1-2007

Multilevel modelling of the incidence of visceral leishmaniasis in Teresina, Brazil

G Werneck
Universidade do Estado do Rio de Janeiro, gwerneck@nesc.ufrj.br

C Costa
Universidade Federal do Piaui

Alexander Walker
Harvard University, amwalker@hsph.harvard.edu

John David
Harvard University

M. P. Wand
University of Wollongong, mwand@uow.edu.au

See next page for additional authors

Multilevel modelling of the incidence of visceral leishmaniasis in Teresina, Brazil

Abstract

Epidemics of visceral leishmaniasis (VL) in major Brazilian cities are new phenomena since 1980. As determinants of transmission in urban settings probably operate at different geographic scales, and information is not available for each scale, a multilevel approach was used to examine the effect of canine infection and environmental and socio-economic factors on the spatial variability of incidence rates of VL in the city of Teresina. Details on an outbreak of greater than 1200 cases of VL in Teresina during 1993-1996 were available at two hierarchical levels: census tracts (socio-economic characteristics, incidence rates of human VL) and districts, which encompass census tracts (prevalence of canine infection). Remotely sensed data obtained by satellite generated environmental information at both levels. Data from census tracts and districts were analysed simultaneously by multilevel modelling. Poor socio-economic conditions and increased vegetation were associated with a high incidence of human VL. Increasing prevalence of canine infection also predicted a high incidence of human VL, as did high prevalence of canine infection before and during the epidemic. Poor socio-economic conditions had an amplifying effect on the association between canine infection and the incidence of human VL. Focusing interventions on areas with characteristics identified by multilevel analysis could be a cost-effective strategy for controlling VL. Because risk factors for infectious diseases operate simultaneously at several levels and ecological data usually are available at different geographical scales, multilevel modelling is a valuable tool for epidemiological investigation of disease transmission.

Keywords

brazil, teresina, leishmaniasis, incidence, modelling, multilevel, visceral

Disciplines

Engineering | Science and Technology Studies

Publication Details

Werneck, G. L., Costa, C. H. N., Walker, A. M., David, J. R., Wand, M. P.. & Maguire, J. H. (2007). Multilevel modelling of the incidence of visceral leishmaniasis in Teresina, Brazil. Epidemiology and Infection, 135 (2), 195-201.

Authors

G Werneck, C Costa, Alexander Walker, John David, M. P. Wand, and J Maguire

Epidemiol. Infect. (2007), **135**, 195–201. © 2006 Cambridge University Press
doi:10.1017/S0950268806006881 Printed in the United Kingdom

Multilevel modelling of the incidence of visceral leishmaniasis in Teresina, Brazil

G. L. WERNECK[1,2]*, C. H. N. COSTA[3], A. M. WALKER[4], J. R. DAVID[1], M. WAND[5] AND J. H. MAGUIRE[1]

[1] *Department of Immunology and Infectious Disease, Harvard School of Public Health, Boston, MA, USA*
[2] *Instituto de Medicina Social/IMS, Departamento de Epidemiologia, Universidade do Estado do Rio de Janeiro, NESC/UFRJ, Brazil*
[3] *Instituto de Doenças Tropicais Natan Portella, Universidade Federal do Piauí, Brazil*
[4] *Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA*
[5] *Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA*

(*Accepted 25 April 2006; first published online 7 July 2006*)

SUMMARY

Epidemics of visceral leishmaniasis (VL) in major Brazilian cities are new phenomena since 1980. As determinants of transmission in urban settings probably operate at different geographic scales, and information is not available for each scale, a multilevel approach was used to examine the effect of canine infection and environmental and socio-economic factors on the spatial variability of incidence rates of VL in the city of Teresina. Details on an outbreak of greater than 1200 cases of VL in Teresina during 1993–1996 were available at two hierarchical levels: census tracts (socio-economic characteristics, incidence rates of human VL) and districts, which encompass census tracts (prevalence of canine infection). Remotely sensed data obtained by satellite generated environmental information at both levels. Data from census tracts and districts were analysed simultaneously by multilevel modelling. Poor socio-economic conditions and increased vegetation were associated with a high incidence of human VL. *Increasing* prevalence of canine infection also predicted a high incidence of human VL, as did high prevalence of canine infection *before* and *during* the epidemic. Poor socio-economic conditions had an amplifying effect on the association between canine infection and the incidence of human VL. Focusing interventions on areas with characteristics identified by multilevel analysis could be a cost-effective strategy for controlling VL. Because risk factors for infectious diseases operate simultaneously at several levels and ecological data usually are available at different geographical scales, multilevel modelling is a valuable tool for epidemiological investigation of disease transmission.

INTRODUCTION

Heterogeneity in exposure to risk factors leads to spatial and temporal variability in transmission rates of infectious agents [1, 2]. To understand these patterns of disease spread, it is necessary to realize that not all risk factors are reducible to individual or local attributes. Factors that vary at large ecological levels can be important determinants of infection rates in smaller regions. For instance, unvaccinated persons living in a region where a vaccination programme has been completed enjoy a lower risk of infection than unvaccinated persons living in areas with no intervention [3].

Determinants of the occurrence of zoonotic vector-borne diseases, such as visceral leishmaniasis (VL)

* Author for correspondence: Dr G. L. Werneck – Instituto de Medicina Social/IMS, Departamento de Epidemiologia, Universidade do Estado do Rio de Janeiro (UERJ), Rua São Francisco Xavier 524, 70 andar, Bloco D, Maracanã, Rio de Janeiro, RJ, Brazil 20559-900.
(Email: gwerneck@nesc.ufrj.br)

in the Americas also operate at several levels. For instance, on a broad scale, climate and land cover determine the habitat of the vector *Lutzomyia longipalpis* and the size and longevity of its population [4–6]. At the community level, factors such as land use and quality of housing influence vector populations and their interaction with susceptible persons [7, 8]. At the level of individual persons, young age and malnutrition increase the risk of development of overt VL following infection [7, 9].

The interplay of factors operating at various ecological levels undoubtedly underlies the geographical clustering of cases of VL that has been observed in the Americas and elsewhere [10, 11]. In an earlier report, we demonstrated both large-scale and small-scale variation in the incidence rates of VL during an epidemic in the Brazilian city of Teresina [12]. In this paper, we report a multilevel modelling approach to further examine the effect of socio-economic factors, landscape features and rates of canine infection, each operating at a different geographical scale, on the spatial distribution of human disease.

MATERIALS AND METHODS

Study area

Teresina, the capital of the state of Piauí, Brazil, occupies an area of 176 km² at the confluence of the Parnaíba and Poti rivers, 72 m above sea level and 339 km inland at 05° 05′ latitude South and 42° 48′ longitude West. The climate is tropical with an average temperature of 27 °C and total annual rainfall of 1300 mm. The predominant vegetation within the city consists of grass, shrubs and sparse mango and palm trees. Peri-urban areas are covered by tropical forest and farmland.

Until 1980, infrequent and sporadic cases of VL had occurred in Teresina. Between 1980 and 1985, the first urban epidemic of VL in Brazil occurred in Teresina, when almost 1000 new cases were detected as the population increased from 370 000 to 460 000 inhabitants [13]. The incidence declined and remained at low levels until 1992 when a new epidemic began. By 1996, at which time the city's population had grown to 650 000 inhabitants, there had been more than 1200 new cases, of which over 90% required hospitalization and 5% resulted in death despite treatment.

For administrative purposes, the city is divided into 494 census tracts within 74 districts. At the district level, the National Health Foundation [Fundação Nacional de Saúde (FNS)] is responsible for control activities, such as canine surveys for infection, and insecticide spraying. At the census tract level, the Brazilian Institute of Geography and Statistics (IBGE) collects and reports socio-economic and demographic information.

Human and canine data

The age, date of diagnosis, and geographic location of the residence of 1061 cases of VL that occurred in Teresina from 1993 to 1996 were obtained from FNS and confirmed from clinical and laboratory records from all hospitals in Teresina. This figure represents about 95% of the total VL cases reported to FNS during this period. It is likely that few cases of VL were overlooked, since there is no alternative centre for treating VL close to Teresina, and, by law, all suspect and confirmed cases of VL are reported to FNS, which is the sole distributor of anti-leishmanial drugs in Brazil.

Incidence rates of VL were calculated for each of the city's 494 census tracts, using data from the 1991 and 1996 censuses. Prevalence data on canine infection with *L. chagasi* were available from 1987 to 1994 for 63 of the city's 74 districts; the 11 districts without information of canine infection were excluded from the analysis. For the analysis, the original census tracts were consolidated into 430 areas (consolidated census tracts) so that at least one case of VL would be expected in each tract had cases been distributed uniformly throughout the city. Aggregation of census tracts was based on similarities in socio-economic profiles and spatial proximity. Using a similar aggregation strategy, the 63 districts were aggregated into 39 areas (consolidated districts), each with a minimum of three consolidated census tracts, in order to ensure adequate information on canine infection for analysis. The prevalence of canine infection was grouped into 2-year periods (1987–1988, 1989–1990, 1991–1992, and 1993–1994), and the change in prevalence calculated from each period to the next (Table 1).

A socio-economic status (SES) index was derived for each consolidated census tract by principal-component analysis [14] (SAS®, SAS Institute Inc., Cary, NC, USA) using data obtained during the 1991 Brazilian census on household characteristics such as running water, indoor sanitation, garbage collection, level of education, family income and adequacy of

Table 1. *Variables included in the multilevel analysis*

Variable	Reference	Definition
Level 1 (census tract)		
LINC	—	Natural logarithm of the incidence rates of visceral leishmaniasis
SES*	—	Socio-economic status index
URB†	—	Urbanization index
Level 2 (district)		
NDVI	Minimum	Minimum value of the Normalized Difference Vegetation Index
	Maximum	Maximum value of the Normalized Difference Vegetation Index
	Mean	Average value of the Normalized Difference Vegetation Index
PREV	1987/88	Prevalence of infection in dogs in 1987/88
	1989/90	Prevalence of infection in dogs in 1989/90
	1991/92	Prevalence of infection in dogs in 1991/92
	1993/94	Prevalence of infection in dogs in 1993/94
	1987/88→1989/90	Relative change in the prevalence of infection in dogs from 1987/88 to 1989/90
	1989/90→1991/92	Relative change in the prevalence of infection in dogs from 1989/90 to 1991/92
	1991/92→1993/94	Relative change in the prevalence of infection in dogs from 1991/92 to 1993/94

* Based on a principal components analysis of the following variables: % of households connected to the water supply system; % of households with presence of water taps; % of households connected to the sewage disposal system; % of households with indoor sanitation; % of households with regular garbage collection; % of population of the census tract with basic education; % of the heads of the households with basic education, mean income of the head of the household, mean number of persons per household; and % of population living in *favelas*.

† Based on a correspondence analysis of the number of pixels in each census tract classified as water, forest, riparian vegetation, mixed vegetation, shrub/scrub, secondary growth, asphalt roads, pasture, grass/some bare, commercial/residential, residential trees, bare/some grass, medium density residential, high density residential, new construction, and bare.

housing (Table 1). The SES index was the first principal-component factor, which explained 55% of the total variance. Values of the index ranged from positive (wealthiest census tracts) to negative (poorest).

Environmental data

Landscape features were identified by remote sensing using a Landsat 5 Thematic Mapper (TM) scene (6 bands, 30 m resolution) of Teresina during October 1995. Pixels were assigned to one of 30 clusters using an unsupervised classification algorithm (Isoclust, using Imagine® software; ERDAS Inc., Atlanta, GA, USA). Clusters were then grouped into 16 land cover classes by comparison with georeferenced data collected on the ground and with colour aerial and ground-level photographs [15]. Environmental features were also characterized using the Normalized Difference Vegetation Index (NDVI), defined as [16]:

$$\mathrm{NDVI} = (\mathrm{Ch2} - \mathrm{Ch1})/(\mathrm{Ch2} + \mathrm{Ch1}),$$

where Ch1 is the reflectance from each pixel in the red wavelength band (Landsat band 3) and Ch2 is the reflectance in the near-infrared wavelength band (Landsat band 4). NDVI varies from $-1{\cdot}0$ to $+1{\cdot}0$ with positive values in general indicating green vegetation, and negative values indicating lack of green vegetation. NDVI correlates positively with rainfall and humidity, factors that are related to sandfly abundance [16, 17]. In this study we determined the minimum, the maximum, and the mean NDVI over the pixels in each district.

Digital maps of the consolidated census tracts and districts were produced using CartaLinx® (Clark Labs, Worcester, MA, USA). IDRISI® software (Clark Labs) was used to overlay the digital map on the RS image to extract the land cover and NDVI information for census tracts and district.

An urbanization index was obtained by applying correspondence analysis [18] (SAS Inc.) to the portion of land-cover classes found in each consolidated census tract. The urbanization index was the first correspondence analysis factor, which explained about 39% of the total inertia of the matrix. It is a continuous variable on a scale extending from 'high density residential and commercial areas' to 'heavily vegetated areas with few residences'.

Statistical analysis

The incidence rates of VL and prevalence of canine seropositivity were linked in IDRISI to the

consolidated census tracts and districts respectively. A multilevel model [19] was then used to analyse data simultaneously at the consolidated census tract and district levels (Table 1). A general model for our data can be conceptualized as follows, with the random variables underlined [20]:

$$\ln(\underline{\mathrm{INC}}_{ij}) = \underline{\alpha}_j + \underline{\beta}_{1j}\,\mathrm{SES}_{ij} + \beta_{2j}\,\mathrm{URB}_{ij} + \underline{\varepsilon}_{ij}$$
$$\varepsilon_{ij}\ \mathrm{iid} \sim N(0, \sigma^2). \quad (1)$$

The natural logarithm of the VL incidence rates for the ith census tract in the jth district [$\ln(\mathrm{INC}_{ij})$] is the continuous outcome variable (LINC). The explanatory variables SES_{ij} and URB_{ij} are the SES and urbanization indices in the ith census tract of the jth district. The intercept ($\underline{\alpha}_j$) and slopes ($\underline{\beta}_{1j}$ and $\underline{\beta}_{2j}$) are not deemed to be fixed as they would be in standard linear regression, but are allowed to vary from one district to the other. This is the so-called random coefficients model [20] in which each random coefficient consists of two components. The first component is the overall value of the coefficient, estimated for all census tracts, independently of the districts to which they belong. The second component is the coefficient variance measuring the deviations of districts from that overall effect [20].

Equations (2)–(4) relate the district level variables PREV_j and NDVI_j to the random intercept ($\underline{\alpha}_j$) and random slopes ($\underline{\beta}_{1j}$ and $\underline{\beta}_{2j}$):

$$\underline{\alpha}_j = \gamma_{00} + \gamma_{01}\,\mathrm{PREV}_j + \gamma_{02}\,\mathrm{NDVI}_j + \underline{\sigma}_{0j}, \quad (2)$$

$$\underline{\beta}_{1j} = \gamma_{10} + \gamma_{11}\,\mathrm{PREV}_j + \gamma_{12}\,\mathrm{NDVI}_j + \underline{\sigma}_{1j}, \quad (3)$$

$$\underline{\beta}_{2j} = \gamma_{20} + \gamma_{21}\,\mathrm{PREV}_j + \gamma_{22}\,\mathrm{NDVI}_j + \underline{\sigma}_{2j}, \quad (4)$$

where PREV_j is one of the four 2-year prevalences of infection in dogs or one of the three relative changes in prevalence between periods, NDVI_j represents one of the three NDVI estimates (minimum, maximum or mean) for districts, and $\underline{\delta}_j$ are the error terms at the district level.

By substituting equations (2), (3), and (4) in equation (1):

$$\begin{aligned}\ln(\underline{\mathrm{INC}}_{ij}) &= \gamma_{00} + \gamma_{01}\,\mathrm{PREV}_j + \gamma_{02}\,\mathrm{NDVI}_j + \underline{\sigma}_{0j} \\ &+ (\gamma_{10} + \gamma_{11}\,\mathrm{PREV}_j + \gamma_{12}\,\mathrm{NDVI}_j + \underline{\sigma}_{1j})\mathrm{SES}_{ij} \\ &+ (\gamma_{20} + \gamma_{21}\,\mathrm{PREV}_j + \gamma_{22}\,\mathrm{NDVI}_j + \underline{\sigma}_{2j})\mathrm{URB}_{ij} + \underline{\varepsilon}_{ij}. \end{aligned} \quad (5)$$

Expanding and rearranging terms yields:

$$\begin{aligned}\ln(\underline{\mathrm{INC}}_{ij}) &= \gamma_{00} + \gamma_{01}\,\mathrm{PREV}_j + \gamma_{02}\,\mathrm{NDVI}_j + \gamma_{10}\,\mathrm{SES}_{ij} \\ &+ \gamma_{20}\,\mathrm{URB}_{ij} + \gamma_{11}\,\mathrm{PREV}_j\,\mathrm{SES}_{ij} \\ &+ \gamma_{12}\,\mathrm{NDVI}_j\,\mathrm{SES}_{ij} + \gamma_{21}\,\mathrm{PREV}_j\,\mathrm{URB}_{ij} \\ &+ \gamma_{22}\,\mathrm{NDVI}_j\mathrm{URB}_{ij} \\ &+ (\underline{\sigma}_{0j} + \underline{\sigma}_{1j}\,\mathrm{SES}_{ij} + \underline{\sigma}_{2j}\,\mathrm{URB}_{ij} + \underline{\varepsilon}_{ij}). \end{aligned} \quad (6)$$

The result is a single equation that resembles a traditional regression equation with a complex error term. Equation (6) includes estimates for the overall grand mean effect (γ_{00}), the main effects of the district-level variables (γ_{01} and γ_{02}), the main effects of the census tract level variables (γ_{10} and γ_{20}), and the four cross-level interaction effects (γ_{11}, γ_{12}, γ_{21}, and γ_{22}). The deviation of each district from the overall grand mean is measured by $\underline{\delta}_{0j}$, while $\underline{\delta}_{1j}$ and $\underline{\delta}_{2j}$ measure the deviations of each district from the SES and URB grand slopes respectively, after taking into account the effects of PREV_j and NDVI_j. Model 6 is essentially a mixed-effects model with random intercepts and random slopes for each district, fitted using SAS PROC MIXED (SAS Institute Inc.).

Twenty-one separate models were fitted, one for each of the seven PREV_j variables with each of the three NDVI_j variables (Table 1). All models were adjusted for the census tract level variables SES_{ij} and URB_{ij}. By using backward elimination and comparing the deviances and Akaike's Information Criteria (AIC) we chose as final models the most parsimonious version of each of the 21 saturated models [20, 21]. Models that best fit the data had lower deviance values and/or larger AIC values compared to the other versions. All variables were treated as continuous in the analysis.

RESULTS

Of the 21 final models, only the three that included the minimum NDVI with either the prevalence of infection in dogs in 1991–1992, the prevalence of infection in dogs in 1993–1994, or the relative change in prevalence from 1989–1990 to 1991–1992 significantly explained variability in VL incidence rates over Teresina's census tracts (Table 2).

In Model 1, which included the minimum NDVI and prevalence of infection in dogs in 1991–1992 (P9192) as explanatory variables, the fixed-effects component showed that the more urbanized census tracts had lower LINC, and districts with high minimum NDVI had higher incidence rates of VL.

Table 2. *Random coefficients models for the effects of measures of prevalence of infection in dogs on visceral leishmaniasis incidence rates in Teresina, Brazil, 1993–1996*

	Model 1		Model 2		Model 3	
	Estimate	*P* value	Estimate	*P* value	Estimate	*P* value
Fixed part						
Intercept	−0·523	0·007	−0·622	<0·001	−0·446	0·002
SES (census tract)	0·036	0·218	−0·090	0·002	−0·078	0·004
Urbanization index (census tract)	−0·729	0·006	−0·546	0·026	−0·647	0·010
Minimum NDVI (district)	0·941	0·040	0·999	0·026	0·844	0·051
Urbanization * NDVI	−1·950	0·009	−1·674	0·020	−1·821	0·012
Prevalence of infection in dogs (district)						
1991/92	6·493	0·229				
1993/94			2·582	0·012		
1989/90→1991/92					0·131	0·027
SES * Prevalence of infection in dogs						
1991/92	−6·010	<0·001				
1993/94						
1989/90→1991/92					−0·049	0·007
Random part						
Intercept	0·067	0·007	0·036	0·088	0·058	0·045
SES slope			0·013	0·025	0·011	0·035
Covariance			−0·001	0·887	−0·005	0·638
Residual variance	0·334	<0·001	0·319	<0·001	0·319	<0·001
Deviance	754·41		761·96		774·75	
AIC	−379·21		−384·98		−391·37	

SES, Socio-economic status index; NDVI, Normalized Difference Vegetation Index; AIC, Akaike's Information Criteria.

There was a significant cross-level interaction between URB and NDVI. The effect of a given P9192 level on LINC varied depending on the SES index level. As an illustration of this effect, holding all other variables at the same level, a canine prevalence of 3% (1991–1992) in a poor census tract (SES index of −3·0) was associated with an increase of 0·63 points in LINC, or about 1·9 cases/1000 person-years, beyond that in a wealthy census tract (SES index of 3), for which the canine prevalence of 3% would have produced only 0·8 cases/1000 person-years. The random component of Model 1 shows that the LINC grand mean varied between districts even after controlling for the effects of SES, URB, P9192, NDVI, and interaction terms. In this model, there was no evidence for variation in the association between SES and VL across districts.

Model 2 showed an association between P9394 and LINC but no cross-level interaction between SES and P9394. The random component of the model showed that the SES slope varied among districts after controlling for the effects of SES, URB, P9394, NDVI, and interaction terms.

In addition to results of the fixed and random components which were similar to those of Model 2, Model 3 showed that the SES level in census tracts modulated the relationship between CHANGE2 (relative change in canine prevalence from 1989 to 1992) and LINC. A positive CHANGE2 increased LINC to a greater extent in census tracts that had lower SES.

DISCUSSION

Epidemics of VL in large Brazilian cities are a new phenomenon since 1980. Little is known about the dynamics of transmission of *L. chagasi* in urban and peri-urban centres or the variables that determine the distribution of disease in these settings. In this study, multilevel analysis of the epidemic of VL in Teresina during 1993–1996 demonstrates the importance of the interactions of environmental, human, and canine factors in establishing patterns of disease occurrence.

The independent effect of several factors demonstrated in this study is not unexpected. High rates of disease in neighbourhoods with low socio-economic

standing may result from poor living conditions and location in the periphery of the city where vegetation density is favourable for vector populations and perhaps sylvatic canine reservoir hosts. Malnutrition, a known risk factor for the development of VL among persons infected with *L. chagasi*, is also likely to be more prevalent in the poorer areas of the city. A high or rising prevalence of infection in dogs would lead to an increase in the incidence of human disease if the dog is the principal or sole reservoir of the infection as is commonly believed.

In this study, a high level of infection in dogs, particularly in impoverished areas, just *before* the start of the human epidemic (P9192), or *during* the epidemic (P9394), or a rising prevalence (CHANGE2) of canine infection before the epidemic all strongly predicted a high incidence of VL. These results seem to support the widely held hypothesis that identification and killing of infected dogs can reduce rates of human VL. If this were true, targeting areas with rising rates of canine infection would be more effective than targeting only areas with high absolute prevalence since the latter approach would miss low-prevalence areas experiencing a recent increase.

However, other possible explanations linking high prevalences of infections in dogs and high incidence of VL among humans should also be considered. It is possible, for instance, that dogs and humans are under the same epizootic pressure, but dogs are more susceptible and develop disease before the human population, and there is no direct connection between infection in the two populations. Although a link between canine and human infection seems more likely [22, 23], the failure of the strategy of killing dogs to reduce the incidence of human VL supports the need of further research on the role of the infected dog in the transmission of human VL.

An important new finding of this study is the influence of socio-economic conditions on the relationship between canine infection and human VL. In areas of low socio-economic standing, high or rising rates of infection in dogs were predictive of high rates of human disease, perhaps because the degree of contact among susceptible people, infectious dogs, and infected vectors is greater in impoverished areas than in wealthy areas, even if the prevalence of infection in dogs is similar in both areas. An alternative explanation for the finding is that residents of wealthy areas with high prevalence of infection in dogs before the epidemic may have demanded intervention by the government that resulted in lower rates of human VL during the epidemic. The lack of a significant interaction between SES and the prevalence in dogs in 1993–1994 suggests that implementation of control measures once the epidemic was underway may have been too late to be effective, even in wealthy neighbourhoods. Whatever the explanation, our findings have practical implications for control programmes. Interventions to interrupt transmission should focus upon poor neighbourhoods experiencing high or rising rates of canine infection.

Several criticisms concerning methodological aspects of this study deserve mention. First, because the design of this study is ecological, and both the outcome and explanatory variables were measured at the group level (census tracts and districts), extrapolation of the findings to individual persons or households may not be appropriate. However, even though analysis at the individual level is important to enhance understanding of the epidemiology of VL, ecological effects are particularly relevant when interventions to control disease are executed at the level of entire communities [24].

Second, this study used geographical scales to define the ecological analyses that were not grounded on theory, but on availability of information. Areas with boundaries and shapes defined for administrative reasons may not accurately capture the phenomena under investigation, and a better design might have employed regular blocks with equal numbers of households and separated from each other by a distance sufficiently large to ensure independence of transmission cycles. However, use of ecological levels defined for administrative reasons is convenient, and has the advantage of generating results that are compatible with the geographical scales at which public health interventions are implemented.

Multilevel modelling has not been employed frequently in epidemiological analyses of infectious diseases. However, surveillance data often are gathered at several geographical scales and the planning, execution and design of public health interventions commonly involves multiple layers of data. From the statistical point of view, carrying out an analysis that does not take the hierarchical structure of the data into consideration will generally cause standard errors of regression coefficients to be underestimated [25]. An alternative approach, modelling variation between districts by introducing separate terms for each district or sector, would require estimates of many times more coefficients than the multilevel procedure [25]. Multilevel modelling, used with appropriate

attention to definitions of levels of hierarchy and theory of disease spread, is an attractive approach to the analysis of the type of data available in many epidemiological studies of infection.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Anderson RM, May RM.** *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press, 1992, p. 17.
2. **Woolhouse ME, *et al.*** Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences USA* 1997; **94**: 338–342.
3. **Halloran ME, Longini Jr. IM, Struchiner CJ.** Design and interpretation of vaccine field studies. *Epidemiologic Review* 1999; **21**: 73–88.
4. **Young DG, Lawyer PG.** New World vectors of the leishmaniases. *Current Topics in Vector Research* 1987; **4**: 29–71.
5. **Quinnell RJ, Dye C.** Correlates of the peridomestic abundance of *Lutzomyia longipalpis* (Diptera: Psychodidae) in Amazonian Brazil. *Medical and Veterinary Entomology* 1994; **8**: 219–224.
6. **Dias-Lima AG, Guedes ML, Sherlock IA.** Horizontal stratification of the sand fly fauna (Diptera: Psychodidae) in a transitional vegetation between caatinga and tropical rain forest, state of Bahia, Brazil. *Memórias do Instituto Oswaldo Cruz* 2003; **98**: 733–737.
7. **Tesh RB.** Control of zoonotic visceral leishmaniasis: is it time to change strategies? *American Journal of Tropical Medicine and Hygiene* 1995; **52**: 287–292.
8. **Dye C.** The logic of visceral veishmaniasis control. *American Journal of Tropical Medicine and Hygiene* 1996; **55**: 125–130.
9. **Dye C, Williams BG.** Malnutrition, age and the risk of parasitic disease: visceral leishmaniasis revisited. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 1993; **254**: 33–39.
10. **Magill AJ.** Epidemiology of the leishmaniases. *Dermatologic Clinics* 1995; **13**: 505–521.
11. **Franke CR, *et al.*** Trends in the temporal and spatial distribution of visceral and cutaneous leishmaniasis in the state of Bahia, Brazil, from 1985 to 1999. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2002; **96**: 236–241.
12. **Werneck GL, *et al.*** The urban spread of visceral leishmaniasis: clues from spatial analysis. *Epidemiology* 2002; **13**: 364–367.
13. **Costa CHN, Pereira HF, Araújo MV.** Epidemia de leishmaniose visceral no estado do Piauí, Brasil, 1980–1986. *Revista de Saúde Pública* 1990; **24**: 361–372.
14. **Armitage P, Berry G.** *Statistical Methods in Medical Research*, 3rd edn, London: Blackwell Scientific Publications, 1994, pp. 327–333.
15. **Maguire JH, *et al.*** Application of Remote Sensing and Geographical Information Systems (GIS) technology to study the transmission of Leishmania chagasi in Teresina, Piauí, Brazil: First year report (http://geo.arc.nasa.gov/sge/health/projects/leishb/leishbrpt1.html). Acessed 6 June 2005.
16. **Hay SI, *et al.*** Remotely sensed surrogates of meteorological data for the study of the distribution and abundance of arthropod vectors of disease. *Annals of Tropical Medicine and Parasitology* 1996; **90**: 1–19.
17. **Elnaiem DE, *et al.*** Risk mapping of visceral leishmaniasis: the role of local variation in rainfall and altitude on the presence and incidence of kala-azar in eastern Sudan. *American Journal of Tropical Medicine and Hygiene* 2003; **68**: 10–17.
18. **Greenacre MJ.** *Correspondence Analysis in Practice*. London: Academic Press, 1993.
19. **Goldstein H.** *Multilevel Statistical Models*. London: Edward Arnold, 1995.
20. **Kreft I, de Leeuw J.** *Introducing Multilevel Modelling*. London: Sage, 1998.
21. **Singer JD.** Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behaviural Statistics* 1998; **23**: 323–355.
22. **Ashford DA, *et al.*** Studies on control of visceral leishmaniasis: impact of dog control on canine and human visceral leishmaniasis in Jacobina, Bahia, Brazil. *American Journal of Tropical Medicine and Hygiene* 1998; **59**: 53–57.
23. **Gavgani AS, *et al.*** Effect of insecticide-impregnated dog collars on incidence of zoonotic visceral leishmaniasis in Iranian children: a matched-cluster randomised trial. *Lancet* 2002; **360**: 374–379.
24. **Morgenstern H.** Ecologic studies in epidemiology: concepts, principles, and methods. *Annual Review of Public Health* 1995; **16**: 61–81.
25. **Goldstein H, *et al.*** *A User's Guide to MlwiN*. London: Multilevel Models Project, Institute of Education, University of London, 1998.