Identification of Risk Areas for Visceral Leishmaniasis in Teresina, Piaui State, Brazil

Andréa S. de Almeida, Roberto de Andrade Medronho, and Guilherme L. Werneck*

Instituto de Medicina Social, Departamento de Epidemiologia, e Instituto de Estudos em Saúde Coletiva, Departamento de Medicina Preventiva, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil

Abstract. This study used spatial analysis to identify areas at greatest risk of visceral leishmaniasis (VL) in the urban area of Teresina, Brazil during 2001–2006. The results from kernel ratios showed that peripheral census tracts were the most heavily affected. Local spatial analysis showed that in the beginning of the study period local clusters of high incidence of VL were mostly located in the southern and northeastern parts of the city, but in subsequent years those clusters also appeared in the northern region of the city, suggesting that the pattern of VL is not static, and the disease may occasionally spread to other areas of the municipality. We also observed a spatial correlation between VL rates and all socioeconomic and demographic indicators evaluated (P < 0.01). The concentration of interventions in high-risk areas could be an effective strategy to control the disease in the urban setting.

INTRODUCTION

The urbanization process has raised new challenges for endemic disease control programs, particularly in the case of visceral leishmaniasis (VL). Because urbanization of VL is a relatively recent phenomenon in Brazil, there are few studies on the pattern of occurrence of this disease in cities.1–4 Accelerated urbanization in various regions of Brazil has been paralleled by other factors such as climate and environmental changes, precarious housing and sanitation, and adaptability of the vector Lutzomyia longipalpis to environments altered by humans. Thus, these factors have created favorable conditions for the emergence and persistence of VL in urban areas in the country.5

In the early 1980s, the first major urban VL epidemic in Brazil was reported in the municipality of Teresina, Piaui State.6 Intense population growth and migration were implicated in this phenomenon.7 Rapid and disorganized occupation of the city’s periphery exposed the population to extensive areas covered with tropical forests and dense vegetation, probable places for the wild reproduction of the parasite responsible for the disease, the protozoan Leishmania chagasi. This accelerated and disorganized urbanization also resulted in poor living and housing conditions, which may have also contributed to the emergence of the disease in the urban environment, because the vector Lu. longipalpis adapts easily to peridomestic conditions in impoverished areas.10,11

Dogs, which are considered the domestic reservoir of the protozoan, serve as the immediate source of infection for humans. Stray dogs wandering around the city’s outskirts can enter into direct contact with wild reservoirs and rapidly acquire the infection. When they reenter the city, these animals serve as amplifiers of the infection to other dogs and humans. Foxes, potential wild reservoirs for the parasite, were observed increasingly on the periphery of Teresina, scavenging in uncollected garbage in search of food.1 In this context, the presence of a large number of susceptible animals, infected reservoirs, and abundant vectors provide the basic conditions for the occurrence of autochthonous VL cases in the city of Teresina.12,13

The disease control strategy recommended by the Ministry of Health Visceral Leishmaniasis Control Program is based on reducing in degree of morbidity and case-fatality rates through early diagnosis and treatment of cases and decreasing the transmission risk by controlling the population of domestic reservoirs and the vector.14 However, some operational problems contribute to the insufficient effectiveness of control strategies aimed at reservoirs, e.g., the excessively long time between diagnosis and elimination of the infected dog; insufficient accuracy of the diagnostic tests commonly used to detect infected dogs, leading to the persistence of asymptomatic animals (which are also infective for vectors) in the environment; rapid replenishment of culled dogs by a new canine population subject to acquiring the infection; and high canine infection and infectiveness rates.15–17 In relation to vector control, operational difficulties and the high cost related to sustained large-scale intradomiciliary and peridomestic insecticide spraying, associated with limited knowledge on sand fly ecology and biology in urban areas, and the need for a comprehensive entomologic surveillance system providing qualitative and quantitative data on the vector are other factors that facilitate maintenance of transmission.11,16,19

From this perspective, the identification of high-risk areas for VL can be a useful strategy to increase the effectiveness of these control measures and optimize operational costs in the urban context. Thus, the current study aimed to identify areas at greatest risk of VL in the urban area of Teresina, Piaui State, Brazil, during 2001–2006.

MATERIALS AND METHODS

Study area. Teresina, the capital of Piaui State, occupies an area of 1,756 km² and has a population of 779,939 inhabitants, giving it a population density of 444.16 inhabitants/km², according to the 2007 population count by the Brazilian Institute of Geography and Statistics (IBGE). The city is located between the Paranaíba and Pothi Rivers, at 5°5’S, 42°48’W. The climate is tropical; the mean annual temperature is 27°C and annual rainfall is 1,300 mm. The highest temperatures begin in August and last until December, and the rainy season ranges from January through April. The periphery of the city has areas of pasture and tropical forest, but the predominant vegetation consists of sparse trees and bushes. According to the IBGE, the city has 654 census tracts and five geographic regions or subdivisions (Figure 1).
According to data from the 2000 census, some 90% of the permanent private households in Teresina have running water, only 13% are connected to the sewage system, and 86% have regular garbage collection. In addition, 66% of the heads-of-households earn less than three times the minimum wage and 17% have had no education.

**Study design.** This study was an ecological study in which units of analysis were urban census tracts in Teresina. Of the 654 census tracts, one located in the southern region of the city was excluded because of lack of data on its population.

**Data and variables.** Cases considered for the study were those available on the Brazilian Reportable Diseases Information System (SINAN) and those in the urban area of Teresina during 2001–2006. Because there are no other centers for treatment of VL cases near Teresina and because VL is a serious disease for which diagnosed cases must be reported before health services release the specific medication, which is only supplied by the Municipal Health Department, underreporting is believed to be relatively low.

Six hundred seventy VL cases, or 88.6% of the 756 cases reported to SINAN during 2001–2006, were georeferenced to the household level by using a global positioning system Universal TransverseMercator system. Georeferencing coverage varied from 82.1% in 2006 to 93.7% in 2001. Annual population estimates for each census tract were obtained by interpolation using a geometric growth equation with an annual component for population data from the 2000 Demographic Census and the 2007 Population Count.21

First, the annual incidence rates for each census tract were calculated by dividing the number of VL cases in each year by the estimated mid-year population for the corresponding year. The mean annual VL incidence rate was calculated by dividing the total number of new cases reported during 2001–2006 by the sum of the estimated populations for each of the years in the study.22

Seeking to minimize the problem of instability in the incidence rates calculated for small areas, we opted to correct them by using an empirical Bayesian approach.23,24 Empirical Bayesian procedures yield more reliable estimates because they use information from other areas to estimate the rates in a given region.25 This study used the local empirical Bayesian procedure, which includes the effects of spatial proximity, by using information from areas bordering the geographic area for which one wishes to estimate the incidence rate. In general, this procedure produces a set of incidence rates that when presented on a thematic map yield a less heterogeneous visual appearance than produced by uncorrected incidence rates. Therefore, this pattern is usually referred to as smoothed.

The indicators used in the analysis were constructed with data from the 2000 Demographic Census available on the IBGE website.20 These indicators were developed to reflect some key characteristics of the urban infrastructure and population in the census tracts in the study period, namely: 1) illiteracy rate; 2) children less than five years of age as a percentage of the total population; 3) mean income of heads-of-households; 4) percentage of permanent private households connected to the water supply; 5) percentage of households with regular garbage collection; and 6) percentage of permanent private households connected to the sewage system. The digital database for the urban census tract grid for 2000 was obtained from the IBGE.

**Spatial statistical analysis.** To describe the spatial correlation between the smoothed VL rate and socioeconomic and demographic indicators, the bivariate Moran’s I statistic was used, which tests whether geographically connected areas display greater similarity in terms of the target indicator than would be expected considering a random pattern. The I statistic can take values ranging from –1 to +1, and is positive for direct correlation and negative when inverse. The main idea is to evaluate whether values for a given variable in a region bear a relationship to values for another variable in neighboring regions.26,27

Thus, the bivariate global Moran I statistic (y and x) is expressed by the equation

\[ I^* = \frac{n}{\sum_i W_{ij} \sum_j (x_i - \bar{x})(y_j - \bar{y})}{\sum_i (x_i - \bar{x})^2} \]

where \( n \) represents the number of observations; \( W \) represents the matrix of spatial weights in which the elements \( W_{ij} \) indicate the way by which region \( i \) is spatially connected to region \( j \) and the elements of its main diagonal are equal to zero; \( x, y \) represent target variables; and \( \bar{x}, \bar{y} \) represent means of target variables.

The \( W \) neighborhood matrix, defined by contiguity, was the Queen matrix, which defines two regions as neighboring when they present common borders, in addition to common nodes (vertices).

Thematic maps were used to visualize the spatial distribution of VL incidence. For the maps of incidence rates (crude rates) and local empirical Bayesian incidence rates (smoothed rates) to be compared with each other, it was necessary to guarantee that the incidence range corresponding to a given class was the same in all the maps. Thus, the crude and smoothed incidence rates were discretized in six intervals, contemplating the wide range in these rates.

Areas with greater risk of VL were detected through the kernel ratio, a smoothed estimate of the risk surface, which is calculated by the ratio between the kernel (intensity) of
cases in relation to the kernel of the population distribution, obtained thusly:

$$
\hat{\rho}_m(s) = \frac{\sum_{j=1}^{n} K \left( \frac{s - s_j}{\tau} \right) y_j}{\sum_{j=1}^{n} K \left( \frac{s - s_j}{\tau} \right)}
$$

where $K()$ is the kernel weighting function; $\tau$ is the bandwidth; $s$ is the center of the area to be estimated; $s_j$ is the location of points (cases); $n$ is the total number of points (events); $m$ is the total number of points (population); $y_j$ is the population count for area $j$; and $\hat{\rho}_m(s)$ is the intensity estimator.

Kernel density is an appropriate interpolation technique for individual point positions that is based on a mathematical function that creates a symmetrical surface on each point, evaluating the distance from the point to a reference position, and later adding up the values of all the surfaces for this reference position.\(^{26,29}\)

 Besides estimating the intensity of occurrence of VL cases in the entire surface analyzed, this non-parametric approach enables filtering the variability of a dataset, while maintaining its principal local characteristics.

The smoothing function used was the quartic kernel, which assigns greater weight to closer events and less weight to more distant ones, but with a gradual decrease. The degree of smoothing is controlled by the choice of a parameter known as bandwidth, which is defined to reflect the geographic scale of the target hypothesis. The parameter influences the degree of smoothness on the estimated risk surface: the higher the $\tau$ value, the greater the smoothing. In this study, we opted to use the adaptive radius, which varies according to the density of points and the total extension of the area analyzed.

To determine the pattern of local clustering beyond the large-scale spatial variation in the incidence rates of VL, we used the local indicator of spatial autocorrelation (LISA).\(^{26,28}\) The LISA maps are used to identify five patterns: census tracts showing no statistically significant spatial autocorrelation (not significant); clusters of census tracts with low incidence of VL (Low-Low); clusters of census tracts with high incidence of VL (High-High); and two other showing a pattern of inverse spatial autocorrelation (High-Low and Low-High). The High-High pattern indicates hot-spots of high incidence of disease.\(^{26,28}\)

The software used to calculate crude and Bayesian incidence rates, implement the kernel ratio, and construct the maps was TerraView 3.2.0 (http://www.dpi.inpe.br/terraview_eng/index.php). The bivariate Moran’s I statistic and LISA were calculated by using the GeoDa 0.9.5-i application (http://geodacenter.asu.edu/software/downloads).

### RESULTS

The bivariate global Moran’s I index for 2001–2006 is shown in Table 1. We observed a spatial correlation between the smoothed VL incidence rate and all the indicators evaluated ($P < 0.01$).

Positive spatial correlation was observed for illiteracy rate and children less than five years of age as a proportion of the population, indicating that the higher the incidence in the census tract, the greater the proportion of illiterate persons and children less than five years of age in neighboring tracts. The mean income indicator showed an inverse spatial correlation, indicating that the higher the incidence in the census tracts, the lower the mean income in neighboring tracts. All the household-related indicators showed inverse spatial correlation, i.e., the higher the incidence rates in the census tracts, the worse the basic infrastructure conditions in neighboring tracts.

The maps shown in Figure 2 confirm the presence of higher rates in the peripheral census tracts, located in areas of expansion of the city, and lower risk areas in the central area of the city. As expected, when smoothing was applied to the VL incidence rates in Teresina, there was a major reduction in the extreme values (Figure 2).

Local Bayesian VL rates by census tracts during 2001–2006 are shown in Figure 3. A visual grasp of the thematic maps shows that the tracts with the highest incidence rates were located in the northern, eastern, and southern regions of the city, which also include the census tracts with the lowest percentage of households with running water, garbage collection (except for the southern region), and sewage connections, the highest illiteracy rate, and the lowest mean income (Figure 4). As for the proportion of children less than five years of age in the population, less than 25% of the population belonged to this age bracket in all the tracts, but the highest percentages were observed in the peripheral tracts and had the highest incidence rates (Figure 4).

The map for 2001 shows higher rates in the northern and southern regions, especially in the more peripheral census tracts, and in some tracts in the central and eastern regions. The map for 2002 shows that other tracts located in the eastern, southern, and southeastern regions were also more heavily affected. In 2003, there was an increase in incidence involving peripheral tracts to the north, east, and south. However,

### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Moran’s I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illiteracy rate</td>
<td>0.3427</td>
</tr>
<tr>
<td>Children less than five years of age as % of population</td>
<td>0.5054</td>
</tr>
<tr>
<td>Mean income of heads-of-households</td>
<td>-0.2222</td>
</tr>
<tr>
<td>% of permanent private households connected to water supply</td>
<td>-0.4138</td>
</tr>
<tr>
<td>% of households with regular garbage collection</td>
<td>-0.2091</td>
</tr>
<tr>
<td>% of permanent private households connected to sewage system</td>
<td>-0.4247</td>
</tr>
</tbody>
</table>

* $P < 0.01$ for all six variables.

### Figure 2

Crude incidence rate (A) and local empirical Bayesian rate (B) for visceral leishmaniasis according to census tracts, 2001–2006, Teresina, Piauí State, Brazil.
in 2004, some tracts on the northern, eastern, and southern periphery showed slight decreases in their incidence rates, and some tracts in the southern and central regions were affected for the first time. This situation persisted into 2005, except in the southeastern region. In 2006, various peripheral tracts close to the expansion areas in the eastern region and in the southern and northern regions of the city showed higher incidence rates (Figure 3).

Regions with the highest VL risk, estimated by the kernel ratio, during 2001–2006 are shown in Figure 5. The intensity of gray levels is directly related to the intensity of VL occurrence in the study area, so the darker the area, the greater the risk of VL.

Initially, one observes that the census tracts in the southern region of the city generally show medium occurrence of VL in nearly every year studied; the greatest intensities were in 2003. In 2001 (Figure 5), we observed low-risk VL foci in peripheral census tracts in the north, east, south, and southeast. In 2002, we observed a similar VL spatial distribution as in 2001, but with a slight increase in VL intensity in the eastern and southern census tracts, which showed medium risk. In 2003 and 2004, we observed a sharp increase in VL intensity.
in all the regions of the city, but with a high-risk focus in the southern region, evident in 2003. In 2005, the risk decreased in the urban area as a whole, but there was medium-risk VL foci in southern and southeastern regions. In 2006, there was also an important reduction in VL intensity in the urban area as a whole, but there was a new medium-risk foci in northern and southern regions. Thus, results obtained from the kernel ratio indicated that the peripheral census tracts located in the northern, eastern, and southern regions of the city showed higher risks of VL in every year analyzed, although in 2001 and 2005 the risk was relatively lower (Figure 5).

Maps of LISA are shown in Figure 6. These maps show that during the entire study period, most of the central areas of the city census tracts showed no significant spatial autocorrelation or a pattern of clustering of low incidence rates (Low-Low). In 2001 and 2002, local clusters of high incidence of VL

![Figure 5. Kernel ratio for visceral leishmaniasis by year in the census tracts of Teresina, Piauí State, Brazil, 2001–2006.](image)

![Figure 6. Maps of the local indicator of spatial autocorrelation (LISA cluster maps) of the annual local empirical Bayesian rate for visceral leishmaniasis according to census tracts, Teresina, Piauí State, Brazil, 2001–2006.](image)
(High-High) were mostly located in the southern and northeastern parts of the city, but in the next three years those clusters spread also to the northern region of the city.

**DISCUSSION**

This study showed that VL incidence rates and socioeconomic and demographic indicators are spatially correlated, indicating that disease foci are associated with precarious living conditions. Thus, although the study design did not enable establishing causal nexuses, it is plausible to suppose that the introduction, maintenance, and spread of VL in Teresina are linked to the environmental conditions typically found in poorer areas with less urban infrastructure, including worse sanitation. Such conditions favor not only the proliferation of VL vectors and infected reservoirs, but also the presence of a large number of susceptible persons, in particular, poor, probably malnourished, young persons, thereby forming the basis for cases of the disease to occur.

The results obtained from smoothed VL incidence rates and kernel ratios showed that the peripheral census tracts located in the northern, eastern, and southern regions of the city were the most heavily affected. A more detailed analysis by using a local indication of spatial autocorrelation showed that in the beginning of the study period local clusters of high incidence of VL were mostly located in the southern and northeastern parts of the city, but in subsequent years those clusters also appeared in the northern region of the city. These results suggests that the ideal conditions for establishment and maintenance of transmission are found in these places and that the pattern of VL occurrence is not static and disease may occasionally spread to other areas of the municipality.

An operational spinoff of this study is the indication that the use of the kernel ratio to detect areas at highest risk of occurrence of VL cases may be useful for assisting LV surveillance and control measures. The identification of focal areas at greatest risk can help define priority areas for specific interventions. As already shown in relation to transmissible diseases with heterogeneous spatial distribution, targeted interventions tend to be the most effective.

Two procedures used in this study helped minimize some problems related to the use of aggregate secondary data areas demarcated for political and administrative purposes. For example, the use of the local Bayesian method enabled correcting occasional underreporting of cases in small areas. By means of this procedure, these areas had their rates corrected by the information from neighboring areas. Use of the kernel ratio also proved adequate for differentiating risk areas independently of the political and administrative boundaries, which can be useful for defining areas or neighborhoods in which interventions should be targeted.

Some limitations of this study should also be highlighted. First, despite the relatively high coverage of geocoding of VL cases, it is likely to have been worse in peripheral areas of the city. Likewise, there were variations in coverage over the years. Both situations may have caused distortions in the estimates, possibly leading to either spurious identification of high-risk areas or non-identification of specific foci. Second, even assuming that most of the infections occur in and around households, georeferencing at the household level may not be a good indicator of risk for all transmission situations because one cannot rule out that certain infections occurred far from the home. Third, although from the public health perspective the control of symptomatic cases is a priority, the non-use of markers for infection prevents the demarcation of transmission foci because more infections occur than cases of the disease and the evolution from asymptomatic infected status to symptomatic status is mediated by genetic and nutritional factors. Finally, as mentioned, this study is based on secondary data available in SINAN. Thus, although VL is a serious disease subject to mandatory reporting (and the treatment of which is only provided by government health services), the available data are based on passive case detection, and one cannot rule out the possibility of case underreporting.

Despite these limitations, the study showed that using a combination of techniques for spatial analysis of point patterns and area data, it was possible to identify areas at greatest risk of VL in Teresina. One can thereby infer that occurrence of the disease in the urban census tracts of Teresina is a focal process, which corresponds to the limited flight autonomy of the sand fly. Thus, the concentration of interventions in high-risk areas could be an effective strategy to control the disease in the urban setting, reducing not only the operational costs but also contributing to the sustainability of control programs.