Scientific
Research
Publishing

Analyzing Spatial Patterns of Cardiorespiratory Diseases in the Federal District, Brazil

Weeberb Requia, Henrique Roig

Campus Darcy Ribeiro, University of Brasilia, Brasilia, Brazil
Email: weeberb@gmail.com

Abstract

Cardiorespiratory diseases are a serious public health problem worldwide. Identification of spatial patterns in health events is an efficient tool to guide public policies in environmental health. However, only few studies have considered spatial pattern analysis which is considered the evaluation of spatial autocorrelation, degree of autocorrelation and dependence behavior in terms of distances. Therefore, the objective of this study is to propose a set of procedures to evaluate the spatial patterns of cardiorespiratory diseases in the Federal District, Brazil. Specifically, our proposal will be based on four questions: a) is the spatial distribution of all patients clustered, random or dispersed? b) what is the degree of clustering for either high values or low values of patients? c) what is the spatial dependence behavior? d) considering the spatial variation, at what distance does the type of distribution (cluster, random or disperse) begin to change? We chose four methods to answer these questions Global Moran's I (question "a"); Getis-Ord General G (question "b"); semivariogram analysis (question "c"); and multi-distance spatial cluster-K-function (question "d"). Our results suggest that there is a different behavior for people up to 5 years old (cluster, $p < 0.01$), especially in distances below 2.5 km. For people above 59 years old, cluster is significant just in short distances (<200 m). For other age groups, the spatial distribution is basically random. Our study showed that it was possible to capture evidences of health disparities in the Federal District.

1. Introduction

Cardiorespiratory diseases are a serious public health problem worldwide [1]. According to World Health Or-

ganization [2], in 2012 cardiovascular and respiratory diseases were responsible for 17.5 million and 4 million deaths globally, respectively. Children and seniors are the age groups more susceptive to present cardiorespiratory illness [3] [4].

Ning *et al.* (2012) [5] and Troncoso *et al.* (2012) [6] showed that the identification of spatial behavior in health events was an efficient tool to guide public policies. The benefits of this tool are directly related with public health and government spending [7] [8].

Many studies have applied geostatistical methods to evaluate the cause and effect relationships between two or more variables, such as occurrence of diseases due to environmental factor [9]-[11]. However, only a few studies have considered spatial pattern analysis of the outcomes as a pre-analysis of cause and effect. An example of these limited studies was the work of Zou *et al.* (2014) [12], who analyzed the spatial clustering of air pollution exposure across the United States.

Mitchell (1999) [13] and Kurland *et al.* (2012) [14] highlighted that the spatial pattern was an important analysis before any evaluation of spatial relationships between two or more variables. Identifying patterns is a technique that allows a better understanding of geographic phenomena; in other words, it is a guide to monitor conditions on the ground, to calculate temporal changes and to compare populations [13].

Thus, the objective of this study is to propose a set of procedures to evaluate the spatial patterns of cardiorespiratory diseases. Specifically, our proposal will be based on four questions: 1) is the spatial distribution of all patients clustered, random or dispersed? 2) what is the degree of clustering for either high values or low values of patients? 3) what is the spatial dependence behavior? 4) considering the spatial variation, at what distance does the type of distribution (cluster, random or disperse) begin to change?

2. Materials and Methods

2.1. Study Area and Data

This study used data from Federal District (FD), the Brazilian state in which the city of Brasília, Brazil's capital, is located. FD has an area of 5802 km^2 and an estimated population of 2.7 million people. It is the fourth most populous city in Brazil [15].

Three types of data were used in this study: a) health data, b) geographic database of addressing, and c) a demographic census.

Health data were provided by the Individual Health System [16]. These data included the total number of individuals who were admitted to all FD hospitals between 2008 and 2013 for circulatory and respiratory system illness. This information was classified by the residential address and by age, including 0 to 5, 6 to 17, 18 to 59, and older than 59 years. To preserve the privacy of each patient, Datasus (2013) does not provide the complete address of patients. The address data provided are detailed to the level of the street number (first three levels of the address system in FD: 1—administrative region, 2—sector, 3—street). We did not have access to house number, that is, the fourth and last level of addresses in FD.

The geographic addressing database was provided by the Urban State Secretary of FD [17]. These data are represented by vectors (polygons) and cover the geographic addressing of all FD.

The demographic census data were provided by the Brazilian Institute of Geography and Statistics [18]. These data describe the number of people classified by age groups who live in each census tract. The census tract is the smallest level of aggregation used by the Brazilian government.

2.2. Data Integration

Initially, the health data were summarized with the aim of combining patients who live in the same area. A total of 7066 patients who were admitted for circulatory and respiratory system illness between 2008 and 2013, were grouped a total of 169 address groups.

Subsequently, groups were combined with the geographic addressing database (using address as parameter) and the 7066 patients were integrated into the geographic information system (GIS) database. Finally, based on the principle of proportionality (number of patients per population), the 169 address groups were normalized based on demographic data. **Figure 1** synthesizes the procedures used for data integration.

We used the software ArcGis, version 10.2, as a tool for data integration and to evaluate the spatial patterns.
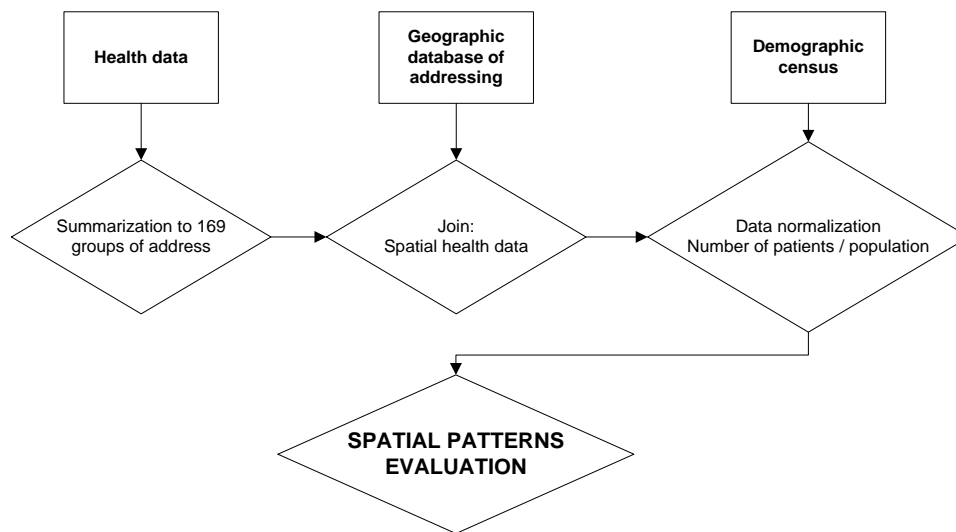
Figure 1. Procedures used for data integration.

2.3. Spatial Pattern Evaluation

We chose four methods to answer the questions which the aim of this study was based (presented on the introduction): Global Moran's I (question "a"); Getis-Ord General G (question "b"); semivariogram analysis (question "c"); and multi-distance spatial cluster—K-function (question "d").

2.3.1. Global Moran's I

The Global Moran's I test measures spatial autocorrelation based on two parameters: location (address of each patient) and values (number of patients living in the same location). The results of this test classify the distribution as clustered, dispersed or random. Equation (1) shows the calculation used in the Global Moran's I test [19]:

$$I = \frac{n}{So} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{i,j} Z_i Z_j}{\sum_{i=1}^{n} Z_i^2} \tag{1}$$

where Z_i is the deviation of the number of patients for polygons that represent each of the 169 address groups; $W_{i,j}$ is the spatial weight between polygons i and j; n is the total number of address groups, which in this study was 169; and So is the aggregate of all the spatial weights, represented by Equation (2).

$$So = \sum_{i=1}^{n}\sum_{j=1}^{n} W_{i,j} \tag{2}$$

A positive Moran's I index value indicates a tendency toward clustering, while a negative value indicates a tendency toward dispersion. However, this classification tendency can only be accepted based on a hypothesis test. The null hypothesis of the Global Moran's I test is that the value of each polygon (in this study, the number of patients) is randomly distributed across the study area. The null hypothesis is rejected in two cases: first, when the p-value is ≤0.10 and z-score is ≥1.65, so the distribution is statistically significantly clustered, and second, when the *p*-value is ≤0.10 and the z-score is ≤−1.65, so the distribution is statistically significantly dispersed.

2.3.2. Getis-Ord General G

The Getis-Ord General G, the second test, identifies the degree of clustering for either high or low values. Equation 3 shows the calculation used in the test [20]:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{i,j} X_i X_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j}, \quad \forall j \neq i \tag{3}$$

where X_i and X_j are the number of patients for each of the polygons I and j; $W_{i,j}$ is the spatial weight between polygons i and j; n is the total number of address groups; and indicates that polygons i and j cannot be the same polygon.

A high General G index indicates that high values for the attribute are clustered, while a low General G index indicates low values for the attribute are clustered. The null hypothesis of the Getis-Ord General G is that there is no spatial clustering. The null hypothesis can be rejected in two cases: first, when the p-value is ≤0.10 and the z-score is ≥1.65, so there is clustering for high values (High Cluster), and second, when the *p*-value is ≤0.10 and the z-score is ≤−1.65, so there is clustering for low values (Low Cluster).

The test Getis-Ord General G is sensitive to a choice of distance which the spatial relationships among features are calculated. In this study we choose the fixed distance band, which each feature is analyzed in terms of the context of neighboring features.

2.3.3. Semivariogram Analysis

According to Chun and Griffith (2013) [21], this method evaluates the spatial dependence in the data. In the semivariogram analysis can be tested the hypothesis that "Everything is related to everything else, but near things are more related than distant things"-first law of geography [22]. Thus, an empirical semivariogram is used, which is the squared difference between the values of two points making up a pair (Equation (4)):

$$\gamma(d) = 0.5 \times \text{Average}\left[\left(A_i - A_j\right)^2\right] \qquad (4)$$

where is the semivariogram value; d is the distance; A_i is the value at location i, and A_j is the value at location j. In this study, these values were the number of patients in each location.

The result from the empirical semivariogram is a chart that shows all pairs of locations that were compared. On the x-axis is the distance between the locations, and on the y-axis is the semivariogram value calculated by the empirical semivariogram (Equation (4)). The interpretation of this chart is: the lower Y value, the closer are the values from the points compared for the location X.

2.3.4. Multi-Distance Spatial Cluster, K-Function

The K-function is used to examine whether the distribution is clustered or dispersed considering a range of distances [13] and is estimated as follows:

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{n}\sum_{j=1}^{n}k_{i,j}}{\pi n(n-1)}} \qquad (5)$$

where $L(d)$ is the K-function; d is the distance; n is the total number of address groups; A is the total area of the address groups; and $k_{i,j}$ is the weight. This weight will be 1 when the distance between i and j is less than d and will be 0 otherwise.

The k value is calculated at several distances and displayed on a chart, which shows the value observed, the value expected and the confidence level. When the observed k value for a particular distance is above the line for the expected value, the distribution is more clustered. When the observed k value is below the line for the expected value, the distribution is more dispersed. When the observed k value is larger than the upper confidence level, clustering for that distance is statistically significant. When the observed k value is smaller than the lower limits of the confidence level, dispersion for that distance is statistically significant [13].

3. Results

3.1. Exploratory Analysis

From the total of 7066 patients 3381 (47.8%) were less than 5 years old, which was the largest group. The smallest group, 177 (2.5%), was the patients between 6 and 17 years old. The groups of patients between 18 and 59 years old and above 59 years of age included 1908 and 1600 patients, respectively (**Table 1**).

The descriptive analysis of the data before normalization showed that the 7066 patients were distributed in 169 address groups in the FD. The average of this distribution was 41.81 (± 76.9) patients per address. The patients up to 5 years old had the highest average, of 20.1 (±42.1) patients per address (**Table 1**).

Table 1. Descriptive statistics—non-normalized data.

Parameter	Up to 5 years	Between 6 and 17 years	Between 18 to 59 years	Above 59 years	All ages
Maximum	272	25	105	128	428
Total	3381	177	1908	1600	7066
Average	20.1	1.1	11.3	9.5	41.8
Standard deviation	42.1	2.7	19.8	17.8	76.9

In the normalized data analysis (patients per population in each address), the age group above 59 years had the highest average, 0.8 (±5.8), followed by the group up to 5 years old, with an average of 0.8 (± 4.2). The lowest average was in the age group between 6 and 17 years old, 0.02 (±0.2), **Table 2**.

Figure 2 shows the spatial distribution considering the patients of all ages and **Figure 3** with regard to the age groups. Patients between 6 and 17 years old were the least affected. A total of 109 address groups (64%) did not present cases of people between 6 and 17 years old who were hospitalized for cardiorespiratory illness. However, patients up to 5 years old were the most affected. This age group had the highest number of address groups with high numbers of patients. Additionally, the spatial distribution of patients up to 5 years old (**Figure 3**) is similar to the spatial distribution of patients of all ages (**Figure 2**).

3.2. Spatial Autocorrelation-Global Moran's I

There was clustering (I value > 0) when the analysis considered the patients of all ages (I = 0.012, $p < 0.10$). However, upon analyzing age groups, only the group up to 5 years old exhibits statistically significant tendency toward clustering (I = 0.024, $p < 0.10$). Other age groups presented a tendency to cluster (I value > 0), but without statistical significance (**Table 3**).

3.3. Degree of Clustering, High and Low Clustering-Getis-Ord General G

There was no high or low clustering for the spatial distribution of patients. Random distributions were identified in all cases by Getis-Ord General G analysis due to the z-score value (near zero) and *p*-value (>0.10), **Table 4**.

3.4. Spatial Dependence-Semivariogram Analysis

There was a concentration of a high number of points with low semivariogram values for all age groups, which means that the number of patients are related (similar) along at the distances. Only for the groups up to 5 years old and above 59 years old there were high semivariogram values, especially between 1500 and 5000 m (**Figure 4**).

3.5. Range of Distances Analysis—K Function

Considering a range of distances, some differences in spatial patterns were found among age groups. For patients up to 5 years old, there is statistically significant clustering until ~2500 m. Above this distance, the spatial distribution is random. For the age groups 6 - 17 years old and 18 - 59 years old, the spatial distribution is random, except for the patients between 6 and 17 at distances between ~5500 and ~8000 m, when the distribution shows statistically significant clustering. For the oldest patients (>59 years old) there is statistically significant clustering at lower distances (~<1000 m) and between ~7500 and ~10,500 m. Above ~10,500 m, the distribution exhibits statistically significant dispersion for the oldest patients. Finally, considering all age groups, there is significant clustering up to ~2500 m (**Figure 5**).

4. Discussion

Our results showed that the greater numbers of patients (patients per population) are children and seniors. We expected this result, since the previous studies have already showed that people up to 5 years old and above 59 years are more vulnerable to cardiorespiratory diseases [12] [23]-[25].

The group of youngest and the group of all ages were the only that presented positive spatial autocorrelation
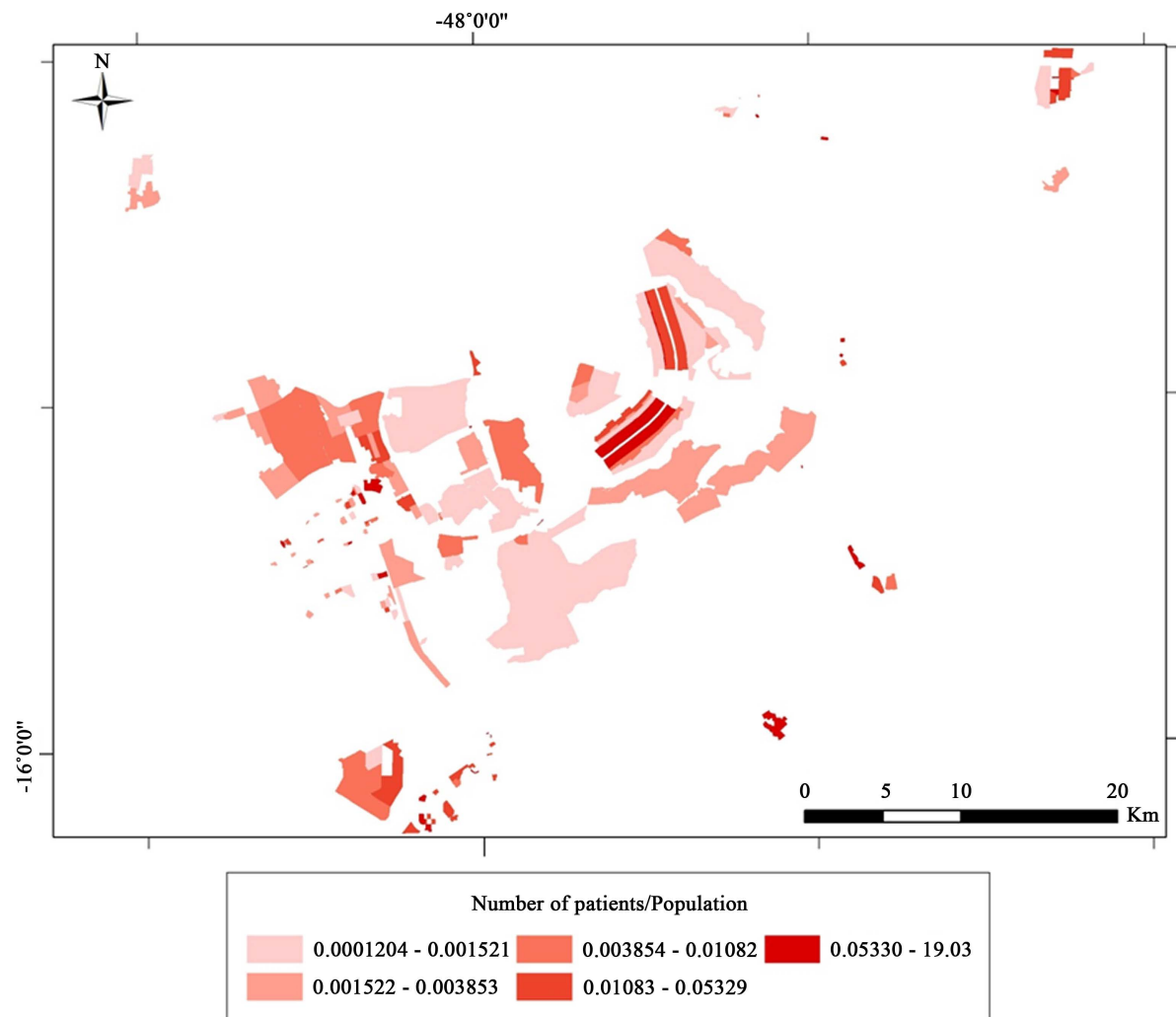
Figure 2. Spatial distribution—patients of all ages per population. Note: classification method—quantile.

Table 2. Descriptive statistics—normalized data (patients per population).

Parameter	Up to 5 years	Between 6 and 17 years	Between 18 and 59 years	Above 59 years	All ages
Minimum	0	0	0	0	0.0001
Maximum	52	2	14	75	19
Total	132	4	22	137	35
Average	0.8	0.02	0.1	0.8	0.2
Standard deviation	4.2	0.2	1.1	5.8	1.5

Table 3. Global Moran's I results.

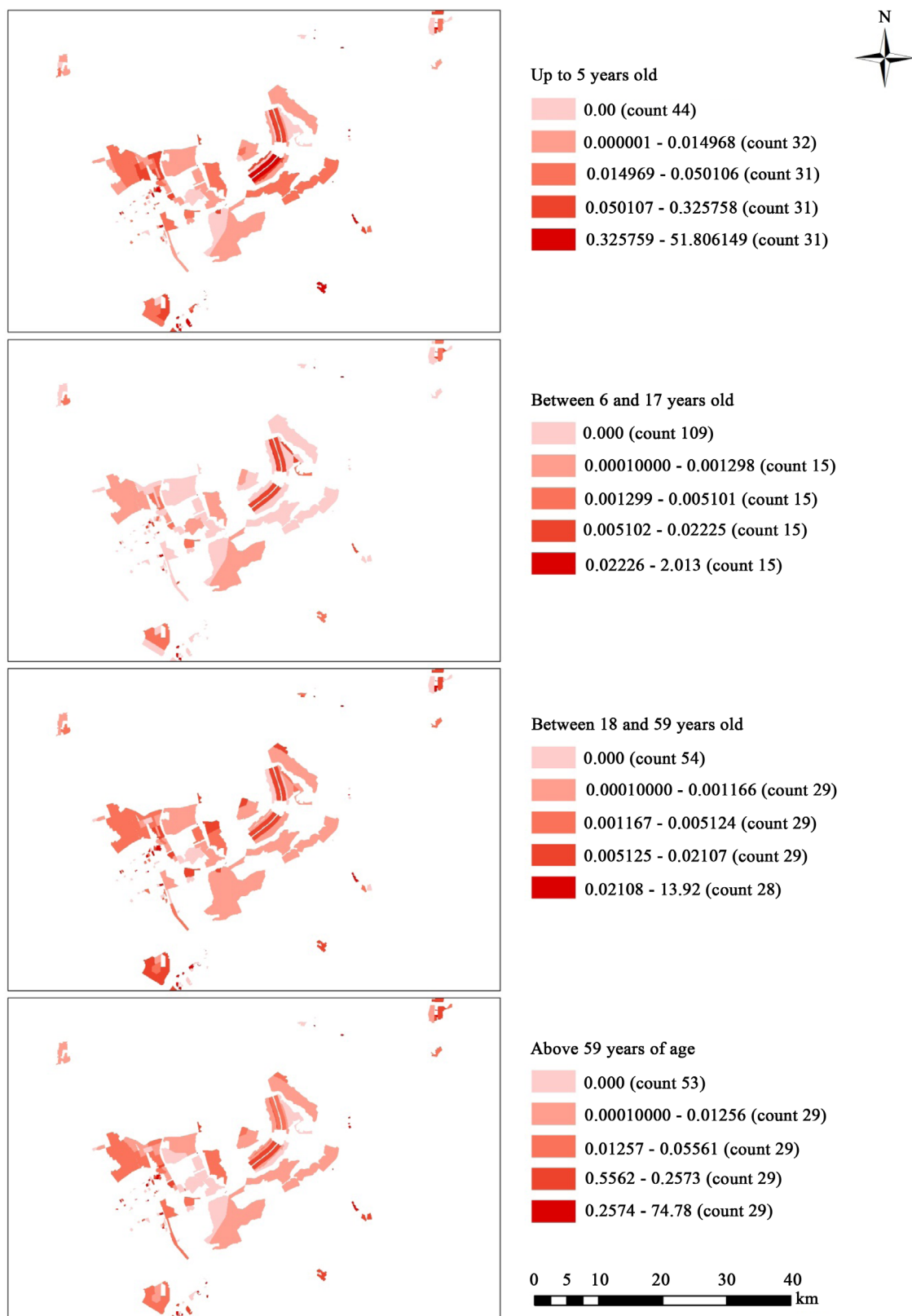| Age	Moran's Index (I)	z-score	p-value	Distribution
Up to 5 years old	0.024	2.055	0.040	Clustered
Between 6 and 17 years old	0.015	1.470	0.141	Random
Between 18 and 59 years old	0.007	1.348	0.178	Random
Above 59 years of age	0.011	1.561	0.118	Random
All ages	0.012	1.678	0.093	Clustered

Figure 3. Spatial distribution-age groups per population. Note 1: count—number of polygons (number of address groups, 169). Note 2: classification method—quantile.
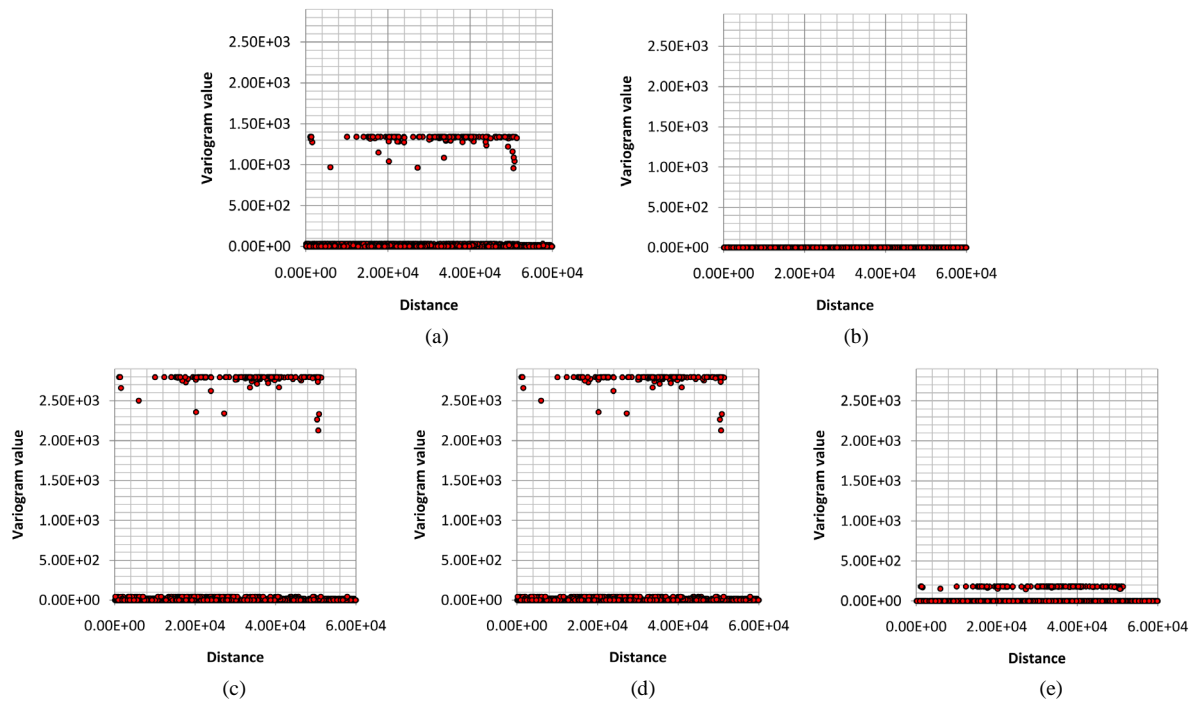
Figure 4. Semivariogram charts. Note: (a) up to 5 years old; (b) between 6 and 17 years old; (c) between 18 and 59 years old; (d) above 59 years old; (e) all ages.
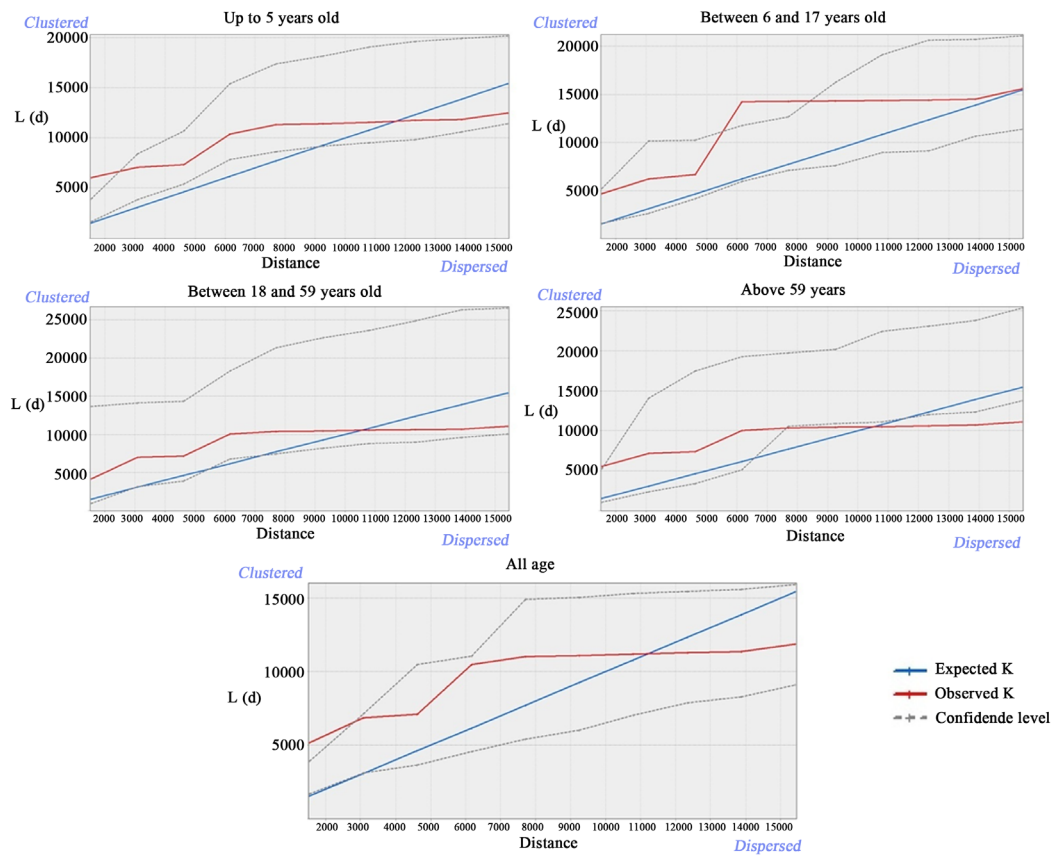

Figure 5. K function charts.

Table 4. Getis-Ord general G results.

Age	General G	z-score	p-value	Distribution
Up to 5 years old	0.000114	0.537	0.591	Random
Between 6 and 17 years old	0.000090	0.044	0.965	Random
Between 18 and 59 years old	0.000070	−0.197	0.844	Random
Above 59 years of age	0.000098	0.161	0.872	Random
All ages	0.000095	0.130	0.897	Random

(cluster). We suggest that there is a similar factor responsible for the cardiorespiratory illness occurrence in the FD for these two groups (<5 years old; all ages). Considering that this factor is related with the urban environment, probably the spatial distribution of air pollution is affecting the cardiorespiratory illness occurrence. Several studies have showed the relation between air pollution and cardiorespiratory diseases [1] [23] [26] [27].

Specifically for the all ages analysis, probably the group with youngest people (up to 5) is influencing the results, because the age group up to 5 years old is predominant in our analysis, which represents 47.8% of our sample.

Although the groups up to 5 and all ages present positive spatial autocorrelation, there was no specifically high or low clustering. This result may be because the extremes and lowest values are considered geographically isolated outliers in FD.

The semivariogram analysis showed that the higher concentration of points has low semivariogram values. So we suggest that the number of patients per residential sector is homogenous. For the group up to 5 years old and above 59 years old we found a low concentration of points with high semivariogram values, especially between 1500 and 5000 m. This result shows that for the children and seniors the spatial distribution of patients is less homogenous.

As the K-function analysis, the group up to 5 years old and the group above 59 years old showed significant cluster at short distances (<2500 m). Above 2500 m the spatial distributions is random or disperse. This result can be compared with the semivariogram, which the higher concentration of high semivariogram values for the children and seniors is between 1500 and 5000 m. The heterogeneity (high semivariogram values) at this range of distances can be affecting the changing of the spatial distribution (from cluster to random or disperse).

There are some limitations in this study. The first one is regarding to the method that we use to estimate the population in each address group where there are patients. It is different the spatial scale of the polygons that represent the health data and the polygons that represent the population data. In other words, the shape and the size of the polygons are not the same. We suggest to future studies to apply the dasymetric method [28] [29] and then compare with the method that we used.

The second limitation is regarding to the shape of the input data used to calculate the spatial patterns. We used polygons, which represent the aggregation of the health data. In this way, was used the centroid of each polygon in order to evaluate the spatial patterns. This is the GIS technique used when the input data is not a point. Therefore, this transformation can cancel some patterns, especially because the polygons are not uniform. However, we highlight that this limitation does not invalidate the results, since it is a limitation linked to the availability of data. Others studies have reported results with the same limitation (centroid of each polygon), such Cook *et al.* (2013) [30], Zou *et al.* (2014) and Tian *et al.* (2010) [31].

Finally, the health data do not representative of all hospital admissions in the FD. The health data provided to us is a subsample of the National Health Database, which includes information about residential addresses. According to the census for the FD, the National Health Database included 399,564 hospital admissions for cardiorespiratory diseases during the period 2008-2013, while in our subsample the total hospital admissions is 7,066.

This study was the first in Brazil with the aim to evaluate the spatial patterns of cardiorespiratory diseases for all age groups (children, teens, adults and seniors). Also, it was the first environmental health study for the Brazilians cities which tested the geostatistical approaches—spatial autocorrelation, degree of clustering, semivariogram and k-function. Most of the studies in Brazil have used only the spatial autocorrelation analysis [32]-[37].

Our study showed that in the FD the spatial occurrence of cardiorespiratory diseases is different among the age groups. Understanding the spatial distribution of diseases, the degree of clustering and the dependence be-

havior in terms of distances are critical to the development of public policies in health, environment, and urban planning.

Acknowledgements

References

[1] Mortimer, K., Gordon, S.B., Jindal, S.K., Accinelli, R.A., Balmes, J. and Martin, W.J. (2012) Household Air Pollution Is a Major Avoidable Risk Factor for Cardiorespiratory Disease. *Chest*, **142**, 1308-1315. http://dx.doi.org/10.1378/chest.12-1596

[2] WHO. Global Health Observatory (2014) http://www.who.int/gho/ncd/mortality_morbidity/en/

[3] Ignotti, E., Valente, J.G., Longo, K.M., Freitas, S.R., Hacon, S. de S. and Netto, P.A. (2010) Impact on Human Health of Particulate Matter Emitted from Burnings in the Brazilian Amazon region. *Revista de Saúde Pública*, **44**, 121-130. http://dx.doi.org/10.1590/S0034-89102010000100013

[4] Xu, Z., Hu, W., Williams, G., Clements, C.A., Kan, H.D. and Tong, S. (2013) Air Pollution, Temperature and Pediatric Influenza in Brisbane, Australia. *Environment International*, **59**, 384-388. http://dx.doi.org/10.1016/j.envint.2013.06.022

[5] Ning, Z., Wubulihairen, M. and Yang, F. (2012) PM, NOx and Butane Emissions from On-Road Vehicle Fleets in Hong Kong and Their Implications on Emission Control Policy. *Atmospheric Environment*, **61**, 265-274. http://dx.doi.org/10.1016/j.atmosenv.2012.07.047

[6] Troncoso, R. and Cifuentes, L.A. (2012) Effects of Environmental Alerts and Pre-Emergencies on Pollutant Concentrations in Santiago, Chile. *Atmospheric Environment*, **61**, 550-557. http://dx.doi.org/10.1016/j.atmosenv.2012.07.077

[7] Brajer, V., Hall, J. and Rahmatian, M. (2012) Air Pollution, Its Mortality Risk , and Economic Impacts in Tehran, Iran. *Iranian Journal of Public Health*, **41**, 31-38.

[8] Tayra, F., Ribeiro, H. and Nardocci, A. de C. (2012) Avaliação Econômica dos Custos da Poluição em Cubatão—SP com Base nos Gastos com Saúde Relacionados às Doenças dos Aparelhos Respiratório e Circulatório. *Saúde e Sociedade*, **21**, 760-775. http://dx.doi.org/10.1590/S0104-12902012000300020

[9] Hsu, H.-H., Adamkiewicz, G., Houseman, E.A., *et al.* (2012) The Relationship between Aviation and Ultrafine Particulate Matter Concentrations near a Mid-Sized Airport. *Atmospheric Environment*, **50**, 328-337. http://dx.doi.org/10.1016/j.atmosenv.2011.12.002

[10] Leiva, M., Santibañez, D., Ibarra, E.S., Matus, C.P. and Seguel, R. (2013) A Five-Year Study of Particulate Matter (PM2.5) and Cerebrovascular Diseases. *Environmental Pollution*, **181**, 1-6. http://dx.doi.org/10.1016/j.envpol.2013.05.057

[11] Gonzalez-Barcala, F.J., Pertega, S., Garnelo, L., *et al.* (2013) Truck Traffic Related Air Pollution Associated with Asthma Symptoms in Young Boys: A Cross-Sectional Study. *Public Health*, **127**, 275-281. http://dx.doi.org/10.1016/j.puhe.2012.12.028

[12] Zou, B., Peng, F., Wan, N. and Mamady, K. (2014) Wilson GJsCD of APEI across the US. Spatial Cluster Detection of Air Pollution Exposure Inequities across the United States. *PLoS One*, **9**, e91917. http://dx.doi.org/10.1371/journal.pone.0091917

[13] Mitchell, A. (1999) The Esri Guide to GIS Analysis: Geographic Patterns & Relationships. 1st Edition, Esri Press; Nova York.

[14] Kurland, K. and Gorr, W. (2012) Gis Tutorial Fo Health. 4th Edition, Esri Press, Nova York.

[15] IBGE. IBGE Cidades (2013) www.cidades.ibge.gov.br

[16] Datasus (2013) Base de dados: Endereço dos pacientes atendidos e internados no Dist Fed.

[17] Sedhab (2012) Base de dados: Organ Territ do Dist Fed.

[18] IBGE (2012) Base de informações geográficas do setor censitário. http://www.ibge.gov.br/home/download/estatistica.shtm

[19] Anselin, L. (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, **27**, 93-115.

http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x

[20] Getis, A. and Ord, J. (1992) The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, **24**, 189-206. http://dx.doi.org/10.1111/j.1538-4632.1992.tb00261.x

[21] Chun, Y. and Griffith, D. (2013) Spatial Statistics & Geostatistics. Vol First. Sage, London.

[22] Tobler, W. (1970) A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, **46**, 234-240. http://dx.doi.org/10.2307/143141

[23] Nandasena, S., Wickremasinghe, A.R. and Sathiakumar, N. (2012) Respiratory Health Status of Children From Two Different Air Pollution Exposure Settings of Sri Lanka: A Cross-Sectional Study. *American Journal of Industrial Medicine*, **55**, 1137-1145. http://dx.doi.org/10.1002/ajim.22020

[24] Hoffman, K., Kalkbrenner, A.E., Vieira, V.M. and Daniels, J.L. (2012) The Spatial Distribution of Known Predictors of Autism Spectrum Disorders Impacts Geographic Variability in Prevalence in Central North Carolina. *Environmental Health*, **11**, 80. http://dx.doi.org/10.1186/1476-069X-11-80

[25] Fan, X., Lam, K. and Yu, Q. (2012) Differential Exposure of the Urban Population to Vehicular Air Pollution in Hong Kong. *Science of The Total Environment*, **426**, 211-219. http://dx.doi.org/10.1016/j.scitotenv.2012.03.057

[26] Young, G.S., Fox, M., Trush, M., Kanarek, N., Glass, T.A. and Curriero, F.C. (2012) Differential Exposure to Hazardous Air Pollution in the United States: A Multilevel Analysis of Urbanization and Neighborhood Socioeconomic Deprivation. *International Journal of Environmental Research and Public Healt*, **9**, 2204-2225. http://dx.doi.org/10.3390/ijerph9062204

[27] Scoggins, A., Kjellstrom, T., Fisher, G., Connor, J. and Gimson, N. (2004) Spatial Analysis of Annual Air Pollution Exposure and Mortality. *Science of the Total Environment*, **321**, 71-85. http://dx.doi.org/10.1016/j.scitotenv.2003.09.020

[28] Poulsen, E. and Kennedy, L.W. (2004) Using Dasymetric Mapping for Spatially Aggregated Crime Data. *Journal of Quantitative Criminology*, **20**, 243-263. http://dx.doi.org/10.1023/B:JOQC.0000037733.74321.14

[29] Freitas, M. de B.C., Xavier, A.M. de S. and Fragoso, R.M. de S. (2012) Redistributing Agricultural Data by a Dasymetric Mapping Methodology. *Agricultural and Resource Economics Review*, **3**, 351-366.

[30] Cook, A.J., Gold, D.R. and Li, Y. (2013) Spatial Cluster Detection for Longitudinal Outcomes Using Administrative Regions. *Commun Stat Theory Methods*, **42**, 2105-2117.

[31] Tian, N., Wilson, G. and Zhan, B. (2010) Female Breast Cancer Mortality Clusters within Racial Groups in the United States. *Health Place*, **16**, 209-218. http://dx.doi.org/10.1016/j.healthplace.2009.09.012

[32] Barrozo, L.V. (2014) Contribuições da cartografia aos estudos de geografia da saúde: Investigando associações entre padrões espaciais. *Revista do Departamento de Geografia—USP*, Volume Especial Cartogeo, 413-425.

[33] Miranda, M.J.De, Costa, C., Santana, P. and Barrozo, L.V. (2014) Associação espacial entre variáveis socioeconômicas e risco relativo de nascimentos pré-termo na Região Metropolitana de São Paulo (RMSP) e na Área Metropolitana de Lisboa (AML). *Saúde e Sociedade*, **23**, 1142-1153. http://dx.doi.org/10.1590/S0104-12902014000400002

[34] Campos, F.G.De, Barrozo, L.V., Ruiz, T., *et al.* (2008) Distribuição espacial dos idosos de um município de médio porte do interior paulista segundo algumas características sócio-demográficas e de morbidade. *Cadernos de Saúde Pública*, **25**, 77-86. http://dx.doi.org/10.1590/S0102-311X2009000100008

[35] Bando, D.H., Moreira, R.S., Pereira, J.C. and Barrozo, L.V. (2012) Spatial Clusters of Suicide in the Municipality of São Paulo 1996-2005: An Ecological Study. *BMC Psychiatry*, **12**, 124. http://dx.doi.org/10.1186/1471-244X-12-124

[36] Muller, E.V., Aranha, S.R.R., Roza, W.S.S.Da and Gimeno, S.G.A. (2012) Distribuição espacial da mortalidade por doenças cardiovasculares no Estado do Paraná, Brasil: 1989-1991 e 2006-2008. *Cadernos de Saúde Pública*, **28**, 1067-1077. http://dx.doi.org/10.1590/S0102-311X2012000600006

[37] Almeida, S.L. de. (2013) Análise espacial das doenças respiratórias e a poluição relacionada ao tráfego no município de São Paulo.