Spatial analysis of distribution of dengue cases in Espírito Santo, Brazil, in 2010: use of Bayesian model

ABSTRACT: **Objective:** To study the relationship between the risk of dengue and sociodemographic variables through the use of spatial regression models fully Bayesian in the municipalities of Espírito Santo in 2010. **Method:** This is an ecological study and exploration that used spatial analysis tools in preparing thematic maps with data obtained from SinanNet. An analysis by area, taking as unit the municipalities of the state, was performed. Thematic maps were constructed by the computer program R 2.15.00 and Deviance Information Criterion (DIC), calculated in WinBugs, Absolut and Normalized Mean Error (NMAE) were the criteria used to compare the models. **Results:** We were able to geocode 21,933 dengue cases (rate of 623.99 cases per 100 thousand habitants) with a higher incidence in the municipalities of Vitória, Serra and Colatina; model with spatial effect with the covariates trash and income showed the best performance at DIC and Nmae criteria. **Conclusion:** It was possible to identify the relationship of dengue with factors outside the health sector and to identify areas with higher risk of disease. **Keywords:** Epidemiology and Biostatistics. Dengue. Linear models. Social determinants of health. Spatial analysis. Bayesian Inference.
INTRODUCTION

Dengue is an infectious disease caused by an arbovirus — the dengue virus (DENV) — with four known serotypes. It mainly affects tropical developing countries, where sociocultural conditions favor the spread of the disease and the proliferation of its main vector, the female mosquito *Aedes aegypti*.1

For its endemic and epidemic character of great magnitude, it has an important social and economic impact, since a significant proportion of affected individuals are of working age.2

Several strategies for more effective approaches to disease control have been adopted in Brazil. The training of health professionals and the structuring of services appropriate to the care of individuals with dengue are examples of important measures to reduce mortality from this condition. However, the primary means of control of a dengue epidemic is the control of its vector, especially in places where there are more cases reported.

In Brazil, since 1975, dengue is a disease with mandatory reporting, which is obligatory for all health professionals, as well as to those responsible for public and private health and education organizations and facilities.3

The data contained in the written report form can provide information for an assessment of the population strata at risk. The identification of these strata aggregated to social groups and well defined geographical areas allows the establishment of prevention and control actions for specific areas.4 Allied to this, the development of techniques for spatial...
analysis and Geographic Information Systems (GIS) is considered an important tool in mapping the risk of dengue.

Previous studies have postulated that the emergence of large urban areas with inadequate housing, water supply and garbage collection are important social determinants for the increased risk of incidence and maintenance of the number of dengue cases.

Recently, the fully Bayesian approach model is proposed as an alternative to the traditional models for the study of spatial analysis of diseases. The advantage of this model was the assumption that the Bayesian modeling could minimize the variance of the estimators, especially in places where the population is small.

In this approach, variables that aim to capture influences of factors specific to the areas are introduced, as well as the dependency between these areas, thereby generating more accurate estimates of associations between disease incidence and factors related to the areas under investigation.

Thus, in studies on social determinants of health, where heterogeneity is large, understanding if the variability that occurs in the data analysis is due to random fluctuations or spatial effect itself is of utmost importance.

Given the above, this study aimed to investigate the relationship between the risk of dengue and sociodemographic variables through the use of fully Bayesian spatial regression models in the municipalities in the state of Espírito Santo in 2010.

**METHODOLOGY**

**DATA UTILIZED**

An ecological study of the spatial distribution of reported cases of dengue in the state of Espírito Santo in 2010, according to the municipality of residence, was conducted. Data regarding the disease were obtained from the Ministry of Health through the Department of Technology of Unified Health System (DATASUS), and the cartographic database was obtained from the website of the Jones dos Santos Neves Institute (IJSN).

The data on the covariates considered were obtained from the Brazilian Institute of Geography and Statistics (IBGE) based on the results of the Demographic Census of 2010.

**STATISTICAL ANALYSIS**

The data used in the study were aggregated by municipality. On the map, each area (municipality) is shaded according to the respective risk values. One way to do this is through gross rates, given by: $GR_i = \frac{cases_i}{pop_i} \times (1,000)$, where $cases_i$ and $pop_i$ represent the observed number...
of cases and the population of municipality \( i \), respectively. However, these gross rates have high instability when associated with small populations.

Regarding the proposal of a model that better estimates the risk of dengue, the methodology of a fully Bayesian hierarchical model was applied. Based on this methodology, covariates were incorporated into the model in order to find factors that contribute to the occurrence of the event. In this proposed modeling structure, the inclusion of random variables that reflect any local (\( \theta_i \)) and spatial (\( \phi_i \)) effects, aiming to find better risk estimates for each municipality.

Although no consensus exists on what factors really influence the increase or decrease in the risk of dengue in a given area or region\(^4\), based on the variables provided by the 2010 census, we opted to use the following covariates: literacy (proportion of illiterate individuals above 15 years of age); water (ratio of permanent private households whose form of water supply does not occur by the general water supply network); garbage (proportion of permanent private households where garbage is burned, buried, thrown in a vacant lot or on the street, river, lake, sea, or other destination); and income (proportion of people aged 10 years or more of age and monthly income of less than three minimum wages).

In this model, dengue was considered the response variable, with the following factors being evaluated in the fully Bayesian regression model: the spatial effect and the covariates garbage, water and literacy (including in isolation and combined).

We first consider \( Y_i \) as the number of dengue cases in municipality \( i \), \( i = 1,2,...,78 \), where \( Y_i \) has binomial distribution. However, in epidemiological studies, the binomial model can be approximated by distribution\(^1\).

The Bayesian spatial hierarchical model\(^1\) has three levels of hierarchy. The first is given basically by:

\[
Y_i \mid \Psi_i, E_i \sim \text{Poisson}(\mu_i = E_i \Psi_i), \quad i = 1,2,...,78
\]

Where \( \Psi_i \) and \( E_i \) are, respectively, the relative risk of dengue and the expected number of dengue cases in municipality \( i \). The calculation performed to obtain the expected number of dengue cases in each county number was given by:

\[
E_i = \frac{\sum j Y_{ij}}{\sum j \text{pop}_{ij}}, \quad i = 1,2,...,78
\]

It should be noted that it is not always easy to obtain mathematical solutions to calculate the risk of dengue in each municipality; thus, the natural logarithm of the risks considered was \( \Psi_i \):

\[
\log(\Psi_i) = \alpha_0 + \alpha X + \theta_i + \phi_i, \quad \Psi_i > 0
\]
Where $\Psi_i$, as mentioned before, is the risk in municipality $i$, $\alpha_i$ is the model intercept, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)$ represents the coefficients of the covariates, $\theta_i$ is the specific random effect for each municipality, and $\phi_i$ is the spatial effect of neighboring municipalities.

The spatial effect $\phi_i$, as can be seen, takes into account the spatial dependence among risks. The Bayesian approach considers each parameter as a random variable, and this is the second level of hierarchy, where distributions are specified a priori for each one with the objective of obtaining the a posteriori distribution of risks (where risk estimates are obtained). Thus, it is assumed that $\alpha$ has normal distribution with mean equal to 0 and covariance matrix $\Sigma = \sigma^2_{\alpha} I$, where $\sigma^2_{\alpha} = 1,000$. The random effect $\theta_i$ is specified as follows: $\theta_i \sim N(0, \tau_\theta)$. And $\phi_i$ has conditional auto regressive (CAR) a priori distribution, which also depends on a $\tau_\phi$ parameters. Both $\tau_\theta$ and $\tau_\phi$ are called hyperparameters, because they are a priori parameters.

The third level is the specification of a priori distributions for each hyperparameter, called hiperpriors: $\tau_\theta \sim \text{Gamma}(0.5; 0.005)$ and $\tau_\phi \sim \text{Gamma}(0.5; 0.005)$, which are uninformative distributions.

The neighborhood structure adopted in this article is called binary neighborhood, which takes into account the fact of one municipality bordering another.

The criterion for selecting the best model adopted is known as deviance information criterion (DIC) and, based on that information, the model with the lowest DIC was selected as the model that best estimates the risk of dengue. After the comparison of models with best fit from the DIC, the middle normalized absolute error (NMAE) was calculated for the three models with lower DICs, thus providing another indicator of adjustment.

The trajectory of three chains was considered, starting from different initial values, and only one value in every 100 values generated was included in the a posteriori sample. The convergence was analyzed by the method of Gelman and Rubin. To ensure convergence, 250,000 interactions were generated, disregarding the first 50,000.

The whole inferential process was based on computational methods for Monte Carlo Markov chain (MCMC) in WinBugs software, version 1.4. While for statistical analysis and risk mapping, we used the R software, version 2.15.0.

RESULTS

The State of Espírito Santo, according to the 2010 Census, has 78 municipalities, 3,514,952 inhabitants and an area of 46,095.583 km$^2$, which represents a population density of 76.25 inhabitants/km$^2$.

In 2010, Espírito Santo had 21,933 cases of dengue, with the highest concentration in the metropolitan region of Vitória and in Colatina. Of the 78 municipalities, 6 showed no dengue cases reported in 2010, which meant that the gross rate was 0. They are: Águia Branca, Brejetuba, Divino de São Lourenço, Fundão, Ibitirama and Vila Pavão.
Table 1 shows the correlation matrix between the variables, and a strong association between the covariates water and garbage can be observed. Therefore, the combination of these covariates between the proposed models was not considered.

Thus, in the regression model, 14 models were evaluated with their respective values of DIC and NMAE (Table 2): (1) no spatial effect, (2) with spatial effect, (3) with spatial effect, with covariate trash (4) with spatial effect, with covariate water, (5) with spatial effect, with covariate literacy, (6) with spatial effect, with covariate income, (7) with spatial effect, with covariates water and literacy (8) with spatial effect, with covariates water and income, (9) with spatial effect, with covariates income and literacy (10) with spatial effect, with covariates trash and literacy (11) with spatial effect, with covariates trash and income (12) without spatial effect, with covariates trash and income (13) with spatial effect, with covariates trash, income and literacy, and (14) with spatial effect, with covariates, water, income and literacy.

Among the 14 models proposed, model 11 showed the smallest value of DIC (517.504). In this model, the spatial effect and the presence of covariates garbage and income were observed. The second best model was model 13, with spatial effect and composed by the

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>NMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No spatial effect</td>
<td>519.065</td>
</tr>
<tr>
<td>2</td>
<td>With spatial effect</td>
<td>519.076</td>
</tr>
<tr>
<td>3</td>
<td>With spatial effect, with covariate trash</td>
<td>519.658</td>
</tr>
<tr>
<td>4</td>
<td>With spatial effect, with covariate water</td>
<td>519.022</td>
</tr>
<tr>
<td>5</td>
<td>With spatial effect, with covariate literacy</td>
<td>518.297</td>
</tr>
<tr>
<td>6</td>
<td>With spatial effect, with covariate income</td>
<td>518.916</td>
</tr>
<tr>
<td>7</td>
<td>With spatial effect, with covariates water and literacy</td>
<td>518.515</td>
</tr>
<tr>
<td>8</td>
<td>With spatial effect, with covariates water and income</td>
<td>518.277</td>
</tr>
<tr>
<td>9</td>
<td>With spatial effect, with covariates income and literacy</td>
<td>518.576</td>
</tr>
<tr>
<td>10</td>
<td>With spatial effect, with covariates trash and literacy</td>
<td>519.316</td>
</tr>
<tr>
<td>11</td>
<td>With spatial effect, with covariates trash and income</td>
<td>517.502</td>
</tr>
<tr>
<td>12</td>
<td>Without spatial effect, with covariates trash and income</td>
<td>518.161</td>
</tr>
<tr>
<td>13</td>
<td>With spatial effect, with covariates trash, income and literacy</td>
<td>517.779</td>
</tr>
<tr>
<td>14</td>
<td>With spatial effect, with covariates, water, income and literacy</td>
<td>517.985</td>
</tr>
</tbody>
</table>

DIC: deviance information criterion; NMAE: normalized mean absolute error.
covariates garbage, income and literacy. Following that, model 14, with spatial effect and composed by the covariates water, income and literacy. Then, the inclusion of the effect of the combination of variables garbage and income to model 11 has been evaluated; however, there was no improvement in the adjustment (DIC = 518.220).

The NMAE was calculated for these three models (11, 13 and 14), and model 11 was also the one with the lowest NMAE value, being, therefore, the model that best explains the risk of dengue.

The number of dengue cases estimated from the selected model (model 11) is given by:

\[ Y_i \sim \text{Poisson}(\mu_i = E_i \Psi_i) \]

\[ \log(\Psi_i) = -41.9 + 0.2486 \text{ (garbage)} + 0.4983 \text{ (income)} + \theta_i + \phi_i \]

\[ \mu_i = E_i \exp(\log(\Psi_i)) = E_i + \exp(-41.9 + 0.2486 \text{ (garbage)} + 0.4983 \text{ (income)} + \theta_i + \phi_i) \]

Figure 1A shows the map of the gross rate of dengue, and Figure 1B shows the map of estimates from model 11. Out of the 78 municipalities, five of them have values of crude rate higher than 14.7 per one thousand inhabitants - Itarana, Guaçuí, São Gabriel da Palha, Colatina and Marilândia. However, when considering estimates from model 11 (with spatial effect and covariates garbage and income), the municipality of Itarana was no longer included in this group. Regarding the gross rate maps and model 11, it is highlighted in Figure 1B that there was a homogenization of rates, decreasing the effect of “patchwork” of the gross
rate image. Also with respect to the model found, which shows the significance of specific effects of the municipalities, as well as neighborhood effects, it can be concluded that the inclusion of these contributed to the correction of overdispersion and underdispersion of the Poisson model, allowing, from the results, a better assessment of the incidence of dengue correlated with characteristics of each municipality and its surroundings.

DISCUSSION

In 2010, dengue was distributed spatially in the state of Espírito Santo in a non-homogeneous manner, with a higher prevalence of cases in the capital, Vitória, and in the municipalities of Serra and Colatina. In the municipalities of Águia Branca, Brejetuba, Divino de São Lourenço, Fundão Ibitirama and Vila Pavão, no cases were notified. In the fully Bayesian approach, it was observed that the covariates garbage and income were the best to estimate the risk of dengue in the state in 2010.

Some limitations, however, should be mentioned. First, the inclusion of data only from 2010 could result in a reporting bias for the fluctuation of disease incidence. The previous assessment shows that, in Espírito Santo, incidences of 740.06 and 952.84 per hundred thousand inhabitants in 2008 and 2009, respectively, were higher than in 2010 (623.99/100,000). It is also important to emphasize that the use of a set of data for a single year does not impose obstacles to the implementation of the modeling proposed here, considering the information available in the year of the census in the country. Another limitation would be underreporting, arising mainly from private services, already reported in other studies. This underreporting may interfere with covariates found in the model studied (proper garbage collection and income), as populations insured by private plans have different socioeconomic conditions than those assisted by the public health system. Another issue was the choice of the covariates included in the model. Given the availability, only socioeconomic variables contained in the census of 2010 were included in this study; therefore, covariates that might be associated with the risk of dengue, but which were not part of this database, cannot be analyzed.

Our finding of a higher incidence of dengue in areas with high population density agrees with the results of other studies. This association has been explained by social inequality in these urban centers, with varying socioeconomic strata. There are several indicators of socioeconomic conditions that can affect health. In this study, the covariates that best explained the risk of dengue were inadequate destination of the garbage and monthly income below three minimum wages, corroborating other studies. In São José do Rio Preto (SP), dengue was related to inadequate garbage collection and lack of sewage system. In Porto Alegre (RS), the variable income was the only one that satisfactorily explained the spatial distribution of the disease. On the other hand, in a study conducted in Nova Iguaçu (RJ), the authors found a low correlation between the
living conditions studied (garbage collection, literacy, income, population density etc.) and the risk of dengue.

Regarding the application of the fully Bayesian model, we emphasize the homogenization of rates between neighboring municipalities in relation to the gross ratio, which facilitates the understanding of the spatial pattern of dengue in the state, because it reduces random noise, especially in municipalities with small populations, as is the case of the municipalities of Espírito Santo. Thus, the method allowed the identification of municipalities with more priority for control and, therefore, it would have a potential for improved surveillance. The results of this study and the aforementioned studies seem to reinforce the thesis that social inequalities are related to the risk of dengue in urban areas.

Recently, different initiatives and control programs have sought alternatives for dengue prevention and health promotion through community mobilization and public intervention in areas of difficult access for garbage collection and vector surveillance. Although such interventions do not target significantly the social structure that generates socioeconomic inequalities, they have the potential to reduce current rates of disease by acting mainly on vector control, which could minimize the risks for groups more vulnerable to disease.

CONCLUSION

This study found that the incidence of dengue in the state of Espirito Santo in 2010 is mainly related to the covariates: inadequate garbage collection and income below three minimum wages. Moreover, it is observed that the municipalities that had higher risk are concentrated in metropolitan Vitória, highlighting the importance of these social determinants of health on the risk of dengue. These determinants, in turn, are associated with housing, urban infrastructure, sociocultural profile of the population, among others, that determine the conditions of life in these places and that contribute with sickening by dengue. Given the above, the most effective control measures should prioritize action strategies that could change the social structure generating these socioeconomic inequalities.

REFERENCES


Received on: 05/16/2012
Final version presented on: 03/14/2013
Accepted on: 06/12/2013