

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

Análise espacial do risco de dengue no Espírito Santo, Brasil, 2010: uso de modelagem completamente Bayesiana

Taizi Honorato^I, Priscila Pagung de Aquino Lapa^I, Carolina Maia Martins Sales^{II}, Barbara Reis-Santos^{III}, Ricardo Tristão-Sá^{IV}, Adelmo Inácio Bertolde^V, Ethel Leonor Noia Maciel^{VI}

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.

^ISchool of Statistics, *Universidade Federal do Espírito Santo* – Vitória (ES), Brazil.

^{II}School of Nursing, *Universidade Federal do Espírito Santo* – Vitória (ES), Brazil.

^{III}Center for Infectious Diseases, *Universidade Federal do Espírito Santo* – Vitória (ES); *Universidade Federal de Pelotas* – Pelotas (RS), Brazil.

^{IV}School of Medicine of *Santa Casa de Misericórdia* – Santa Luíza (ES), Brazil.

^VCenter for Exact Sciences, *Universidade Federal do Espírito Santo* – Vitória (ES), Brazil.

^{VI}Graduate Program in Public Health, *Universidade Federal do Espírito Santo* – Vitória (ES), Brazil.

Corresponding author: Adelmo Inácio Bertolde. *Universidade Federal do Espírito Santo*, Centro de Ciências Exatas, Departamento de Estatística. Avenida Fernando Ferrari, s/n, Goiabeiras, CEP: 29060-090, Vitória, ES, Brasil. E-mail: adelmoib@gmail.com

Conflict of interests: nothing to declare – **Financing source:** none.

RESUMO: *Objetivo:* Estudar a relação entre o risco de dengue e as variáveis sociodemográficas através da utilização de modelos de regressão espacial completamente Bayesianos nos municípios do Estado do Espírito Santo no ano de 2010. *Métodos:* Trata-se de um estudo ecológico e exploratório que utiliza as ferramentas de análise espacial na elaboração de mapas temáticos, com dados obtidos do SinanNet. Foi feita uma análise por área, tomando-se como unidade os municípios do Estado. Os mapas temáticos foram construídos pelo programa computacional R 2.15.00 e o *Deviance Information Criterion* (DIC), calculado no WinBugs, e o *Normalized Mean Absolut Error* (NMAE) foram os critérios utilizados para a comparação dos modelos. *Resultados:* Foram georreferenciados 21.933 casos de dengue (taxa de 623,99 casos/100 mil habitantes) com maior incidência nos municípios de Vitória, Serra e Colatina; o modelo com efeito espacial, com as covariáveis lixo e renda foi o que apresentou melhor desempenho, segundo os critérios DIC e NMAE. *Conclusão:* Foi possível identificar a relação do dengue com fatores externos ao setor saúde e identificar locais com maior risco da doença.

Palavras-chave: Epidemiologia e Bioestatística. Dengue. Modelos lineares. Determinantes sociais da saúde. Análise espacial. Inferência Bayesiana.

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 age².

Several strategies for more effective approaches to disease control have been adopted in Brazil¹. 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 facilities³.

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 areas⁴. 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 dengue⁵⁻⁷.

Previous studies^{8,9} 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 small¹⁰.

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^{10,11}.

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)¹², and the cartographic database was obtained from the website of the Jones dos Santos Neves Institute (IJSN)¹³.

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 2010¹⁴.

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: $GR_i = \frac{cases_i}{pop_i} (1,000)$, 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 (θ_i) and spatial (ϕ_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 region⁴, 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 , $i = 1, 2, \dots, 78$, where Y_i has binomial distribution. However, in epidemiological studies, the binomial model can be approximated by distribution¹¹.

The Bayesian spatial hierarchical model¹¹ has three levels of hierarchy. The first is given basically by:

$$Y_i \mid \Psi, E_i \sim \text{Poisson}(\mu_i = E_i \Psi_i), \quad i = 1, 2, \dots, 78$$

Where Ψ_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:

$$E_i = \frac{\sum y_i}{\sum pop_i} (pop_i), \quad i = 1, 2, \dots, 78$$

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 Ψ_i :

$$\log(\Psi_i) = \alpha_0 + \alpha X + \theta_i + \phi_i, \quad \Psi_i > 0$$

Where Ψ_i , as mentioned before, is the risk in municipality i , α_0 is the model intercept, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$ represents the coefficients of the covariates, θ_i is the specific random effect for each municipality, and ϕ_i is the spatial effect of neighboring municipalities.

The spatial effect ϕ_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 α has normal distribution with mean equal to 0 and covariance matrix $\Sigma = I\sigma_\alpha^2$, where $\sigma_\alpha^2 = 1,000^{12}$. The random effect θ_i is specified as follows: $\theta_i \sim N(0, \tau_\theta)$. And ϕ_i has conditional auto regressive (CAR) *a priori* distribution, which also depends on a τ_ϕ parameters. Both τ_θ and τ_ϕ are called hyperparameters, because they are *a priori* parameters.

The third level is the specification of *a priori* distributions for each hyperparameter, called *hyperpriors*: $\tau_\theta \sim \text{Gamma}(0,5; 0,005)$ 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)¹¹ 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 Rubin¹⁵. 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 km², which represents a population density of 76.25 inhabitants/km².

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 1. Matrix of correlations between the covariates included in the spatial regression model for dengue in Espírito Santo, 2010.

	Income	Literacy	Water	Garbage
Income	1			
Literacy	0.418	1		
Water	0.646	0.447	1	
Garbage	0.550	0.642	0.808	1

Table 2. Results of deviance information criterion and normalized mean absolute error of the spatial regression models for dengue in Espírito Santo, 2010.

	Model	DIC	NMAE
1	No spatial effect	519.065	
2	With spatial effect	519.076	
3	With spatial effect, with covariate trash	519.658	
4	With spatial effect, with covariate water	519.022	
5	With spatial effect, with covariate literacy	518.297	
6	With spatial effect, with covariate income	518.916	
7	With spatial effect, with covariates water and literacy	518.515	
8	With spatial effect, with covariates water and income	518.277	
9	With spatial effect, with covariates income and literacy	518.576	
10	With spatial effect, with covariates trash and literacy	519.316	
11	With spatial effect, with covariates trash and income	517.502	0.502
12	Without spatial effect, with covariates trash and income	518.161	
13	With spatial effect, with covariates trash, income and literacy	517.779	0.604
14	With spatial effect, with covariates, water, income and literacy	517.985	0.578

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

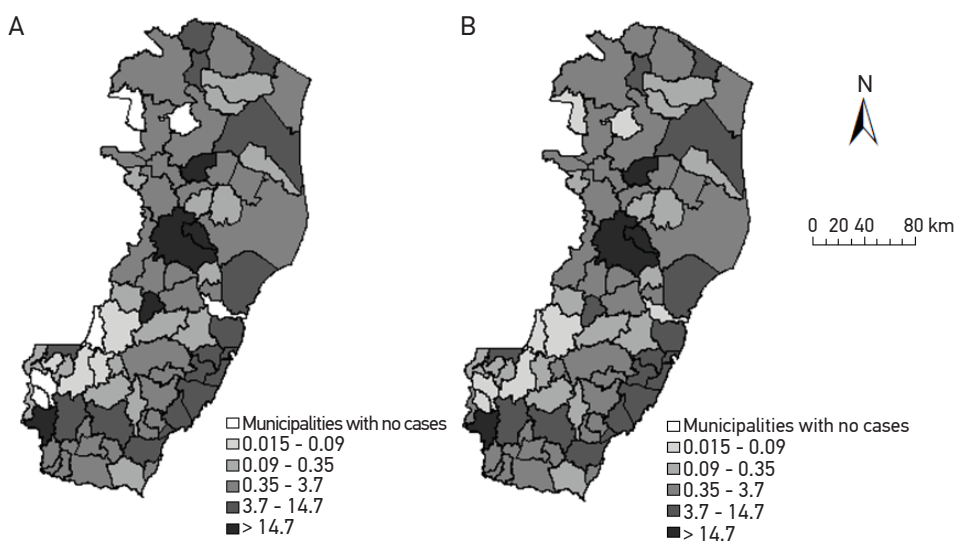


Figure 1. (A) Map of the gross rate of dengue and (B) map from the fully Bayesian model estimates with spatial effect and covariates garbage and income for Espírito Santo in 2010.

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^{16,17}. 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^{18,19}. This association has been explained by social inequality in these urban centers, with varying socioeconomic strata^{20,21}.

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^{8,22}. In São José do Rio Preto (SP), dengue was related to inadequate garbage collection and lack of sewage system⁸. In Porto Alegre (RS), the variable income was the only one that satisfactorily explained the spatial distribution of the disease²². 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 dengue²³.

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 Santo¹⁰. 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 Espírito 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

1. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Guia de Vigilância Epidemiológica. 6ª edição. Brasília: Ministério da Saúde; 2005. p. 231-53
2. World Health Organization. Dengue: guidelines for diagnosis, treatment, prevention and control. Geneva: WHO Library Cataloguing-in-Publication Data; 2009.
3. Brasil. Lei nº 6.259, de 30 de outubro de 1975. Dispõe sobre a organização das ações de Vigilância Epidemiológica, sobre o Programa Nacional de Imunizações, estabelece normas relativas à notificação compulsória de doenças, e dá outras providências. Brasília, 30 de outubro de 1975, 154º da Independência e 87º da República.

4. Carvalho MS, Santos RS. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cad Saúde Pública* 2005; 21(2): 361-78.
5. Ferreira GS. Análise espaço-temporal da distribuição dos casos de dengue na cidade do Rio de Janeiro no período de 1986 a 2002 [dissertação de mestrado]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2004.
6. Barcellos C, Pustai AK, Weber MA, Brito MRV. Identificação de locais com potencial de transmissão de dengue em Porto Alegre através da técnica de geoprocessamento. *Rev Soc Bras Med Trop* 2005; 38(3): 246-50.
7. Flauzino RF, Souza-Santos R, Oliveira RM. Dengue, geoprocessamento e indicadores socioeconômicos e ambientais: um estudo de revisão. *Rev Panam Salud Publica* 2009; 25(5): 456-61.
8. Costa AIP, Natal D. Distribuição espacial da dengue e determinantes socioeconômicos em localidade urbana no Sudeste do Brasil. *Rev Saúde Pública* 1998; 32(3): 232-6.
9. Machado JP, Oliveira RM, Souza-Santos R. Análise espacial da ocorrência de dengue e condições de vida na cidade de Nova Iguaçu, Estado do Rio de Janeiro, Brasil. *Cad Saúde Pública* 2009; 25(5): 1025-34.
10. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. *J R Stat Soc Ser C Appl Stat* 1991; 40(2): 283-94.
11. Spiegelhalter DJ, Best NG, Carlin BP, Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 2002; 64(4): 583-639.
12. Epidemiológicos e morbidade. Disponível em: <http://datasus.saude.gov.br/informacoes-de-saude>. (Acessado em 21 agosto de 2012).
13. Bases Geográficas. Disponível em: <http://www.ijsn.es.gov.br/Sitio/#>. (Acessado em 21 de agosto de 2012).
14. Instituto Brasileiro de Geografia e Estatística. Censo 2010. Disponível em: <http://www.censo2010.ibge.gov.br>. (Acessado em 21 de agosto de 2012).
15. Gelman A, Rubin D. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; 7(4): 457-511.
16. Teurlai M, Huy R, Cazelles B, Duboz R, Baerh C, Vong S. Can human movements explain heterogeneous propagation of dengue fever in Cambodia? *Plos Negl Trop Dis* 2012; 6(12): e1957.
17. Gomes AF, Nobre AA, Cruz OG. Temporal analysis of the relationship between dengue and meteorological variables in the city of Rio de Janeiro, Brazil, 2001-2009. *Cad Saúde Pública* 2012; 28(11): 2189-97.
18. Hu W, Clements A, Williams G, Tong S, Mengersen K. Spatial patterns and socioecological drivers of dengue fever transmission in Queensland, Australia. *Environ Health Perspect* 2012; 120(2): 260-6.
19. Schmidt WP, Suzuki M, Thiem VD, White RG, Tsuzuki A, Yoshida LM et al. Population density, water supply, and the risk of dengue fever in Vietnam: cohort study and spatial analysis. *PLoS Med* 2011; 8(8): e1001082.
20. Flauzino RF, Souza-Santos R, Oliveira RM. Dengue, geoprocessamento e indicadores socioeconômicos e ambientais: um estudo de revisão. *Rev Panam Salud Publica* 2009; 25(5): 456-61.
21. Carvalho RM, Nascimento LF. Spatial distribution of dengue in the city of Cruzeiro, São Paulo State, Brazil: use of geoprocessing tools. *Rev Inst Med Trop São Paulo* 2012; 54(5): 261-6.
22. Barcellos C, Pustai AK, Weber MA, Brito MRV. Identificação de locais com potencial de transmissão de dengue em Porto Alegre através de técnicas de geoprocessamento. *Rev Soc Bras Med Trop* 2005; 38(3): 246-50.
23. Machado JP, Oliveira RM, Souza-Santos R. Análise espacial da ocorrência de dengue e condições de vida na cidade de Nova Iguaçu, Estado do Rio de Janeiro, Brasil. *Cad Saúde Pública* 2009; 25(5): 1025-34.

Received on: 05/16/2012

Final version presented on: 03/14/2013

Accepted on: 06/12/2013