Systematized Literature Review on Spatial Analysis of Environmental Risk Factors of Malaria Transmission

Tiago Canelas¹,²,³, Carlos Castillo-Salgado²,³, Helena Ribeiro¹

¹School of Public Health, University of São Paulo, São Paulo, Brazil
²Johns Hopkins Bloomberg School of Public Health, Baltimore, US
³Global Public Health Observatory, Johns Hopkins Bloomberg School of Public Health, Baltimore, US

Received 18 March 2016; accepted 13 June 2016; published 16 June 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Abstract

Malaria is still the major parasitic disease in the world, with approximately 438,000 deaths in 2015. Environmental risk factors (ERF) have been widely studied, however, there are discrepancies in the results about their influence on malaria transmission. Recently, papers have been published about geospatial analysis of ERF of malaria to explain why malaria varies from place to place. Our primary objective was to identify the environmental variables most used in the geospatial analysis of malaria transmission. The secondary objective was to identify the geo-analytic methods and techniques, as well as geo-analytic statistics commonly related to ERF and malaria. We conducted a systematized review of articles published from January 2004 to March 2015, within Web of Science, Pubmed and LILACS databases. Initially 676 articles were found, after inclusion and exclusion criteria, 29 manuscripts were selected. Temperature, land use and land cover, surface moisture and vector breeding site were the most frequent included variables. As for geo-analytic methods, geostatistical models with Bayesian framework were the most applied. Kriging interpolations, Geographical Weighted Regression as well as Kulldorff’s spatial scan were the techniques more widely used. The main objective of many of these studies was to use these methods and techniques to create malaria risk maps. Spatial analysis performed with satellite images and georeferenced data are increasing in relevance due to the use of remote sensing and Geographic Information System. The combination of these new technologies identifies ERF more accurately, and the use of Bayesian geostatistical models allows a wide diffusion of malaria risk maps. It is known that temperature, humidity vegetation and vector breeding site play a critical role in malaria transmission; however, other environmental risk factors have also been identified. Risk maps have a tremendous potential to enhance the effectiveness of malaria-control programs.

How to cite this paper: Canelas, T., Castillo-Salgado, C. and Ribeiro, H. (2016) Systematized Literature Review on Spatial Analysis of Environmental Risk Factors of Malaria Transmission. Advances in Infectious Diseases, 6, 52-62.
http://dx.doi.org/10.4236/aid.2016.62008
Keywords
Malaria, Spatial Analysis, Environmental Risk Factors, Malaria Transmission, Malaria Determinants

1. Introduction

Despite substantial progress having made, malaria was still a major global health problem in 2015 and the cause of about 438,000 deaths, mostly in the World Health Organization (WHO) African Region (90%) [1]. Environmental risk factors (ERF) of malaria have been widely studied, however, there are discrepancies on the results about their influence on malaria transmission, especially at the local level [2]. Since malaria varies from place to place and understanding certain ecological conditions may lead to the presence of malaria, in the recent years, there has been an increasing amount of literature on spatial analysis of ERF of malaria transmission in order to explain this variation. New spatial analysis methods have been used taking into account the space itself as a factor that contributes to malaria heterogeneity between households, communities or neighborhoods, among other spatial units. So far, however, there has been little discussion about the geo-analytic methods and techniques used to model the ERF. Notwithstanding, papers about the importance of the malaria ERF have lately been published to claim that little attention has been giving to environmental variables selection [2] [3]. Advances in technology specially in Geographic Information System (GIS) and remote sensing facilitate researchers mapping the results of their models and thus, help in the control, planning, implementation, and evaluation of malaria risk, at geographic scales ranging from local to global as well as to identify where most vulnerable populations are.

The primary objective of this study was the identification of the environmental variables in the spatial analysis of malaria transmission. The secondary objective was to identify geo-analytic methods and techniques, as well as geostatistics commonly related to ERF and malaria transmission.

2. Methodology

To assess the current state of knowledge we conducted a systematized literature review. This type of review seeks to include some elements of the systematic review process however, due to the resources needed for a full systematic review is much shorter [4]. In order to improve the quality and accuracy of the review the authors followed the guideline created by Costa et al. 2015 [5] of a list of fourteen items. Our review did not include the two items related to peer reviews because we did not have additional resources to include them.

The systematized literature review included free and pay articles within PUBMED, WEB OF SCIENCE (WoS) and the Latin American and Caribbean Health Sciences Literature (LILACS) databases from January 2004 to April 2015 and we accepted papers written in English, Portuguese and Spanish. More than 45 keywords combinations were used based on papers related with the topic, systematic reviews and authors’ prior knowledge (Figure 1, Figure 2).

For PUBMED and WoS databases (Figure 1) the following keywords and Boolean operators were used: (malaria [text word] OR malaria [MeSH Major Topic]) AND (environmental* risk* factor* [text word] OR environmental* determinant* [text word] OR ecologic* risk* factor* [text word]) AND (spatial [text word] OR “spatial analysis” [text word] OR “space-time” [text word] OR spatial temporal analysis [text word] OR gis [text word] OR “geographic* information system*” [text word] OR geostat* [text word] OR spatial cluster* [text word] OR spatial autocorrelation [text word]). The same keywords and operators were used in LILACS database (Figure 2) but no results were found. Due to the lack of results, the search was slightly modified by: (malaria) AND (environment OR ecology) AND (spatial OR “spatial analysis” OR “space-time” OR spatial temporal analysis OR gis OR “geographic* information system*” OR geostat* OR spatial autocorrelation OR cluster).

Firstly, all titles and abstracts were read to determine if they were included in the second step. Once selected, a full comprehensive and application of criteria of inclusion and exclusion was performed in order to select articles for the final selection (Figure 3). Among all the papers only one paper written in Chinese was discarded.

For this review, the environmental risk factors of malaria transmission were defined as all of those factors related to the physical environment without human activity. However, it is well known that social determinants...
play a major role in malaria transmission, especially when spatial analysis is applied. Therefore population, population density, household density, distance from health facilities as well as land use were included in our analysis.

3. Results

After applying the criteria of exclusion we were left with 29 out of 74 papers (Figure 3). Table 1 illustrates some of the main characteristics of the selected papers. Firstly, ERF were checked in order to extract only those that were statistically significant (by the authors) since the authors usually started with a large number of ERF and after applied some statistical analysis, to check for collinearity and for statistical significance, a reduced number of variables least were left in the final spatial model. Finally, we extracted those ERF that were statistical significant in the spatial model. We observed that slightly more than half of the prior ERF selected by the authors in the first step were eliminated from the final spatial model. To check for correlation between malaria and environmental variables most authors used statistical data analysis, they usually applied regression analysis.
Figure 2. Flow chart from the keywords combinations in LILACS.

[6]-[24], sometimes using the Bayesian approaches [25] [26]. However, others assessments used as multilevel models [9] or they only mention (uni- or bi-) variate analysis [27]-[30]. Some authors included the ERF directly in the spatial model without listing any previous statistical analysis in their study [31]-[34] based in previous work or authors knowledge.

When extracting the ERF from the selected papers we had to group them in order to reduce the number of variables. For example, distance to water bodies is the combination of distance to water bodies, rivers and lagoons. Coverage, forest cover, land cover, distance to swamp or forest are together the category of land cover. After that, as shown in Table 2, we aggregated different ERF in categories following the methodology of Weiss, D.J. et al. [3] in order to simplify the reading and to be able to categorize the ERF. We considered and combined the EFR as one whole group but some authors prefer to divide the risk factors in many categories as climatic, environmental or topographical [30].

We excluded other risk factors used by the authors if they do not match our definition of ERF, thus we removed; level of poverty, local migration patterns, age, number of households, house construction, household...
Figure 3. Flow chart of criteria of exclusion.

Table 1. Summary of environmental risk factors and spatial methodology to assessed malaria transmission from 2005 to 2014.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Country</th>
<th>ERF significant</th>
<th>Spatial method used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mabaso et al. 2005 [6]</td>
<td>Zimbabwe</td>
<td>Rainfall, vapor pressure, temperature, NDVI</td>
<td>Thematic map; Bayesian spatial-temporal CAR</td>
</tr>
<tr>
<td>Gosoniu et al. 2006 [7]</td>
<td>Mali</td>
<td>NDVI, rainfall, temperature, distance to water bodies</td>
<td>Thematic map; Bayesian stationary and non-stationary model; Bayesian kriging</td>
</tr>
<tr>
<td>Hernández-Avila et al. 2006 [8]</td>
<td>Mexico</td>
<td>Climate, distance to river, elevation, evaporation, distance to road, population</td>
<td>Thematic map; Spatial logistic model; Bayesian spatio-temporal filter</td>
</tr>
<tr>
<td>Ernst et al. 2006 [9]</td>
<td>Kenya</td>
<td>Altitude, distance to forest/swamp/rivers /health services, pop. density</td>
<td>Thematic map; Spatial scan statistics; Moran’s I statistic</td>
</tr>
<tr>
<td>Gemperli et al. 2006 [27]</td>
<td>West and Central Africa</td>
<td>Temperature, rainfall, NDVI, distance water bodies, length of season</td>
<td>Bayesian lineal geostatistical model; Bayesian kriging</td>
</tr>
<tr>
<td>Silué et al. 2008 [11]</td>
<td>Côte d’Ivoire</td>
<td>None of the covariates are significant in the spatial model</td>
<td>Thematic map; Bayesian stationary and non-stationary model</td>
</tr>
<tr>
<td>Gosoniu et al. 2009 [25]</td>
<td>West Africa</td>
<td>Temperature, rainfall, distance to water bodies, land use, season length</td>
<td>Thematic map; Bayesian kriging; Bayesian nonparametric non-stationary model</td>
</tr>
<tr>
<td>Raso et al. 2009 [12]</td>
<td>Côte d’Ivoire</td>
<td>Rainfall, distance to river</td>
<td>Thematic map; Bayesian kriging; Bayesian stationary and non-stationary model</td>
</tr>
<tr>
<td>Gosoniu et al. 2010 [13]</td>
<td>Angola</td>
<td>None of the covariates are significant in the spatial model</td>
<td>Thematic map; Bayesian kriging; Bayesian geostatistical model with location-specific random effects</td>
</tr>
<tr>
<td>Grillet et al. 2010 [14]</td>
<td>Venezuela</td>
<td>Number of breeding sites</td>
<td>Thematic map; Kulldorff’s scan; local Getis statistic; Moran’s I; Geographical weighted regression</td>
</tr>
<tr>
<td>Riedel et al. 2010 [15]</td>
<td>Zambia</td>
<td>None of the covariates are significant in the spatial model</td>
<td>Thematic map; Bayesian kriging</td>
</tr>
<tr>
<td>Reid et al. 2010 [16]</td>
<td>Bangladesh</td>
<td>Forest cover, elevation, temperature</td>
<td>Thematic map; Bayesian geostatistical model; Interpolation; Semivariogram</td>
</tr>
<tr>
<td>Haque et al. 2010 [17]</td>
<td>Bangladesh</td>
<td>Forest cover</td>
<td>Thematic map; Bayesian geostatistical model; Kriging</td>
</tr>
</tbody>
</table>
Continued

<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Country</th>
<th>Variables</th>
<th>Spatial Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zacarias and Majlender 2011 [28]</td>
<td>Mozambique</td>
<td>Rainfall, temperature, humidity,</td>
<td>Thematic map; Bayesian hierarchical models</td>
</tr>
<tr>
<td>Zayeri et al. 2011 [18]</td>
<td>Iran</td>
<td>Temperature, humidity, altitude, rainfall</td>
<td>Empirical Bayesian smoothing; Bayesian random effects model</td>
</tr>
<tr>
<td>Ashton et al. 2011 [19]</td>
<td>Ethiopia</td>
<td>Distance to water bodies, EVI, LST</td>
<td>Thematic map; Bayesian spatial multivariate models; Kulldorff's spatial scan</td>
</tr>
<tr>
<td>Manh et al. 2011 [31]</td>
<td>Vietnam</td>
<td>Temperature, rainfall, forest cover</td>
<td>Thematic map; Bayesian zero-inflated Poisson regression</td>
</tr>
<tr>
<td>Zacarias and Andersson 2011 [29]</td>
<td>Mozambique</td>
<td>Temperature, relative humidity</td>
<td>Bayesian kriging; Bayesian model with random effects and space-time interactions</td>
</tr>
<tr>
<td>Haque et al. 2011 [20]</td>
<td>Bangladesh</td>
<td>Forest cover, altitude, household density</td>
<td>Thematic map; Semivariogram</td>
</tr>
<tr>
<td>Achar et al. 2011 [32]</td>
<td>Brazil</td>
<td>Deforestation</td>
<td>Thematic map; Bayesian spatiotemporal Poisson model</td>
</tr>
<tr>
<td>Giardina et al. 2012 [26]</td>
<td>Senegal</td>
<td>LST, NDVI, region (urban/rural)</td>
<td>Thematic map; Bayesian kriging; Bayesian Zero-inflated binomial geostatistical model</td>
</tr>
<tr>
<td>Magalhães et al. 2012 [21]</td>
<td>Angola</td>
<td>Distance to health care, distance to river</td>
<td>Thematic map; Semivariogram; Bayesian geostatistical model</td>
</tr>
<tr>
<td>Gosoni et al. 2012 [22]</td>
<td>Tanzania</td>
<td>Region (urban/rural)</td>
<td>Thematic map; Bayesian hierarchical geostatistical model; Bayesian kriging</td>
</tr>
<tr>
<td>Townes et al. 2013 [23]</td>
<td>Malawi</td>
<td>Land cover</td>
<td>Thematic map; Getis-Ord Gi*</td>
</tr>
<tr>
<td>Alegana et al. 2014 [33]</td>
<td>Afghanistan</td>
<td>Temperature suitability index</td>
<td>Thematic map; Bayesian hierarchical spatio-temporal model</td>
</tr>
<tr>
<td>Kumar et al. 2014 [34]</td>
<td>India</td>
<td>Temperature, rainfall, altitude, density population</td>
<td>Thematic map; Getis-Ord Gi*</td>
</tr>
<tr>
<td>Chirombo et al. 2014 [30]</td>
<td>Malawi</td>
<td>Region (rural/urban), altitude</td>
<td>Thematic map; Bayesian structured additive logistic regression model</td>
</tr>
</tbody>
</table>

Table 2. Composition of the main ERF by variables aggregation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Temperature</th>
<th>Land cover</th>
<th>Land use-land cover (LULC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance from services</td>
<td>Distance to road</td>
<td>Distance to health services</td>
<td>Land use/land cover</td>
</tr>
<tr>
<td>Distance to water bodies</td>
<td>EVI</td>
<td>NDVI</td>
<td>Surface moisture and vector breeding site</td>
</tr>
<tr>
<td>Breeding site</td>
<td>Soil water holding capacity</td>
<td>Climate</td>
<td>Temperature</td>
</tr>
<tr>
<td>Surface moisture and vector breeding site</td>
<td>Humidity</td>
<td>Vapor pressure</td>
<td>Land use</td>
</tr>
<tr>
<td>Population density</td>
<td>Population</td>
<td>Evaporation</td>
<td>Region (urban/rural)</td>
</tr>
<tr>
<td>Household density</td>
<td>Household density</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be seen from the data in Figure 4 that the category of temperature and land use and land cover are the ERF more statistically significant in the final spatial model. Although, a group of three variables (surface moisture and breeding site, rainfall and altitude) was also statistically significant in the model. When we took into account only the sub-variable then temperature, rainfall and altitude are the most significant ERF. Although not all papers revealed whether co-variables were statistically positive or negative significant in the final spatial model. For those papers who mentioned positive or negative influence in malaria risk we found discrepancies
between articles, thus, temperature appeared positively and also negatively associated (9/4), as well as Land Use and Land Cover (11/3), surface moisture and vector breeding site (4/4), altitude (5/2) and rainfall (6/4) respectively. Only humidity (4/0) suggested positive association at all the times.

In addition to the information provided in Table 1, we analyzed which were the spatial scales and the spatial units most used. Authors chose health facilities twice and households in eleven papers as a spatial scale, both are a point unit and by far households were the individual data most used. Among the aggregated data, district and province, areal unit, and villages and surveys, point unit, were selected in the analysis.

This review also has the aim to identify which spatial methods were used to model the ERF for malaria transmission. In order to gather this information, we listed all the spatial methods in Table 1. We started to recognize cartographic methods and we noticed that 26 out 29 used some type of thematic map. These might be dot or areal maps to show prevalence or incidence, others used the map to show malaria clusters. Thematic maps were used also to map the ERF [6] [8] [9] [11]-[15] [24] [26] [27] [34]. The ability to elaborate a malaria risk map is one of the main objectives of spatial analysis. The following papers used the spatial methods to predict incidence, parasitaemia or prevalence risk maps [7] [12] [13] [15]-[17] [21] [22] [25]-[27] [30].

Malaria risk maps were created after applying data in the final spatial model. These models were conducted under the Bayesian approach in 21 out 29 papers. Depending on the type of data, the authors used different settings in the Bayesian framework but they always introduced the spatial variation of the data in the model formulation. Some authors stated the settings, in contrast others only mentioned the Bayesian geostatistical model as a
final model \[13\] \[16\] \[17\] \[21\] \[27\]. Kriging is one of the techniques used to create malaria risk maps and was used in 10 out of 12 malaria risk maps in this review and as a step before kriging the authors \[8\] \[16\] \[20\] \[21\] checked to determine if the semiovariogram fitted properly with the data and looked for anisotropy.

Investigating malaria clusters over space and time has been a research concern since the beginning of spatial analysis. In order to address this topic, authors \[9\] \[14\] \[19\] used spatial scan statistics that search by means of different size circles and calculates the likelihood of malaria risk inside the circle compared to outside. Others authors instead preferred to apply Getis and Ord’s G statistics to measure the spatial dependence to find local clusters \[14\] \[23\] \[34\]. Local statistics as Moran’s I were also applied \[9\] \[14\]. To know where non-stationarity is taking place on the map they used Geographic Weight Regression \[12\] \[22\]. To deal with small number problem \[18\] several studies used the empirical Bayes smoothing where rates are adjusted according with the size of population on which they are based. Finally, in order to remove spatial autocorrelation \[6\] some studies applied a spatial logistic regression.

4. Discussion

Although largely studied ERF were still a source of discussion among authors. Comparing Figure 4 with previous ERF (data not showed) reveals that as expected, environmental determinants could not fully explain malaria transmission. When addressing a subject as malaria under an holistic approach, both environmental and social determinants, are required to fully understand this important public health problem. However, exclusively concerning the ERF this review produced results which corroborate the findings of previous work in this field. It is well established that certain temperature, rainfall, altitude and humidity threshold may lead to an increase or decrease of malaria transmission. Likewise, different land use and land cover influence the transmission of malaria. Nevertheless, these studies are mainly on the African continent 19 out 29 and predominantly with Plasmodium falciparum, therefore we suggest having additional studies about ERF in malaria transmission outside of the African continent and with others species of Plasmodium. These results therefore need to be interpreted with caution.

Nonetheless, the findings of the current review support the selection of geostatistical variables as an important factor in malaria mapping and ERF modeling. Variable selection is often based on regression models that ignore spatial correlation, leading to incorrect estimates of covariates effects and their significance \[26\]. The results of this research support the idea that different variable selection steps could improve the accuracy of the final model.

In order to take into account the ERF in the study site some authors opted for the geostatistical approach which assume that the observed data are a sample of one realization of a continuously indexed spatial stochastic process. These methods focus on estimating the global and the local trend structure and predicting or interpolating the values at the non-sampled locations \[35\]. Taken together, these outcomes suggest that the Bayesian geostatistical approach allows the researcher a better modeling than the traditional statistic approach of the ERF. The spatial variation in disease risk is one of the most important functions of spatial analysis. To take into account the spatial autocorrelation authors handled different methodologies among them the semivariogram as the first step in this process \[16\]. After performing the spatial model the results have to be extrapolated and the authors used mainly kriging, which is a method used to represent measurements taken at discrete set of control points as a continuous surface. Kriging considers not only the distances to control points, but also the spatial autocorrelation of measurements among control points. In kriging accuracy depends on the semivariogram which is used to generate a set of spatial weights. It is important to check the semivariogram is properly specified and fits well the data being modeled \[36\]. Kriging was used to extrapolate the results of the spatial model to households non-sampled.

Other spatial methods have taken place such as Geographical Weighted Regression that is an exploratory technique mainly intended to indicate where non-stationary is taking place on the map, that is where locally weighted regressions coefficients move away from their global values \[37\]. Malaria varies from place to place even in the same region hence it is important to apply local spatial analysis as Getis and Ord’s G* or Moran’s I in order to understand why malaria is heterogeneous in the same environment. The use of the Geographic Information System as well as the improvement in the last three decades in computer power allows researchers to perform exploratory analysis easily and quickly. This fact might explain the limited number of studies in the last years using spatial clustering.
5. Conclusions

There are a small number of ERF on the geospatial models that play an important role in understanding malaria risk transmission. There are discrepancies and a limited knowledge about the specific positive or negative influence of ERF in malaria transmission. The results of the studies included in the review showed different directions of their influence for malaria risk, with the exception of humidity showing a positive association in all studies. The current review was not specifically designed to evaluate factors related to social determinants and spatial analysis of malaria transmission. However, these factors need to be included in further reviews.

In the last 10 years, the Bayesian geostatistical approach has been used to model environmental risk factors of malaria transmission due to its capacity to handle complex modelling frameworks and the ability to create malaria risk maps that are crucial in economically constrained countries to allow efficient allocations of limited resources.

Acknowledgements

The author Tiago Canelas acknowledges a predoctoral grant from the Government of Andorra, ACT014-AND-2015/2016 and the scholarship by the Coordination for the Improvement of the Higher Level Personnel (CAPES), Ministry of Education, Brazil.

References


http://dx.doi.org/10.1016/j.amjtmh.2010.10.0154

http://dx.doi.org/10.1186/1475-2875-9-120

http://dx.doi.org/10.1016/S1995-7645(11)60231-9


http://dx.doi.org/10.1186/1475-2875-10-367


http://dx.doi.org/10.1371/journal.pone.0023966


http://dx.doi.org/10.1016/j.csda.2009.02.022

http://dx.doi.org/10.1371/journal.pone.0032625


