Spatial modeling of malaria incidence rates in Sistan and Baluchistan province, Islamic Republic of Iran

Masoud Salehi, PhD, Student, Kazem Mohammad, PhD, Mahmud M. Farahani, PhD, Hojjat Zeraati, PhD, Keramat Nourijelyani, PhD, Farid Zayeri, PhD

ABSTRACT

Objectives: To identify the effect of environmental factors on malaria risk, and to visualize spatial map of malaria standard incidence rates in Sistan and Baluchistan province, Islamic Republic of Iran.

Methods: In this cross-sectional study, the data from 42,162 registered new malaria cases from 21 March 2001 (Iranian new year) to 21 of March 2006 were studied. To describe the statistical association between environmental factors and malaria risk, a generalized linear mixed model approach was utilized. In addition, we used the second ordered stationary Kriging, and a variogram to determine the appropriate spatial correlation structure among the malaria standard incidence rates, and provide a proper malaria risk map in the area under study.

Results: The obtained results from the spatial modeling revealed that humidity ($p=0.0004$), temperature ($p<0.0001$), and elevation ($p<0.0001$) were positively, and precipitation ($p=0.0029$) was inversely correlated with the malaria risk. Moreover, the malaria risk map based on the predicted values showed that the south part of this province (Baluchistan), has a higher risk of malaria, compared to the northern area (Sistan).

Conclusion: Since the effective environmental factors on malaria risk are out of human’s control, the health policy makers in this province should pay more attention to the areas with high temperature, elevation, and humidity, as well as, low rainfall districts.


From the Department of Biostatistics and Epidemiology (Salehi, Mohammad, Farahani, Zeraati, Nourijelyani), School of Public Health, Tehran University of Medical Sciences, and the Department of Biostatistics (Zayeri), Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Received 13th July 2008. Accepted 8th November 2008.

Address correspondence and reprint request to: Professor Kazem Mohammad, Department of Biostatistics and Epidemiology, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. Tel. +98 (21) 88973901, Fax. +98 (21) 88974462. E-mail: mohamadm@tums.ac.ir / Salehi74@yahoo.com
Malaria is one of the most important health problems that affect over 100 million persons, and kills approximately one million people, especially in children under 5 years. This infectious communicable disease is more prevalent in semiarid areas such as the Middle East, and African countries. Therefore, measuring malaria burden, and determining its related factors in a population, is a challenge in most developing countries. The malaria incidence rate is strongly associated with factors such as massive displacements, famine, lack of efficient disease surveillance, and environmental conditions. Such as other Middle East countries, malaria is an important health problem in Iran, particularly in the eastern and south-eastern areas. According to the report of the Malaria Office of the Ministry of Health of Iran, the annual malaria cases were at least 15,000 during 21 of March 2001 (Iranian new year), and 21 of March 2006. Among these cases, approximately 60% were reported from a vast province in the east (Sistan), and southeast of Iran (Baluchistan). The analysis of the geographical distribution of the incidence of the disease, and its relationship to potential risk factors, has an important role to play in various kinds of public health and epidemiological studies. In this context, using proper statistical models is a helpful strategy in the analysis of geographical data, and describing the statistical relationship between potential risk factors and disease incidence/prevalence. Additionally, spatial modeling is a useful statistical tool for producing a proper incidence/prevalence map of the area illustrating the variation in malaria risks. Many standard statistical models assume independence of observations. However, the majority of infectious diseases, such as malaria cases, cluster due to underlying common environments. When spatially correlated data are analyzed, this independent assumption leads to overestimation of the statistical significance of covariates. Spatial models incorporate the spatial correlation according to the way the geographical information is available. For rate or count data, the spatial correlation is defined by a neighborhood structure. For geo-statistical data, the spatial correlation is usually considered as a function of the distance between locations. In the malaria rates, due to its strong spatial correlation, the regression modeling results should be adjusted using convenient generalized linear mixed models (GLMMs), and variograms. Recently, Bithell, Diggle, Lawson, and Lawson et al provided brilliant reviews of disease mapping using spatial modeling. Despite the high incidence rate of malaria in Sistan and the Baluchistan province (SBP), we found no published article on spatial modeling of the incidence rate of malaria and its related factors in this part of Iran. In this study, therefore, we aim to explore the effective environmental factors of malaria such as precipitation, humidity, elevation, and temperature in SBP, Islamic Republic of Iran.

**Methods.** This epidemiologic cross-sectional research was carried out in SBP in southeastern Iran as a part of a PhD dissertation in the Department of Epidemiology and Biostatistics of Tehran University of Medical Sciences, Tehran, Iran. This study was approved by the Ethics Committee of Tehran University of Medical Sciences.

**Study area.** Covering a surface area of 181,470.9 square kilometers, Sistan and Baluchistan province is twice the size of England. Sistan and Baluchistan provinces consist of 2 main parts. Sistan, in the northern part, is in the neighborhood of Afghanistan, and Baluchistan, in the southern part, is in the neighborhood of Pakistan and Oman Sea. Sistan and Baluchistan has 10 provinces (Zahedan, Zabol, Chabahar, Iranshahr, Saravan, Khash, Konarak, Zahak, Sarbaz and Nikshahr), 37 districts, and 97 subdistricts. Zahedan is the capital province. Khash, the highest city is situated 1,394 meters above sea level, and Chabahar, is 7 meters above sea level, and is the lowest city of the province. Sistan and Baluchistan provinces’ has a common border of 1,265 km with southeastern neighboring countries.

**Data source.** In this study, we used registered malaria data from urban and rural health centers in SBP. Variables such as patients’ and their parents’ names, age, gender, type of accommodation (permanent or temporary), nationality (Iranian or immigrant), location (name of village or city), and the date of obtaining and reading blood films were gathered from 21 of March 2001 (Iranian new year) to 21 of March 2006. A MicroSoft-Access data bank was utilized to locate the recurrent malaria cases. In the Islamic Republic of Iran, the national census is conducted every 10 years by the Statistical Center of Iran (2 last censuses were in 1996 and 2006). The adjusted midyear population size for the period between the 2 censuses is calculated, and represented by the Statistical Center of Iran. We used these adjusted midyear population size for calculating the incidence rates of malaria. Overall, the information from 37 districts was collected. For each district a standard incidence rate (SIR) was calculated for the described period. To calculate these SIRs, it

*The full text including Appendix is available in PDF format on Saudi Medical Journal website (www.smj.org.sa)*
was assumed that the entire population of a district was exposed to the risk of malaria during this period, that is, each person contributed exactly one person-year of exposure. In computing the malaria SIRs, the repeated episodes of the subject were considered just as an observed subject for each year. The environmental explanatory variables such as monthly precipitation, average of monthly humidity, elevation, and average of maximum daily temperature in a month for the above mentioned period, were obtained from the registered climate data in Islamic Republic of Iran Meteorological Organization (IRIMO).22 By using Geographical Information System (GIS), the average value of each variable for each district under study was calculated by averaging the pixel variable over the area in the district.23

**Statistical analysis.** For the epidemiological description of the data, the SIR of malaria was computed separately, for the total population and by gender, in different districts of SBP.24 The spatial correlation between district and malaria SIRs, was assessed using Moran’s I statistics.16 This analysis showed that the malaria SIR are severely spatially correlated. In addition, since we had unobserved malaria cases in some areas, the Kriging approach was used for predicting these unobserved values. In the next step, a variogram was utilized to find the appropriate autocorrelation structure (sill, nugget and range).16,25 Finally, a spatial generalized linear mixed modeling approach was utilized for describing the relationship between environmental and climatic explanatory variables, and malaria SIRs. Appendix A shows more details on the structure of the data, and the utilized model for data analysis. We used R?? codes for computing the variogram and the SAS software, version 9.1, for spatial modeling of the data. The interested reader can refer to Appendix B for R?? codes. ArcGIS software, version 9, was also used to visualize the geographical map of malaria in SBP.

**Results.** In general, a 42,162 malaria new cases were registered from March 2001 to March 2006 in SBP. Among them, 27700 cases (65.7%) were male, and 14461 cases (34.3%) were female. This study also showed that 2530 (6%) of these patients were children less than 5 years old, and 35078 cases (83.2%) were adults older than 15 years. Plasmodium vivax was observed in 33097 (78.5%), Plasmodium falciparum in 8601 (20.4%), and mixed infections were observed in 464 (1.1%). Among the detected new cases, 3837 cases (9.1%) were from the urban, and 38,325 cases (90.9%) were from the rural areas. Table 1 shows more description on the population and malaria SIRs (Table 1). Malaria SIRs showed significant positive spatial correlation with Moran’s I statistics (p<0.001). To describe the relationship between climatic and environmental covariates and observed malaria SIRs while accounting for spatial correlation of the data, an iterative approach of variogram and GLMMs was utilized. Figure 1 displays the variogram of deviance residuals of the final model of malaria SIRs in SBP. Regarding the presented variogram in Figure 1, the GLMM model with convenient correlation structure for this data set was fitted. Table 2 shows the obtained results. According to the presented results in Table 2, one can conclude that temperature

---

**Table 1** - Malaria SIR in Sistan and the Baluchistan province between March of 2001 and March of 2005.

<table>
<thead>
<tr>
<th>Year</th>
<th>Gender</th>
<th>Total population</th>
<th>Malaria new cases</th>
<th>Age meansSD</th>
<th>SIR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Male</td>
<td>1011570</td>
<td>4103</td>
<td>24.2±18.07</td>
<td>0.041</td>
<td>0.039 – 0.044</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1007606</td>
<td>1759</td>
<td>23.7±17.64</td>
<td>0.018</td>
<td>0.017 – 0.019</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2017176</td>
<td>5862</td>
<td>23.9±17.98</td>
<td>0.029</td>
<td>0.028 – 0.031</td>
</tr>
<tr>
<td>2002</td>
<td>Male</td>
<td>1045081</td>
<td>3733</td>
<td>22.7±16.93</td>
<td>0.036</td>
<td>0.034 – 0.038</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1038921</td>
<td>1758</td>
<td>21.9±18.76</td>
<td>0.017</td>
<td>0.016 – 0.018</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2084002</td>
<td>5491</td>
<td>22.2±17.35</td>
<td>0.027</td>
<td>0.025 – 0.028</td>
</tr>
<tr>
<td>2003</td>
<td>Male</td>
<td>1078172</td>
<td>9935</td>
<td>22.9±17.16</td>
<td>0.093</td>
<td>0.088 – 0.099</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1071827</td>
<td>4464</td>
<td>21.9±17.49</td>
<td>0.042</td>
<td>0.040 – 0.045</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2150009</td>
<td>14399</td>
<td>22.5±17.28</td>
<td>0.067</td>
<td>0.064 – 0.072</td>
</tr>
<tr>
<td>2004</td>
<td>Male</td>
<td>1112977</td>
<td>4178</td>
<td>23.4±16.29</td>
<td>0.038</td>
<td>0.036 – 0.040</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1106416</td>
<td>2058</td>
<td>22.7±17.79</td>
<td>0.019</td>
<td>0.018 – 0.020</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2219339</td>
<td>6236</td>
<td>23.2±16.83</td>
<td>0.028</td>
<td>0.027 – 0.030</td>
</tr>
<tr>
<td>2005</td>
<td>Male</td>
<td>1148423</td>
<td>6308</td>
<td>23.5±16.21</td>
<td>0.055</td>
<td>0.053 – 0.059</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1141653</td>
<td>3866</td>
<td>24.0±18.31</td>
<td>0.034</td>
<td>0.032 – 0.037</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>2290076</td>
<td>10174</td>
<td>23.6±16.96</td>
<td>0.045</td>
<td>0.042 – 0.047</td>
</tr>
</tbody>
</table>

SD - standard deviation, SIR - standard incidence rate, CI - confidence interval.
Figure 1 - Variogram of deviation residuals of the model of malaria standards incidence rates for the population of the districts of Sistan and Baluchistan province, Islamic Republic of Iran.

Table 2 - Spatial modeling results for assessing the environmental and climatic factors on malaria standard incidence rates.

<table>
<thead>
<tr>
<th>Explanatory factor</th>
<th>Est</th>
<th>SE</th>
<th>P-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>-0.0105</td>
<td>0.0033</td>
<td>0.0029</td>
<td>-0.0171 – -0.0039</td>
</tr>
<tr>
<td>Temperature</td>
<td>0.4629</td>
<td>0.0621</td>
<td>&lt;0.0001</td>
<td>0.3364 – 0.5894</td>
</tr>
<tr>
<td>Elevation</td>
<td>0.0033</td>
<td>0.0005</td>
<td>&lt;0.0001</td>
<td>0.0023 – 0.0043</td>
</tr>
<tr>
<td>Humidity</td>
<td>0.0403</td>
<td>0.0101</td>
<td>0.0004</td>
<td>0.0197 – 0.0609</td>
</tr>
</tbody>
</table>

Est - estimate of the model parameter, SE - standard error of the estimate, CI - confidence interval for the estimate

(p<0.0001), humidity (p=0.0004), elevation (p<0.0001) and precipitation (p=0.0029) had a significant effect on malaria SIR in SBP. In addition, the sign of the estimates tells us that temperature, humidity and elevation had a direct effect, and precipitation has an inverse effect on malaria incidence rates. In the next step, we provided the geographical map of malaria in SBP based on the predicted SIR values by the described GLMM. Our findings show that Zabol, Zahak, and Khash had the lowest, and Nikshahr, Sarbaz, and Iranshahr had the highest malaria SIRs among the SBP counties. In general, these results show that Baluchistan has a considerably higher malaria incidence rate compared to Sistan.

Discussion. Generally, transmission of malaria occurs at 640 North and 320 South of the earth, where there are favorable conditions for the life cycle of malaria parasite.\(^\text{26}\) Sistan and Baluchistan provinces are also located in this limited area. Accurate maps of malaria incidence are important tools in malaria control as they can guide interventions, and assess their effectiveness. These maps rely on predictions of risk at locations without observed prevalence data. Malaria is an environmental disease, and environmental factors mainly temperature, rainfall, and humidity are good predictors of transmission, but the relation between these environmental factors, mosquito abundance, and malaria incidence/prevalence is not linear. This relation can be established only by means of adequate spatial statistical models. By determining the relations between malaria and the environment, the burden of malaria can be estimated at places where data on transmission are not available, and high-risk areas can be identified. Accurate maps of malaria transmission can guide intervention strategies, and thus optimize the use of limited and financial resources to areas of utmost need. In addition, early warning systems can be developed to predict epidemics of malaria from environmental changes.\(^\text{27}\)

While there is a growing body of literature on the use of spatial modeling for malaria research and control, there have been a few issues on malaria indexes in Iran,\(^\text{28,29}\) and there is no comprehensive survey and published article on malaria incidence, or prevalence rate in SBP. In this study, we used a GLMM to assess the relationship between malaria SIRs and environmental factors and provide an accurate malaria incidence map in SBP. Elevation has long been recognized to be associated with malaria due to its association with cooler temperatures and humidity.\(^\text{30-32}\) In the present study, our findings revealed that the elevation had a positive association with malaria SIRs. Due to the large variation in elevation distribution in SBP (from 7 meters to 1394 meters above the sea level), this factor provides a useful measure of transmission risk, and malaria incidence rate. A number of research has found a strong correlation between the malaria incidence rate and variations in the environmental variables during several preceding months,\(^\text{33}\) or with inter-annual variations in these variables.\(^\text{34}\) In many studies, humidity, temperature, and rainfall are considered major risk factors that affect the life cycle and breeding of mosquitoes.\(^\text{35}\) The model derived in the present study uses monthly average of maximum daily temperature, humidity, and total monthly precipitation. This model showed a strong correlation between malaria SIRs and precipitation. We know that reasonable amounts of rainfall create additional breeding sites for mosquitoes, thereby, increasing its population. Thus, we expect a positive association between malaria SIRs in the region and precipitation, however, in our survey we found a negative correlation between precipitation and malaria risk. This negative correlation can be explained with huge amounts of rain that wash the ground and kill the eggs.\(^\text{35}\)

In addition, our statistical modeling revealed a positive
effect of temperature and humidity on malaria SIRs. These findings confirm the obtained results from other studies in different countries.\textsuperscript{27,34,36} Factors related with population vulnerability are also critically important in malaria transmission. The presence of parasite resistance to the usual anti-malaria and to insecticides, population movements, and the presence of other underlying infections (namely, HIV) are responsible for a large part of the variability in the incidence of malaria. Considering all these factors, in addition to environmental variables in statistical models, to predict malaria incidence is complex, and not yet well understood. This has led some researchers to develop models, in which incidence rates are standardized with respect to non-climatic variables, so that the influence of climate on fluctuations in the malaria rate can be seen more clearly.\textsuperscript{33,37} In this study, we also used the described model to obtain the predicted values of SIRs in different districts of SBP. Differences between the observed and predicted malaria SIRs, are highly due to other above-mentioned effective factors, which were not considered in our statistical modeling. Lack of information on some important factors such as patients’ educational level, socio-economic status, occupation, traditional bedroom type was the major limitation of this study.

In general, the present study showed significant association between environmental factors and malaria risk in SBP. Since these environmental factors are out of human’s control, the health policy makers in this province should pay more attention to the areas with high temperature, elevation, and humidity, as well as, low rainfall districts.

References

4. Ministry of Health and Medical Education of Iran. Malaria Center. [update 2008 Oct 7]. Available at URL: \url{http://www.mohme.gov.ir/health/index.htm}
21. Statistical Center of Iran. Index of publications. [update 2008 October]. Available at URL: \url{http://amar.sci.org.ir/}

### Related topics

Appendix A

Spatial Kriging: Assume suppose $Z(s_1), Z(s_2), \ldots, Z(s_n)$ are the outcome data associated with spatial locations $s_1, s_2, \ldots, s_n$. We are interested in predicting an unobserved variable $Z(s)$. An optimal predictor of $Z(s)$ can be written as

$$\hat{Z}(s) = \sum_{i=1}^{n} w_i Z(s_i)$$

Where the weights $w_i$ are chosen to minimize the mean-squared prediction error. We chose a second order stationary process with marginal moments

$$E(Z(s)) = m(s) = x'(s) \beta$$
$$cov(Z(s_i), Z(s_j)) = \sigma_e^2 + \sigma^2 \rho(s_i - s_j)$$

Where $x(s_i)$ is a $p \times 1$ vector of explanatory variables, $\beta$ is a $p \times 1$ vector of unknown model parameters, $\sigma_e^2 + \sigma^2$ is the variance of the data, $\sigma_e^2$ is a residual or nugget variance and $\rho(s_i - s_j)$ is the correlation function of the process.

GLM for spatial data: Linear models and linear predictors may not be the best choices for many applications, particularly when the data are binary or counts. In this context, GLMs are the most common alternative. We assume that some function of marginal mean, called the link function, related to the spatial surface, so that

$$g[m(s)] = x'(s) \beta$$

The two most common link functions are log for count data and the logit for binomial data. Since the SIR data has a count nature, the log transformation was used for modeling the malaria SIRs in this paper.
Thus, we used the following model for describing the statistical relationship between environmental explanatory variables and SIR values:

\[ \log(z) = \log(n) + x'(s)\beta + \varphi_i \]

Where \( \log(z) \) is the logarithm of the number of observed malaria cases, \( \log(n) \) is an offset, \( \beta \) consist of \( \beta_0 \) as an intercept and \( \beta_i \) as regression coefficient vector corresponding to the environmental variables (precipitation, humidity, elevation and temperature), and \( \varphi_i \) is the spatial random effect for district \( i \). In particular, if \( d_{ij} \) denotes the distance between points \( i \) and \( j \), where observations \( z_i \) and \( z_j \) were made \( V = 1\sigma^2 + F\sigma^2 \), and \( F_{ij} = \exp \left(-\frac{d_{ij}}{\rho} \right) \). The unknown parameters in this model, \( \sigma^2 \) (the nugget), \( \sigma^r \) (the sill) and \( \rho \) (the range), can be estimated from the variogram of the data \( \theta = \begin{bmatrix} \sigma^2 \\ \sigma^r \\ \rho \end{bmatrix} \) as a vector of. For non-normally distributed data \( \theta \) is specified in the macro call to enable GLIMMIX to estimate the appropriate correlation matrix for spatially correlated Poisson model.
Appendix B

Attach (SBP)
library (gstat)
sir.var<- variogram(SIR~1, loc=-X+Y, SBP)
log.sir.var<- variogram(log(SIR+0.000001)-1, loc=-X+Y, SBP)
print (sir.var)
print (log.sir.var)
par (mfrow=c(2,2))
plot (sir.var)
plot (log.sir.var)

log.sircr.var<- variogram(log(SIR+0.000001)-1, loc=-X+Y, SBP)
print (log.sircr.var)
plot log.sircr.var
plot (log.sircr.var$dist, log.sircr.var $gamma, xlab="distance", ylab="Varigram")
plot (variogram(log(SIR+0.000001)-1, loc=-X+Y, data=SBP, cloud=TRUE))

model.1 <- fit.variogram(log.sircr.var, vgm(psill=1,model="Sph", range=100,nugget=1))
plot(log.sircr.var, model=model.1)
title("Best Fit Spherical")

model.2 <- fit.variogram(log.sircr.var, vgm(psill=12,model="Gau", range=45000,nugget=1))
model.3 <- fit.variogram(log.sircr.var, vgm(psill=5,model="Exp", range=20000,nugget=1))
model.4 <- fit.variogram(log.sircr.var, vgm(psill=10,model="Mat", range=20000,nugget=1))

m <- vgm(psill=130.02,"Exp", range=1324948,nugget=3.85)
log.sircr.krige <- krige(log(SIR+0.000001)-1, ~X+Y, model = m, data = SBP, newd = loc)
log.sircr.krige