

## **SPATIAL DISTRIBUTION OF PULMONARY TUBERCULOSIS IN URBAN AREAS: A CASE FROM BELÉM, BRAZIL**

**Marco Aurélio Arbage LOBO\***, **Hermógenes de Carvalho PAIVA NETO** &  
**Leonardo Augusto Lobato BELLO**

Universidade da Amazônia, Belém, PA, Brazil.

\*Corresponding author: lobo2502@hotmail.com

<http://dx.doi.org/10.18571/acbm.105>

### **ABSTRACT**

The precise location of areas with high incidence of pulmonary tuberculosis (TB) is important to improve public health actions. Official data records of the addresses and neighborhoods where the infected people live allow the mapping of the disease on this spatial scale. However, great socioeconomic diversity often exists inside neighborhoods, wherein high- and low-income families reside. This situation hampers the location of those areas that require close attention. **Objective:** This study aimed to estimate the risk of pulmonary TB infections in census tracts in Belém City (Brazil) from data on neighborhoods. **Methods:** A partial least-squares regression model was constructed in the scale of neighborhoods based on the record of addresses of TB-infected people and socioeconomic data from official sources. The model was then slightly modified and used to estimate the risk of TB prevalence in urban census tracts. The results were mapped using a geographical information system. **Results:** The percentages of explained variance of the set of independent variables and dependent variable were 86.4% and 30.2%, respectively. These values indicated that the model is acceptable for its purpose. **Conclusion:** The model's results were consistent with the spatial distribution of socioeconomic and environmental characteristics of Belém City.

**Keywords:** Tuberculosis Pulmonary; Urban Population; Public Health.

### **RESUMO**

A localização precisa das áreas com alta incidência de tuberculose pulmonar (TB) é importante para melhorar as ações de saúde pública. Os dados oficiais registram endereços e bairros onde moram pessoas infectadas, permitindo o mapeamento da enfermidade nesta escala espacial. Comumente, porém, há grande diversidade socioeconômica no interior dos bairros, onde coexistem áreas com famílias de alta e baixa renda. Isso dificulta a localização das áreas prioritárias para ação. **Objetivo:** estimar o risco de infecções por tuberculose pulmonar (TB) nos setores censitários na cidade de Belém (PA), a partir de dados por bairros. **Métodos:** construção de um modelo de regressão múltipla de mínimos quadrados parciais, na escala dos bairros, com base no cadastro de endereços de pessoas infectadas e dados socioeconômicos de fontes oficiais. O modelo, com uma pequena modificação, foi então usado para estimar o risco de incidência de TB nos setores censitários urbanos. Em seguida, os resultados foram mapeados por meio de um sistema de informações geográficas. **Resultados:** o percentual de variância explicada do conjunto de variáveis independentes foi de 86,4% e o da variável dependente, 30,2%; valores que tornam o modelo aceitável para o fim proposto. **Conclusão:** o modelo mostrou resultados coerentes com a distribuição espacial das características socioeconômicas e ambientais da cidade de Belém.

**Palavras-chave:** Tuberculose Pulmonar; População Urbana; Saúde Coletiva.

## 1 Introduction

Several cities in non-developed countries have become favorable environments to the spread of respiratory diseases, not only because of the intense presence of solid particles in suspension originating from various activities (industries, transport, and burning of vegetation and garbage), but also to the great number of overcrowded dwellings without ventilation and with excess moisture indoors (SANTOS, 2011).

Among the respiratory diseases, pulmonary tuberculosis (TB) poses a serious public health problem with deep social roots, being one of the most important causes of mortality and morbidity of human beings (ROBERTS; BUIKSTRA, 2003). The disease is presently regarded as one of the most serious threats to human health (WORLD HEALTH ORGANIZATION, 2016). A deep relationship exists between TB prevalence and the social development of a country, because of factors connected with situations of poverty, misery, and social exclusion, mainly in the urban peripheries and slums (TEIXEIRA; COSTA, 2011; BARBOSA; COSME, 2013; ACOSTA; BASSANESI, 2014). TB also constitutes a serious problem of health in native Indian (ESCOBAR et al., 2001) and imprisoned populations (OLIVEIRA; CARDOSO, 2004; SÁNCHEZ et al., 2007) in Brazil. The growth of TB and HIV co-infection has increased disease dissemination (JAMAL, 2007).

The case study carried out by Vendramini et al. (2005) concluded that men have a greater risk of contracting TB than women, whereas the age group of 50 years and above is the most sensitive to the disease. This finding is important considering Brazilian's aging population. The study also identified high co-infection between HIV and TB. Furthermore, the authors attracted attention for the sub-notification of cases.

San Pedro and Oliveira (2013) performed a systematic literature review on the relation between socioeconomic factors (individual and collective) and occurrence of pulmonary TB. Regarding the collective studies, the authors identified the important influence of variables related to the gross national product (GNP) per capita, human development index, and level of access to basic sanitation in the country scale. Other factors were identified in different spatial scales: average number of persons per room, density of poor persons, schooling, decline of familiar income, and residences earning government monetary help. Wu and Dalal (2012) concluded that TB occurrence is related to the rate of human development, rate of corruption perception, GNP per capita, and percentage of people with insufficient food supply. The same authors also stressed the government expenses per capita in health, availability of hospital beds, and access to good sanitation conditions.

In spite of the large presence of the illness in non-developed countries and areas of low-income families worldwide, the World Health Organization pointed out the important progress made in the reduction of the disease. From 1990, the mortality for TB has been reduced by 47%, which suggested that the death of nearly 43 million individuals was avoided (WHO, 2016). The decline of TB spread is more remarkable in countries with high human development, low infant mortality, and good access to sanitation services (DYE et al., 2009).

The role of socioeconomic factors directly related to TB or other intermediating factors should be identified to improve disease control (MENEZES et al., 1998). Considering that Brazilian cities are quite uneven in socio-economic and environmental terms, the urban areas where the infection occurs more frequently should be identified to define the priority sites for the governmental programs, such as the Family Health, and the implementation of public services for TB treatment.

Despite the diversity of research relating urban development and population health, pulmonary TB is not among the most studied diseases in relation to the damages caused by environmental factors to human health. In accordance with Queiroga et al. (2012), knowledge about the spatial distribution of TB in urban areas is limited, so understanding how the disease

spreads in the territory and its association with general and local aspects is not clear until now.

At first glance, neighborhoods can be considered a suitable spatial unit to identify the areas with the highest TB prevalence in a particular city. However, these spatial units are often internally heterogeneous areas in relation to social and environmental attributes. Often, areas where high income families live are very close to those of low-income families in the same neighborhood. Such a situation may cause inaccuracy in studies conducted in this spatial scale. Estimation of the prevalence coefficient of pulmonary TB in small territorial units, such as the census tracts defined by the Brazilian Institute of Geography and Statistics (IBGE, 2010a), may reduce this uncertainty.

One possible alternative in this direction could be the geocoding of addresses where at least one TB-infected patient lives. Such information is available on the Notifiable Diseases Information System (SINAN) of the National Program for Tuberculosis Control (BRASIL, 2014) using Google Maps<sup>®</sup>, for example. The addresses are then plotted as points in a georeferenced layer of a geographic information system (GIS). The problem of this method in large cities is that the georeference of those addresses is quite laborious because of the large number of data.

The main objective of this paper was to propose an alternative method that involves the development of a partial least-squares regression model (PLSR) in the spatial scale of neighborhoods. As a secondary objective, in order to test the method, we aimed to identify and analyze the main features of the spatial distribution of TB prevalence in Belém City (Brazil) for determining the relationship between the geographical pattern of disease occurrence and the city's urban space conformation. Belém is a city located north of Brazil, in the Amazon region, in the southern portion of the Amazon River delta, on the confluence of Guajará bay with the Guamá River, having an estimated population of 1.4 million in 2015. Like all large Brazilian cities, it is characterized by strong concentration of wealth and large number of poor families.

## 2 Materials and Methods

This paper describes an epidemiological, ecological, descriptive, cross-sectional, and exploratory study, in which the unit of analysis was the population aged 10 years or older living in the neighborhoods that comprise the urban areas of Belém City.

In the model proposed, the TB prevalence coefficient in persons aged 10 years and above (called "TBCOEF10") was the dependent variable. Three factors frequently cited in the specialized literature were independent variables: agglomeration of people living in reduced spaces, represented by the average of persons per room ("PEROOM"); level of illiteracy, expressed by the illiteracy rate in persons aged 10 years and above ("ILLITER"); and inadequate sanitation ("INADSAN"), defined by the percentage of households in which wastewater is not directed to a general sewage or rainwater drainage system or absence of a septic tank. The model was used to estimate the TBCOEF10 by census tracts from data provided by IBGE (2010a).

The addresses data in which at least one TB patient lives were obtained from SINAN. Among other information, each address gives the neighborhood where it is located, so the frequency of pulmonary TB cases in each neighborhood could be determined. The addresses that omitted this information were disregarded. The authors have taken care to preserve the confidentiality of individual addresses, presenting aggregate quantitative information by neighborhoods to preclude the location of these addresses.

The statistics refer to the years 2008 to 2012 to enable the calculation of the annual average of cases per neighborhood in this period. This procedure was adopted for two reasons. First, this procedure reduced variability in the data, which showed significant annual fluctuations, especially in districts with a small population. Second, the average was centered in 2010, the reference year of socioeconomic data obtained in the Population Census 2010 (IBGE, 2010a). In consideration of the population aged 10 years and above, the TB prevalence

coefficient per 100,000 inhabitants in each neighborhood was calculated.

People under 10 years were not considered because of the very small share of this age group in the total number of TB cases in Belém; only 101 cases exist in the considered valid universe of 7,609 cases, or 1.3%. However, people in the age group cited represent 14.7% of the population of the city. The authors of this study believe that if the prevalence coefficient of pulmonary TB covers the entire population, the most common measure used in the specialized literature, distortions may likely occur in areas marked by large disparities in income such as Belém and other large Latin American cities.

We argue that the percentage of children is usually higher in poor families than in the richest. As the youth population is much less likely to contract the disease, the TB prevalence coefficient in lower income areas decreases in relation to those of higher incomes only because of the different demographic profiles between these areas, and not to some other social or environmental cause that interferes in disease dissemination.

To provide empirical support to this statement, we calculated the correlation coefficient between the household median income per capita of the neighborhoods and the result of the division between two TB prevalence coefficients, that is, the population aged 10 and above coefficient divided by the total population coefficient, which expresses the relative difference between them. The result was the coefficient correlation of  $-0.752$ , which can be considered high according to the classification proposed by Cohen (1992). In other words, the lower the median incomes per capita, the higher the relative difference between the two coefficients, and vice versa. Studies that did not consider this fact may introduce a component of inaccuracy in its results, because the vast majority of TB cases occur in low-income families, that is, the population segment with high relative differences between the two coefficients. Therefore, we argue that the TB prevalence coefficient in people aged 10 years or above is a measure of high accuracy if a study aims to observe different levels of illness occurrence in situations with great disparity of income, because of the interference of the different demographic profiles related to this disparity.

The data analysis was developed considering two complementary aspects: statistical and geographical. Statistical analysis had the initial intention to set up an ordinary least squares multiple regression model based on neighborhood data, with the prevalence rate of pulmonary TB per 100,000 persons aged 10 years and above as the dependent variable and three independent variables: average number of people per household, literacy rate in persons aged 10 years and above, and percentage of households with inadequate sanitation. The choice of these variables took into account their high influence on the occurrence of pulmonary tuberculosis in urban areas and the availability of statistical data by census tracts.

The second problem for the elaboration of the model initially thought by the authors was multicollinearity, which was solved by replacing it with PLSR. This model aggregates the original variables in orthogonal “components” or “factors” that are not correlated (also called “latent variables”),” which are new variables resulting from the linear combination of the original ones.

The construction of the PLSR model in the spatial scale of neighborhoods and the subsequent prediction of TBCOEF10 values for the urban census tracts of Belém were performed using the statistical program R with the “pls” package installed (MEVIK; WEHRENS, 2007). Notably, the said prediction quantified the estimated risk of a person aged 10 years and above being infected by TB in each census tract.

Geographical analysis was conducted with the support of a GIS using the QGIS program. The aim was to identify which features of the urban space organization of Belém influence TB prevalence. A map was constructed in the scale of census tracts to analyze a fundamental aspect of that organization: the spatial distribution of average household income per capita based on data from the Population Census 2010. A map was also created to show the spatial distribution of



TBCOEF10 in the neighborhoods considered in the regression model. Finally, we represented the spatial distribution of the model results by census tract in another map to analyze and evaluate the consistency of the results.

In accordance with the principles contained in the Helsinki Declaration of the World Medical Association, this research has taken due care with exposure to risks related to the identification of the addresses used in the study.

### 3 Results and Discussion

A preliminary check on the prevalence coefficients of TB in the neighborhoods finds that 34 neighborhoods with a population aged 10 years and above, in which less than 10,000 inhabitants lived in 2010, had high variability among their coefficients, even considering that the data on TB prevalence are an average of five years. For this reason, we chose to exclude such neighborhoods from the analysis, which accounted 8.3% of the total population of Belém. Thus, this research considered only 37 of the 71 city neighborhoods. As a consequence, all the neighborhoods of Mosqueiro District were excluded. This is a somewhat expected problem in studies like this. According to Medronho (2009), a recurring problem in ecological studies is the great variability of illness rates in areas where few cases exist.

Table 1 presents the original data of these 37 neighborhoods initially selected for the model construction ordered by the population aged 10 years and above.

It is important to mention that the original idea of the authors was to use in the model the variable “persons per household” to express the agglomeration of people living in reduced spaces, but it was found to be inadequate as a predictor in the geographical scale of census tracts. A preliminary examination demonstrated that this variable was a good predictor for TBCOEF10 on the scale of neighborhood. However, this relationship does not always occur on the scale of census tracts.

The reduction in the average number of persons per household, which mostly occurs with increasing income, is more related to the change in the type of housing. Specifically, the presence of housing type apartment increases as income grows, and the average number of residents in this type of housing is usually lower than that in houses: 2.92 people versus 3.93, according to the Population Census 2010. Thus, in the stratum of average household income per capita up to 1/4 Brazilian minimum wage (BMW; one BMW is worth about US\$291 in average in 2010), only 0.6% of households are apartments and 0.7% is in the stratum of 1/4–1/2 BMW. These percentages grow to 47.4% in the stratum of 20–30 BMW and 61.5% in those of 30 BMW and more.

This phenomenon causes distortions in the analysis of certain situations. For the predicted value of TBCOEF10 for census tracts located in horizontal residential condominiums of high-income households, in which there are large houses, the average number of persons per household was found to be higher than in sectors with lower income but higher presence of apartments.

To solve this problem, the average number of persons per household was replaced by the estimation of the average number of persons per room (PEROOM), which considers an element of the physical size of households, namely, the number of rooms. PEROOM is strongly related to the average household income per capita, a variable with data available in the scale of neighborhoods. Thus, whether the dwelling is a house, apartment, or tenement will not interfere with the model results.

**Table 1:** Socioeconomic and health data, neighborhoods of Belém, 2010.

Neighborhood	Pop. 10 years and above	TBCOEF10 <sup>1</sup>	Persons per household	ILLITER	INADSAN
Guamá	80,362	192.9	4.036	4.1	19.9
Pedreira	61,178	170.0	3.660	2.0	13.1
Marco	58,439	107.8	3.624	1.6	16.4
Marambaia	57,764	150.6	3.800	2.6	21.1
Jurunas	55,239	195.5	4.062	4.0	15.6
Tapanã	55,067	134.4	3.805	3.9	45.6
Terra Firme	51,193	185.6	3.973	4.8	46.2
Coqueiro	44,103	77.1	3.645	1.8	26.5
Sacramento	38,262	209.1	3.861	3.0	16.5
Telégrafo	36,882	176.2	4.061	2.9	8.5
Condor	36,613	161.1	4.103	3.6	16.2
Parque Verde	32,969	39.4	3.634	2.6	30.2
Mangueirão	30,780	78.0	3.592	2.4	18.7
Parque Guajará	28,107	64.0	3.710	5.6	44.3
Cremação	27,661	177.1	3.760	1.8	13.8
Umarizal	27,321	80.5	3.501	1.0	5.4
Maracangalha	25,886	11.6	3.963	3.0	29.3
Tenoné	24,979	108.1	3.683	3.4	44.5
Bengui	24,490	253.2	3.870	4.8	48.3
Cabanagem	22,945	274.6	3.888	4.4	42.3
Campina de Icoaraci	22,714	211.3	3.788	3.7	41.5
Barreiro	21,243	127.1	4.127	5.8	8.9
Castanheira	21,227	84.8	3.635	2.2	41.3
Nazaré	18,851	15.9	3.194	0.4	0.9
Pratinha	18,295	147.6	3.939	5.0	37.3
São Bras	18,238	115.1	3.398	0.9	5.7
Batista Campos	17,346	34.6	3.294	0.8	4.8
Agulha	16,448	73.0	3.817	4.8	54.3
Curió-Utinga	14,426	97.0	3.772	3.2	37.3
Águas Lindas	13,966	186.2	3.716	5.6	46.9
Canudos	12,206	213	3.814	1.9	37.1
Souza	11,692	68.4	3.419	1.2	14.7
Ponta Grossa	11,385	79.1	3.588	2.4	42.6
Maracacuera	11,056	99.5	3.831	6.3	53.1
Fátima	10,933	118.9	4.121	1.8	2.4
Cidade Velha	10,909	100.8	3.932	1.4	6.4
São João do Outeiro	10,035	159.4	3.593	5.0	64.7

Source: Brasil (2014); IBGE (2010a); Authors. <sup>1</sup>Average 2008 to 2012.

To carry out such estimates, we developed a polynomial regression model of second order to link the average household income per capita as an independent variable and the average number of persons per room as a dependent variable. The model was based on data from 44 areas of Belém in which the sample statistics of Population Census 2010 were published (IBGE, 2010b). The regression model was used to estimate PEROOM in each neighborhood based on the average household income per capita.

The first step of the analysis was to calculate the basic statistics of the dependent variable (TBCOEF10) and the three independent variables (PEROOM, ILLITER, and INADSAN). By comparing the minimum and maximum values of the variables, it is clear that there is significant disparity in their values distribution, consequence of the strong socio-spatial differentiation inside the city, similar to many other Brazilian and Latin American cities. These values are: TBCOEF10, 11.6 and 274.6 per 100,000 people aged 10 years and above; PEROOM, 0.486 and 0.927 person per room; ILLITER, 0.4% and 6.3%; and INADSAN, 0.9% and 64.7%.

The skewness coefficients show that all distributions were asymmetrical. TBCOEF10 (0.195), ILLITER (0.218), and INADSAN (0.163) were moderately asymmetrical toward the right (coefficient much lower than 1), whereas PEROOM was heavily skewed toward the left (-1.237, a coefficient lower than -1), indicating a large number of dwellings with high averages or persons per room. The three distributions with positive skewness had most of their cases in the lower values side. This situation was more favorable to TBCOEF10 and ILLITER than if the distribution were normal, on the contrary in relation to INADSAN.

The positive values of kurtosis coefficients in the four variables indicated the presence of several cases around a narrow range of values, thereby making their curves of distribution more peaked than the normal (leptokurtic type). The combination of skewness and kurtosis values (higher than 3) of PEROOM showed a distribution very different from the normal. However, the use of this variable in the partial least-squares regression is not a problem, as the normality of the variables is not an assumption in this statistical technique.

Subsequently, the Pearson correlation matrix of the four variables was constructed. Notably, the transformation of the original values of the variables in natural logarithms increased the correlation coefficients with TBCOEF10, which was favorable for the development of the model. However, such a transformation also increased the correlation between the independent variables (multicollinearity), which was not a problem for the type of regression model used in this study.

The advantages pointed out by Keene (1995) must also be considered when using data transformed to natural logarithm compared with untransformed data. Applying that transformation to the four variables yielded a log-log type model, in which the relationship among the independent and dependent variables ceased to relate amounts of quantities and became a relationship of proportional changes or elasticity (HAIR et al., 2009).

Scatterplots of the dependent variable LN\_TBCOEF10 related to the independent variables LN\_PEROOM, LN\_ILLITER, and LN\_INADSAN (the variables transformed to natural logarithm now have the prefix "LN\_") were then constructed. Scatterplots allow the visual identification of possible extreme values and other problems in the distribution of the transformed data to logarithms. The graphs showed an extreme value in LN\_TBCOEF10, corresponding to the Maracangalha neighborhood. Its TB prevalence coefficient was far below what would be expected by the tendency presented by the other points in the graph. This case, which was excluded from the analysis, should be considered for future research to understand this anomalous behavior. Such behavior may be attributed to a registration error or a phenomenon specific to the neighborhood. Thus, 36 cases remained in the regression analysis.

Considering the aforementioned exclusion, correlations between variables transformed to natural logarithm showed that the largest correlation coefficient of LN\_TBCOEF10 was with LN\_PEROOM (0.578), implying that the average number of people per room was the factor of greatest influence on TB prevalence in Belém. The second highest correlation of LN\_TBCOEF10 was registered with LN\_ILLITER (0.531), and the smallest correlation was observed with LN\_INADSAN (0.399). A significant correlation was noted between the three pairs of independent variables (LN\_PEROOM: 0.922 with LN\_ILLITER and 0.733 with LN\_INADSAN; and 0.732 between LN\_ILLITER and LN\_INADSAN). With the exception of the correlation coefficient of the dependent variable with LN\_INADSAN, which was considered

medium intensity according to the criterion proposed by Cohen (1992), all the other coefficients were classified as high.

The multiple linear regression model of PLSR was then constructed. Similar to principal component analysis (PCA), the partial least square regression is a particularly useful modeling technique when one has strongly correlated variables, as in the present study. PLSR differs from PCA in that the latter creates components that focus on the variation observed in the independent variables without considering the dependent variable, whereas PLSR also considers the variability in the dependent variable to maximize the covariance between them (WOLD et al., 2001). PLSR is a widely used technique in natural sciences, but it is rarely used in social sciences (SAWATSKY et al., 2015).

The cross-validation test with partition of the data in five randomly selected segments was used to set the number of model components. The test found that the lowest adjusted mean square error of prediction (0.516) was observed in the model with only one component among the three possible. The percentage of explained variance of the set of independent variables reached 86.4%, whereas that of the dependent variable was 30.2%. This value stresses the multicausality associated with the disease occurrence. Even though the regression model has considered three factors deemed relevant in the literature, almost 70% of the total variance remains unexplained. The factor loadings of the selected component (the so-called “latent variable”) were 0.602 to LN\_PEROOM; 0.6 to LN\_ILLITER; and 0.536 to LN\_INADSAN.

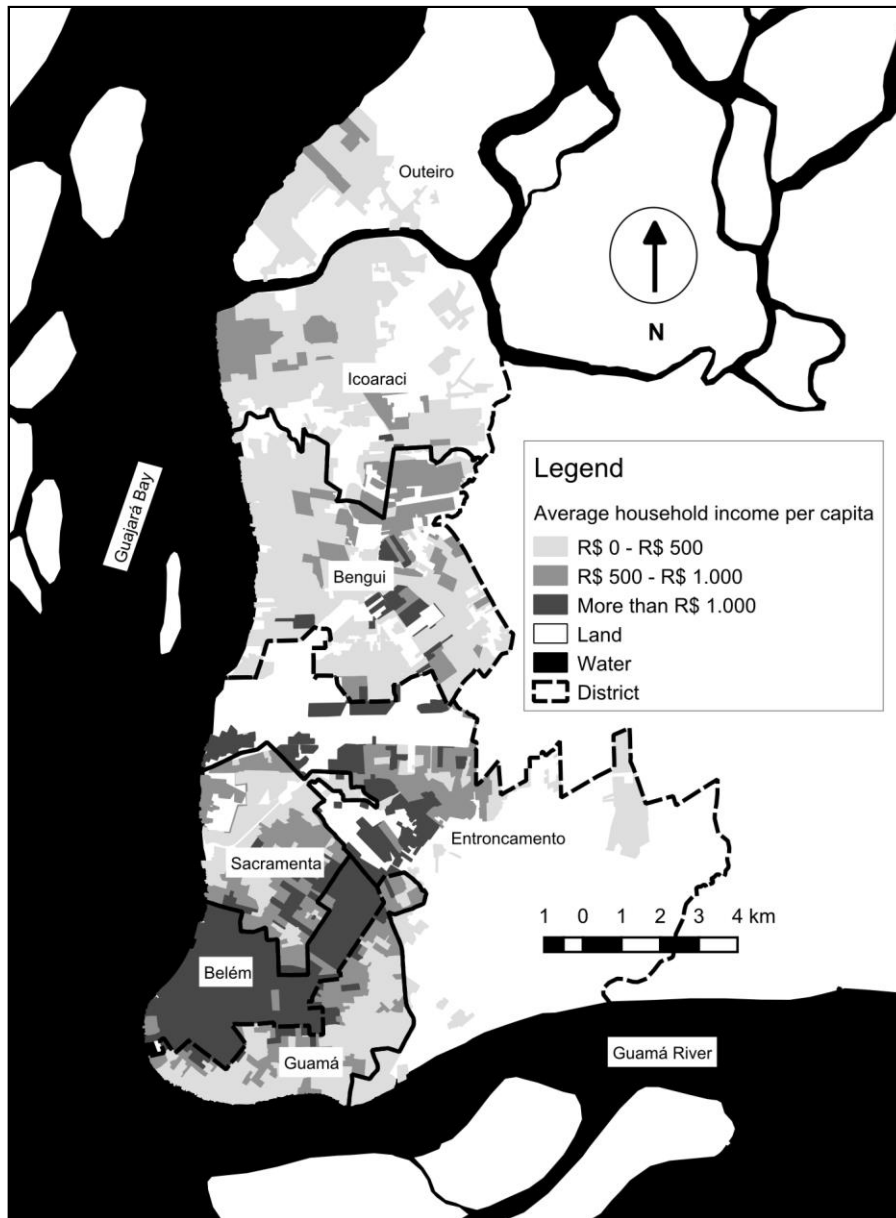
Finally, the model was used to estimate the LN\_TBCOEF10 values of the census tracts, which were then converted to the original scale of the variable TBCOEF10 via exponential function.

With regard to the geographic organization of Belém, Figure 1 shows that the high values of monthly household income per capita were spatially concentrated. Almost all census tracts with higher incomes (R\$1,000 and more) were located in Belém District. This district comprises the city’s most expensive lands and has good environmental conditions and urban services of the best quality. Moreover, it houses a significant part of the economic establishments of the city, including two large shopping malls. Some high-income sectors are found at Entroncamento District, and most of these sectors are places of residence of military families. Several high-income sectors are in Bengui District, which mainly consists of gated communities.

The low-income sectors (up to R\$500) are divided among the four districts of the periphery and the two located in the vicinity of Belém District. An important distinction exists between them. Those in the periphery mostly have low population densities, unlike those located at Guamá and Sacramento Districts, where there is a large presence of high density areas.

A significant part of these two districts consists of the so-called “baixadas”, which are areas of quite low altitudes and characterized by a high presence of low-income families. Some of these areas are at risk of flooding during the rainy season, which occurs from December to May. The “baixadas” are also places with numerous high population density areas, because of the combination of intense concentration of dwellings in a small urban area and high average persons per household. Among the ten neighborhoods having a high average of persons per household in Belém, seven are composed wholly or partly by “baixadas” areas, namely, Fátima (4.12 persons per household), Condor (4.1), Jurunas and Telégrafo (4.06), Guamá (4.04), Terra Firme (4), and Cidade Velha (3.9). The “baixadas” represent the local version of slum areas, wherein low-income families live in unhealthy environmental conditions in exchange for proximity to job sites. In Belém and its metropolitan area, the largest concentration of jobs is located in the area formed by Belém, Sacramento, and Guamá Districts.

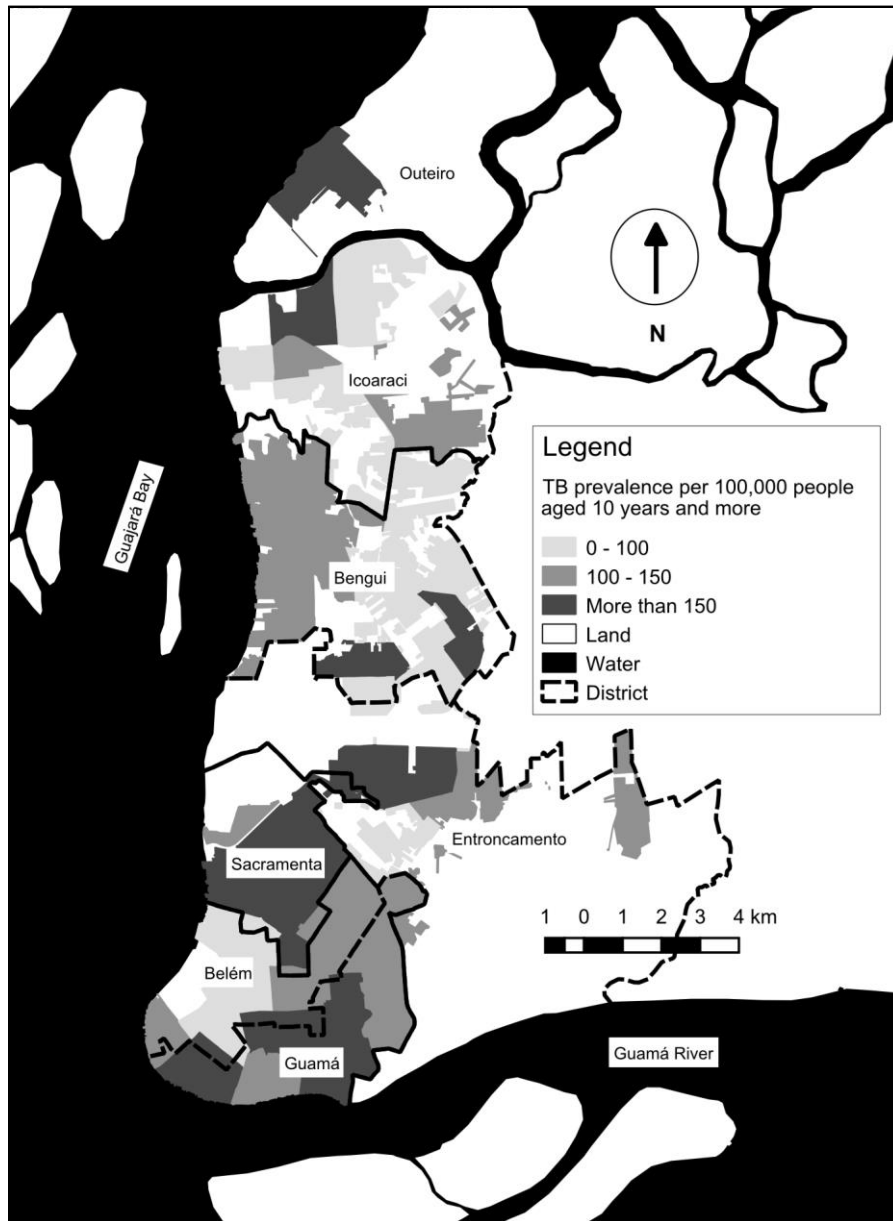




**Figure 1:** Monthly average household income per capita, census tracts of Belém.  
Sources: IBGE (2010a); Authors.

Figure 2 shows the map of TB prevalence coefficients in the areas effectively urbanized of the 36 neighborhoods considered in statistical modeling, based on data provided by the official source. Visual comparison between Figures 1 and 2 showed that the spatial distribution of TB prevalence was inversely related to income level.

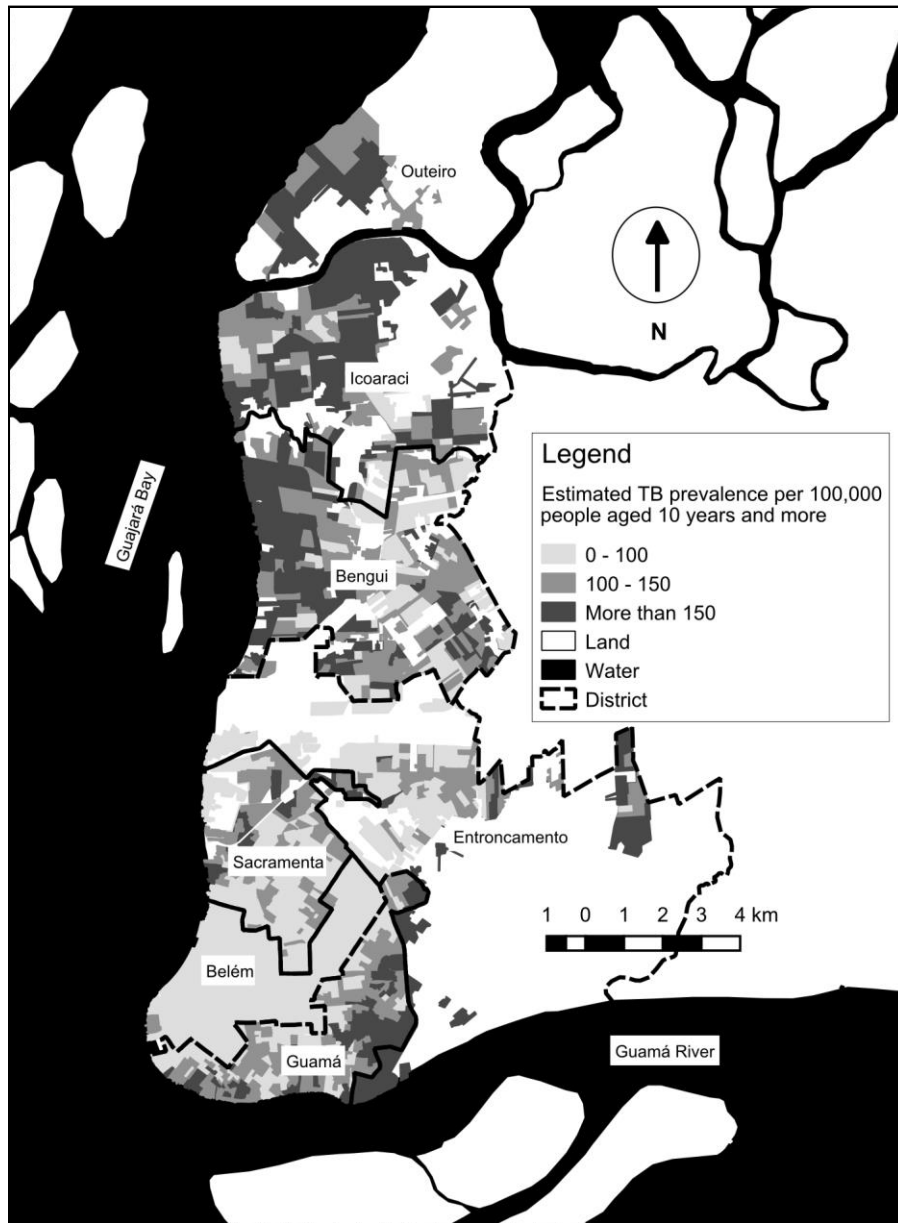
Three large areas consisted of high TB prevalence neighborhoods, and these areas were located in Sacramento and Guamá Districts. Among the ten neighborhoods of high TBCOEF10 values, six were located wholly or partly in “baixadas” areas: Canudos (213 cases per 100,000 inhabitants), Sacramento (209.1), Jurunas (195.5), Guamá (192.9), Terra Firme (185.6), and Cremação (177.1).



**Figure 2:** TB prevalence coefficients per 100,000 people aged 10 years and more, 36 neighborhoods of Belém. Sources: Brasil (2014); Authors.

Neighborhoods of high TB prevalence were also noted in the urban outskirts, including the two highest rates recorded in Belém, both at Bengui District: Cabanagem (274.6) and Bengui (253.2). This finding was expected, because of the large number of low-income families living at Bengui, Entroncamento, Icoaraci, and Outeiro Districts.

The spatial distribution of the statistical modeling results - the estimated prevalence of pulmonary tuberculosis in persons aged 10 years and above by census tract - is presented in Figure 3. This distribution was measured in terms of expected coefficient of cases per 100,000 inhabitants aged 10 years and above in 2010.



**Figure 3:** Estimated TB prevalence coefficients per 100,000 people aged 10 years and more, census tracts of Belém. Sources: Authors.

By comparing Figures 2 and 3, the model results were generally consistent with the incidence coefficients recorded in the neighborhoods. However, some discrepancies were noted between the model predictions and what was actually recorded in the data of neighborhoods. For example, in part of Sacramento District, the data indicated a high-risk area, but the model estimated low and medium risk for some census tracts. The same phenomenon occurred in the central portion of Entroncamento District. On the northern right part of Icoaraci District, the data recorded low prevalence, whereas the model predicted high values. Other minor discrepancies were also recorded, but such differences were insufficient to invalidate the model.

Figures 1 and 3 show that the modeling results were consistent with the inverse relationship between incidence level of the disease and income level, which further proved that the model had good approximation of reality.

#### 4 Conclusion

Because of the major existing socio-spatial inequalities in Brazilian and Latin American cities, the most vulnerable areas to major health disorders, among them pulmonary TB, must be identified and located in a detailed spatial scale as census tracts. This requirement is important to improve the effectiveness of public actions toward sick people. Georeferencing of addresses where infected people live, which are available at official sources, and subsequent plotting in a GIS could be a path in this direction. However, this procedure is quite laborious in urban areas with a large population.

This study proposed an alternative method to avoid this difficulty via a multivariate partial least squares regression, a statistical technique still little-known in social studies on health. This model aggregates data provided by official sources on the scale of neighborhoods. The model was then used to estimate the prevalence of pulmonary TB in the scale of census tracts. This method is relatively inexpensive in terms of time, but it has two disadvantages. It is less accurate than the plotting of addresses on a GIS and may not be valid, depending on parameters of the statistical model. On the other hand, the statistical modelling effort increases, by itself, the available knowledge on the local spatial distribution of pulmonary TB.

Another contribution of this paper is the proposal to model the spatial distribution of pulmonary tuberculosis prevalence, in urban areas marked by high levels of social inequalities, considering only persons aged 10 years and above and not the population as a whole. For the authors, this procedure provides a more accurate picture of the uneven geography of TB occurrences in Brazilian and Latin American cities.

In terms of strengths and weaknesses, the proposed method demonstrated good potential for modeling the spatial distribution of pulmonary TB prevalence in urban areas. This approach can be evaluated by conducting new case studies.

#### 5 References

ACOSTA, L.; BASSANESI, S. The Porto Alegre paradox: social determinants and tuberculosis incidence. **Rev Bras Epidemiol**, São Paulo, v.17, n.2, p.88-101, 2014.

BARBOSA, I. R.; COSME, C. L. Distribuição espacial dos casos novos de tuberculose pulmonar nos municípios da Região Norte do Brasil. **Caminhos de Geografia**, Uberlândia, v.14, n.47, p.110-21, Sept. 2013.

BRASIL. Ministério da Saúde. **SINAN**- Sistema de Informação de Agravos de Notificação. Brasília, 2014.

COHEN, J. A power primer. **Psych Bull**, Washington DC, v.112, n.1, p.155-9, Jul. 1992.

DYE, C. et al. Trends in tuberculosis incidence and their determinants in 134 countries. **Bull World Health Organ**, Geneva, v.87, n.9, p.683-91, 2009.

ESCOBAR A. L. et al. Tuberculose em populações indígenas de Rondônia, Amazônia, Brasil. **Cad Saúde Pública**, Rio de Janeiro, v.17, p.285-98, 2001.

HAIR, J. F. et al. **Análise multivariada de dados**. 6th ed. São Paulo: Pearson, 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 2010**. Rio de Janeiro, 2010a.

---

. **Censo Demográfico 2010:**  
Resultados gerais da amostra por áreas de ponderação. Rio de Janeiro: IBGE; 2010b.

JAMAL, L. F.; MOHERDAUI, F. Tuberculose e infecção pelo HIV no Brasil: magnitude do problema e estratégias para o controle. **Rev Saúde Pública**, São Paulo, v.41, Suppl.1, p.104-10, Sept. 2007.

KEENE, O. N. The log transformation is special. **Stat Med**, Malden (MA, USA), v.14, n.8, p.811-9, 1995.

MEDRONHO, R. A. Estudos ecológicos. In: MEDRONHO, R. A.; BLOCH, K. V.; WERNECK, K. V. **Epidemiologia**. 2 ed. São Paulo: Atheneu, 2009.

MENEZES, A. M. et al. Incidência e fatores de risco para tuberculose em Pelotas, uma cidade do Sul do Brasil. **Rev Bras Epidemiol**, São Paulo, v.1, n.1, p.50-60, Apr. 1998.

MEVIK, B.; WEHRENS, R. The pls Package: Principal component and partial least squares regression in R. **Journal of Statistical Software**, Los Angeles, v.18, n.2, p.1-23, Jan. 2007.

OLIVEIRA, H. B.; CARDOSO, J. C. Tuberculose no sistema prisional de Campinas, São Paulo, Brasil. **Rev Panam Salud Publica**, Washington DC, v.15, n.3, p.185-93, Mar. 2004.

QUEIROGA, R. P. et al. Distribuição espacial da tuberculose e a relação com condições de vida na área urbana do município de Campina Grande - 2004 a 2007. **Rev Bras Epidemiol**, São Paulo, v.15, n.1, p.222-32, Mar. 2012

ROBERTS, C. A.; BUIKSTRA, J. E. **The Bioarchaeology of tuberculosis: A global view on a reemerging disease**. Gainesville: University Press of California, 2003.

SÁNCHEZ, A. R. et al. A tuberculose nas prisões do Rio de Janeiro, Brasil: uma urgência de saúde pública. **Cad Saúde Pública**, Rio de Janeiro, v.23 n.3, p.545-52, 2007.

SANTOS, F. O. Saúde ambiental e as doenças respiratórias na cidade de Caldas Novas (GO). **Hygeia. Revista Brasileira de Geografia Médica e da Saúde**, Uberlândia, v.7, n.12, p.127-44, 2011.

SAN PEDRO, A.; OLIVEIRA, R. M. Tuberculose e indicadores socioeconômicos: revisão sistemática da literatura. **Rev Panam Salud Publica**, Washington DC, v.33, p.4, p.294-301, Apr. 2013.

SAWATSKY, M. L. et al. Partial least squares regression in the social sciences. **Quant Meth Psych**, Ottawa, v.2, n.11, p.52-62, 2015.

TEIXEIRA, E. C.; COSTA, J. S. O impacto das condições de vida e da educação sobre a incidência de tuberculose no Brasil. **Revista de Economia**, Curitiba, v.37, n.2, p.106-23, May./Aug. 2011.

VENDRAMINI, S. H. F. et al. Tuberculose em município de porte médio do Sudeste do Brasil: indicadores de morbidade e mortalidade, de 1985 a 2003. **J Bras Pneumol**, Brasília, v.31, n.3,



p.237-43, 2005.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v.58, n.2, p.109-30, 2001.

WORLD HEALTH ORGANIZATION. **Global tuberculosis report 2015**. Geneva, 2016. Available from: <[http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/)>. Access on 15 Feb. 2016.

WU, J.; DALAL, K. Tuberculosis in Asia and the Pacific: the role of socioeconomic status and health system development. **Int J Prev Med**, Mumbai (India), v.3, n.1, p.8-16, 2012.