detected a large number of infections and results reflect this season and may not be representative of malaria dynamics in the dry season. The RightSign Biotest<sup>®</sup> test detects the histidine rich protein 2 antigen of the *P. falciparum* parasite which can last over a month in the blood among patients recently treated with malaria.

### Conclusion

This study evaluated the sociodemographic factors that affect malaria prevalence in Sussundenga Village, Mozambique. Recent diagnosis and treatment, population density and age category were found to be significant predictors. The model accuracy was 72.3% implying that the model is robust. Targeting malaria control at the community level can contribute to decreased transmission that may be more impactful than passive case detection and treatment alone. With the age shift in malaria cases, targeting malaria control at the community level and, involving the entire community, not only children and pregnant women.<sup>48</sup> This model indicates that 13.5% of malaria cases can be attributed to sociodemographic factors while previous studies indicated that environmental conditions are attributed to approximately 73% of malaria cases. Further studies are needed especially in the dry season and in other areas of the district to fully understand the malaria transmission dynamics in this region and inform efficient control measures.

### Data availability

#### Underlying data

Harvard Dataverse: Replication Data for: Modelling sociodemographic factors that affect malaria prevalence in Sussundenga, Mozambique: a cross-sectional study. <https://doi.org/10.7910/DVN/BUMDEM>.<sup>49</sup>

This project contains the following underlying data:

- [Additional file -F1000Research.tab] (raw data from questionnaires).

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

### Ethical consideration

This study is part of the Malaria Risk, Prevention, and Health Seeking Behaviors in Sussundenga, Mozambique Project. All participants, or the guardians provided informed written assent and consent prior to participation. Ethical review and approval for the study was completed by the Institutional Review Board (IRB) at the University of Minnesota [STUDY00007184] and from A Comissão Nacional de Bioética em Saúde (CNBS) at the Ministry of Health of Mozambique [IRB00002657]. The images taken from google and other sites were properly cited and referenced.

### Acknowledgments

We would like to thank the provincial and district directorates of health to grant permission to carry out this study especially the Dr. Firmino Jaqueta, Dr. Serafina Benesse, Dr. Filipe Murgorgo and Mrs. Elsa Trabuco. We also thank Mr. Gabriel Viegas for editing figures. A preprint version of this article can be found on Research Square (DOI: 10.21203/rs.3.rs-614728/v1).

## References

1. Centers for Disease Control and Prevention: *About Malaria-Biology*. Centers for Disease Control and Prevention; 2012. Accessed July 14, 2020.  
[Reference Source](#)
2. World Health Organization: *The World malaria report 2018*. WHO; 2018. Accessed July 27, 2020.  
[Reference Source](#)
3. World Health Organization: *Making strides against malaria in Mozambique: Control, prevention and treatment in action*. WHO; 2020. Accessed July 14, 2020.  
[Reference Source](#)
4. World Health Organization: *Mozambique 2018*. WHO; 2018. Accessed July 14, 2020.  
[Reference Source](#)
5. DPS: *Relatório Do Primeiro Semestre Do Ano 2020*. Governo da Província de Manica; 2020.
6. Chuang TW, Soble A, Ntshalintshali N, et al.: **Assessment of climate-driven variations in malaria incidence in Swaziland: toward malaria elimination**. *Malar. J.* 2017; **16**: 232.  
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Manh BH, Clements ACA, Thieu NQ, et al.: **Social and environmental determinants of malaria in space and time in Viet Nam**. *Int. J. Parasitol.* 2011; **41**: 109–116.  
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Instituto Nacional de Estatística: *Inquérito Demográfico e de Saúde*. INE; 2011. Accessed July 14, 2020.  
[Reference Source](#)
9. Edwin P, Msengwa A: **Prevalence and socio-demographic factors associated with malaria infection among children under five years in Tanzania**. *Jour. Pub. Helt. Epid.* 2018; **10**: 387–394.  
[Publisher Full Text](#)
10. Gomez-Elipe A, Otero A, van Herp M, et al.: **Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997–2003**. *Malar. J.* 2007; **6**: 129.  
[Publisher Full Text](#)
11. Kazembe LN, Kleinschmidt I, Sharp BL: **Patterns of malaria-related hospital admissions and mortality among Malawian children: an example of spatial modelling of hospital register data**. *Malar. J.* 2006; **5**: 93.  
[Publisher Full Text](#)
12. Global Fund. **Invest. In the Future, Defeat Malaria**. World Malaria. 2015. Accessed 10 April 2022.  
[Reference Source](#)
13. Ferrão JL, Mendes JM, Painho M: **Modelling the influence of climate on malaria occurrence in Chimoio Municipality, Mozambique**. *Parasit. Vectors.* 2017; **10**: 260.  
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Cenacarta: **Cartography and Remote Sensing Center**. Cenacarta, 2020.  
[Reference Source](#)
15. INE: **Projeções Anuais da População Total, Urbana e Rural, dos Distritos, 2007-2040 (volumes correspondentes às 11 províncias)**. INE. 2010.  
[Reference Source](#)
16. Ministério da Administração Estatal: *Perfil Do Distrito De Sussundenga*. INE; 2005. Accessed July 12, 2020.  
[Reference Source](#)
17. Google: **Google Earth Pro (Version 7.3)**. CA, USA, 2019.
18. Harris PA, Taylor R, Thielke R, et al.: **Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support**. *J. Biomed. Inform.* 2009, **42**.
19. Center for Disease Control: *Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction*. CDC; editor. CDC; Third 2006.  
[Reference Source](#)
20. de Oliveira EC, dos Santos ES, Zeilhofer P, et al.: **Geographic information systems and logistic regression for high-resolution malaria risk mapping in a rural settlement of the southern Brazilian Amazon**. *Malar. J.* 2013; **12**.  
[Reference Source](#) | [Publisher Full Text](#)
21. Laerd Statistics: **Binomial Logistic Regression using SPSS Statistics**.  
[Reference Source](#)
22. Trevethan R: **Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice**. *Front. Public Health.* 2017; **5**.  
[PubMed Abstract](#) | [Publisher Full Text](#)
23. IBM, SPSS, version 20. **IBM SPSS software**. IBM Corporation, Armonk, New York, USA.
24. Ferrão JL, Mendes JM, Painho M, et al.: **Spatio-temporal variation and socio-demographic characters of malaria in Chimoio municipality, Mozambique**. *Malar. J.* 2016; **15**: 329.  
[PubMed Abstract](#) | [Publisher Full Text](#)
25. Sande S, Zimba M, Chinwada P, et al.: **A review of new challenges and prospects for malaria elimination in Mutare and Mutasa Districts, Zimbabwe**. *Malar. J.* 2016; **15**: 360.  
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Bhondokhan FRP, Searle KM, Hamapumbu H, et al.: **Improving the efficiency of reactive case detection for malaria elimination in southern Zambia: A cross-sectional study**. *Malar. J.* 2020; **19**: 175.  
[PubMed Abstract](#) | [Publisher Full Text](#)
27. Mathanga DP, Tembo AK, Mzilahowa T, et al.: **Patterns and determinants of malaria risk in urban and peri-urban areas of Blantyre, Malawi**. *Malar. J.* 2016; **15**: 590.  
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Chilanga E, Collin-Vézina D, MacIntosh H, et al.: **Prevalence and determinants of malaria infection among children of local farmers in Central Malawi**. *Malar. J.* 2020; **19**: 308.  
[PubMed Abstract](#) | [Publisher Full Text](#)
29. Mundagowa P, Chimberengwa PT: **Poor housing construction is associated with contracting malaria in a rural area south of Zimbabwe: A Case-control Study**. *Malar. J.*  
[Publisher Full Text](#)
30. Paintsil EK, Omari-Sasu AY, Addo MGBM: **Analysis of Haematological Parameters as Predictors of Malaria Infection Using a Logistic Regression Model: A Case Study of a Hospital in the Ashanti Region of Ghana 2019**. *Malar. Res. Treat.* 2019; 2019: 1–7.  
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Zeleke GT, Egide H, Francois T: **Application of Logistic Regression Model to Identify Potential Risk Factors of Malaria in Rwanda using 2010 Demographic and Health Survey**. *IJASM.* 2015; **2**(3): 50–54.  
[Reference Source](#)
32. Sultana M, Sheikh N, Mahumud RA, et al.: **Prevalence and associated determinants of malaria parasites among Kenyan children**. *Trop Med Health.* 2017 Oct 23; **45**: 25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
33. Nankabirwa J, Brooker SJ, Clarke SE, et al.: **Malaria in school-age children in Africa: an increasingly important challenge**. *Tropical Med. Int. Health.* 2014 Nov; **19**(11): 1294–1309.  
[PubMed Abstract](#) | [Publisher Full Text](#)
34. World Bank: **Inquérito Nacional sobre Indicadores de Malária 2018**. 2019.  
[Reference Source](#)
35. Di Gennaro F, Marotta C, Pizzol D, et al.: **Prevalence and predictors of malaria in human immunodeficiency virus infected patients in beira, mozambique**. *Int. J. Environ. Res. Public Health.* 2018; **15**(9): 2032.  
[PubMed Abstract](#) | [Publisher Full Text](#)
36. Ndong IC, Okyere D, Enos JY, et al.: **Prevalence of asymptomatic malaria parasitaemia following mass testing and treatment in Pakro sub-district of Ghana**. *BMC Public Health.* 2019; **19**: 1622.  
[PubMed Abstract](#) | [Publisher Full Text](#)
37. Artadji A: **Recul et persistance du paludisme en Union des Comores: une approche géographique pour déterminer l'importance des facteurs environnementaux et sociaux dans son maintien**. *Géographie Univ la Réunion.* 2019.  
[Publisher Full Text](#)
38. ERNST KC, Lindblade KA, Koech D, et al.: **Environmental, socio-demographic and behavioural determinants of malaria risk in the western Kenyan highlands: A case-control study**. *Tropical Med. Int. Health.* 2009; **14**(10): 1258–1265.  
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Hasyim H, Dale P, Groneberg DA, et al.: **Social determinants of malaria in an endemic area of Indonesia**. *Malar. J.* 2019; **18**(1): 134.  
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Ranjbar M, Shoghli A, Kolifarhood G, et al.: **Predicting factors for malaria re-introduction: An applied model in an elimination setting to prevent malaria outbreaks**. *Malar. J.* 2016; **15**(1): 138.  
[PubMed Abstract](#) | [Publisher Full Text](#)
41. Mioto LD, Ligia Carla Faccin Galhardi MKA: **Aspectos parasitológicos e imunológicos da malária**. *Bios.* 2012; **14**(1): 42–55.  
[Reference Source](#)
42. Campos PA, Valente B, Campos RB, et al.: **Plasmodium falciparum infection in pregnant women attending antenatal care in Luanda, Angola**. *Rev Soc Bras Med Trop.* 2012; **45**(3): 369–374.  
[PubMed Abstract](#) | [Publisher Full Text](#)

43. Kabaria CW, Gilbert M, Noor AM, *et al.*: **The impact of urbanization and population density on childhood *Plasmodium falciparum* parasite prevalence rates in Africa.** *Malar. J.* 2017; **16**: 49.  
[Publisher Full Text](#)
44. Bouma MJ, SantosVega M, Yeshiwondim AK, *et al.*: **Temperature and population density determine reservoir regions of seasonal persistence in highland malaria.** *Proc. R. Soc. B.* 2015; **282**: 20151383.  
[Publisher Full Text](#)
45. Instituto Nacional de Estatística: *Inqueriro Sociodemografico e de Saúde 2011*. INE; 2012.  
[Reference Source](#)
46. Degarege A, Fennie K, Degarege D, *et al.*: **Improving socioeconomic status may reduce the burden of malaria in sub-Saharan Africa: A systematic review and meta-analysis.** *PLoS One.* 2019; **14**.  
[Publisher Full Text](#)
47. Kabaghe AN, Chipeta MG, Gowelo S, *et al.*: **Fine-scale spatial and temporal variation of clinical malaria incidence and associated factors in children in rural Malawi: a longitudinal study.** *Parasit. Vectors.* 2018; **11**: 129.  
[Reference Source](#) | [Publisher Full Text](#)
48. Galatas B, Saúte F, Martí-Soler H, *et al.*: **A multiphase program for malaria elimination in southern Mozambique (the Magude project): A before-after study.** *PLoS Med.* 2020; **17**(8).  
[Publisher Full Text](#)
49. Ferrao J: **Replication Data for: Modelling sociodemographic factors that affect malaria prevalence in Sussundenga, Mozambique: a cross-sectional study.** Harvard Dataverse, V1, UNF:6:AOiZHf6kq1bQfkIDHoh+MA=[fileUNF]. 2021.  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 19 May 2022

<https://doi.org/10.5256/f1000research.133640.r136728>

© 2022 Laporta G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Gabriel Zorello Laporta** 

Graduate Research and Innovation Program, Centro Universitario FMABC, Santo André, Brazil

I want to thank the authors by the responses that clarified my doubts about this intellectual piece.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Epidemiology of vector-borne diseases

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 17 May 2022

<https://doi.org/10.5256/f1000research.133640.r136727>

© 2022 Cameron E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Ewan Cameron**

School of Public Health, Faculty of Health Sciences, Curtin University, Perth, WA, Australia

I thanks the authors for their detailed and considerate responses to my first round reviewer questions/comments. I am happy to recommend the revised article for indexing.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Malaria risk stratification; statistics; mechanistic disease modelling; model calibration

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

Version 1

Reviewer Report 14 April 2022

<https://doi.org/10.5256/f1000research.79031.r129856>

© 2022 Laporta G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Gabriel Zorello Laporta**

Graduate Research and Innovation Program, Centro Universitario FMABC, Santo André, Brazil

Ferrao *et al.* provide us with a cross-sectional study aimed at estimation of malaria falciparum prevalence in Mozambique, Africa. The study is well written and has good data, but some parts are not clear enough, as follows:

1. Study sample size and figure 2. Please specify the randomization process to select 125 households in the study area. What is the statistical power of the selection of 100 houses? If the study area's landscape is heterogenous in terms of risk of malaria infection, is a random selection the best approach for selection? Should a stratified sample selection approach be used instead?
2. To estimate sensitivity and specificity, it is essential to have a training dataset and a testing dataset. The training dataset is to build a statistical model and the testing dataset is to evaluate the fitted model. Please specify the training and the testing data.
3. The built model shows that access to treatment and age are the only important predictors. Please explain the lack of importance of social and economic predictors in the built model. The built model shows a coefficient (1.289) lacking a variable. Please revise.
4. RDT based on HRP2 has issues to detect falciparum lacking HRP2 genes. This is an important limitation and indicates that the estimates of prevalence may be underestimated.
5. In data source, please include labels to the variables' levels per variable, i.e., 1=, 0=.
6. Conclusions have several issues:

"Recent diagnosis and treatment, population density and age category were found to be significant predictors". Issue: pop density was not significant predictor.

"The model accuracy was 72.3% implying that the model is robust". Issue: it depends on the approach that it was calculated.

“This model indicates that 13.5% of malaria cases can be attributed to sociodemographic factors while previous studies indicated that environmental conditions are attributed to approximately 73% of malaria cases”. Issue: be specific of which sociodemographic factor, and environmental conditions were not studied in this study.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

Partly

**Are all the source data underlying the results available to ensure full reproducibility?**

Partly

**Are the conclusions drawn adequately supported by the results?**

No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Epidemiology of vector-borne diseases

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 29 Apr 2022

**Joao Ferrao**, UnISCED, Beira, Mozambique

Dear Reviewer. thank you very much for your precious comments. They are very useful and they were used to improve the manuscript.

- Ferrao *et al.* provide us with a cross-sectional study aimed at estimation of malaria falciparum prevalence in Mozambique, Africa. The study is well written and has good data, but some parts are not clear enough, as follows:

Study sample size and figure 2.

Please specify the randomization process to select 125 households in the study area., what is the statistical power of the selection of 100 houses? If the study area's landscape is heterogenous in terms of risk of malaria infection, is a random selection

the best approach for selection? Should a stratified sample selection approach be used instead?

**Response:** Thanks for the important question raised. Indeed, this was a pilot study to determine malaria prevalence, risk factors, and health seeking behaviors. The sample size was determined by feasibility for the study team and study design of the community based cross-sectional survey.

All households in the study area were digitized and enumerated using Google Earth Pro. A random sample of 125 households was taken, with the aim of enrolling 100 to account for errors in the digitizing process.

The village is relatively small (156.9 Km<sup>2</sup>) and we added the area of the village in the test. The Sussundenga village is within an area of 156.9 Km<sup>2</sup>.

- To estimate sensitivity and specificity, it is essential to have a training dataset and a testing dataset. The training dataset is to build a statistical model and the testing dataset is to evaluate the fitted model. Please specify the training and the testing data.

**Response:** Thank you very much for the raised question. Usually the prediction of classes for data classification are based on finding the optimum boundary between classes.

For this case where we used an accuracy with cut-off=0.5, we don't see the need of training data.

After data imputation and engendering feature the third step was to split data into train and test. This was carried out by the software.

- The built model shows that access to treatment and age are the only important predictors. Please explain the lack of importance of social and economic predictors in the built model.

**Response:** Thank you very much for the question. Several studies indicated that the major causes of malaria occurrence are the climatic conditions specially temperature bellow 20°C the sporogony cease, and humidity below 50 and over 90 %.

- The built model shows a coefficient (1.289) lacking a variable. Please revise.

**Response:** Thank you very much for the observation. This was revised, the variable is Age category (5 to 14 years).

- RDT based on HRP2 has issues to detect falciparum lacking HRP2 genes. This is an important limitation and indicates that the estimates of prevalence may be underestimated.

**Response:** This is true and a limitation of the current HRP2 based RDTs. There are limited data on HRP2 deletions throughout Mozambique and specifically in Manica Province. However, in a study published in 2019 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6711899/>) the authors found very few cases of HRP2 deletion and impacts on the efficacy of the current RDTs. Of those infections not



detected by the RDTs, not all were *P. falciparum*. It is unlikely in our study setting that HRP2 deletions impacted the efficacy of the RDT and biased our prevalence estimates.

- In data source, please include labels to the variables' levels per variable, i.e., 1=, 0=,II)

- **Conclusions have several issues:**

"Recent diagnosis and treatment, population density and age category were found to be significant predictors". Issue: pop density was not significant predictor.

**Response:** Thank you for the observation. Rescaled to 3 decimal places its significant. This was corrected in table.

- "The model accuracy was 72.3% implying that the model is robust". Issue: it depends on the approach that it was calculated.

**Response:** Thanks for the observation.

- "This model indicates that 13.5% of malaria cases can be attributed to sociodemographic factors while previous studied indicated that environmental conditions are attributed to approximately 73% of malaria cases". Issue: be specific of which sociodemographic factor, and environmental conditions were not studied in this study.

**Response:** Thank you very much for the observation. In the manuscript, more text was added to address this issue.

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 01 April 2022

<https://doi.org/10.5256/f1000research.79031.r127464>

© 2022 Cameron E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Ewan Cameron

School of Public Health, Faculty of Health Sciences, Curtin University, Perth, WA, Australia

In the manuscript entitled, "Modelling sociodemographic factors that affect malaria prevalence in Sussundenga, Mozambique: a cross-sectional study", the authors present the results of a survey and statistical analysis designed to uncover predictors of positive parasite status by rapid diagnostic test in this area of endemic malaria transmission. Malaria is a significant contributor to disease burden in Mozambique, so this topic is of high importance. In my report below I raise a number of questions for the authors regarding aspects of the analysis that were unclear to me (i.e., perhaps where the clarity of presentation could be improved) or where I felt that further analysis might strengthen the conclusions.

### Context and Aims:

- The framing of the study in terms of context and aims as it currently stands deserves considerable review. Owing precisely to its status as one of the most malarious countries in the world, Mozambique has attracted a great deal of attention from malaria researchers over many decades and there have been many studies investigating environmental and socio-demographic factors behind malaria transmission in the country. For example, Temu *et al.* (Plos One, 2012)<sup>1</sup> identified IRS use, pig-keeping, and house construction as key factors influencing malaria risk in Zambezia province. See also, Brentlinger *et al.* (Am J Trop Med Hyg, 2007)<sup>2</sup>. Likewise, there are many studies examining factors related to explanatory factors for malaria prevalence such as treatment seeking (Cassy *et al.* Mal J, 2019)<sup>3</sup> and ITN use (Moon *et al.*, Mal J, 2016)<sup>4</sup>. Factors shaping malaria risk in Mozambique have also recently been studied in the context of measuring the impact of malaria control strategies in Mozambique by Galatas *et al.* (Plos Medicine, 2020)<sup>5</sup>.

This wealth of prior research and understanding contrasts strikingly with the authors' proposition that "little is known about the epidemiology to inform appropriate and effective interventions". Importantly, it also raises the question of why information on two well-known factors shaping malaria transmission in Mozambique, namely IRS and ITN use, do not seem to have been captured or included in this study? Another key known factor, household construction, may have been included but there is no detail that I could see in the manuscript to illuminate the meaning of the "household category" variable used in the model?

- Regarding the statement "ITNs are only available at antenatal clinics, indicated for pregnant women and children under five": I'm not sure that this is well phrased, since the WHO recommends universal net use in high transmission areas. While ITN distribution campaigns focus on the highest risk groups of young children and pregnant women these are not the only groups who should be advised to use bed nets; likewise, ITNs would generally be available commercially at markets. (See e.g. Scott *et al.* Mal J, 2021)<sup>6</sup>.

### Statistical Analysis:

- The statistical analysis method used to derive the primary results of this study, namely the identification of key factors behind malaria prevalence in the study area, is a stepwise logistic regression, which is indeed appropriate for this objective. Some minor details require clarification or revision:

(i) there are some minor formatting errors in the equations: e.g. on page 5, equation 2 uses lower case  $g$  and then an upper case  $G$  for seemingly the same function, and  $\beta_{ai}$  should have a subscript  $i$  as in  $\beta_{a_i}$ ; on page 8 the last equation should have 1.289 "x" Age category (5 to 14 years) instead of "+"

(ii) "To evaluate potential confounders and, effect modifiers between the final model variables, the Hosmer-Lemeshow (1989) test was performed." This doesn't make sense to me: the HM test is for model specification / acceptable fit, rather than for breaking variables down into their roles in the causal hierarchy.

(iii) I was confused by the chi square test reports in some places: for the distribution by sex

the chi-squared statistic of 0.081 doesn't sound like the right order of magnitude and in fact I get 0.081 as the p-value for a binomial exact test on this sample so perhaps this is a typo?; for the tests by age category, since this is a four x two table I would have thought we're looking at 3 degrees of freedom rather than 6?

(iv) The population density variable must have a large dynamic range because it is assigned a slope ("constant"?) of 0.000 in table 8: could this be rescaled so that we can see its slope within the 3 decimal places?

(v) I was confused why the age categories changed from four in the earlier discussion to three in Table 4? Also, it would help to nominate one age group as the reference group so that the odds ratios can be understood as relative to that group.

(vi) I'm confused by the focus on understanding the predictive accuracy of the model in terms of specificity and sensitivity(\*), which are appropriate for a diagnostic tool, but which may not be particularly relevant to the use of a risk factor model such as this one for field epidemiology. I.e., if the end use is to prioritise the delivery of a particular intervention such as seasonal malaria chemopraxis then identifying that a certain age group has twice the parasite prevalence of another could be of substantial value even where the sensitivity was low because prevalence itself was very low across both strata. This comes back to the context and aims of the study, in the sense that the value of the fitted model (or more precisely the knowledge discovered through it) is ultimately something that exists only relative to the way in which it is intended to be used.

(\*also: are these defined according to a thresholding of the predictive prevalence above and below 50%?)

## References

1. Temu EA, Coleman M, Abilio AP, Kleinschmidt I: High prevalence of malaria in Zambezia, Mozambique: the protective effect of IRS versus increased risks due to pig-keeping and house construction. *PLoS One*. 2012; **7** (2): e31409 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Brentlinger PE, Dgedge M, Correia MA, Rojas AJ, et al.: Intermittent preventive treatment of malaria during pregnancy in central Mozambique. *Bull World Health Organ*. 2007; **85** (11): 873-9 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Cassy A, Saifodine A, Candrinho B, Martins MDR, et al.: Care-seeking behaviour and treatment practices for malaria in children under 5 years in Mozambique: a secondary analysis of 2011 DHS and 2015 IMASIDA datasets. *Malar J*. 2019; **18** (1): 115 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Moon TD, Hayes CB, Blevins M, Lopez ML, et al.: Factors associated with the use of mosquito bed nets: results from two cross-sectional household surveys in Zambézia Province, Mozambique. *Malar J*. 2016; **15**: 196 [PubMed Abstract](#) | [Publisher Full Text](#)
5. Galatas B, Saúte F, Martí-Soler H, Guinovart C, et al.: A multiphase program for malaria elimination in southern Mozambique (the Magude project): A before-after study. *PLoS Med*. **17** (8): e1003227 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Scott J, Kanyangarara M, Nhama A, Macete E, et al.: Factors associated with use of insecticide-treated net for malaria prevention in Manica District, Mozambique: a community-based cross-sectional survey. *Malar J*. 2021; **20** (1): 200 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the work clearly and accurately presented and does it cite the current literature?**

Partly

**Is the study design appropriate and is the work technically sound?**

Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Malaria risk stratification; statistics; mechanistic disease modelling; model calibration

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 29 Apr 2022

**Joao Ferrao**, UnISCED, Beira, Mozambique

Dear Reviewer. thank you very much for your precious comments. They are very useful and they were used to improve the manuscript.

- In the manuscript entitled, "Modelling sociodemographic factors that affect malaria prevalence in Sussundenga, Mozambique: a cross-sectional study", the authors present the results of a survey and statistical analysis designed to uncover predictors of positive parasite status by rapid diagnostic test in this area of endemic malaria transmission. Malaria is a significant contributor to disease burden in Mozambique, so this topic is of high importance.

**Response:** Thank you very much for the comment.

- In my report below, I raise a number of questions for the authors regarding aspects of the analysis that were unclear to me (i.e., perhaps where the clarity of presentation could be improved) or where I felt that further analysis might strengthen the conclusions.

**Response:** Thank you very much. All comments will be taken in consideration accordingly.

**Context and Aims:**

- The framing of the study in terms of context and aims as it currently stands deserves considerable review.

**Response:** Thank you very much for the comment. The review will be done according to the reviewer's suggestions.

- Owing precisely to its status as one of the most malarious countries in the world, Mozambique has attracted a great deal of attention from malaria researchers over many decades and there have been many studies investigating environmental and socio-demographic factors behind malaria transmission in the country. For example, Temu *et al.* (Plos One, 2012)<sup>1</sup> identified IRS use, pig-keeping, and house construction as key factors influencing malaria risk in Zambezia province. See also.

**Response:** Thank you very much for the comment. These studies were considered in our discussion. Considering the size of the country 800,000 Km<sup>2</sup>, more than 2,700 Km long, its heterogeneity in terms of landscape, sociodemographic and culture and poor resources, I would consider that the number of studies "investigating environmental and socio-demographic factors behind malaria transmission in the country" very few for local programmatic decisions making policies.

- See also Brentlinger *et al.* (Am J Trop Med Hyg, 2007).

**Response:** Thanks for the recommendation. We find two studies from the recommended authors. However, we feel that they are specific for HIV patients and are outdated.

- Likewise, there are many studies examining factors related to explanatory factors for malaria prevalence such as treatment seeking (Cassy *et al.* Mal J, 2019) and ITN use (Moon *et al.*, Mal J, 2016).

**Response:** Thanks very much for the observations. Two articles (2019 and 2022 on care seeking were found and include in our discussion. As stated in the paper "This study is part of the Malaria Risk, Prevention, and Health Seeking Behaviors in Sussundenga, Mozambique Project". There is another study dealing specifically with this issue.

- Factors shaping malaria risk in Mozambique have also recently been studied in the context of measuring the impact of malaria control strategies in Mozambique by Galatas *et al.* (Plos Medicine, 2020)<sup>5</sup>.

**Response:** Thank you very much for the comment. The studies were considered in our discussion.

- This wealth of prior research and understanding contrasts strikingly with the authors' proposition that "little is known about the epidemiology to inform appropriate and effective interventions". Importantly, it also raises the question of why information on two well-known factors shaping malaria transmission in Mozambique, namely IRS and ITN use, do not seem to have been captured or included in this study?

**Response:** Thank you very much for rising a very important issue of IRS and ITN use. We do consider IRS and ITN use suiting better as "health seeking behaviors" and due to its high relevance, a separate paper was produced:

[https://www.researchgate.net/publication/357008627\\_P\\_Falciparum\\_Community\\_Prevalence\\_and\\_Health\\_Se](https://www.researchgate.net/publication/357008627_P_Falciparum_Community_Prevalence_and_Health_Se)

- Another key known factor, household construction, may have been included but there is no detail that I could see in the manuscript to illuminate the meaning of the "household category" variable used in the model?

**Response:** Thank you very much for the question.

Household size and household construction are very important variables. In the methodology we rephrase the variables and their meaning. We hope that now is clear that household category means type of house or type of construction. In this study, household category or household construction was found as a predictor variable.

- Regarding the statement "ITNs are only available at antenatal clinics, indicated for pregnant women and children under five": I'm not sure that this is well phrased, since the WHO recommends universal net use in high transmission areas. While ITN distribution campaigns focus on the highest risk groups of young children and pregnant women these are not the only groups who should be advised to use bed nets; likewise, ITNs would generally be available commercially at markets. (See e.g. Scott *et al.* Mal J, 2021)<sup>6</sup>.

**Response:** Thank you for the comment. We do agree that "young children and pregnant women are not the only groups who should be advised to use bed nets; Indeed, recent studies are indicating an age shift in Malaria due this situation.

As for the statement: "likewise, ITNs would generally be available commercially at markets. We would agree in a "normal" market driven country. For the Mozambican case where, most people are living bellow the poverty line, buying a mosquito net for prevention can be a luxury.

For example, in 2021, a mosquito factory in Chimoio, Manica closed it is doors and the major reason was lack of clients to purchase the nets.

We added this useful contribution in our discussion.

#### **Statistical Analysis:**

- The statistical analysis method used to derive the primary results of this study, namely the identification of key factors behind malaria prevalence in the study area, is a stepwise logistic regression, which is indeed appropriate for this objective. Some minor details require clarification or revision:

**Response:** Thank you very much for the observation.

- there are some minor formatting errors in the equations: e.g. on page 5, equation 2 uses lower case g and then an upper-case G for seemingly the same function,

**Response:** Thank you very much for observation. This was corrected.

- and betai should have a subscript i as in beta<sub>i</sub>;

**Response:** Thank you for the observation. Correction was made.

- o on page 8 the last equation should have 1.289 "x" Age category (5 to 14 years) instead of "+"

**Response:** Thank you for the observation.

- o "To evaluate potential confounders and, effect modifiers between the final model variables, the Hosmer–Lemeshow (1989) test was performed." This doesn't make sense to me: the HM test is for model specification / acceptable fit, rather than for breaking variables down into their roles in the causal hierarchy.

**Response:** Thanks for a very good observation We agree, to avoid confusing we rephrased the sentence.

- o (iii) I was confused by the chi square test reports in some places: for the distribution by sex the chi-squared statistic of 0.081 doesn't sound like the right order of magnitude and in fact I get 0.081 as the p-value for a binomial exact test on this sample so perhaps this is a typo?;

**Response:** Thank you for the observation.

As stated in the methodology, "Sussundenga has an estimated population of 31,429 inhabitants, 47% males and 53% females". In the present study, the enrolled participants among sex, 55% were female and 45% males. Using the Biostat 5.3 software we find the following out put

	Results
Contingncy table =	2 x 2
Chi - square	1.281
Degrees of freedon	1
(p) =	0.2578
Yates correction=	0.98
(p) =	0.3221

The table was corrected.

- o For the tests by age category, since this is a four x two table I would have thought we're looking at degrees of freedom rather than 6?

**Response:** Thanks for the observation.

We believe that is more appropriate to check the age category compared also to sample results and National Institute of Statistics projections for accuracy, giving us a 4 x 4 contingency table. The following recalculations are presented and were corrected in the manuscript.

	Results
Contingency table	4 x 4
Chi square	25.857
Degrees of freedom	9
(p) =	0.0022

The table was corrected.

- The population density variable must have a large dynamic range because it is assigned a slope ("constant"?) of 0.000 in table 8: could this be rescaled so that we can see its slope within the 3 decimal places?

**Response:** This was very important observation. We increased one decimal place and we rewrite the equation.

- I was confused why the age categories changed from four in the earlier discussion to three in Table 4?

**Response:** Thank you for your valuable observation.

We stated four age categories, 0 to 4, 5 to 14, 14 to 24 and > 24 (Additional file 1). The software did not find difference between age categories 14 to 24 and > 24 and grouped them as the same category. This seems to be right.

- Also, it would help to nominate one age group as the reference group so that the odds ratios can be understood as relative to that group.

**Response:** Thank you for the observation. We fill that a priori, it would be difficult to select a reference age group.

- I'm confused by the focus on understanding the predictive accuracy of the model in terms of specificity and sensitivity(\*), which are appropriate for a diagnostic tool, but which may not be particularly relevant to the use of a risk factor model such as this one for field epidemiology. I.e., if the end use is to prioritise the delivery of a particular intervention such as seasonal malaria chemoprophylaxis then identifying that a certain age group has twice the parasite prevalence of another could be of substantial value even where the sensitivity was low because prevalence itself was very low across both strata. This comes back to the context and aims of the study, in the sense that the value of the fitted model (or more precisely the knowledge discovered through it) is ultimately something that exists only relative to the way in which it is intended to be used.

**Response:** Thanks for the observation.

Since sensitivity and specificity are measures of performance of a binary model, is pertinent to access them in the logistic model. We included an explanation in the methodology and we also discussed the results.

- (\*also: are these defined according to a thresholding of the predictive prevalence above and below 50%?)

**Response:** Yes



**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**