

**MODELAGEM ESPACIAL, TEMPORAL E LONGITUDINAL:
DIFERENTES ABORDAGENS DO ESTUDO DA LEPTOSPIROSE
URBANA**

Wagner de Souza Tassinari

Tese apresentada à Escola Nacional de Saúde
Pública, Fundação Oswaldo Cruz, para a
obtenção do título de Doutor em Saúde
Pública.

RIO DE JANEIRO
Rio de Janeiro - Brasil
Março - 2009

**MODELAGEM ESPACIAL, TEMPORAL E LONGITUDINAL:
DIFERENTES ABORDAGENS DO ESTUDO DA LEPTOSPIROSE
URBANA**

Wagner de Souza Tassinari

Orientadora: Dra. **Marilia Sá Carvalho**

Co-Orientador: Dr. **Albert Icksang Ko**

Tese apresentada à Escola Nacional de Saúde
Pública, Fundação Oswaldo Cruz, para a
obtenção do título de Doutor em Saúde
Pública.

RIO DE JANEIRO
Rio de Janeiro - Brasil
Março - 2009

“To err is human, to forgive divine, but to include errors in your design is statistical”

(Leslie Kish)

“Não use a Estatística como a arte de torturar os dados até que ele confesse, mas
como o bêbado usa o poste: mais por apoio do que para iluminação”

(Lang)

DEDICATÓRIA

Aos meus pais Venerando e Iris, a minha esposa Erika, a minha filha Eloah, e a todos os meus familiares e amigos pelas horas subtraídas de nosso convívio.

Dedico também a todos àqueles que acreditam que a ousadia e o erro são caminhos para as grandes realizações.

AGRADECIMENTOS

À Deus por ter me ajudado a concluir mais essa tarefa.

Aos meus pais, minha esposa, minha filha, minha irmã e meus parentes em geral, pela paciência nas horas mais insólitas.

A minha professora, orientadora, amiga ‘Mãerilia’, pela paciência para lidar com os meus erros, pela dedicação e transmissão de conhecimentos em todas as horas, pela confiança depositada em mim, pela sua presteza em me atender sempre quando batia em sua porta, etc, etc, etc, etc, ... , infelizmente aqui não existe espaço suficiente para escrever tudo o que gostaria.

Ao meu co-orientador Albert Ko, pela oportunidade de me deixar fazer parte de seu grupo de pesquisa no Centro de Pesquisa Gonçalo Moniz (CPqGM).

Aos doutores Oswaldo G. Cruz, Reinaldo Souza Santos, Guilherme Werneck, Antônio Miguel V. Monteiro, Virgínia Ragoni, Cláudio Bustamante Pereira de Sá (in memorium) e ao doutorando Daniel Skaba, pelos subsídios materiais e interpessoais nas diferentes etapas do desenvolvimento deste trabalho.

Aos meus companheiros do PROCC (Aline, Carlos, Ernesto, Franklin, Luciane, Marcel, Ronaldo, e todos os outros), pelos ótimos momentos de trocas de experiências e descontração.

Aos amigos ‘Alexandres’, Hugo e Tiago pelo compartilhamento de experiências, tanto profissionais quanto de vida.

Aos membros da banca examinadora, pela predisposição em analisar este trabalho e pelas sugestões recebidas.

E finalmente a todos aqueles que contribuíram de maneira direta ou indireta para a realização desse trabalho.

SUMÁRIO

	Página
LISTA DE FIGURAS	ix
LISTA DE TABELAS	xi
RESUMO	xii
SUMMARY	xv
1 INTRODUÇÃO	2
2 EPIDEMIOLOGIA DA LEPTOSPIROSE URBANA	5
3 TÉCNICAS ESTATÍSTICAS	8
3.1 Identificação de aglomerados espaciais e espaço-temporais	8
3.2 Modelos de Regressão	10
3.2.1 Regressão Linear	10
3.2.2 Modelo Linear Generalizado	11
3.2.3 Extensões do Modelo Linear Generalizado	12
4 OBJETIVOS	15
5 MATERIAL E MÉTODOS	16
5.1 Artigos	17
5.1.1 Artigo 1 - Detection and Modeling of Case Clusters for Urban Lep- tospiriosis	17

5.1.2	Artigos 2 e 3 - Spatial Modeling of Leptospirosis in a Urban Slum Area and Spatial-Longitudinal Models Applied to Leptospiral Seroconversion Incidence	18
5.2	Ferramentas Computacionais Utilizadas	20
6	ARTIGO 1 - DETECTION AND MODELING OF CASE CLUS- TERS FOR URBAN LEPTOSPIROSIS	22
6.1	INTRODUCTION	24
6.2	METHODS	25
6.3	RESULTS	29
6.4	DISCUSSION	32
7	ARTIGO 2 - SPATIAL MODELING OF LEPTOSPIROSIS IN A URBAN SLUM AREA	40
7.1	INTRODUCTION	42
7.2	METHODS	43
7.3	RESULTS	46
7.4	DISCUSSION	48
8	ARTIGO 3 - SPATIAL-LONGITUDINAL MODELS APPLIED TO LEPTOSPIRAL SEROCONVERSION INCIDENCE	54
8.1	INTRODUCTION	56
8.2	METHODS	57
8.3	RESULTS	60
8.4	DISCUSSION	62
9	COMENTÁRIOS FINAIS	69

LISTA DE FIGURAS

	Página
1 Fluxograma do censo até os estudos de soroprevalência e a coorte de soroconversão em Pau da Lima - Salvador/BA	19
2 Distribution of areas with altitude greater than 100 m, slums areas and regions of flood risk in Rio de Janeiro, Brazil.	26
3 Distribution of leptospirosis cases and Voronoi polygons associated with each of the 32 meteorological stations in Rio de Janeiro, Brazil.	28
4 Distribution of six leptospirosis case clusters in Rio de Janeiro from 1997 to 2002, which were identified in spatial scan statistics. The spatial distribution of cluster events is shown according to the census tract in which cluster cases resided. Cluster events in 1999, 2000 and 2001 involved few census tracts while cluster events in 1997, 1998 and 2002 involved more widespread areas of the city. All cluster events occurred in census tracts that were situated in the city's periphery.	38
5 Generalized additive models (GAM) of the association between the risk of acquiring <i>Leptospira</i> antibodies and continuous variables of (A) Individual age (years), (B) Distance in metres to the nearest open sewer, and (C) Distance in metres to the trash collection. The adjusted odds ratio, in the GAM model is a measure for the risk of acquiring <i>Leptospira</i> antibodies. Solid lines represent the point estimate; dotted lines represent upper and lower 95% confidence band.	51

6	Risk maps for the adjusted odds ratio of logistic spatial regressions for the of prevalence of Leptospirosis in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004. Black and white lines represent upper and lower 95% confidence bands, respectively. And a common legend for odds ratio surface adjusted.	53
7	GAMM of the association between the risk of leptospiral seroconversion and continuous variables of (A) Individual age (years), (B) Individual income (R\$), (C) Distance in meters to the nearest open sewer, (D) Distance in meters to the trash collection, and (E) Altitude in meters of the domicile sea level. The adjusted odds ratio, in the GAMM model is a measure for the risk of acquiring <i>Leptospira</i> antibodies. Solid lines represent the point estimate; dotted lines represent upper and lower 95% confidence band.	66
8	Risk maps for the adjusted Odds Ratio of logistic spatial regressions (GAMs) for the of incidence of leptospiral seroconversion in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004. Black and white lines represent upper and lower 95% confidence bands, respectively.	68

LISTA DE TABELAS

	Página
1	Leptospirosis Cases and Rainfall in Rio de Janeiro, Brazil from 1997 to 2002. 30
2	Characteristics of Leptospirosis Case Clusters Identified between 1997 and 2002. 31
3	Generalized Linear Mixed Model Estimates of Risk Factors for Leptospirosis Custers Cases. 39
4	Descriptive analysis of categorical variables available for the univariate and bivariate analysis for the prevalence of Leptospirosis study in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004. 47
5	Adjusted Odds Ratio and others fit measurements of de logistic regression for the of prevalence of Leptospirosis in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004. 52
6	Table of the frequency and the infection, re-infection and new infections incidence of leptospiral seroconversion cohort study in Pau da Lima, Salvador, Bahia, Brazil, 2003-2007. 64
7	Descriptive analysis of categorical variables available for the univariate analysis for the incidence and bivariate GEE modeling analysis for the odds ratio estimation in leptospiral seroconversion cohort study in Pau da Lima, Salvador, Bahia, Brazil, 2003-2007. 65
8	Odds Ratio and others fit measurements of de spatial logistic regression (GAM) for the of Incidence of Leptospirosis in Pau da Lima, Salvador, Bahia, Brazil, 2003-2007. 67

**MODELAGEM ESPACIAL, TEMPORAL E LONGITUDINAL:
DIFERENTES ABORDAGENS DO ESTUDO DA LEPTOSPIROSE
URBANA**

Autor: WAGNER DE SOUZA TASSINARI
Orientadora: Dra. MARILIA SÁ CARVALHO
Co-Orientador: Dr. ALBERT ICKSANG KO

RESUMO

A leptospirose, enfermidade causada por uma espiroqueta patogênica do gênero *Leptospira*, é uma das zoonoses mais difundidas no mundo, considerada um importante problema de saúde pública estando associada a falta de saneamento e a pobreza. Sendo endêmica no Brasil, dados provenientes da vigilância epidemiológica apontam que surtos de leptospirose ocorrem como epidemias cíclicas anuais durante intensas chuvas. O objetivo desta tese foi modelar os fatores de risco associados à ocorrência de leptospirose urbana em diferentes contextos, com especial atenção para aspectos espaciais e temporais. Foram utilizadas técnicas de modelagem tais como, modelos generalizados aditivos e mistos. Também explorou-se técnicas de detecção de aglomerados espaço-temporais. Nesta tese foi priorizado o

uso de softwares livres - R, sistema operacional linux ubuntu, LaTeX, SatScan (este sendo não livre porém gratuito). Esta tese foi elaborada sob forma de três artigos.

No primeiro artigo é apresentada uma análise espaço-temporal da ocorrência de casos de leptospirose no município do Rio de Janeiro entre 1997 e 2002. Utilizando o método de detecção de aglomerados espaço-temporais - “surto” - foram estatisticamente significativos somente nos anos de 1997 e 1998. Modelos Lineares Generalizados Mistos foram utilizados para avaliar fatores de risco associados a ocorrência dos casos que pertenceram aos surtos em relação aos casos endêmicos. Os casos pertencentes aos surtos estão associados com a ocorrência de chuvas acima de 4mm (OR, 3,71; 95 % CI, 1.83-7.51). Não foram encontradas associações significativas com as covariáveis socioeconômicas, ou seja, sendo endêmica ou epidêmica a leptospirose ocorre na mesma população.

No segundo e terceiro artigos analisou-se um inquérito de soroprevalência e uma coorte de soroconversão realizados na comunidade em Pau da Lima, Salvador, Bahia. Em ambos foram utilizados Modelos Generalizados Aditivos para modelar variáveis de exposição tanto no nível dos indivíduos quanto no nível do contexto do peridomicílio, e bem como estimar a superfície espacial de risco de contrair leptospirose. Foram significativas as variáveis: sexo, idade, presença de ratos no peridomicílio, proximidade da residência de um aglomerado de lixo ou de um esgoto aberto e altitude do domicílio em relação ao nível do mar. Os estudos mostram que as variáveis individuais e contextuais explicam grande parte da variabilidade espacial da leptospirose, porém ainda existem fatores que não foram mensurados nos estudos mas que merecem ser investigados. Os mapas de risco de soroprevalência e de soroconversão apontam distintas regiões onde o efeito espacial é significativamente diferente da média global.

Ainda falta uma integração mais sólida entre os profissionais que desenvolvem e operam os SIGs, epidemiologistas e os bioestatísticos. Essa integração representa um avanço importante viabilizando o desenvolvimento e a utilização dessas técnicas em prol da Saúde Pública. O estudo da prevalência e da incidência das

endemias urbanas, no caso a leptospirose, tem grande complexidade e muito ainda a avançar. A reunião de expertises oriundas de várias áreas do conhecimento humano (ex: clínicos, epidemiologistas, geógrafos, biólogos, estatísticos, engenheiros, etc.) é indispensável para avançar no conhecimento sobre as doenças e suas relações com a desigualdade social e ambiental assim a contribuir para na criação de medidas eficazes e efetivas no controle de endemias.

Palavras chaves: *Epidemiologia Ambiental, Análise Espacial, Métodos de Detecção de Aglomerados Espaço-Temporais, Modelos Lineares Generalizados Mistos, Modelos Aditivos Generalizados*

SPACE, TIME AND LONGITUDINAL MODELING : DIFFERENT APPROACHES FOR THE URBAN LEPTOSPIROSIS STUDY

Author: WAGNER DE SOUZA TASSINARI
Adviser: Prof. Dr. MARILIA SÁ CARVALHO

SUMMARY

Leptospirosis, a disease caused by pathogenic spirochete of the genus *Leptospira*, is one of the most widespread zoonoses in the world, considered a major public health problem associated with the lack of sanitation and poverty. It is endemic in Brazil, data from surveillance show that outbreaks of leptospirosis occur as cyclical annual epidemics during rainfalls. The aim of this thesis was modeling the risk factors associated with the occurrence of leptospirosis in different urban contexts, with particular attention to spatial and temporal aspects. We used some modeling techniques such as generalized additive and mixed models. Techniques for detection space-time clusters were also explored. This thesis has prioritized the use of free softwares - R, ubuntu linux operating system, L^AT_EX , SatScan (this is not open source but free). This thesis was prepared in the form of three articles.

In the first article is presented a spatio-temporal analysis of leptospirosis cases occurrence in Rio de Janeiro between 1997 and 2002. Using the

detection of space-time clusters - “outbreaks” method - were statistically significant only cluster occurred in 1997 and 1998. Generalized Linear Mixed Models were used to evaluate the risk factors associated with the occurrence of cases that belonged to outbreaks in endemic cases. The cases belonging to the outbreaks are associated with the occurrence of rainfall over 4 mm (OR, 3.71; 95% CI, 1.83 - 7.51). There were no significant associations with socioeconomic covariates, in other words, being endemic or epidemic leptospirosis occurs in the same population.

The second and third articles examined a seroprevalence survey and seroconversion cohort conducted in Pau da Lima community, Salvador, Bahia. In both Generalized Additive Models were used to fit the exposure variables both in individuals and peridomicile context, as well as to estimate the spatial area of leptospirosis risk. The significant variables were: gender, age, presence of rats in the peridomicile, domicile near a trash collectin or an open sewer and domicile altitude above sea level. Studies show that individual and contextual variables explain much of the spatial variability of leptospirosis, but there are still factors that were not measured in the studies but which should be investigated. The maps of risk of seroprevalence and seroconversion show distinct regions where the spatial effect is significantly different from the global average.

It is still lack for a more robust integration between the professionals who develop and operate the GIS, epidemiologists and biostatistics. This integration represents an important advance enabling the development and use of these techniques in Public Health support. The study of prevalence and incidence of endemic areas, in the leptospirosis context, it is very complex and still grow up. The reunion of professional specialists from several areas of human knowledge (eg, clinicians, epidemiologists, geographers, biologists, statisticians, engineers, etc.), it is essential to advance the knowledge about the disease and their relationship to social inequality and environmental well to contribute to the creation of efficient and effective

measures to control endemic diseases.

Keywords: *Environmental Epidemiology, Spatial Analysis, Methods for Detection of Space-Time Clusters, Generalized Linear Mixed Models, generalized additive models*

APRESENTAÇÃO

Esta tese segue o modelo proposto pela Escola Nacional de Saúde Pública incluindo três artigos, estando o primeiro artigo já publicado.

Na introdução são abordados os aspectos conceituais relacionados à transição demográfica e epidemiológica e a importância da aplicação de técnicas estatísticas de análise espacial, temporal e longitudinal no contexto epidemiológico.

No capítulo 2 apresentamos brevemente a epidemiologia da leptospirose urbana, incluindo os fatores de risco e possíveis formas de prevenção e controle, bem como uma breve descrição do panorama da leptospirose nas cidades do Rio de Janeiro e Salvador, áreas alvo de nosso estudo.

No capítulo 3 descreveremos os modelos estatísticos utilizados neste trabalho, discutindo suas aplicações.

A seguir, nos capítulos 4 e 5, encontram-se os objetivos e a metodologia desta tese. A metodologia com detalhamento específico é encontrada no corpo de cada um dos artigos.

Após a exposição dos artigos, nos comentários finais, indicamos algumas conclusões e limitações de cada estudo. Ao final, são apresentados as referências bibliográficas utilizadas em toda tese.

Em anexo encontram-se os questionários utilizados nos estudo em Pau da Lima, Salvador/BA.

1 INTRODUÇÃO

O rápido e intenso processo de urbanização do Brasil têm colocado dificuldades adicionais ao sistema de saúde do país. Se, por um lado, essa condição implica em maior acesso aos bens e serviços ligados à saúde das populações, por outro, implica no aumento de riscos físicos (ex: poluição) e sócio-ambientais (ex: hiper-adensamento populacional e violências). As cidades apresentam um perfil epidemiológico mais complexo, demandando maior gasto e maior complexidade na assistência. Como consequência deste contexto a análise da saúde neste ambiente impõe grandes desafios metodológicos.

No Brasil, o quadro epidemiológico atual caracteriza-se pela coexistência de doenças degenerativas e o retorno de antigas doenças infecciosas, como a malária, leishmaniose, dengue, leptospirose, hanseníase, tuberculose, entre outras. O final do século passado foi marcado pela re-emergência de velhas doenças e o aparecimento de novas, resultantes do crescimento desigual das cidades, movido por uma urbanização excludente e fragmentária de territórios e populações [4] que desafiam, de forma radical a saúde pública do século XXI na busca de soluções democráticas, eqüitativas e integrais.

A partir da década de 50 ocorreram rápidas mudanças no perfil demográfico em função do grande fluxo migratório rural-urbano. As populações fixaram-se em áreas periféricas aos grandes centros urbanos - industriais, dando origem ao surgimento de habitações e aglomerados subnormais, cortiços e favelas, sem a correspondente infra-estrutura necessária para garantir-lhes minimamente a salubridade da moradia e do seu entorno [85]. Algumas doenças passíveis de prevenção, como a poliomielite e o sarampo, passam por períodos de redução significa-

tiva em decorrência de coberturas vacinais, enquanto outras, como a febre amarela e a doença de Chagas, em decorrência do controle de vetores. Outras doenças não têm apresentado sinais de redução, pelo contrário, expandem-se, como é o caso da malária, hepatite B, hepatite C, tuberculose, hanseníase e leishmaniose. Associa-se a este cenário a reintrodução do cólera e da dengue [63].

Neste contexto o estudo da distribuição geográfica de endemias nos grandes centros urbanos, e da sua relação com potenciais fatores de risco, vem constituindo um terreno fértil para a aplicação e desenvolvimento de métodos e modelos estatísticos. Nos últimos anos, foram desenvolvidas técnicas cada vez mais poderosas e versáteis nessa. A popularidade desses estudos deve-se, em parte, à disponibilidade de sistemas de informações geográficas (SIGs) de baixo custo e com interfaces amigáveis. Esses permitem a visualização espacial de, por exemplo, número de casos de uma determinada doença numa região. Para tanto basta dispor de um banco de dados e de uma base geográfica (como um mapa de bairros), e o SIG é capaz de apresentar um mapa colorido permitindo não só a visualização do padrão espacial do fenômeno, mas também o padrão de potenciais fatores de exposição [18].

Os métodos de análise espacial têm sido empregados na área da saúde nas seguintes situações [19, 7]:

- *“Quando o evento em estudo é gerado por fatores ambientais de difícil detecção no nível do indivíduo;*
- *Na delimitação de áreas homogêneas segundo intervenção pretendida;*
- *Quando o evento em estudo e os fatores relacionados têm distribuição espacialmente condicionada;*
- *No estudo de trajetórias entre localidades”.*

As técnicas de análise espacial podem ser sistematizadas, a partir do objeto e do tipo de dados disponíveis, são eles:

- Processo pontual (“point pattern”) - quando a disposição espacial dos eventos é a variável aleatória a ser modelada;
- Geoestatística - quando o fenômenos a ser investigado está distribuídos continuamente no espaço a partir de pontos amostrados aleatoriamente;
- Dados de áreas - representado por quantidades aleatórias de um indicador sumário de cada área de estudo;
- Deslocamento - caracterizado por dados do formato “origem - destino”.

A relação entre um evento de interesse epidemiológico e o espaço geográfico pode se dar em diversas escalas. Por exemplo, a epidemia de cólera pode ser olhada em um quarteirão ou na sua dispersão mundial, conforme o interesse do pesquisador. Definido então o objetivo cabe a seleção da escala de representação e análise que, evidentemente, deve ser compatível com a escala de ocorrência. Um princípio básico é que não se pode inferir sobre a ocorrência de fenômenos observados em determinada escala, para níveis de maior detalhamento. Isso pode levar ao viés de agregação, também denominado de falácia ecológica, decorrente da suposição de que atributos de uma determinada área aplicam-se de forma homogênea a todos os indivíduos dessa mesma área [80, 81, 35, 84].

Na pesquisa de causas de doenças o desenho paradigmático é o estudo longitudinal, caracterizado por medidas repetidas de um mesmo indivíduo. A modelagem desses dados deve respeitar a estrutura de dependência, das observações aninhadas por indivíduo, utilizando técnicas estatísticas específicas para lidar com tal tipo de problema.

E quando o interesse é compreender o fenômeno de interesse variando no tempo e no espaço, faz-se uso de técnicas estatísticas mais sofisticadas que levem em conta as estruturas espaciais e temporais interagindo simultaneamente em um só modelo.

2 EPIDEMIOLOGIA DA LEPTOSPIROSE URBANA

A leptospirose, considerada uma das zoonoses mais difundidas no mundo, é uma doença que tem como agentes etiológicos espiroquetas do gênero *Leptospira*. A infecção por *Leptospira* produz uma diversidade de sinais e sintomas clínicos que podem variar desde uma infecção sub-clínica até a doença de Weil, que é uma forma clínica de grande gravidade [30]. A taxa de letalidade geral varia de 5 a 20%, nas formas mais graves, que evoluem com disfunção de múltiplos órgãos e sistemas, incluindo hemorragia pulmonar, a letalidade pode chegar a 50% [42].

O gênero *Leptospira* é atualmente classificado em oito espécies genômicas de espiroquetas sendo o *Leptospira interrogans* o mais difundido[36]. A *L. interrogans* é subdividida em vários sorogrupos que, por sua vez, são divididos em diversos sorotipos denominados também sorovares [11, 28]. Cada sorovar pode estar associado a uma característica grave ou doença, afinidade por um hospedeiro ou distribuição geográfica distinta.

A *Leptospira* pode ser identificada na urina, em fluidos corporais e tecidos através de observação microscópica (diagnóstico direto), podendo ser isolada em meios de cultivo e inoculada em animais de laboratório (hamster, cobaia jovem). A confirmação do diagnóstico pode ser feita através de testes laboratoriais, sendo que a microaglutinação (MAT) é considerada o método de referência ou padrão-ouro na confirmação do diagnóstico de leptospirose [53]. Outro teste utilizado é o ELISA IgM (imunoenzimático) que detecta anticorpos na primeira semana da doença, pois os anticorpos aglutinantes IgM. Este teste é menos laborioso que a microaglutinação

e mais fácil de interpretar especialmente quando não se possui amostras pareadas. Porém, é aconselhável uma posterior confirmação pela MAT pois o resultado do ELISA é considerado preliminar. Os testes ELISA IgM e antígenos recombinantes são mais sensíveis que a microaglutinação porém não o suficiente para decidir as avaliações e intervenções clínicas [53].

A leptospirose é adquirida através do contato com reservatórios animais ou ambientes contaminados pela urina destes [71]. Quase todos os mamíferos e marsupiais podem ser reservatórios para a *Leptospira*, porém os principais são roedores domésticos, das espécies *Rattus norvegicus* que é o roedor urbano de maior porte, cujo habitat é o solo das redes de esgoto e terrenos baldios onde escava túneis subterrâneos como tocas, alimentando-se de lixo [28].

Em países temperados, a leptospirose humana ocorre predominantemente de forma esporádica, principalmente em grupos ocupacionais que estão particularmente expostos, como agricultores, fazendeiros, magarefes, tratadores de animais, veterinários, militares e outras profissões que tenham contato com animais ou água contaminada. Entre os mamíferos envolvidos nas exposições ocupacionais estão bovinos, suínos, roedores, cães e mamíferos silvestres. Atividades recreativas que se encontram associadas à transmissão de leptospirose incluem natação, caça, canoagem, passeios por trilhas em locais com acúmulo de água [30, 48, 3].

Observou-se aumento da notificação nos últimos anos em vários países: Nicarágua, Índia, Sudeste da Ásia, Estados Unidos, Malásia, e Brasil, onde cerca de 10 mil casos são notificados por ano em todas as grandes metrópoles. Detecta-se alterações no padrão epidemiológico desta doença, com a ocorrência de grandes epidemias urbanas na América Latina. Entretanto, a ocorrência de surtos de leptospirose após enchentes não é um fenômeno novo e nem restrito a regiões tropicais [46].

O intenso e desordenado processo de urbanização criou ambientes físicos e sociais extremamente insalubres. A falta de saneamento básico nos grandes centros urbanos, principalmente nas favelas, somada à freqüente exposição à contam-

inação ambiental durante as fortes chuvas e enchentes, são considerados os fatores que contribuem com maior magnitude na ocorrência das epidemias de leptospirose. Além disso, a alta densidade demográfica contribui para o aspecto explosivo das epidemias gerado em grandes contingente submetidos simultaneamente a condições ambientais propícias [42, 59].

No Brasil, durante o período de 1985 a 1997, foram notificados 35.403 casos da doença, variando desde 1.594 em 1987 a 5.576 em 1997, com 3.821 óbitos registrados. A taxa de letalidade média foi de 12,5% [15]. Além disso, nas apresentações clínicas graves como a Síndrome de Weil e Síndrome Hemorrágica Pulmonar, a letalidade pode exceder a 50% [71]. Ainda assim a leptospirose não vem recebendo a devida prioridade. Poucos trabalhos apresentam propostas preventivas capazes de diminuir o impacto da doença, seja em situação endêmica ou em enchentes. São ainda mais raros os estudos que contribuem para compreender o perfil da população atingida em cada uma das situações epidemiológicas que fogem ao tradicional grupo ocupacional de risco.

Dos poucos estudos existentes, Barcellos e Sabroza (2001) [8] analisando o contexto ambiental de um surto de leptospirose em 1996 na Zona Oeste do Rio de Janeiro, verificaram que as maiores taxas de incidência ocorreram nas regiões sujeitas à inundação e ao redor das zonas de acumulação de lixo, apontando para a combinação de fatores sociais e ambientais, ainda que a relação não seja direta.

Em Salvador/BA, alguns estudos [17, 20] mostram que a leptospirose ocorre em epidemias cíclicas anuais, relacionadas à pobreza, durante o período de chuvas acompanhadas por alagamento de zonas urbanas de baixas condições sanitárias. Cerca de 15% das infecções resultam em complicações graves, como icterícia, insuficiência renal aguda e hemorragia pulmonar e os índices de letalidade variam de 10-15% entre os casos graves.

3 TÉCNICAS ESTATÍSTICAS

3.1 Identificação de aglomerados espaciais e espaço-temporais

Um aglomerado consiste em um grupo de ocorrências em um espaço geograficamente limitado em tamanho e concentração tais que seja improvável de ocorrer por mero acaso. Identificar a ocorrência de aglomerados não casuais ainda é um problema metodológico em discussão. Os testes estatísticos para detecção de aglomerados dividem-se em duas categorias: focados e genéricos [22]. Os testes focados se caracterizam por verificarem a existência de aglomerados em uma ou algumas poucas regiões definidas e delimitadas antes da observação dos eventos. Isto é, os testes focados para hipóteses que determinam *a priori* onde os aglomerados poderiam estar, sendo úteis quando há uma forte suspeita de um foco, por exemplo, uma fábrica poluidora, gerando um possível aglomerado de casos de câncer de pulmão ao seu redor. Os testes genéricos ou globais distinguem-se por não suporem de antemão um local específico como possível aglomerado de risco mais elevado. Isto é, eles testam a hipótese de que não existe aglomerado na região de estudo contra a hipótese alternativa de que há algum aglomerado na região sem especificar onde esse possível aglomerado estaria [6].

Já os testes genéricos objetivam identificar um padrão geral de distribuição da doença em uma grande região subdivididas em várias áreas, em cada área do estudo apresenta um número de casos ou eventos que seguem uma distribuição de Poisson. O número esperado de eventos em dada área é igual a θ vezes a população sob risco. Esta constante θ refere-se à taxa per capita de ocorrência de

eventos na área que, sob a hipótese nula, possui o mesmo valor em toda e qualquer área ou sub-região do mapa. Já na hipótese alternativa, admite-se que alguma sub-região do mapa tenha um valor de θ maior dentro do que fora dela. O teste genérico mais utilizado no âmbito da epidemiologia atualmente foi proposto por Kulldorff & Nagarwalla [43], e consiste basicamente em uma estatística de varredura. Este método prioriza uma região formada pelas áreas cujos centróides caem dentro de um círculo. Variando o raio e o centro do círculo, os possíveis aglomerados são formados. E em cada círculo, é calculado a razão entre o máximo da verossimilhança sob a hipótese alternativa de que θ é maior dentro do que fora do círculo e o máximo da verossimilhança sob a hipótese nula de que o valor encontrado é casual. Em seguida, é calculado o máximo dessas razões para todos os possíveis círculos. Esse valor máximo da estatística do teste da razão da máxima verossimilhança é denominado T . A região associada com esse máximo é denominada aglomerado mais verossímil ou primário. A distribuição deste e o p -valor associado são obtidos através da simulação de conjuntos de dados gerados sob a hipótese nula. A hipótese nula é rejeitada ($\alpha=0,05$) quando menos do que 5% dos valores simulados de T são maiores do que o valor realmente observado de T obtido dos dados não simulados. Este método não fornece apenas o aglomerado mais verossímil, como também os aglomerados secundários, compostos para todas as áreas onde rejeita-se a hipótese nula.

A estatística espaço-tempo na qual o programa SaTScan se baseia é definida por uma janela cilíndrica com uma base geográfica circular e peso correspondente ao tempo. A base é centrada nos vários centróides da região de estudo, com os raios variando constantemente em tamanho. O peso é determinado por um intervalo de tempo menor ou igual a metade do período total de estudo; entretanto, o período de estudo como um todo também pode ser utilizado. A janela é então movida no espaço e no tempo para cada localização geográfica e intervalo de tempo possíveis. Como resultado, obtêm-se um número infinito de cilindros sobrepostos de diferentes tamanhos e formas, cobrindo conjuntamente a região do estudo como um todo. Cada cilindro reflete um possível aglomerado. A estatística supõe que os

casos possuem a distribuição de Poisson com risco constante no espaço e no tempo sob a hipótese nula, e com risco diferente dentro de pelo menos um dos cilindros sob a hipótese alternativa. Para cada cilindro o número de casos da doença dentro e fora do cilindro são verificados juntamente com o número de casos esperado, o que reflete a população à risco e as covariáveis mais relevantes. Baseado nestes números, a verossimilhança é calculada para cada cilindro. Da mesma forma que na estatística T , o cilindro com a máxima verossimilhança e com mais que o número de casos esperados é denominado aglomerado mais verossímil [43].

3.2 Modelos de Regressão

3.2.1 Regressão Linear

Para estudar a relação entre um desfecho (variável dependente ou resposta) e um conjunto de potenciais fatores de risco (variáveis independentes ou explicativas), utiliza-se modelos estatísticos de regressão, com o objetivo de determinar um modelo matemático que descreve esta relação.

Na maior parte das situações pode-se pensar na variável de desfecho consistindo de duas partes distintas: um componente sistemático (μ) e um componente aleatório (ϵ). Tem-se então em um modelo linear clássico de regressão: $Y = \mu + \epsilon$, onde, Y é o vetor, de dimensões $n \times 1$, da variável de desfecho, $\mu = E(Y) = X\beta$, o componente sistemático, X a matriz, de dimensões $n \times p$, do modelo, $\beta = (\beta_1, \dots, \beta_p)^T$ o vetor dos parâmetros, $\epsilon = (\epsilon_1, \dots, \epsilon_n)^T$, o componente aleatório com $\epsilon_i \rightarrow N(0, \sigma^2)$, $i = 1, \dots, n$. O método de estimação mais comumente usado neste caso é baseado na minimização dos quadrados do componente aleatório ϵ^2 , e por isso chamado de mínimos quadrados (MMQ). É importante ressaltar alguns de seus pressupostos básicos para o ajuste de modelos de regressão linear:

1. A ausência de autocorrelação entre os erros (componentes aleatórios), $cor(\epsilon_i, \epsilon_j) \rightarrow 0$;
2. Variáveis independentes não correlacionadas (colinearidade), $cor(x_1, \dots, x_p) \rightarrow$

0 ;

3. A existência de homocedasticidade, ou seja, variância constante dos resíduos, $var(\epsilon_i) = \sigma^2$.

3.2.2 Modelo Linear Generalizado

É possível utilizar métodos análogos àqueles desenvolvidos para o modelo de regressão linear, em situações em que a variável resposta obedece a outras distribuições que não a Normal, ou em que a relação entre a variável resposta e as variáveis explicativas não é linear. Isto se deve, em parte, ao conhecimento de que muitas das boas propriedades da distribuição Normal são partilhadas por uma larga classe de distribuições denominado de família exponencial [24].

Nelder e Wedderburn [60] propuseram uma extensão dos modelos lineares clássicos, denominado Modelos Lineares Generalizados (GLM). As principais características desses modelos são:

- A variável resposta, componente aleatório do modelo, tem uma distribuição pertencente à família exponencial na forma canônica: distribuições normal, gama e normal inversa para dados contínuos; binomial para proporções; Poisson e binomial negativa para contagens;
- As variáveis explicativas, entram na forma de um modelo linear (componente sistemático);
- A ligação entre os componentes aleatório e sistemático é feita através de uma função de ligação (por exemplo, logarítmica para os modelos log-lineares), conforme a fórmula abaixo.

$$f(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right), \quad (1)$$

sendo θ é o parâmetro natural e $a(\phi)$ é o fator de dispersão. Tendo como componentes básicas:

- A variável desfecho y , cuja distribuição de probabilidade pertencente à família exponencial, com valores esperados $E(y_i) = \mu_i$;
- Um preditor linear baseado nas variáveis explicativas $x_{i1}, \dots, x_{i(p-1)}$ denotado por $x_i\beta = \eta_i$;
- A função de ligação g relacionada ao preditor linear do valor esperado do desfecho: $\eta = g(\mu_i)$.

Entre estes modelos, os mais usados na área de epidemiologia são: a regressão logística, tendo uma variável binária como desfecho, e a de regressão de poisson, tendo como variável desfecho contagens de casos ou óbitos de uma determinada patologia.

Tradicionalmente o ajuste destes modelos é baseado no método de estimação da máxima verossimilhança, pelo qual os estimadores são obtidos a partir da maximização da função de verossimilhança, e os cálculos envolvem um procedimento iterativo [13].

3.2.3 Extensões do Modelo Linear Generalizado

Uma extensão dos modelos lineares generalizados são os Modelos Aditivos Generalizados(GAM). Neste, Hastie & Tibshirani [37] propuseram a utilização de funções, usualmente não-paramétricas, sobre as variáveis independentes de forma a linearizar a relação com a variável resposta. O parâmetro estimado, neste caso, não relaciona diretamente a quantidade x à quantidade y , mas uma função de x a y . Na verdade, esta idéia é uma extensão da transformação de variáveis já muito utilizada, que tem sua maior aplicação quando o tipo de relação entre as variáveis é de forma complexa. Uma particularidade das funções não paramétricas é a capacidade de ajustar mesmo nos extremos. Temos então:

$$\eta = f_1(x_1) + \dots + f_k(x_k) + \epsilon \quad (2)$$

onde $k = 1, \dots, p$ e f_k são as funções de alisamento (suavização) das covariáveis x_k .

Essa abordagem além de possuir a vantagem de permitir a estimação do risco espacial controlado por fatores individuais e contextuais de forma simples e de fácil interpretação, através da inclusão das coordenadas geográficas via funções de suavização bivariadas (ex: “thin plate splines” e “tensor product”) [93], também permite a construção de contornos de tolerância que auxiliam na identificação de áreas de alto e baixo risco.

Porém quando os dados apresentam uma estrutura de dependência, seja ela espacial e/ou temporal, é necessário considerar que esta dependência faz com que o erro associado às estimativas dos parâmetros sejam subestimados, pois a informação de cada observação não é independente das demais. Assim, a primeira questão a ser incorporada na análise é a correção dessas estimativas. Entre as diversas possibilidades de corrigir as estimativas estão os Modelos Lineares Generalizados Mistos (GLMM), também chamados de modelos hierárquicos, multinível ou de efeitos aleatórios, são considerados uma ferramenta poderosa e flexível para a análise de dados com qualquer tipo de dependência, pois permite estimar os efeitos da própria estrutura de dependência, espacial e/ou temporal, além de corrigir as estimativas dos efeitos das covariáveis conhecidas, que passam a ser denominados de efeitos fixos [62, 52].

Já os Modelos Aditivos Generalizados Mistos (GAMM) além de incorporarem efeitos fixos e aleatórios nas covariáveis, pode incorporar em sua estrutura preditores semi-paramétricos aditivos, ou seja, funções não-paramétricas (funções de alisamento ou suavização) muito utilizadas para estimar efeitos temporais (tendência e sazonalidade) e espaciais.

Dados provenientes de estudos longitudinais se caracterizam pela seqüência temporal de duas ou mais observações em cada indivíduo, observadas ao longo do tempo, podendo haver uma correlação entre elas. Ao ignorar a correlação existente entre as observações repetidas no mesmo indivíduo, não se obtém inferências confiáveis. Em particular, as estimativas dos erros-padrão dos coeficientes do modelo são subestimados [23]. Além dos GLMM, os modelos baseados em Equações

de Estimação Generalizada (GEE) são utilizados para análise de dados longitudinais.

Os modelos GLMM permitem que os coeficientes da regressão variem entre os indivíduos. Esses modelos têm dois componentes: um intra-individual (uma mudança longitudinal intraindividual é descrita pelo modelo de regressão com um intercepto e inclinação populacional) e outro entre-indivíduos (variação no intercepto e inclinação individual). Tais modelos permitem não somente descrever a tendência temporal levando em conta a correlação que existe entre medidas sucessivas como também estimar a variação na medida basal e a taxa de mudança ao longo do tempo [23]. Já os modelos GEE são utilizados quando a inferência sobre a média populacional é o principal objetivo de estudo. Eles tem por finalidade modelar os efeitos das covariáveis na esperança marginal, ou seja, a resposta média das observações partilhando das mesmas covariáveis [47].

4 OBJETIVOS

A motivação pessoal dessa tese é a exploração de várias técnicas de análise espacial, temporal e longitudinal aplicadas no contexto da leptospirose urbana nas cidades do Rio de Janeiro e Salvador, modelando seus principais fatores de risco de natureza ambiental, individual e socioeconômica. Neste sentido foram traçados os objetivos específicos que serão apresentados nos artigos compreendidos nesta tese.

- Identificar a presença de aglomerados (*clusters*) espaço-temporais de indivíduos acometidos com leptospirose, e estudar fatores ambientais e sócioeconômicos associados à ocorrência desses agregados de casos;
- Investigar o padrão espacial da infecção por leptospirose em um grande estudo de soroprevalência realizado na comunidade de Pau da Lima situada em Salvador/BA, modelando simultaneamente fatores individuais e ambientais associados à soropositividade;
- Investigar o padrão espacial da soroconversão para leptospirose em uma coorte populacional na comunidade de Pau da Lima em Salvador/BA, modelando simultaneamente fatores individuais e sócio-ambientais associados.

5 MATERIAL E MÉTODOS

Esta tese esta organizada em três artigos. A metodologia completa de cada trabalho poderá ser apreciada no corpo dos artigos. Neste capítulo apresentamos algumas informações não contempladas nos artigos.

O primeiro artigo é produto do Projeto SAUDAVEL (<http://saudavel.dpi.inpe.br/>), cuja finalidade foi o desenvolvimento de instrumentos de Tecnologia da Informação Espacial – métodos, algoritmos e produtos de software – para dotar os sistemas de vigilância epidemiológica e de controle de endemias, de capacidade de análise e previsão. A rede SAUDAVEL inclui departamentos das Universidades Federais de Minas Gerais e do Paraná e diversas unidades e centros regionais da Fundação Oswaldo Cruz. Sendo este projeto financiado pelas instituições FINEP e CNPq, a partir do qual diversos sub-projetos e respectivos financiamentos foram estruturados. Este artigo foi publicado no ano de 2008 na revista “Tropical Medicine and International Health” [84].

O segundo e o terceiro artigos são frutos do trabalho em conjunto com o Projeto “Emerging Infectious Diseases and Urbanization”, coordenado pelo Dr. Albert Icksang Ko, da Universidade de Cornell e executado no Centro de Pesquisa Gonçalo Moniz/FIOCRUZ, com financiamento do programa “Global Infectious Disease Research Training Program (GID)” promovido pelo NIH e CDC.

5.1 Artigos

5.1.1 Artigo 1 - Detection and Modeling of Case Clusters for Urban Leptospirosis

O universo do estudo foi formado pelos 488 casos notificados de leptospirose, disponibilizados pelo Sistema de Nacional de Agravos de Notificação (Sinan), ocorridos no Município do Rio de Janeiro por semana epidemiológica, entre os anos de 1997 e 2002.

Os dados sócioeconômicos foram obtidos através do Censo Demográfico de 2000 por setores censitários foram adquiridos da Fundação Instituto Brasileiro de Geografia [31] em formato digital e as áreas sujeitas a inundações definidas pela Empresa Municipal de Informática e Planejamento (IplanRio) [39].

O georreferenciamento por setor censitário dos casos de leptospirose se deu em duas fases. A primeira fase foi feita no laboratório de Geoprocessamento (LABGEO/CICT/FIOCRUZ), em duas etapas: automática e manual. A etapa automática se deu utilizando um programa baseado em um sistema de localização através de cadastros e mapas existentes em formato digital [45]. A manual foi a geocodificação do restante dos casos com o auxílio de cartas topográficas e do Guia Quatro Rodas [2]. A segunda fase do georreferenciamento foi feita visando testar o algoritmo para geocodificação que está sendo desenvolvido por Skaba [76].

Para visualização e análise espacial foi utilizado o Sistema de Informação Geográfica TerraView versão 3.0.3 [21]. E para as análises estatísticas foi utilizado o pacote estatístico R, versão 1.7.1 [86], em conjunto com a biblioteca Splanacs [69].

5.1.2 Artigos 2 e 3 - Spatial Modeling of Leptospirosis in a Urban Slum Area and Spatial-Longitudinal Models Applied to Leptospi- ral Seroconversion Incidence

Os estudos de soroprevalência e de soroconversão foram conduzidos em Pau da Lima, uma comunidade pobre de Salvador, uma cidade com 2.443.107 habitantes no Nordeste do Brasil. Pau da Lima é uma região de colinas e vales, contendo uma área de aproximadamente 0,46 km², que era uma zona escassamente povoada de Mata Atlântica na década de 1970 e posteriormente transformada em uma favela densamente povoada.

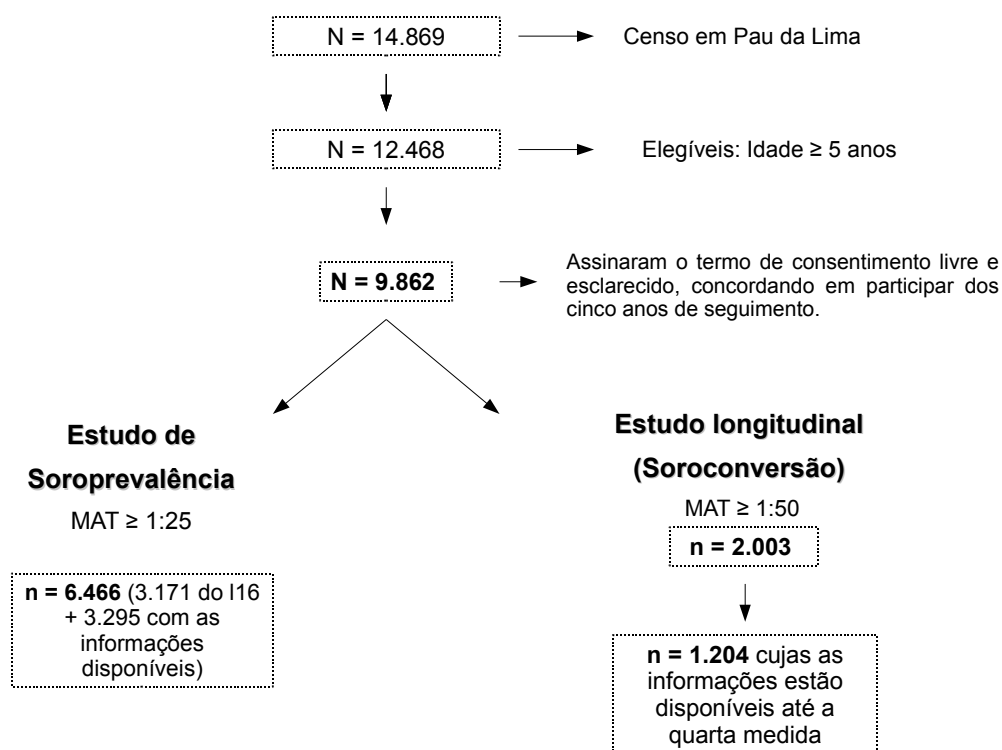
No censo realizado em 2003 na área do estudo em Pau da Lima, foram identificados 14.869 indivíduos residentes em 5.110 domicílios, destes, 12.468 (85%) tinham idade igual ou superior a cinco anos e eram, portanto, elegíveis. Dos indivíduos elegíveis, 9.862 (78%) pessoas assinaram um termo de consentimento livre e esclarecido, concordando em participar dos cinco anos de seguimento. Uma amostra de 3.171 (32%) indivíduos foi selecionada randomicamente. Todos os indivíduos que dormiam no domicílio selecionado três noites ou mais por semana, possuíam cinco anos de idade ou mais e que forneceram consentimento formal foram incluídos no estudo.

No estudo sorológico além dos 3.171 indivíduos alocados aleatoriamente para o inquérito de soroprevalência (L16), elaborado por Reis [64], foram analisados mais 3.295 indivíduos cujos exames sorológicos já estavam prontos, totalizando 6.466 indivíduos que foram investigados neste estudo (figura 1). Isso equivale a aproximadamente 67% de toda a população elegível. Os indivíduos foram recrutados para o estudo entre abril de 2003 e Maio de 2004.

Já no estudo da coorte de soroconversão, tendo como linha de base o inquérito de soroprevalência (L16) descrito acima [64], foi conduzido entre os anos de 2003 e 2007. Uma amostra de 684 (18%) domicílios foi selecionada randomicamente a partir dos indivíduos que assinaram o termo de consentimento. Deste, 2.003 (83%) consentiram em participar durante o recrutamento da coorte. A maior parte de per-

das (288 indivíduos) foram por mudança de domicílio (figura 1). Estes indivíduos foram seguidos por quatro anos após o inquérito de recrutamento (2003 até 2007). Apenas os indivíduos que tinham as informações completas em todos seguimentos estudados foram incluídos no estudo. Portanto, neste artigo foram analisados 1.204 indivíduos presentes nos quatro seguimentos. Verificou-se um total de aproximadamente 40% (801 indivíduos) de perdas durante todo o período de estudo.

Figura 1 - Fluxograma do censo até os estudos de soroprevalência e a coorte de soroconversão em Pau da Lima - Salvador/BA



Durante o recrutamento dos indivíduos, um questionário foi administrado para obter informações demográficas, socioeconômicas, ocupacionais, exposições a fontes ambientais de contaminação e presença de reservatórios no domicílio e no trabalho (Anexo). Inspeções foram realizadas para identificar localização de esgotos abertos, drenagem de água pluvial e depósito de lixo. Sistema de Informação Geográfica (SIG) foi utilizado para obter a distância tridimensional do domicílio ao

local de drenagem e depósito de lixo e em uma dimensão para cota mais baixa do vale.

Amostras sorológicas também foram obtidas anualmente durante o período do estudo. O teste de microaglutinação (MAT) foi realizado para cada amostra coletada. O desfecho de interesse no estudo de soroprevalência foi a soropositividade do indivíduo ($\text{MAT} \geq 1:25$). Já no estudo longitudinal sobre a soroconversão o desfecho foi a infecção por *Leptospira* definida como soroconversão ($\text{MAT} \geq 1:50$).

5.2 Ferramentas Computacionais Utilizadas

O desenvolvimento desse trabalho tem por princípio a utilização de ferramentas computacionais livres, definida como aquela na qual *“os usuários têm total liberdade de executar, copiar, distribuir estudar, modificar e aperfeiçoar o software”* (<http://www.gnu.org/philosophy/free-sw.pt.html>), como são o R [86] e o TerraView [21]. Foi utilizado também o software SatScan [43], que apesar de não ser livre, é de domínio público.

A linguagem R, foi utilizada para fazer toda a análise estatística, ajuste dos modelos lineares generalizados e suas extensões, e a elaboração dos gráficos e mapas dos artigos.

O Terraview é um aplicativo de visualização de dados geográficos utilizando uma biblioteca pública de geoprocessamento. Essa biblioteca tem como objetivo dar suporte ao desenvolvimento de aplicativos GIS baseado na tecnologia de banco de dados espaciais e manipulação de dados. Ao TerraLib estão sendo incorporados métodos para análise espacial, temporal e espaço-temporal para eventos de saúde. Todo o banco de dados desse projeto ficará disponível nesse aplicativo onde os resultados serão visualizados.

O software SatScan, foi utilizado para obtenção de aglomerados no espaço-temporais. A versão atual 7.0.3 se encontra gratuitamente disponível na internet (www.satscan.org).

Para edição de toda a tese foi utilizado o editor de textos \LaTeX ver-

são $\text{\LaTeX} 2_{\epsilon}$ (www.latex-project.org/). Este editor é definido como um conjunto de macros para o processador de textos, é utilizado amplamente para a produção de textos matemáticos e científicos por causa de sua alta qualidade tipográfica. Entretanto, também é utilizado para produção de cartas pessoais, artigos e livros sobre assuntos muito diversos [55]. Além disso o \LaTeX fornece ao usuário um conjunto de comandos de alto nível, sendo, dessa forma, mais fácil a sua utilização por pessoas nos primeiros estágios de utilização desse sistema. Possui abstrações para lidar com bibliografias, citações, formatos de páginas, referência cruzada e tudo mais que não seja relacionado ao conteúdo do documento em si. O modelo em \LaTeX adotado nesta tese foi o da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ) (www.esalq.usp.br).

E como sistema operacional para a elaboração da tese, foi utilizado o Ubuntu versão 8.10 (www.ubuntu.com). Ubuntu é um sistema operacional baseado em Linux desenvolvido por pessoas voluntárias que visam contribuir com o sistema e com seus usuários, buscando interagir umas com as outras prestando suporte, divulgando, participando de eventos. Esse sistema é eficiente para notebooks, desktops e servidores. Ele contém todos os aplicativos que precisamos - um navegador web, programas de apresentação, edição de texto, planilha eletrônica, comunicador instantâneo e muitos outros.

6 ARTIGO 1 - DETECTION AND MODELING OF CASE CLUSTERS FOR URBAN LEP- TOSPIROSIS

Wagner de Souza Tassinari^{1,2}

Debora C. P. Pellegrini¹

Renato Barbosa Reis³

Albert Ko^{3,4}

Marilia Sá Carvalho¹

*National School of Public Health, Oswaldo Cruz Foundation, Brazilian Ministry of Health, Rio
de Janeiro, Brazil¹*

*Department of Mathematics, Federal University Rural of Rio de Janeiro, Rio de Janeiro,
Brazil²*

*Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Brazilian Ministry of Health,
Salvador, Brazil³*

*Division of International Medicine and Infectious Disease, Weill Medical College of Cornell
University, New York, USA⁴*

Resumo

Leptospirosis is a potentially fatal zoonotic disease which has emerged to become an urban health problem in developing countries due to spatially disorganized process of urbanization and consequent unhealthy urban environment. The aim of this work is to analyse the epidemiological profile of 488 cases of leptospirosis in Rio de Janeiro, Brazil between 1997 and 2002, using a variety of methods of spatial epidemiology, to establish alert guidelines in general hospitals, which might be a tool to improve diagnosis and treatment of leptospirosis to reduce lethality rates. Scan statistics identified six space-time clusters, which comprised a range of 2 to 28 cases per cluster. Generalized linear mixed models were used to evaluate risk factors for a cluster case which incorporated individual characteristics and spatial information on environmental and climatic factors in a single model frame. Cluster case events were associated with heavy rainfall (OR 3.71; 95% CI 1.83–7.51). The model did not identify socioeconomic or environmental covariates that significantly influence the risk of developing a cluster rather than non-cluster case. Clustering of leptospirosis in this urban setting appears to be due to transmission during heavy rainfall.

Keywords: leptospirosis, geographic information systems, spatial epidemiology, generalized linear mixed model

6.1 INTRODUCTION

Leptospirosis is a globally distributed, life-threatening zoonosis [25, 28, 46, 11, 51]. Infection occurs during direct contact with animal reservoirs or indirectly during contact with water and/or soil contaminated with the urine of reservoirs [30, 46]. Severe disease develops in 5–10% of symptomatic infections and causes multisystem complications such as acute renal failure and pulmonary haemorrhage. Overall case fatality is high and varies from 5% to 15% [25, 30, 28], depending on the geographic region.

Leptospirosis is now recognized as an emerging infectious disease due to changes in its epidemiology. In developed countries, leptospirosis was traditionally a sporadic disease associated with risk occupations such as farming and animal husbandry, abattoir work and veterinarians [30, 28, 46]. More recently it has been increasingly associated with recreation and water sports [40] and travel [12] and has become the cause of outbreaks during athletic events, in disaster situations and in adventure tourism [56, 91]. However, the major burden of leptospirosis is borne by developing countries [28, 41, 46, 11], where disease incidence ranges between 10 and 100 per 100 000 inhabitants [27, 58, 96, 77, 82, 94]. Leptospirosis is a major public health problem in rural communities in developing countries, where it affects poor subsistence farmers and herders. In addition to endemic transmission of leptospirosis, large outbreaks occur in these settings [82, 94], as has been reported during post-monsoon seasons.

Moreover, leptospirosis has emerged to become a health threat in urban centres [42, 51]. Rapid and spatially disorganized process of urbanization throughout the developing world has created unhealthy physical and social urban environments [73]. At present more than 1 billion of the world's population resides in slum settlements [88]. The lack of adequate sewage systems, trash deposits and poor housing favour high rodent densities which in turn lead to environmental contamination with pathogenic *Leptospira* and high level transmission of leptospirosis in these communities [95, 42, 8, 44, 71, 67, 83].

Leptospirosis is a major public health problem in Brazil. More than 35 000 confirmed cases were identified between 1985 and 1997, among which case fatality was 12.5% [32]. The majority of these cases were reported from large urban centres [42]. Leptospirosis cases occur throughout the year in this setting [71], indicating that there is endemic transmission. However, large outbreaks have been reported during seasonal periods of heavy rainfall and flooding [42, 8, 59, 71, 83]. Leptospirosis is well-known to occur in disaster situations such as hurricanes and monsoons [91] and is increasingly recognized as an emerging infectious disease with cyclic climatic events [38].

We analysed cases identified during surveillance for leptospirosis in the city of Rio de Janeiro between 1997 and 2002 to detect space-time clusters and identify factors that influence endemic and epidemic transmission in this urban setting. Refined identification of case clusters in urban leptospirosis and elucidation of the environmental, climactic and social factors which influence these cluster events are required to understand the behaviour of the disease. Furthermore, timely detection of outbreaks and identification of their determinants may help in establishing alert guidelines for surveillance and health care professionals and in turn, may improve diagnosis and treatment of leptospirosis which is necessary to reduce the high fatality rates associated with urban epidemics.

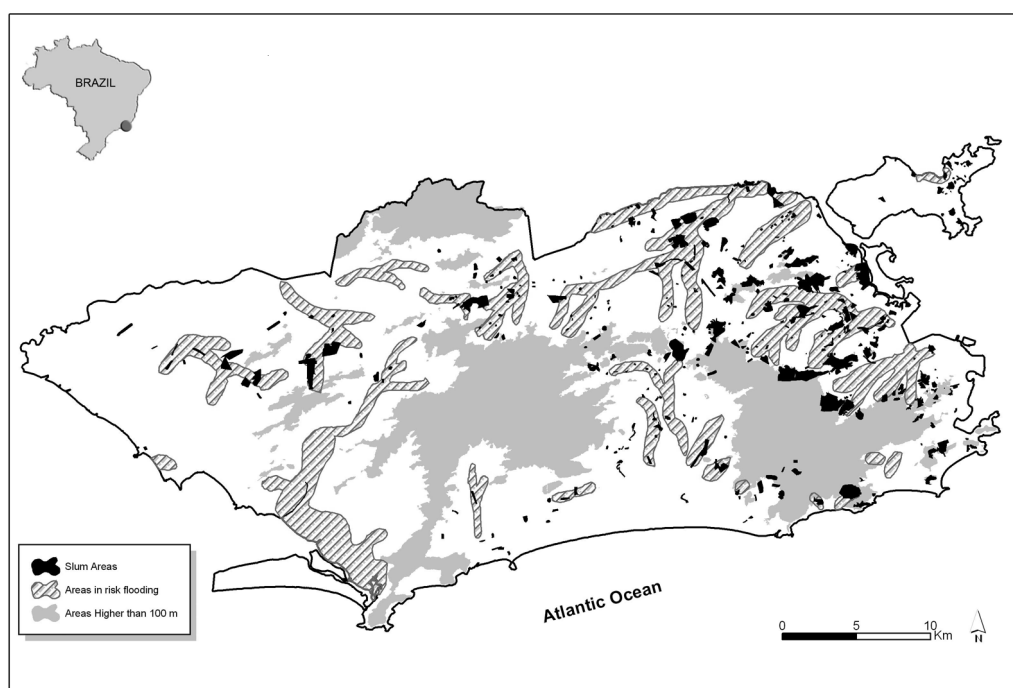
6.2 METHODS

Area of Study

The city of Rio de Janeiro (population 5.8 million [31]) has a large diversity of geographic, environmental and socioeconomic characteristics. The city boundaries include swamps and mountains as high as 800 m; densely populated areas as well as unpopulated forests and slum communities in close proximity to upper and middle class neighbourhoods. The urban plan of Rio de Janeiro was defined by decades of public investment in urban infrastructure, prioritizing neighbourhoods in

the southern areas adjoining the ocean beaches, while neglecting the poorer regions in the north and west sectors of the city [1] (Figure 2). Slum communities (“favelas”) are distributed throughout the city and occupy diverse geographic settings, which include most mountains and swamp regions in the city. Seasonal heavy tropical rain and flooding occur during the summer period between December and March and affects regions with inadequate water drainage.

Figura 2 - Distribution of areas with altitude greater than 100 m, slums areas and regions of flood risk in Rio de Janeiro, Brazil.



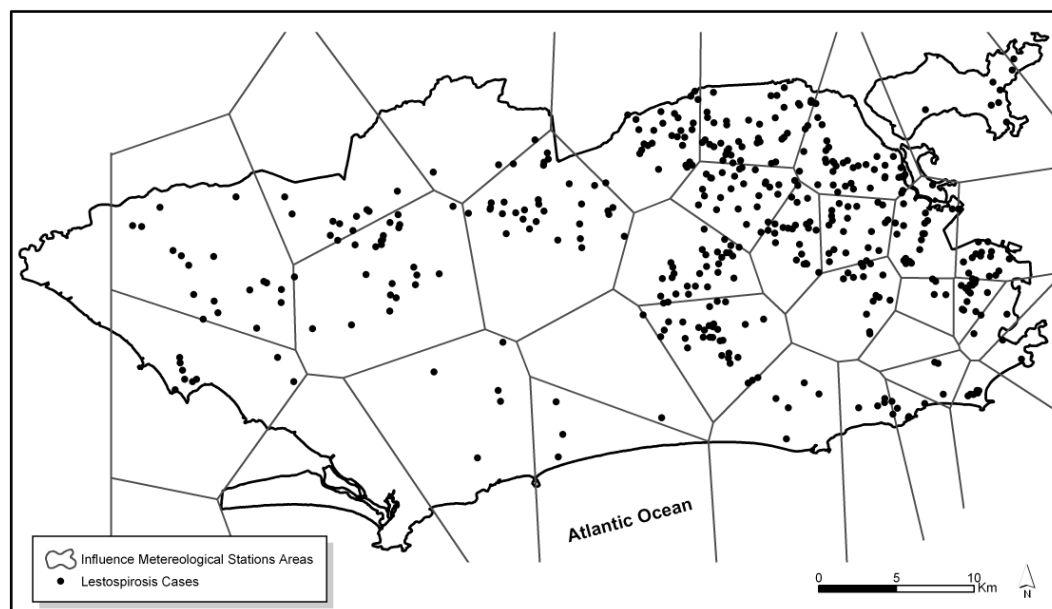
Data Sources and Indicators

Between 1997 and 2002, 514 leptospirosis cases were reported to the Municipal Health Secretary of Rio de Janeiro according to clinical, epidemiological and laboratory criteria of the Brazilian Ministry of Health [32]. Cases are reported on the basis of having signs and symptoms compatible of leptospirosis, such as jaundice,

acute renal insufficiency and haemorrhage; reported history of contact with potential risk factors such as flooding and reservoirs and laboratory evidence for the diagnosis obtained during microscopic agglutination test, culture isolation evaluations.

Automatic geocoding [45] localized only about 64% of the case residence according to census tract. A manual search algorithm [76] identified an additional 31% cases. Thus, the total geocodification process located the residence of 488 (95%) cases in 446 census tract polygons (Figure 2). No differences were detected between geocoded and the small number (26) of non-geocoded cases. Digital maps in 1:5000 scale were obtained from the Geoprocessing Laboratory/DIS/CICT/FIOCRUZ [45] and were used to create databases in the publicly-available geographic information system TerraView version 3.1.2 [21]. Socioeconomic indicators, such as residents per households, years of education of the head of household, numbers of inhabitants residing in slum areas, per capita household income, access to potable water and closed sewage systems, were obtained from the year 2000 national census [31]. Information was aggregated in 8145 census tracts. The Civil Defense Authority of Rio de Janeiro performs routine surveillance of flooding and provided digital maps of flood regions for the city. High risk areas for flooding were defined as the area within a buffer of 1 km surrounding the Civil Defense Authority defined flood regions (Figure 2). A network of 32 meteorological stations provided daily rainfall data for the city for the study period [33]. Voronoi tessellation was used to define the area of influence of the dataset generated from each station (Figure 3). This technique divides a plane with n points into n convex polygons ('Voronoi or Thiessen polygons'). Each point in a given polygons is closer to its central point than to the central point of other polygons [14]. Bartlett's test was used to evaluate the rainfall variability among the study years and Voronoi polygons that corresponded to the areas of influence surrounding meteorological stations [78].

Figura 3 - Distribution of leptospirosis cases and Voronoi polygons associated with each of the 32 meteorological stations in Rio de Janeiro, Brazil.



Statistical methods

SatScan software system was used to perform spatial scan statistics and identify space and time clusters among leptospirosis cases [43]. The algorithm is based on building a series of moving cylindrical windows in which the circular base and height correspond to a geographic area and time span, respectively. An infinite number of overlapping cylinders of different size and shape are generated which together encompass the entire study area and time span. The number of cases observed in each defined window is compared with the expected number, calculated based on the at-risk population in the study area. The maximum likelihood ratio is used to detect windows where the number of observed cases is significantly greater than expected [43]. The size of the moving windows was restricted to $< 10\%$ of the population (585,790 inhabitants). The incubation period of leptospirosis varies from 2 to 30 days, yet the usual range is 5–14 days [28]. We, therefore, evaluated windows with a maximum time span of 30 days in the models that incorporated the presumed incubation period and the possibility that exposures associated with an

outbreak event occurred over a 2– to 3–week period. The SatScan software system was used to identify clusters.

A cluster case was defined as a leptospirosis case, which belonged to a cluster. A generalized linear mixed model was used to evaluate risk factors for a cluster case in comparison with non-cluster case. A multilevel analysis was performed with two spatial levels: individual level and the 32 Voronoi polygon surrounding the meteorological stations. The census tract socioeconomic indicators were used to define the socioeconomic level of each case residence. It was not included as a level in the multilevel analysis because almost no census tract presented more than one case. Each case was related to the mean daily rainfall (measured at the closest meteorological station) that occurred during the preceding 3–20 days before the date of initiation of the symptoms. To identify the threshold of mean daily rainfall associated with the risk for developing cluster cases, several models were fitted, which evaluated different cut-points for mean daily rainfall.

Since information on covariates was obtained for spatial areas, not to the individual, a random effect term (intercept) was included in the logistic multilevel model, assuming a multivariate normal distribution with mean of zero. The variance partition coefficient (VPC) measures the proportion of variance explained by the higher level Voronoi polygon. Values for the VPC, which approach zero, provides an indication that the variability among areas does not affect the estimated parameter [79]. Akaike’s corrected information criterion was used to select the best fit model [52]. Models were fitted in the statistical package R version 2.2.1 [86].

6.3 RESULTS

The incidence of leptospirosis in Rio de Janeiro ranged from 1.06 to 2.05 cases per 100,000 population between 1997 and 2002 (Table 1). The highest incidence was observed in 1997 and 1998 and then decreased by 50% to the end of the period. Cases were distributed throughout the populated areas of the city but were concentrated in the poorer northern region of lowlands and spared the wealthier

southeastern sectors of the city 2).

Tabela 1: Leptospirosis Cases and Rainfall in Rio de Janeiro, Brazil from 1997 to 2002.

Year	1997	1998	1999	2000	2001	2002
Total Cases	114	111	64	65	71	63
Incidence ¹	2.05	1.99	1.14	1.16	1.20	1.06
Total Rainfall (mm)	28,202	50,698	31,010	33,105	33,134	32,913
Days with < 4mm of rain	180	191	177	142	142	151
Variation coefficient for annual rainfall ²	211%	259%	218%	286%	286%	251%

¹cases per 100,000 population. ²variation coefficient is estimate by standard deviation / mean ratio

Scan statistic analysis identified six space-time clusters, with one cluster occurring in each of the six surveillance years (Table 2). Significant clusters of 13 and 19 cases were detected in 1997 and 1998, respectively. Attack rates associated with the 1997 and 1998 clusters were 5.10 and 5.62 per 10 000 person-years, respectively. The four clusters identified between 1999 and 2002, albeit not significant due to the small number of associated cases (2–5), were responsible for high attack rates (144.80 and 1.52 per 100 000 person-years) in the cluster population. The six clusters had time spans between 14 and 25 days. Among clusters, four out of six occurred during the summer season associated with heavy rainfall and flooding. The large 1997 and 1998 clusters occurred in the same regions, but the 1998 cluster encompassed a geographical area twice the size of that for the 1997 cluster (Figure 4). The four clusters identified between 1999 and 2001 were small, both in area and case counts. The 2002 cluster was localized over a swamp region occupied by *favelas*, as defined by the Brazilian census bureau [31] (Figures 2 and 4). More than 20% of the population in the six cluster areas lived in *favelas*, whereas 4% of the population in non-cluster areas of the city resided in such conditions.

The temporal association with the summer season suggested the in-

Tabela 2: Characteristics of Leptospirosis Case Clusters Identified between 1997 and 2002.

Cluster	1	2	3	4	5	6
Time Span (days)	21	24	15	14	18	25
Time Frame	04/01/97 – 28/01/97	07/01/98 – 30/01/98	04/03/99 – 20/03/99	23/09/00 – 06/10/00	28/04/01 – 25/05/01	03/01/02 – 27/01/02
Cluster area (km ²)	24.96	50.26	0.20	0.05	0.14	17.69
No. of cases	13	19	2	2	2	5
Population in cluster area	402,325	566,208	3,361	1,811	5,906	478,952
Cluster attack rate (cases per 10,000 person-years)	5.10	5.62	144.80	287.92	68.67	1.52
Relative risk ¹	24.50	29.45	867.05	1393.24	446.42	12.68
p-value	0.001	0.001	0.291	0.161	0.590	0.973

¹Relative risk was calculated as observed/expected ratio.

fluence of rainfall on leptospirosis case clustering. A large variability was observed with respect to the spatial and temporal distribution of rainfall. The coefficient of variation was more than 200% in each of the surveillance years (Table 1), indicating significant variation in daily rainfall throughout the year. Furthermore rainfall, as measured by the 32 meteorological stations, was significantly heterogeneous across the city (P-value < 0.001, Bartlett’s test).

Multilevel models were used to evaluate the spatial and temporal influence of rainfall and spatial influence of socioeconomic and environmental characteristics on the risk of a leptospirosis case belonging to a cluster vs. noncluster (Table 3). Initial analyses did not identify a significant association between cluster cases and mean values for daily rainfall which occurred during the 3–20 day period preceding onset of the case’s illness. However, a threshold of heavy rainfall may be required to precipitate flooding cluster event. Subsequent analyses found that a threshold of mean daily rainfall >4 mm was significantly associated (OR 3.71; 95% CI 1.83–7.51) with leptospirosis cluster events. Higher threshold values for mean daily rainfall (i.e. >5, >6 mm) were significantly associated with leptospirosis cluster events while such associations were not found when threshold values lower than 4 mm were used in the analyses. Significant associations were not observed for demographic, socioeconomic and environmental available covariates such as flooding risk areas and slum

settlements, indicating that leptospirosis cases, either cluster or non-cluster, have a similar environmental and socioeconomic risk profile (Table 3). The best fit model included a single covariate, mean daily rainfall >4 mm, along with random effects. The high VPC (51%) indicates that incorporation of random effects in the model adequately accounted for the variability associated with the spatial level of Voronoi polygons.

6.4 DISCUSSION

This study addressed two questions: the identification of space-time clusters of leptospirosis cases and the effects of the climactic, socioeconomic and environmental variables on outbreaks. The ability to distinguish outbreaks from background endemic events is critical for mounting rapid and focused public health responses. Health education campaigns may be used in a targeted manner to identify cases early in the illness and therefore reduce the high case fatality (5–40%) associated with leptospirosis [25, 30, 28, 46, 11, 51]. Furthermore, an understanding of environmental risk factors for cluster events provides the basis to identify and implement interventions aimed at preventing future outbreaks.

We identified six distinct cluster events of leptospirosis during a 6-year surveillance period in Rio de Janeiro. Cluster events occurred in regions that comprised *favela* communities during a 14– to 25–day period and were associated with high attack rates (1.52–287.92 per 100,000 person-years). Most of the cluster events occurred during the summer, which is the season of heavy rainfall and flooding in the city. Detection of disease clusters has been a focus within the field of spatial epidemiology [26]. An advantage of the scan statistic approach used in this study [43], in comparison with other methods for identifying space-time clusters [50], is that the scan statistic approach takes into account the differences in the population at risk, while correcting for problems associated with multiple comparisons, and therefore avoids potential selection bias. Furthermore, the scan statistic allows an estimation of the relative risk attributed with the cluster event, and therefore serves

as a powerful epidemiological tool.

The spatial scan statistic approach identified clustering of leptospirosis cases despite limitations inherent with passive surveillance information. At present the performance of passive surveillance systems has not been evaluated in Rio de Janeiro or other cities in Brazil, where epidemics of leptospirosis occur. Cases reported to health authorities significantly underestimate the disease burden, since case ascertainment relies on identification of classic severe manifestations [42, 71, 51]. A minority (5–15%) of symptomatic infections develop such manifestations [30, 28]. Furthermore, case confirmation is achieved in a small proportion of suspected cases because of the low sensitivity of current serologic methods [46, 51]. It is likely that additional cases were associated with these clusters which were not identified by passive surveillance. In total, 43 (8.8%) of the 488 leptospirosis cases identified during surveillance occurred during a cluster event. Additional clusters may have occurred during the study period but were not detected because they were associated with small numbers of reported cases.

A major challenge in using the scan statistic approach is the difficulty in geocoding cases with information obtained from passive surveillance. Although the task of localizing case residence according to large areas such as administrative regions or neighbourhoods is relatively easier, it does not provide a sufficient degree of precision, especially when slum settlements are interspersed with wealthier communities in small regions. The use of postal code regions is limited by the lack of information on socioeconomic and environmental attributes for these polygons [9]. Census tracts, as used in this study, are an attractive alternative since they are relatively small (mean area of 0.063 km² in Rio de Janeiro) and the national census bureau has standardized databases of population counts and socioeconomic indicators for them. However, 31% of the cases required manual ascertainment of the location of their residence, since many had irregular addresses in slum communities which were not represented in official databases. Progress has been made in Brazil to register addresses in marginalized communities [70], which in turn may facilitate

application of more precise geocodification procedures in the future.

Space-time clustering of leptospirosis cases was based on the geographical location of case residence. This finding suggests that epidemic transmission occurs in the communities where high-risk populations reside. Identification of the place of exposure is critical to formulating effective control interventions for slum communities. Leptospirosis is traditionally considered a sporadic rural-based disease associated with risk occupations. Ecological and case-control studies found that household determinants such as poor sanitation infrastructure, exposure to environmental sources of contamination and high rodent populations, were risk factors for acquiring leptospirosis [42, 71]. In this study, information was not available to ascertain the proportion of cases that worked in the same geographical location of their residence or to evaluate potential clustering based on the location of the workplace. In fact, a case-control investigation found that exposure to environmental sources of contamination in the workplace was also a risk factor for acquiring severe leptospirosis [71]. More refined epidemiological investigations will therefore be needed to determine the contribution of household and workplace transmission for urban leptospirosis.

Multilevel modelling identified rainfall to be a significant risk determinant for a leptospirosis case belonging to a cluster vs. non-cluster event. We evaluated rainfall that occurred 3–20 days prior to the onset of illness since this period is the generally accepted range for the incubation period [28, 46]. The risk of developing a cluster case was three times greater (OR 3.71; 95% CI 1.83–7.51) than the risk of developing a non-cluster case for a given period when mean daily rainfall in the preceding 3–20 days was >4 mm. The association was not observed with lower cut-off values, thus indicating that a threshold level of rainfall is required to precipitate outbreaks in this urban setting.

Rainfall is a well-recognized risk factor for leptospirosis outbreaks [28, 46, 38], especially in disaster situations [91]. It also influences epidemics in urban settings [42, 9, 44, 10]. Heavy rainfall may influence the risk for acquiring

leptospirosis in different ways. It may affect the normal demographic cycle of rodent reservoirs by altering the reproductive periods and peak population densities [54]. In addition, flooding may increase transmission to humans by either driving reservoirs into human dwellings or by facilitating the dissemination of pathogenic *Leptospira* excreted from rodent urine [66]. In Rio de Janeiro, rainfall above 4 mm measured in any meteorological station should be used as a threshold for alerting health professionals working in public hospitals, especially in the influence area of the station. Rainfall threshold values associated with outbreaks need to be determined in urban settings where leptospirosis is an endemic disease and has the potential to cause outbreaks.

Generalized linear mixed models were used to evaluate risk determinants for leptospirosis cluster events to incorporate demographic, socioeconomic, environmental and climactic covariates encoded at individual and spatial unit levels. Voronoi polygons are often used in environmental studies to evaluate the effect of spatially distributed rainfall measurements [61]. Census tracts were used since standardized datasets on socioeconomic indicators for populations residing within tracts were available from the national census bureau and information on these factors was incomplete or not reliably collected from cases during routine surveillance. Our approach did not address the hierarchical structure of the spatial data and therefore may have underestimated standard errors for regression coefficients. We assumed that the random effects for geographical areas were independent. The use of a spatially correlated random effect would have been more appropriate, as geographical proximity usually infers a degree of similarity. However, as we were dealing with two different kinds of geographical units, the model would need to use two different neighbourhood matrices, and such an algorithm is not integrated in available statistical software packages.

We did not identify significant determinants, other than rainfall, which influence the risk of developing a cluster rather than non-cluster case. This may reflect the finding that leptospirosis cases, irrespective of whether they occur in

cluster or non-cluster events, are predominantly urban slum dwellers. In Rio de Janeiro, census tracts have a population of approximately 800 inhabitants and are relatively homogeneous, thus limiting the potential for zoning effects related to use of area data. Our study may have not had sufficient precision or power to detect differences in socioeconomic level or environmental exposures that influence the risk of leptospirosis clustering within this poor population. Furthermore, Rio de Janeiro is a city with a complex topology where widely disparate socioeconomic communities are often geographically juxtaposed. Urban slums (*favelas*) are distributed in a mosaic pattern throughout the city.

In summary, we found that urban outbreaks of leptospirosis occur as rainfall surpasses a specific threshold value. Monitoring of rainfall may thus be used to alert health services and communities of outbreak threats and in turn promote rapid responses aimed at early case identification and prevention of mortality due to severe leptospirosis. This study was performed in one city in Brazil, which has specific characteristics of urban poverty, climate and geography. Our findings need to be confirmed in other urban centres where endemic transmission of leptospirosis occurs. Urban epidemics of leptospirosis are likely to become an increasingly important public health problem due to global climate changes that are predicted for the future (IPCC 2007). Urban leptospirosis is a consequence of disorganized urbanization and lack of investment in adequate housing, sewage systems and refuses collection services. The most effective interventions will therefore be those that directly address the underlying conditions of poverty, such as lack of access to proper sanitation, which are responsible for the emergence of this urban health problem (Ko et al. 1999; McBride et al. 2005).

ACKNOWLEDGMENTS

We appreciate the contributions of Dr Oswaldo G. Cruz for writing the R function to relate rainfall and cases; Dr Reinaldo Souza Santos for assistance in the dataset organization and Daniel Skaba for helping with the manual algorithm for

geocoding. We acknowledge Dr Cláudio Bustamante Pereira de Sá (*in memoriam*) contributions: beyond author and friend, he was present during the conception, design and analysis of the study. This research was supported by Brazilian National Research Council (Project SAUDAVEL) and National Institutes of Health, USA.

Figura 4 - Distribution of six leptospirosis case clusters in Rio de Janeiro from 1997 to 2002, which were identified in spatial scan statistics. The spatial distribution of cluster events is shown according to the census tract in which cluster cases resided. Cluster events in 1999, 2000 and 2001 involved few census tracts while cluster events in 1997, 1998 and 2002 involved more widespread areas of the city. All cluster events occurred in census tracts that were situated in the city's periphery.

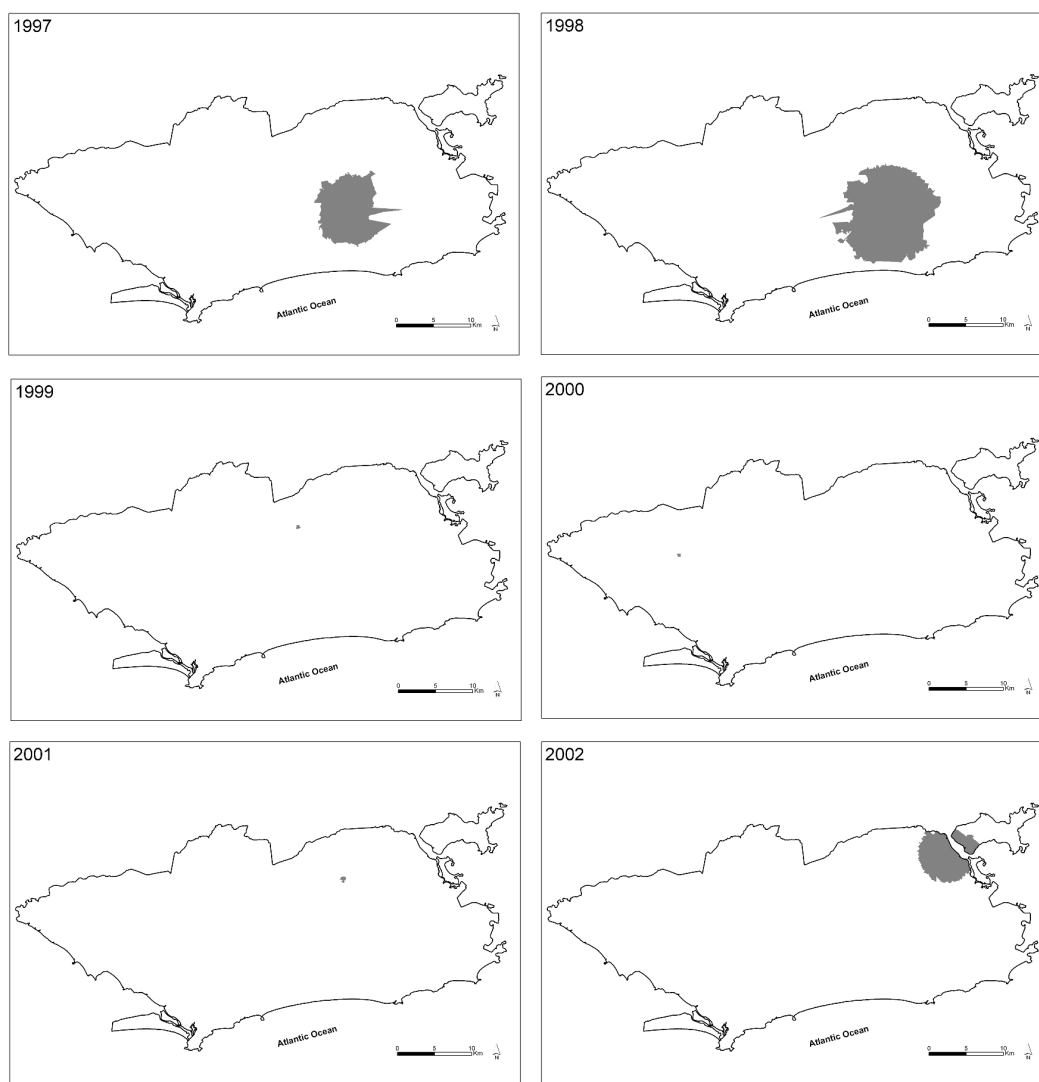


Tabela 3: Generalized Linear Mixed Model Estimates of Risk Factors for Leptospirosis Clusters Cases.

Variables	Full Model		Final Model	
	OR ¹	95 % CI ²	OR ¹	95 % CI ²
Individual level				
Age group				
< 14 y. (ref.)	1.00	—	—	—
15 – 24 y.	0.84	0.05 – 15.12	—	—
25 – 34 y.	0.85	0.05 – 14.61	—	—
35 – 44 y.	0.83	0.05 – 13.69	—	—
> 44 y.	0.61	0.04 – 9.96	—	—
Sex				
Male	0.64	0.23 – 1.80	—	—
Level of Voronoi polygon				
Mean daily rainfall > 4mm ³	4.67	1.87 – 11.70	3.71	1.83 – 7.51
Level of census sector				
> 3 inhabitants per household	4.48	0.88 – 22.84	—	—
> 53% family heads with > 8 years of schooling	0.57	0.22 – 1.49	—	—
Slum region (Favela)	0.64	0.13 – 3.07	—	—
Residing < 1 km from a flood region	1.77	0.48 – 6.46	—	—
Random effects variance	2.83	—	—	—
AIC	209	—	—	—

¹OR, odds ratio. ²CI, confidence intervals. ³Mean daily rainfall was calculated for the 2-30

day period prior on set of symptoms for the case.

7 ARTIGO 2 - SPATIAL MODELING OF LEP- TOSPIROSIS IN A URBAN SLUM AREA

Wagner de Souza Tassinari ^{1,2}

Renato Barbosa Reis ³

Ridalva Dias Martins Felzemburgh³

Francisco Santana⁴

Mitermayer Reis ³

Albert Icksang Ko ^{3,4}

Marilia Sá Carvalho ¹

National School of Public Health, Oswaldo Cruz Foundation, Brazilian Ministry of Health, Rio de Janeiro, Brazil ¹

Department of Mathematics, Federal University Rural of Rio de Janeiro, Rio de Janeiro, Brazil ²

Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Brazilian Ministry of Health, Salvador, Brazil ³

Health Secretary of Bahia State, Bahia, Brazil ⁴

Division of International Medicine and Infectious Disease, Weill Medical College of Cornell University, New York, USA ⁵

Resumo

Leptospirosis is an infectious disease of global importance, which has emerged to be a major urban health problem due to rodent-borne transmission in urban slums. This study aimed to model the spatial distribution and, simultaneously, the individual and environmental factors related to *Leptospira* infection in an urban slum community in Brazil. A household survey was performed on 6,466 residents of Pau da Lima, a slum community in the periphery of Salvador, Brazil. Serological evaluation was used to identify subjects with prior infection. Generalized Additive Models were used to fit individual and contextual covariates, at the same time estimating the spatial risk of acquiring *Leptospira* antibodies. The odds ratio of acquiring *Leptospira* antibodies was significant for: sex, age, race and socioeconomic variables, and contextual factors, such as proximity of open sewers, sighting rats, animals and altitude. The spatial variation in risk was significantly larger for households located at the bottom of the valleys. The range of spatial odds ratio decreased with the inclusion of fixed covariates, but still keeping the same pattern. The methods chosen were reliable than allow us to identify areas that inhabitants actions should be focused to reduce the human contact to pathogenic *Leptospira*.

Keywords: Leptospirosis, spatial epidemiology, generalized additive models

7.1 INTRODUCTION

Leptospirosis is a severe infectious disease of worldwide distribution caused by a pathogenic spirochaete bacteria of the genus *Leptospira* [46]. It is able to infect a range of wild and domestic mammalian species, increasing their diffusion potential [87]. Disease transmission begins when a susceptible person has contact with water, soil, or other media contaminated with leptospire.

According to the World Health Organization [92], the incidence ranges from 0.1 to 1 cases per 100,000 population/year in temperate climates and from 10 to 100 cases per 100,000 population/year in humid tropical climates. In Brazil more than 35,000 confirmed cases were identified between 1985 and 1997, among which case fatality was 12.5% [32]. The majority of cases reported happened in large urban centres [42]. Leptospirosis cases occur throughout the year in this setting [71], indicating that endemic transmission occurs.

However, the severity of this disease varies from asymptomatic or mildly symptomatic to rapidly fatal or severe manifestations, with only a small fraction of infected individuals with pathogenic *leptospira* progressing to severe presentations [89]. Reliable data on its incidence and prevalence in different areas is scarce, because leptospirosis infection is generally underdiagnosed and underreported [75].

Urban epidemics are the main concern and have long been associated with poor sanitation, slum housing and flooding [71, 42]. Flooding facilitates exposure to environments contaminated with rat urine. Individual behaviour, such as unprotected contact with sewer and mud cleaning, is related to higher infection rates [46], possibly due to contact with contaminated soil. Nevertheless, not much is known about the risk factors to seroconversion and to progression from mild to severe clinical manifestations. Certainly the size of the inoculum should account for differences in clinical presentation. In observational studies, however, this variable is only measurable using some proxy approach, and exploring the factors associated with soil contamination, responsible potentially for the presence of viable *leptospira*. Although environmental and behavioural factors being universally acknowledged to

relate to human leptospiral contamination [28], no previous study, to our knowledge, have examined the environmental determinants of sub-clinical infection. Uncovering the factors associated to sub-clinical leptospirosis can help to enlighten the natural history of the disease. A recent study reported that low socioeconomic status, residents of households that were located near to a refuse deposit or an open sewer, working in direct contact with sewerage or waste and flooding around the household were risk factor for acquiring urban *Leptospira* antibodies [64].

The aim of this study is to investigate the spatial pattern of environmental and individual factors associated with leptospirosis infections. We focused in explain possible non linear effect of continuous covariates such as age, domicile altitude and distance from pollution sources. The geographical coordinates of the household was the spatial component included in the models.

7.2 METHODS

Study Site

The study was conducted in the Pau da Lima community which is situated in the periphery of Salvador, a city of almost 2.5 million of inhabitants [31] in Northeast Brazil. Pau da Lima is a region comprised of four valleys in an area of 0.46km², transformed into a densely-populated slum settlement due to immigration of squatters in the seventies [88, 74]. In total, 36% of the population of Salvador and 28% of the population in Brazil reside in slum communities with equal or greater levels of poverty as that found in Pau da Lima [88, 74].

Study Population and Data Collection

A household survey was conducted in 2003, composed of interviews and questionnaires about socioeconomic status, occupation and employment and exposures to sources of environmental contamination in the household and workplace. Subjects were enrolled ($n = 6,466$) according to written informed consent approved

by the Institutional Review Boards of the Oswaldo Cruz Foundation, Brazilian National Commission for Ethics in Research, and Weill Medical College of Cornell University. More details elsewhere [64].

Coordinates of each household were localized and included in a Geographical Information System (GIS). A digital terrain model of topographic data was used (ArcGIS 3D Analyst Extension software) to obtain continuous estimates of altitude for the study area and extract potential GIS predictor variables. The distances, calculated in three dimensional space, of households to nearest open drainage systems and refuse deposits were evaluated as proxies of exposure to these sources of environmental attributes. Elevation of households with respect sea level in which they were situated was used as a surrogate for flood risk.

Study Variables

The response variable leptospirosis infection (yes or no) was defined by a serovar titter greater to or equal to 1:25 to define the presence of *Leptospira* antibodies [64]. The variables belong to two different levels: individual and contextual (domiciliary or peri-domiciliary). Individual covariate were: sex (male or female), age, race (black or non black), schooling (complete or incomplete primary school education), the daily income (in US dollar), works (or worked) with sewage in the last years (yes or no) and individual contact with mud, trash or sewer (yes or no). The contextual covariates were: time living in the same household (\geq or $<$ 15 years), the house distance from open sewer and from trash, maximum number of rats seen in the peri-domicile ($>$ or \leq 2 rats) and presence of cats or dogs in the peri-domicile (yes or no) and altitude of the domicile above sea level. The geographic coordinates of the household were used to estimate the spatial effects.

Data Analysis

Crude prevalence rates and univariate odds ratio (OR) with 95% confidence intervals (CI) were reported for the categorical covariates. A univariate generalized additive model (GAM) [37] were used to evaluated the functional form of the association between continuous covariates and the risk of acquiring *Leptospira* antibodies.

Initially, a semiparametric non-spatial model, a generalized additive model (GAM), was fitted. However, these models assume that the samples are statistically independent. As our observations are spatially related, models that ignore spatial dependence are inappropriate, as they overestimate the effect of covariates (e.g., environmental variables) and underestimate the standard errors [26]. To incorporate spatial dependence in the model we used thin plate smoothing spline logistic regression [93]. Hence, the model was of the form:

$$Y \sim \text{Bernoulli}(p)$$

$$\text{logit}(p) = \mathbf{X}\beta + s(\mathbf{Z}) + s(e, n)$$

where Y is the response variable, \mathbf{X} is a matrix of explanatory categorical covariates and $s(\mathbf{Z})$ is a unidimensional smooth function to estimate possible variations on the effect of the \mathbf{Z} continuous covariates effects. The additional term, $s(e, n)$, indicate the bi-dimensional smooth function of the East and North coordinate. It models the spatial structure still present after controlling for the known risk factors. This smoothing parametre is selected by minimising the unbiased risk estimator, which was equivalent to minimising the expected mean square error [93]. The maps present the spatial OR adjusted by all covariates included in each model. There is only one scale for the spatial effects comparison.

In multivariate analysis, all variables which had a p -value below 0.10 in univariate analyses were included in the regressions. Akaike's corrected informa-

tion criterion (AIC) were used to select the model [52, 29]. All statistics analysis were done in the statistical package R, version 2.7.0 [86], and with mgcv library [93] to fit the GAM models.

7.3 RESULTS

Table 4 presents the observed overall prevalence of *Leptospira* antibodies (16.08%). The prevalence in males (18.55%) is 1.32 times larger than in females. Individuals who work with sewerage or waste and had contact with mud, trash or sewer the leptospira antibodies present prevalence of 29.35% and 24.21%, respectively. Among contextual covariates the prevalence of people who live in the same household for more than 15 years is 23.96%; having seen more than two rats around the peri-domicile in recent weeks is 20.72%; dog or cat in household is approximately 19% and 21%, respectively.

Daily income and altitude above sea level of the domicile presented a linear relationship with the outcome, therefore they entered the model in the original scale. All other continuous covariates effects were non-linear (Figure 5). Age increased the odds of acquiring *Leptospira* antibodies until approximately twenty five years old, decreasing thereafter. The distance to the nearest open sewers and trash collection sites decreased the odds until approximated sixty and one hundred metres respectively.

Table 5 presents the results of the GAM and the spatial GAM models for three approaches: without fixed covariates, just individual covariates and both individual and contextual covariates. The last line of the table shows the AIC, always smaller for the spatial models, indicating that the place of the household is determinant to seroconversion.

The inclusion of the spatial smooth term in the models improved the fit significantly ($p\text{-value} < 0.001$). The effects of individual and contextual covariates were similar in all models, except for race and presence of dog in the household, that lost significance when the spatial effect was incorporated. Comparing the individual

Tabela 4: Descriptive analysis of categorical variables available for the univariate and bivariate analysis for the prevalence of Leptospirosis study in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004.

	N	Positive Results	Prevalence (%)
All Subjects	6466	1040	16.08
Individual level			
Gender			
Males	2927	543	18.55
Females (ref.)	3539	497	14.04
Primary school education			
Incomplete	1536	212	16.79
Complete (ref.)	4930	828	13.8
Contact with mud, trash, sewer, etc			
Yes	818	198	24.21
No (ref.)	5648	842	14.91
Race			
Black	1916	350	18.27
Non-black (ref.)	4550	690	15.16
Work with sewerage/waste			
No (ref.)	6098	932	15.28
Yes	368	108	29.35
Contextual level			
Living in the same household			
≥ 15 years	1398	335	23.96
< 15years (ref.)	5068	705	13.91
See rats near home			
≤ 2 rats (ref.)	3782	484	12.80
> 2 rats	2684	556	20.72
Dog in the household			
No (ref.)	4004	574	14.34
Yes	2462	466	18.93
Cat in the Household			
No (ref.)	5424	823	15.17
Yes	1042	217	20.83

models, the variables with the largest effects were sex (OR 1.39, 95% IC 1.20 - 1.62) and working with sewage (OR 1.48, 95% IC 1.13 - 1.93). The effects of the contextual covariates “time of living in the same household” (OR 1.40, 95% IC 1.18 - 1.65) and “the presence of rats” (OR 1.34, 95% IC 1.15 - 1.55) were the largest. As for the smooth terms, the age was significant ($p\text{-value} < 0.001$) in all models, but open sewer lost significance ($p\text{-value} = 0.193$) when the spatial term was included. In both full models, spatial and non spatial, the altitude above sea level of the domicile and the daily income covariates were included as continuous and thus measured the linear protection effect for leptospirosis infection: for each metre above sea level, the odds of

acquiring *Leptospira* antibodies decreased 1%; for each daily dollar income increase, the odds decreased approximately 11% (full models). The best fit, according to AIC (5360.10), was the full model with individual, contextual covariates and the spatial term, which was significant ($p\text{-value} < 0.001$) (Table 5).

In figure 6 white lines depict significantly higher risk areas and black lines significantly lower risk areas considering 95% confidence intervals. The spatial pattern in all maps was similar: the region of increased risk of acquiring *Leptospira* antibodies was in the north, spreading towards the east. However, controlling for individual and contextual covariates the amplitude of the adjusted odds ratio decreased. The range of the odds ratio varies from 0.19 to 5.75, in the first spatial model, without fixed covariates, to 0.37 to 3.98 in the full model, indicating the explanatory power of the fixed individual and contextual covariates included.

7.4 DISCUSSION

This work addressed two questions: the identification of potential individual and contextual factors that contribute to leptospirosis infection and the localization of high risk areas of leptospirosis contamination. The results of this work are consistent with those observed in other studies of risk factors associate with the urban leptospirosis [16, 8, 9].

The larger number of women (55% of the total sample), due to absence of men from household during the field work time, could change the effect of some environmental factors even after controlling for sex covariate in the model. Another potential source of confounding, that was also controlled in the models, is the time of living in the same place. Althout this covariate was collected it was not validated.

There are several flexible approaches that could be used to verify the dose-responce relationships and to evaluate the functional form besides the models used [34, 68]. Our choice, however was driven by the flexibility of the model. First of all, the generalized linear approach allows any distribution to be modelled including

counts or continuous variables, as long as they belong to the exponential family of probability distributions [52]. Besides the inclusion of a smooth function to deal with effects that vary according to the covariate level is a powerful tool for epidemiological understanding. Applied to spatial analysis, specifically in point pattern modelling, the bi-dimensional smooth function is an intuitive way to estimate the spatial effects and at the same time plot tolerance contours of significant high and low risk areas. The intra-domiciliary sample cluster was not a specific term in the model because the inclusion of a household random effects smooth out the more wiggly component of the spatial term, that was our main focus.

Although our study is based on the serological survey done by Reis et. al. [64], it was possible to validate their results with a larger sample. Besides, the statistical approach was able to deal with the spatial dependence and estimate, at the same time, the effects of several individual and environmental exposures and its effect on the spatial pattern. However, those covariates, either from the questionnaire or derived from the GIS were not sufficient to explain all spatial variation on the leptospirosis risk. The OR maps indicated smaller spatial effects following the inclusion of individual and afterwards contextual covariates (Figure 6).

Despite of the overall poverty, the spatial distribution of the population in the “favelas” is not homogeneous: the poorest people live at the lowest areas, populated by rats, where waste and flood waters accumulate. To survive in the soil, the leptospire require neutral pH and temperate (around 25°C) temperature, while salinity and pollution are inhibiting factors. The stability of the spatial pattern across models is certainly due to non-measured spatially distributed variables, suggesting the presence of unknown aspects of the ecology of this disease, including particularly the population of hosts and possible concentration areas of *leptospira*. However, little is know about the relationship between seroconversion and the development of severe disease and which aspects trigger the incidence of diseases clusters, in relation to environmental and socioeconomic characteristics. The dynamics of leptospirosis transmission in an urban environment is still not completely understood. Never-

theless we believe to have enough evidence to support the development of focused environmental sanitation programs to controll urban rodent populations and their habitats, which in turn would reduce the contact between pathogenic *Leptospira* and humans.

Figura 5 - Generalized additive models (GAM) of the association between the risk of acquiring *Leptospira* antibodies and continuous variables of (A) Individual age (years), (B) Distance in metres to the nearest open sewer, and (C) Distance in metres to the trash collection. The adjusted odds ratio, in the GAM model is a measure for the risk of acquiring *Leptospira* antibodies. Solid lines represent the point estimate; dotted lines represent upper and lower 95% confidence band.

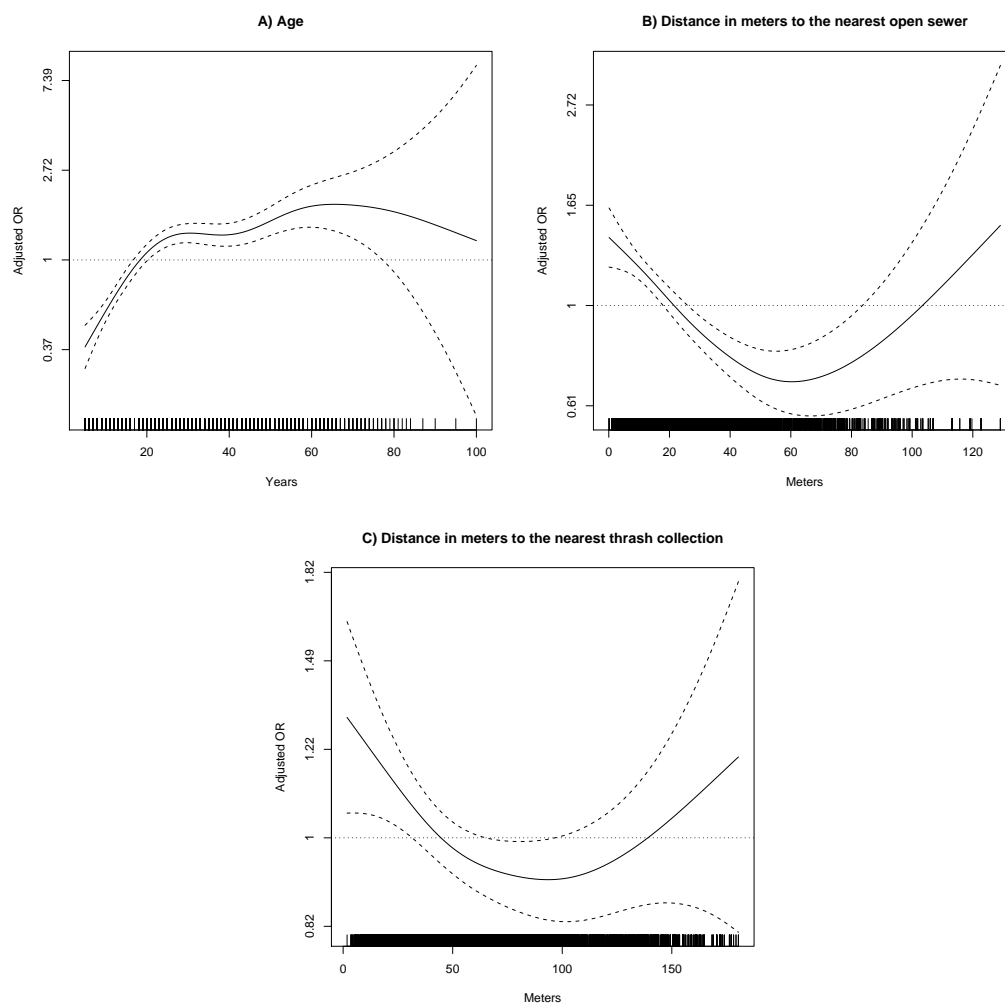
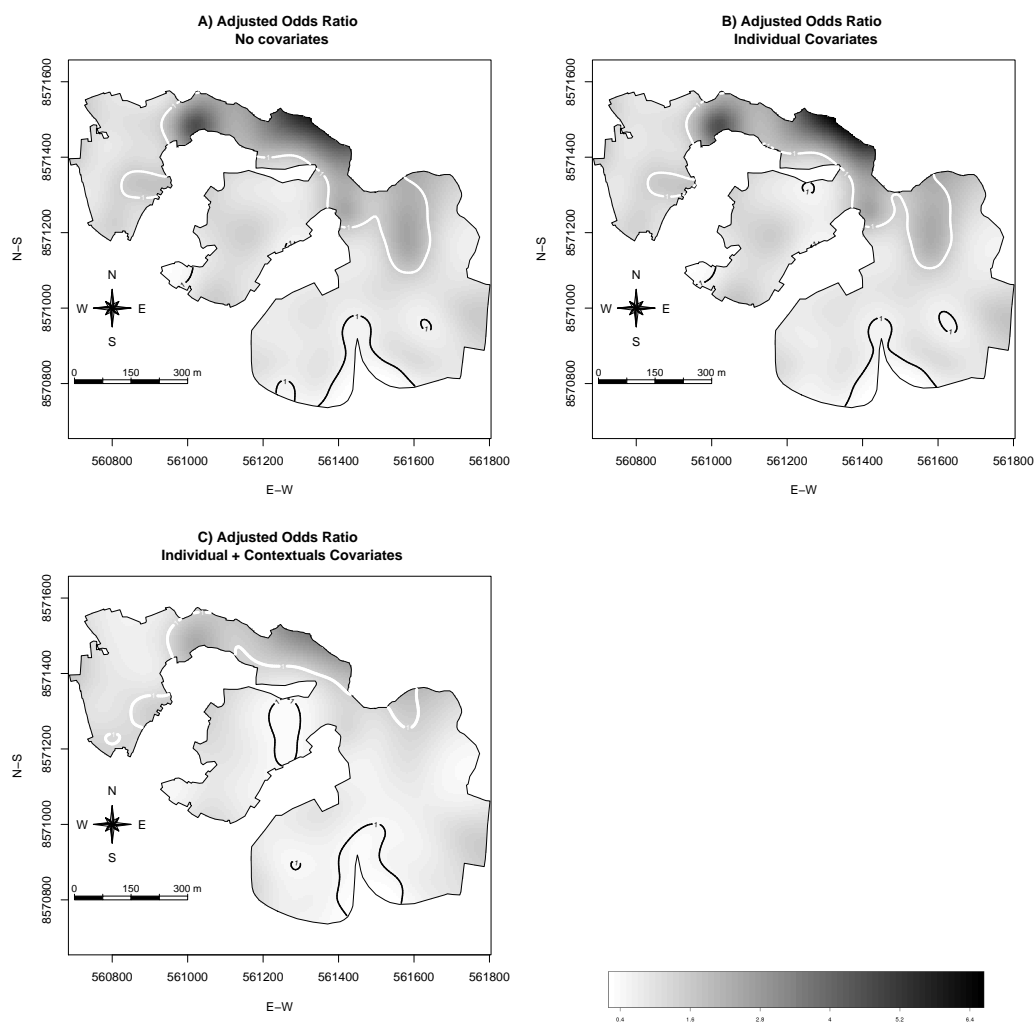


Tabela 5: Adjusted Odds Ratio and others fit measurements of de logistic regression for the of prevalence of Leptospirosis in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004.

Logistic Models Variables	No covariates models		Individual Models		Full Models	
	Empty model OR [95% IC]	Spatial GAM OR [95% IC]	GAM OR [95% IC]	Spatial GAM OR [95% IC]	GAM OR [95% IC]	Spatial GAM OR [95% IC]
Individual Variables						
Sex (Male)			1.41 [1.22; 1.63]	1.42 [1.22; 1.64]	1.40 [1.21; 1.62]	1.39 [1.20; 1.62]
Primary school education (Incomplete)			1.32 [1.10; 1.57]	1.25 [1.04; 1.50]	1.23 [1.02; 1.47]	1.22 [1.02; 1.47]
Daily Income (1 dollar)			0.83 [0.79; 0.88]	0.86 [0.82; 0.91]	0.87 [0.83; 0.92]	0.89 [0.84; 0.93]
Contact with mud, trash, sewer, etc (Yes)			1.30 [1.07; 1.58]	1.30 [1.07; 1.59]	1.24 [1.02; 1.51]	1.26 [1.03; 1.54]
Race (Black)			1.22 [1.05; 1.41]	1.12 [0.96; 1.30]	1.16 [1.01; 1.34]	1.10 [0.94; 1.28]
Scaven Work (Yes)			1.53 [1.18; 1.98]	1.51 [1.15; 1.97]	1.46 [1.12; 1.90]	1.48 [1.13; 1.93]
Age smooth terms (p-value)	-	-	< 0.001	< 0.001	< 0.001	< 0.001
Contextual variables						
Living in the same household (≥ 15 years)					1.46 [1.24; 1.72]	1.40 [1.18; 1.65]
See rats (> 2 rats)					1.41 [1.22; 1.62]	1.34 [1.15; 1.55]
Dog in the household (Yes)					1.23 [1.06; 1.42]	1.16 [0.97; 1.34]
Cat in Household (Yes)					1.26 [1.06; 1.51]	1.22 [1.01; 1.46]
Altitude above sea level (metres)					0.99 [0.98; 0.99]	0.99 [0.97; 0.99]
Open Sewer smooth terms (p-value)					0.04	0.19
Fit measures						
Spatial smooth terms (p-value)	-	< 0.001	-	< 0.001	-	< 0.001
AIC	5882.86	5800.37	5456.28	5351.59	5360.10	5302.35

Figura 6 - Risk maps for the adjusted odds ratio of logistic spatial regressions for the of prevalence of Leptospirosis in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004. Black and white lines represent upper and lower 95% confidence bands, respectively. And a common legend for odds ratio surface adjusted.



8 ARTIGO 3 - SPATIAL-LONGITUDINAL MODELS APPLIED TO LEPTOSPIRAL SE- ROCONVERSION INCIDENCE

Marilia Sá Carvalho¹

Wagner de Souza Tassinari^{1,2}

Ridalva Dias Martins Felzemburgh³

Renato Barbosa Reis³

Mitermayer Galvão dos Reis³

Albert Ko^{3,4}

*National School of Public Health, Oswaldo Cruz Foundation, Brazilian Ministry of Health, Rio
de Janeiro, Brazil¹*

*Department of Mathematics, Federal University Rural of Rio de Janeiro, Rio de Janeiro,
Brazil²*

*Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Brazilian Ministry of Health,
Salvador, Brazil³*

*Division of International Medicine and Infectious Disease, Weill Medical College of Cornell
University, New York, USA⁴*

Resumo

Leptospirosis is a bacterial disease which has emerged to become an urban health problem due to rodent-borne transmission in urban slums. The major known risk factors are poor sanitation, poor housing and flooding which facilitate exposure of slum residents to environments contaminated with rat urine. Floods facilitate exposure to rat-urine-contaminated environments. Brazil reports about 10,000 cases annually in the major cities, with 10-15% mortality during outbreaks. A prospective community-based cohort study was performed with 1,204 residents from Pau-da-Lima, a slum community in the periphery of Salvador, Brazil. The microagglutination test (MAT) was used to identify subjects with leptospiral prior infection defined as seroconversion (MAT titer from zero to $\geq 1:50$). GEE, GAM and GMM approaches were used to fit individual and contextual covariates, estimating the spatial risk of leptospiral seroconversion at the same time. The odds ratio was significant for: sex, age and sighting of rats in the peridomiliary environment. The spatial pattern of leptospiral seroconversion in Pau da Lima is different from seroprevalence, the variation in risk was significantly larger in the border of the middle valley. Leptospirosis is expected to become an increasingly important slum health problem as predicted global climate change and growth of the world slum population evolves.

Keywords: leptospirosis, spatial epidemiology, generalized estimation equations, generalized additive mixed models

8.1 INTRODUCTION

Leptospirosis is a worldwide zoonosis and usually human contamination occurs after contact with water and soil containing urine of infected rats and other animals [28], affecting people on all continents [11, 46]. The severity of this disease varies from asymptomatic or mildly symptomatic to rapidly fatal or severe manifestations. But only a fraction of individuals infected with pathogenic leptospira progress to develop severe leptospirosis [89]. Because leptospirosis has protean clinical manifestations, biological tests are essential for diagnosis, such as microagglutination test (MAT) and polymerase chain reaction (PCR) assay [57].

Leptospirosis occurs all over the world but reliable data on its incidence and prevalence in different areas is scarce, because leptospirosis infection is generally underdiagnosed and underreported [75]. According to currently available reports, incidences of severe form range from approximately 0.1-1 per 100,000 per year in temperate climates to 10-100 per 100,000 in the humid tropics, where the warm and wet climate provides a favourable environment for the survival of the leptospira [27, 58, 72]. During outbreaks and in high-exposure risk groups, disease incidence may reach over 100 per 100,000 [92]. In Brazil, about 10,000 cases are reported annually in the major cities, with 10-15% mortality during outbreaks [71].

Urban epidemics are the main concern and have long been associated with poor sanitation, slum housing and flooding [71, 42]. Flooding facilitates exposure to environments contaminated with rat urine. Individual behaviour, such as unprotected contact with sewer and mud cleaning, is related to higher infection rates [46], possibly due to contact with contaminated soil. Nevertheless, not much is known about the risk factors to seroconversion and to progression from mild to severe clinical manifestations. Certainly the size of the inoculum should account for differences in clinical presentation. In observational studies, however, this variable is only measurable using some proxy approach, and exploring the factors associated with soil contamination, responsible potentially for the presence of viable *leptospira*. Although environmental and behavioural factors being universally acknowledged to

relate to human leptospiral contamination [28], no previous study, to our knowledge, have examined the environmental determinants of sub-clinical infection. Uncovering the factors associated to sub-clinical leptospirosis can help to enlighten the natural history of the disease.

The aim of this study was to investigate the spatial epidemiology of leptospiral seroconversion in a large prospective community-based cohort study performed in a Brazilian slum community (favela), by simultaneously estimating individual and environmental factors associated with this infection process, in order to develop adequate community level intervention measures and contribute to the evaluation of candidate vaccines [64].

8.2 METHODS

Study Area

The study was conducted in the Pau da Lima community which is situated in the periphery of Salvador, a city of 2,443,107 inhabitants [31] in the Northeast of Brazil. Pau da Lima is a region of hills and valleys, which was a sparsely inhabited area of Atlantic rain forest in the 1970s and subsequently transformed into a densely-populated slum settlement due to immigration of squatters. In total, 36% of the population of Salvador and 28% of the population in Brazil reside in slum communities with equal or greater levels of poverty as that found in Pau da Lima [88, 74].

Study Population

The incidence data comes from a prospective community-based cohort study, GIS (Geographic Information Systems) and environmental data description are described elsewhere [64]. Seroconversion, defined as MAT $>$ 1:50, was followed-up in 2004, 2005, 2006 and 2007. In this article we analysed 1204 subjects containing complete data for the four measurements.

Study Variables

The response variable used was leptospiral seroconversion (yes or no), defined as a serovar titter equal to or greater than 1:50. The covariates related to the individual were: gender (male or female), age (continuous), race (black or non black), schooling (never studied, primary and high school or further education as reference categories), monthly income (“reais”), contact with mud, flooding, trash or sewer (yes or no). The peridomiciliary contextual covariates were: from the domicile to an open sewer and a trash deposit (continuous), maximum number of rats seen in the peridomiciliary environment ($>$ or ≤ 3 rats), presence of cat, dog and/or chickens in the domicile (yes or no), altitude above the sea level and geographic coordinates.

Data Analysis

In this longitudinal study three different modeling approaches were used: generalized estimating equations (GEE) models [47], generalized additive models (GAM) [37] and mixed models [93].

GEE methods extend the Generalized Linear Model (GLM) to include adjustment for within subject correlation, modeled separately as a nuisance parameter. In GLM the mean μ_i of the response y_i for the i^{th} subject is related to a vector of covariates \mathbf{x}_i , via a link function $g(\cdot)$, so that $g(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta}$, $\boldsymbol{\beta}$ is the vector of parameters to be estimated. The specification of the marginal GEE model departs from that stated above due to the inclusion of several measures of each individual at t points in time, resulting in a correlation between observations no longer assumed to be zero, but rather characterized by a known variance function $var(y_{it})$. Among various possible specifications of the correlation function – independent, auto-regressive, exchangeable – we used the last one. The parameters affecting the mean response have the same population averaged interpretation as in the standard GLM and are estimated using quasi-likelihood methods.

The Generalized Additive Models (GAM) extend GLMs to include non-linear relationship between response and independent variables [93] via non-

parametric smoothing functions. GAMs are more flexible than linear models, but still interpretable since the function $s(\cdot)$ can be plotted to give a sense of the marginal relationship between the predictor and the response. The advantage of the GAM approach is that the best transformations are determined simultaneously and without parametric assumptions regarding their form [29]. The smoothing parameter is selected by minimizing the unbiased risk estimator, which is equivalent to minimizing the expected mean square error [93]. Considering $s(\cdot)$ the smoothing function of some z covariates, the model can be expressed as:

$$g(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{s}(z_i) \quad (3)$$

The smoother can be uni or bivariate, and among various possibilities, we used a smoothing spline function.

Mixed (or random effects) models are proposed to correlated data either from correlated data, either generated by repeated measures or clustered sample designs. The basic idea is to model additional sources of variability in the response by introducing terms into the relationship between mean response and explanatory variables which are random quantities (i.e. subject to a probability distribution) rather than fixed parameters. For example, the inclusion of appropriately specified subject-specific random effects will induce within-subject correlation and also allow for heterogeneity between subjects. Important dissimilarities between individuals, not explicitly measured by explanatory variables, are thus incorporated into the model. Similarly, correlations due to clustered sample design, such as ours, in which all subjects within each household were examined. Generalized additive mixed models (GAMM) combine both the non-linearity and the random effects associated with the repeated measures and sample design.

We used GAM models to evaluate the functional form of the continuous covariates: age, distance from various environment leptospira potential sources, altitude. The spatial patterns were modelled using a bivariate thin plate smoother on the pair of geographical coordinates. As sample cluster and the spatial coordinates belongs to the household, the smoother function and the random effect are related to

the same variables, resulting in a problem of identifiability. Although intra-individual correlation could be fitted to account for the repeated measures design, it is almost superimposed with the coordinates as well. Random effects in this case tend to capture all small scale spatial variability [93] over-smoothing spatial aspects which were the main focus of this study. Therefore our choice privileged the spatial components instead of the random effects. In brief, our modeling strategy was:

- fit a univariate marginal GEE model to pick the variables with a *p-value* below 0.10 to include in the multivariate regression;
- apply a univariate GAMM model to decide which of continuous variables, and respective functional form, should be included in the multivariate models;
- fit three multivariate models: with just individual covariates, with contextual ones and a combined model;
- use Akaike’s corrected information criterion (AIC) to select the model [52].

All statistics analysis was done in the statistical package R, version 2.7.0 [86], and with mgcv library [93] to run the GAM and GAMM models.

8.3 RESULTS

The overall prevalence of *Leptospira* antibodies (MAT titer $\geq 1:25$) was 15.4% [64]. Table 8.3 shows the incidence of leptospiral seroconversion in the longitudinal study. In the first three years, the infection incidence was approximately 3.5% and in the last year the incidence rose to 10.7%.

The incidence in males (6.98%) is 1.69 times larger than in females. For individuals who had contact with 1) mud or flood waters, 2) sewers or 3) trash, the leptospira incidence were 5.45%, 6.90% and 8.31%, respectively. In bivariate analysis, using GEE modeling, male gender (OR = 1.75), black race (OR = 1.56), contact with sewer (OR = 1.45) and trash (OR = 1.84) were all significant odds ratio to leptospiral seroconversion. The only contextual level covariate statistically

significant was the report of more than two rats around the peri-domicile (OR = 1.51) (table 7).

Figure 8 shows the smooth function of continuous explanatory variables via GAMM analysis. Individual income, altitude above sea level of the domicile, the distance to the nearest open sewers and trash collection sites presented a linear relationship with the outcome, therefore they entered the model in the original scale. Only the age effect was non-linear (Figure 8), increasing the risk of leptospiral seroconversion until approximately thirty years and decreasing thereafter.

Table 8 presents the results of the spatial GAMs for the four models: without fixed covariates, just individual covariates, only contextual and both individual and contextual covariates. The AIC, presented in the last line, decreases with the introduction of the covariates in each model. In the individual models, the variables with the largest effects were sex (OR = 1.89) and individual contact with trash in peri-domicile (OR = 1.76). The age, as a smooth term, was significant (p-value < 0.001) in all models. The individual report of more than two rats around the peri-domicile (OR = 1.38) presented the largest effect. But the altitude of the domicile above sea level, the house distance from open trash and from an open sewer demonstrated a protective effect, although only borderline significant. However, those covariates improved the overall AIC, that dropped from 1571.36 to 1401.99.

In figure 8 the white lines depict the significantly higher risk areas and the black lines the lower risk ones considering 95% confidence intervals. The spatial pattern of all the maps is similar, but less patchy as more covariates were included: the highest risk places were similar in all maps, but the significant areas decreased and some smaller patches disappeared in the complete model. The highest risk area, the black spot on the border of the middle valley, was due to a household with two cases that seroconverted twice in cohort. Yearly incidence maps were explored, but the small number of positive case was insufficient to depict meaningful patterns.

8.4 DISCUSSION

This work addressed two main questions: What are the effects of individual and environmental factors on seroconversion? Where are the hot spots of asymptomatic leptospirosis infection?

In this work all individual and contextual covariates were not sufficient to explain the spatial variation on the risk of leptospiral seroconversion. The individual risk factors were quite similar to those associated with severe leptospirosis: sex, age, presence of rats in the peri-domicile [42].

Severe leptospirosis results from a combination of infecting serovars, size of inoculum, individual susceptibility and previous immune state. Flooding is important to severe cases, because as the water goes inside the rats niches, it brings to the surface, and concentrates in the mud afterwards large number of leptospira, where people live. In our study, the borderline significant effect of the altitude from sea level could indicate either places with larger number of rats or an increase in soil contamination due to an interaction between low altitude and the sea tides affecting the soil humidity.

The spatial pattern of leptospiral seroconversion in Pau da Lima is different from the accumulated seropositive individuals. The seroprevalence is more localized in the really low areas, whereas the incidence fluctuates more. One possible reason is the small number of new cases each year. Besides the intervention of the project in the area, year after year, giving advice to people on leptospirosis, could impact on behaviour and change the spatial patterns. All severe cases known in the area, however, were localized in the lowest places.

The identification of prognostic factors for mild or asymptomatic leptospirosis is essential for the development and testing of successful vaccines. However, much more information is needed, particularly relating to prevalent serovars and cross-reactivity between different serovars and the persistence of leptospiral antibodies [46].

Nowadays nonparametric smoothing methods are widely used for

dependent and independent data [37, 93]. A nonparametric regression model assumes that the regression function belongs to a class of smoothing functions with infinite dimensions. Such nonparametric methods would allow the data to determine the regression curve, and the resultant fitted curve is a balance between the goodness of fit and the smoothness of the curve. Our modelling approach could handle both the spatial dependence and the structure of correlation within individuals. In this study GAMs approach should be more efficient because the individuals clusters are always located in their same homes throughout the cohort study. The methods chosen were reliable and allow us to identify areas that inhabitants actions should be focused to reduce the human contact to pathogenic *Leptospira*. Models which are adequate to estimate the impact of both individual and environment variables should be incorporated in the epidemiologists toolbox, such that effective community based interventions can be identified and implemented.

At present, one billion of the worlds population reside in slum settlements, marginalised from basic services and imposing global health challenges [65]. Leptospirosis is expected to become an increasingly important slum health problem as predicted global climate change and growth of the world slum population evolves. Therefore, besides discussing ways to decrease global, environmental measures to minimise the impact over susceptible populations should be undertaken.

Tabela 6: Table of the frequency and the infection, re-infection and new infections incidence of leptospiral seroconversion cohort study in Pau da Lima, Salvador, Bahia, Brazil, 2003-2007.

	Infection		Re-infection		New infection	
	Positive Results	Incidence (per 100 hab.)	Positive Results	Incidence (per 100 hab.)	Positive Results	Incidence (per 100 hab)
All Measures (n = 4816)	257	5.34	144	2.99	113	2.35
1 st Follow-up (n = 1204)	48	3.99	20	1.66	28	3.33
2 nd Follow-up (n = 1204)	42	3.49	22	1.83	20	1.66
3 rd Follow-up (n = 1204)	38	3.16	24	1.99	14	1.16
4 th Follow-up (n = 1204)	129	10.71	78	8.48	51	4.24

Tabela 7: Descriptive analysis of categorical variables available for the univariate analysis for the incidence and bivariate GEE modeling analysis for the odds ratio estimation in leptospiral seroconversion cohort study in Pau da Lima, Salvador, Bahia, Brazil, 2003-2007.

All subjects in all measures	n = 4816	Positive Results	Incidence (per 100 hab)	OR [95% IC]
Individual level				
Gender				
Males	2048	143	6.98	1.75 [1.36; 2.25]
Females (ref.)	2768	114	4.12	1.00 [- ; -]
Race				
Black	1372	97	7.07	1.56 [1.20; 2.03]
Non-black (ref.)	3444	160	4.65	1.00 [- ; -]
Schooling				
Never study	233	11	4.72	1.07 [0.52; 2.21]
Primary school education	3943	216	5.48	1.25 [0.82; 1.91]
High school or more (ref.)	564	25	4.43	1.00 [- ; -]
Contact with mud or flooding in peri-domicile				
Yes	2550	139	5.45	1.05 [0.81; 1.35]
No (ref.)	2266	118	5.21	1.00 [- ; -]
Contact with sewer in peri-domicile				
Yes	1189	82	6.90	1.45 [1.10; 1.90]
No (ref.)	3591	175	4.87	1.00 [- ; -]
Contact with trash in peri-domicile				
Yes	915	76	8.31	1.84 [1.40; 2.44]
No (ref.)	3865	181	4.68	1.00 [- ; -]
Contextual level				
Domestic animals (dogs, cats or chickens)				
Yes	2472	137	5.54	1.09 [0.85; 1.40]
No (ref.)	2344	120	5.12	1.00 [- ; -]
See rats in peri-domicile				
> 2 rats	1658	112	6.76	1.51 [1.17; 1.94]
<= 2 rats (ref.)	1636	88	5.38	1.00 [- ; -]

Figura 7 - GAMM of the association between the risk of leptospiral seroconversion and continuous variables of (A) Individual age (years), (B) Individual income (R\$), (C) Distance in meters to the nearest open sewer, (D) Distance in meters to the trash collection, and (E) Altitude in meters of the domicile sea level. The adjusted odds ratio, in the GAMM model is a measure for the risk of acquiring *Leptospira* antibodies. Solid lines represent the point estimate; dotted lines represent upper and lower 95% confidence band.

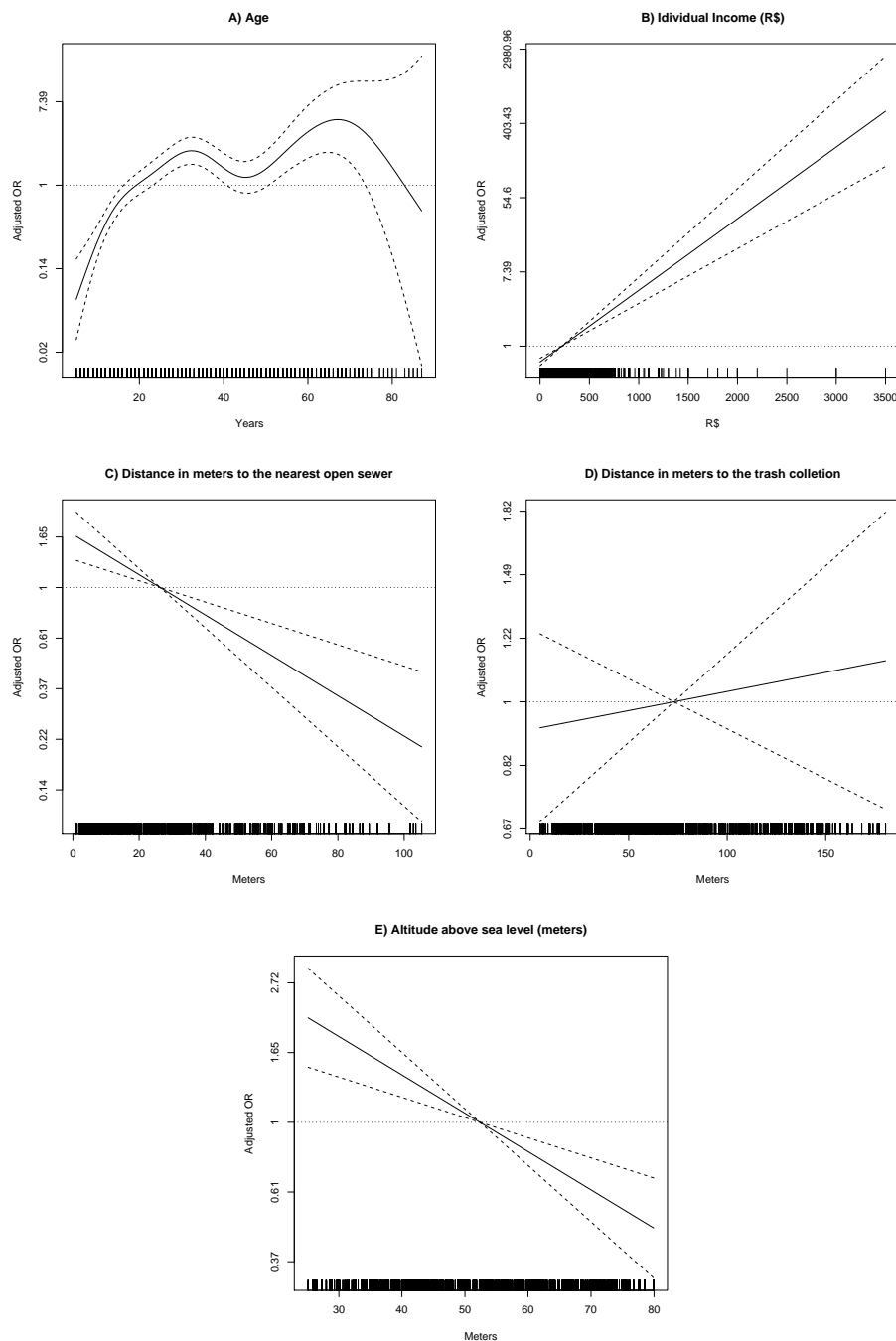
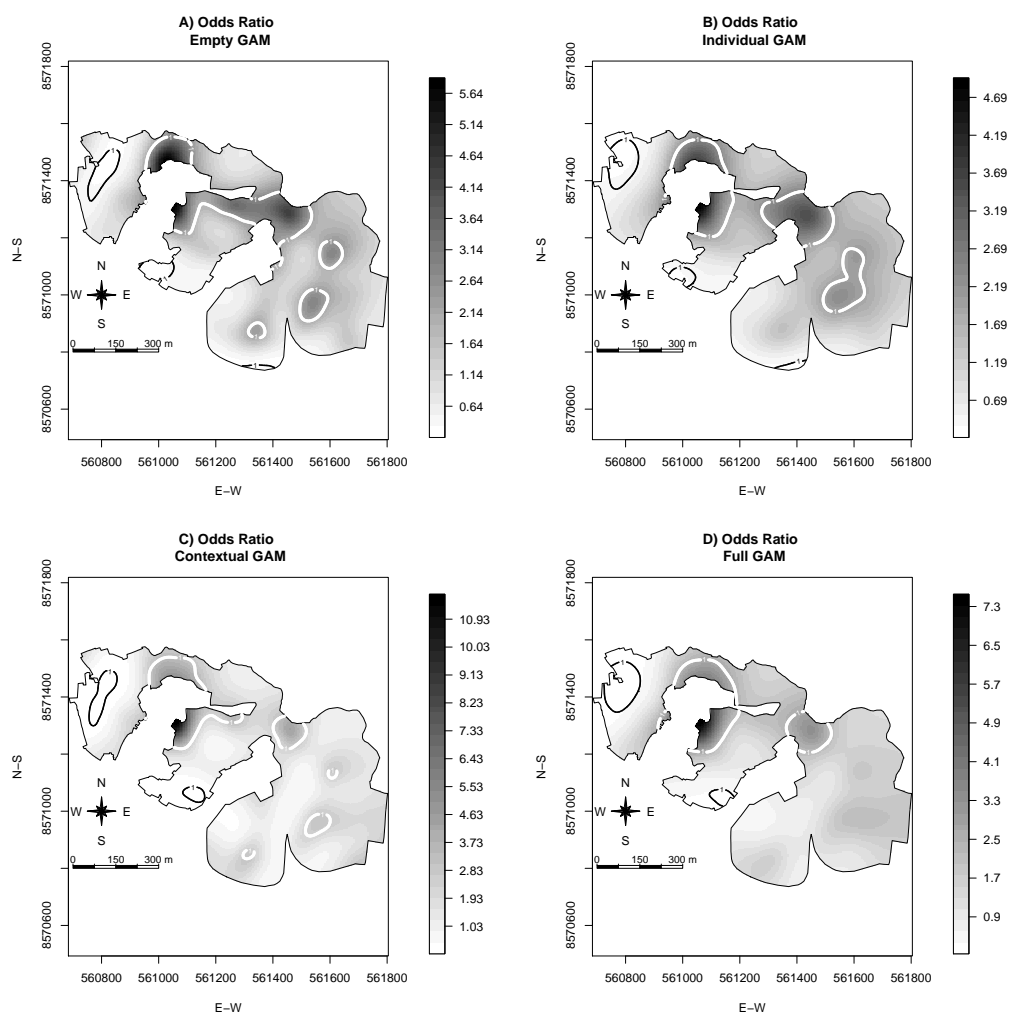


Tabela 8: Odds Ratio and others fit measurements of de spatial logistic regression (GAM) for the of Incidence of Leptospirosis in Pau da Lima, Salvador, Bahia, Brazil, 2003-2007.

Semiparametric Spatial Logistic Models	Odds Ratio [95% Confidence Interval]			
	No Covariates	Individual	Contextual	Full
Individual Variables				
Sex (Male)	–	1.89 [1.39; 2.56]	–	1.70 [1.20; 2.41]
Race (Black)	–	1.20 [0.89; 1.63]	–	–
Income (R\$)	–	1.00 [0.99; 1.00]	–	1.00 [0.99; 1.00]
Contact with sewer in peri-domicile (Yes)	–	0.86 [0.60; 1.23]	–	–
Contact with trash in peri-domicile (Yes)	–	1.76 [1.23; 2.51]	–	1.40 [0.97; 2.01]
Age smooth terms (p-value)	–	< 0.001	–	0.006
Contextual Variables				
See rats (> 2 rats)	–	–	1.38 [1.05; 1.81]	0.99 [0.69; 1.42]
House distance from open sewer (meters)	–	–	0.98 [0.97; 1.00]	0.99 [0.97; 1.00]
House distance from open trash (meters)	–	–	0.99 [0.99; 1.01]	1.00 [0.99; 1.01]
Altitude above sea level (meters)	–	–	1.00 [0.97; 1.02]	0.98 [0.96; 1.02]
Fit Measures				
Spatial smooth terms (p-value)	<0.001	<0.001	<0.001	<0.002
AIC	1935.84	1571.36	1447.29	1401.99

Figura 8 - Risk maps for the adjusted Odds Ratio of logistic spatial regressions (GAMs) for the of incidence of leptospiral seroconversion in Pau da Lima, Salvador, Bahia, Brazil, 2003-2004. Black and white lines represent upper and lower 95% confidence bands, respectively.



9 COMENTÁRIOS FINAIS

- Uma importante limitação para estudos envolvendo análise espacial é o georreferenciamento dos dados. O campo referente ao endereço é de baixa qualidade, com endereços incompletos, erros de digitação e ortografia, ocasionando muitas perdas. No primeiro artigo na primeira etapa do georreferenciamento, utilizando os arquivos e programas disponíveis, só foi possível georreferenciar 64% dos casos notificados residentes na cidade do Rio de Janeiro. Na segunda etapa, porém, utilizando o algoritmo proposto por Skaba (2004) foi possível geocodificar mais 31% dos casos, atingindo portanto 95% (488 indivíduos) das notificações geocodificados com sucesso. Cabe observar que apesar da diferença nos métodos de georreferenciamento, ao fazer as análises estatísticas utilizando os dois conjunto de dados, tanto as localizações dos aglomerados quanto os efeitos das covariáveis foram bastante semelhantes. Isso favorece o uso do dado secundário, mesmo diante da baixa qualidade e das limitações do georreferenciamento. O próprio uso poderá contribuir para a melhoria do dado.
- A invasão de uma população por um agente patogénico pode dar origem a dois regimes: a manutenção da doença após a infecção de um pequeno número de indivíduos da população (endemia), ou então a infecção de uma fração significativa da população durante um período relativamente curto de tempo (epidemia). O limiar que separa a extinção da doença na população em estudo destas duas classes de possíveis evoluções finais chama-se limiar epidémico, ou limiar endêmico se a doença persiste na população. No primeiro artigo com o uso do software SatScan foi possível identificar de maneira satisfatória os

períodos endêmicos, os dois primeiros anos de estudo, e epidêmicos, os demais anos, da leptospirose na cidade do Rio de Janeiro. A técnica utilizada pode trazer grande contribuição para a vigilância epidemiológica, se implantada na rotina das secretarias de saúde.

- No primeiro artigo é observado que o efeito da chuva por si só explicou toda a variabilidade da ocorrência dos agregados de casos notificados por leptospirose na cidade do Rio de Janeiro. Já as variáveis socioeconômicas e ambientais não foram significativas. Ou seja, a leptospirose endêmica e epidêmica acontece nos mesmos grupos populacionais, sendo a epidemia um resultado direto da pluviosidade acima de um determinado ponto de corte, 4mm no caso do Rio de Janeiro. Outros estudos, em outras regiões são necessários para confirmar essa relação entre pluviosidade e casos graves de leptospirose, e identificar o ponto de corte acima do qual a chuva aumenta a concentração de leptospira no solo e conseqüente aumento de casos.
- A utilização do SIG vem se tornando cada vez mais presente nos estudos epidemiológicos, principalmente nos estudos ecológicos, e agora mais recentemente nos inquéritos e também nos estudos de coorte. Sua utilização é de fundamental importância não somente para a visualização das áreas de risco, mas como também para a identificação detalhada de cada área de estudo e principalmente na criação de variáveis de exposição. O SIG fornece informações que não seriam possíveis através dos questionários dos inquéritos. Variáveis como altitude do domicílio em relação ao nível do mar, distância de lixo e esgoto aberto, na maioria das vezes se apresentaram com efeitos significativos nas análises estatísticas deste trabalho, e mesmo quando não sendo significativas a presença delas foi de fundamental importância para a calibração dos modelos estatísticos. Apesar disso, ainda falta uma integração mais sólida entre os profissionais que desenvolvem e operam os SIGs, epidemiologistas e os bioestatísticos. O SIG TerraView mostrou-se inteiramente adequado, com licença livre e código

aberto, o aplicativo geográfico permite a fácil visualização e análise de informações georreferenciadas. Uma de suas vantagens é a integração em um único ambiente das funções do SIG e da estatística (<http://www.dpi.inpe.br/>). Essa integração representa um avanço importante viabilizando o desenvolvimento e a utilização dessas técnicas em prol da Saúde Pública.

- Um dos problemas discutidos entre epidemiologistas é a definição de critério para seleção de variáveis. Este procedimento deve ser feito de forma bastante cuidadosa, o uso indevido de uma variável pode gerar resultados espúrios. É necessário estudar a forma como a variável vai ser analisada, pois frequentemente o efeito não significativo de uma variável deve-se a forma de inclusão no modelo. Dentre as diversas técnicas exploradas neste trabalho optou-se pela utilização de uma função de suavização nas variáveis contínuas para verificar a melhor forma de análise. Uma vantagem do uso de funções não-paramétricas é a possibilidade de observar o efeito da variável em relação ao desfecho em qualquer ponto do plano cartesiano, ou seja, verificar o efeito em vários pontos da escala da variável independente. Em particular a idade sempre deveria ter esse tipo de tratamento, pois raramente seu efeito é constante ao longo de toda a vida.
- A grande maioria das doenças é resultante de uma combinação de fatores que interagem entre si. A determinação da causalidade passa por níveis hierárquicos distintos, sendo que alguns desses fatores causais estão mais próximos do que outros em relação ao desenvolvimento da doença. Desta forma podemos muitas das vezes observar que mesmo após uma seleção criteriosa de variáveis de exposição, as variáveis não dão conta de explicar toda a variabilidade do fenômeno estudado. Em nosso estudo os fatores individuais e ambientais estudados não foram suficientes para explicar toda a variabilidade espacial da soroprevalência e da soroconversão da leptospirose. Provavelmente outros fatores não disponíveis (ex: população e habitat dos ratos, informações sobre o relevo

do peri-domicílio, se é uma área de encosta ou qual o tipo de vegetação, etc.) devem estar envolvidos. Mais inquéritos de soropositividade e estudos longitudinais sobre soroconversão da leptospirose devem ser elaborados para que possamos entender melhor a dinâmica da transmissão da leptospira no meio urbano.

- O último artigo da tese ainda se encontra em fase de amadurecimento, pois os dados da quarta revisita ainda estão em processo de validação laboratorial. Essa confrência é necessária particularmente para verificar o aumento na incidência no último período. Ainda assim, as áreas de maior risco espacial (Figura xxxx) para a soroconversão por leptospirose em Pau da Lima são diferentes daquelas observadas no analisando a soroprevalência (Figura zzz).
- Nessa tese foram exploradas diversas técnicas de modelagem espacial, temporal e longitudinal, sendo possível perceber que não existe o “melhor” modelo mas sim o modelo mais adequado para cada fenômeno que estamos estudando. Muitas das vezes nos deparamos com várias alternativas de modelos a serem utilizados, mas para a escolha do modelo mais adequado sempre utilizamos o princípio da parcimônia e da simplicidade de uso. Na maioria das análises, utilizando diferentes abordagens, os efeitos das covariáveis foram bastante similares. Todos os modelos apropriados para estimar o impacto de variáveis de exposição em diferentes níveis de hierarquia (variáveis individuais, ambientais, climáticas e até no nível molecular), devem ser incorporados na caixa de ferramentas dos epidemiologistas.
- Cabe ressaltar que um aspecto importante deste trabalho é o uso de sistemas operacionais e de softwares de código aberto, disponíveis gratuitamente pela internet. Uma opção de baixo custo no contexto da implementação de métodos de análise que possam ser utilizados por secretarias de saúde.
- Apesar de ser considerada uma doença endêmica no Brasil como um todo e de apresentar manifestações clínicas graves, infelizmente a leptospirose ainda não

é vista como um problema que necessite de estudos e de medidas de intervenção sistemática. No entanto, a inclusão do conceito de “endemicidade” na delimitação das áreas de “risco” e na implementação de políticas de promoção da saúde, prevenção e bem-estar tem sido assunto bastante discutido [92], sendo necessário o desenvolvimento de métodos adequados à implementação da vigilância em saúde com base territorial;

- O impacto da leptospirose é de grande importância social e econômica, pois apresenta elevada incidência em determinadas áreas especialmente na população economicamente ativa, epidemias urbanas cíclicas com elevado custo hospitalar e medidas terapêuticas de alto custo e complexidade. Atualmente, um bilhão de pessoas pelo mundo residem em favelas, marginalizados de serviços básicos, impondo desafios globais de saúde [65]. A leptospirose pode se tornar um problema de saúde cada vez mais importante, sofrendo também o impacto das mudanças climáticas globais e o crescimento das populações em favelas. Portanto, é essencial a necessidade de discutir formas e medidas ambientais, tais como medidas de saneamento e controle dos vetores, para minimizar o impacto sobre tais populações suscetíveis. O estudo da prevalência e da incidência das endemias urbanas, no caso a leptospirose, tem grande complexidade e muito ainda a avançar. A reunião de expertizes oriundas de várias áreas do conhecimento humano (ex: clínicos, epidemiologistas, geógrafos, biólogos, estatísticos, analistas de sistemas, etc.) é indispensável para avançar no conhecimento sobre as doenças e suas relações com a desigualdade social e ambiental assim a contribuir para a criação de medidas eficazes e efetivas no controle de endemias.
- Os modelos estatísticos utilizados nesse trabalho, embora muito flexíveis e adequados, não permitem incorporar aspectos como alterações nos níveis de anticorpos potencialmente indicativas de reinfecções. Um desdobramento do trabalho apresentado nessa tese é implementar modelos de transição de Markov com uma estrutura espacial na modelagem longitudinal da soroconversão em Pau

da Lima. Tal classe de modelo tem como vantagem a estimação das probabilidades de transição entre os distintos estados de saúde (estados de Markov) ao longo do tempo, e conseqüentemente o mapeamento da probabilidade de transição de toda área de estudo. Assim seria possível avaliar o efeito da localização espacial na imunidade.

Referências Bibliográficas

- [1] M. Abreu. *Evolução Urbana do Rio de Janeiro. Rio de Janeiro. IPLANRIO, 1987.*
- [2] E. Abril. *Guia quatro rodas do rio de janeiro. Rio de Janeiro, 2002.*
- [3] P. Acha and B. Szyfres. Zoonosis y enfermedades transmisibles comunes al hombre y a los animales. *Bacteriosis y Micosis. Washington: OPAS, 1, 2001.*
- [4] M. Albuquerque and H. Morais. Decentralization of endemic disease control: an intervention model for combating bancroftian filariasis. *Revista Panamericana de Salud Pública/ Pan American Journal of Public Health, 1:155–163, 1997.*
- [5] R. Assunção, A. Tavares, and M. Costa. Terrastat: Uma biblioteca de algoritmo para análise estatística espacial. In *XVI Simpósio Nacional de Probabilidade e Estatística, Caxambu, 2004.*
- [6] R. Assunção. *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia Associação Brasileira de Estatística, 2001.*
- [7] T. Bailey. Spatial statistical methods in health. *Cadernos Saúde Pública, 17(5):1083–1098, 2001.*
- [8] C. Barcellos and P. Sabrosa. The place behind the case: leptospirosis risks associated environmental conditions in a flood-related outbreak in rio de janeiro. *Cadernos de Saude Publica, 17:1–13, 2001.*
- [9] C. Barcellos and P. Sabroza. Social-environmental determinantes of the leptospirose outbreak of 1996 in western rio de janeiro: a geographical approach.

- international journal of environmental health research. *International Journal of Environmental Health Research*, 10:301–313, 2000.
- [10] R. Bharadwaj, A. Bal, S. Joshi, and et al. An urban outbreak of leptospirosis in mumbai, india. *Japanese Journal of Infectious Diseases*, 55:194–196, 2002.
- [11] A. Bharti, J. Nally, J. Ricaldi, M. Matthias, M. Diaz, and et al. Leptospirosis: A zoonotic disease of global importance. *Lancet Infectious Disease*, 3:757–771, 2003.
- [12] M. Boland, G. Sayers, T. Coleman, and et al. A cluster of leptospirosis cases in canoeists following a competition on the river liffey. *Epidemiology and Infection*, 132:95–200, 2004.
- [13] H. Bolfarine and M. Sandoval. *Introdução à Inferência Estatística*. Sociedade Brasileira de Matemática, 2001.
- [14] K. Brasel and D. Reif. A procedure to generate thiesen polygons. *Geographical Analysis*, 11:289–303, 1979.
- [15] Brasil. *Guia de Vigilância Epidemiológica*. Brasília: Ministério da Saúde, Fundação Nacional da Saúde (FUNASA), 1998.
- [16] E. Caldas, E. Costa, and M. Sampaio. Leptospirosis in salvador (brazil). *Clinical and laboratory aspects*. *Rev Inst Med Trop*, 20:164–176, 1978.
- [17] E. Caldas, M. Sampaio, E. Costa, and G. Miranda. Estudo epidemiológico de um surto de leptospirose ocorrido na cidade do salvador, bahia em maio e junho de 1978. *Rev. Inst. Adolfo Lutz*, 39:85–94, 1979.
- [18] M. Carvalho. *Aplicação de Métodos de Análise Espacial na Caracterização de Áreas de Risco à Saúde*. PhD thesis, Coordenação de Programas em Pós-graduação em Engenharia, Universidade Federal do Rio de Janeiro, 1997.

- [19] M. Carvalho and R. Santos. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cadernos Saúde Pública*, 21(2):361–378, 2005.
- [20] E. Costa, Y. Costa, A. Lopes, E. Sacramento, and J. Bina. Formas graves de leptospirose: aspectos clínicos, demográficos e ambientais. *Rev. Soc. Brás. Méd. Trop.*, 34:1–12, 2001.
- [21] G. Câmara and et al. Terralib: Technology in support of gis innovation. In *II Workshop Brasileiro de Geoinformática, GeoInfo2000. São Paulo*, 2000.
- [22] DCP. Pellegrini. *Análise Espaço-Temporal da Leptospirose no Município do Rio de Janeiro (1995-1999)*. Master thesis, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, 2002.
- [23] P. Diggle. *Statistical Analysis of Spatial Point Patterns*. Number 2nd edition. Edward Arnold, 2002.
- [24] A. Dobson. *An introduction to generalized linear models*. Chapman and Hall, New York, 2 edition, 1990.
- [25] G. Edwards and B. Domm. Human leptospirosis. *Medicine*, 39:117–156, 1960.
- [26] P. Elliott, J. Wakefield, and N. Best. *Spatial Epidemiology : Methods and Applications*. 2000.
- [27] C. Everard, C. Edwards, and D. Carrington. A twelve year study of leptospirosis on barbados. *Eur J Epidemiol*, 11(3):311–320, 1995.
- [28] S. Faine, B. Adler, C. Bolin, and P. Perolat. *Leptospira and Leptospirosis*, volume 2nd Edition. Melbourne, Australia, 2nd edition edition, 1999.
- [29] J. Faraway. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, volume 66. 2005.
- [30] R. Farr. Leptospirosis. *Clinical Infectious Diseases*, 21:1–8, 1995.

- [31] FIBGE. Demographic census of rio de janeiro - brazil. Technical report, Foundation Brazilian Institute of Geography and Statistics, 2001.
- [32] FUNASA. Epidemiologist surveillance guide (in portuguese). Technical report, Health department, National Foundation of the Health, Brazil, Brazil, 2005.
- [33] GEORIO. *Rainfall Data*, 2004.
- [34] S. Gilboa, A. Correa, and C. Alverson. Use of spline regression in an analysis of maternal prepregnancy body mass index and adverse birth outcomes: Does it tell us more than we already know ? *Annals of Epidemiology*, 18(3):196–205, 2007.
- [35] S. Greenland and J. Robins. Ecologic studies: biases, misconceptions and counterexamples. *American Journal Epidemiology*, 61:139–147, 1994.
- [36] D. Haake, M. Dundoo, R. Cader, B. Kubak, and R. Hartskeerl. Leptospirosis, water sports, and chemoprophylaxis. *Clinical Infectious Diseases*, 34:e40–e43, 2002.
- [37] T. Hastie and R. Tibshirani. *Generalised Additive Models*. 1990.
- [38] P. Hunter. Climate change and waterborne and vectorborne disease. *Journal of Applied Microbiology*, 94:37–46, 2003.
- [39] IPLAN. Anuário estatístico da cidade do rio de janeiro. Technical report, 1993.
- [40] A. Jansen, I. Schoneberg, C. Frank, and et al. Leptospirosis in germany 1962–2003. *Emerging Infectious Diseases*, 11:1048–1054, 2005.
- [41] R. Kariv, R. Klempfner, A. Barnea, Y. Sidi, and E. Schwartz. The changing epidemiology of leptospirosis in israel. *Emerging Infectious Diseases*, 7:990–992, 2001.
- [42] A. Ko, M. Reis, C. Dourado, W. Johnson, and L. Riley. Urban epidemic of severe leptospirosis in brazil. *The Lancet*, 353:820–825, 1999.

- [43] M. Kulldorf and N. Nagarwalla. Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14:799–810, 1995.
- [44] E. Kupek, M. Faversoni, and J. Philippi. The relationship between rainfall and human leptospirosis in Florianópolis, Brazil 1991–1996. *The Brazilian Journal of Infectious Diseases*, 4:131–134, 2001.
- [45] LABGEO. Sistema de localização (sisloc). Technical report, Departamento de Informações para a Saúde, Centro de Informação Científica e Tecnológica, Fundação Oswaldo Cruz, 2001.
- [46] P. Levett. Leptospirosis. *Clinical Microbiology Reviews*, 13:296–326, 2001.
- [47] K. Liang and S. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.
- [48] A. Lomar, D. Diament, and J. Torres. Leptospirosis in Latin America. Emerging and re-emerging diseases in Latin America. *Infect Dis Clin N.A.*, 14(1):23–38, 2000.
- [49] E. Maciel, A. Carvalho, S. Nascimento, R. Matos, E. Gouveia, M. Reis, and A. Ko. Household transmission of leptospirosis infection in urban slum communities. *PLoS Neglected Tropical Diseases*, 2(1):e154, January 2008.
- [50] N. Mantel. The detection of disease clustering and a generalised regression approach. *Cancer Research*, 27:209–220, 1967.
- [51] A. McBride, D. Athanazio, M. Reis, and A. Ko. Leptospirosis. *Current Opinion in Infectious Diseases*, 18:376–386, 2005.
- [52] C. McCulloch and S. Searle. *Generalized, Linear, and Mixed Models*. 2000.
- [53] E. Meites and M.T. Jay and S. Deresindki and W.J. Shieh and S.R. Zaki and L. Tompkins and D.S. Smith. Reemerging leptospirosis, California. *Emerging Infectious Diseases*, 10:406–412, 2004.

- [54] J. Mills and J. Childs. Ecologic studies of rodent reservoirs: their relevance for human health. *Emerging Infectious Diseases*, 4:529–537, 1998.
- [55] F. Mittelbach. *The LaTeX Companion*. Addison-Wesley Professional, 2004.
- [56] J. Morgan, S. Bornstein, and A. Karpati. Outbreak of leptospirosis among triathlon participants and community residents in springfield, illinois, 1998. *Clinical Infectious Diseases*, 34:1593–1599, 2002.
- [57] F. Mérien, P. Amouriaux, P. Perolat, G. Baranton, and G. Saint. Polymerase chain reaction for detection of leptospira spp. in clinical. *Journal of Clinical Microbiology*, 30:2219–2224, 1992.
- [58] F. Mérien and P. Perolat. Public health importance of human leptospirosis in the south pacific: a five-year study in new caledonia. *American Journal of Tropical Medicine and Hygiene*, 55:174–178, 1996.
- [59] M. Navarro, A. Filgueiras, H. Asensi, E. Lemos, M. Sidoni, M. Soares, and T. Cardoso. *Doenças Emergentes e Reemergentes, Saúde e Ambiente*. 2002.
- [60] J. Nelder and R. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society A*, 135:370–84, 1972.
- [61] A. Okabe, B. Boots, K. Sugihara, and S. Chiu. *Spatial Tessellations - Concepts and Applications of Voronoi Diagrams*. Number 2nd edition. John Wiley, 2000.
- [62] J. Pinheiro and D. Bates. *Mixed-Effects Models in S and SPLUS*. 2000.
- [63] M. Pinheiro, M. Araújo, E. Branco, A. Barata, R. Dantas, and W. Fernandes. Malária durante a gestação e baixo peso ao nascer em áreas endêmicas da amazônia. *Revista Paraense de Medicina*, 16, 2002.
- [64] R. Reis, G. Ribeiro, R. Felzemburgh, F. Santana, S. Mohr, A. Melendez, A. Queiroz, A. Santos, R. Ravines, W. Tassinari, M. Carvalho, M. Reis, and

- A. Ko. Impact of environment and social gradient on leptospira infection in urban slums. *PLoS Neglected Tropical Diseases*, 2(4):e228, April 2008.
- [65] L. Riley, A. Ko, A. Unger, and M. Reis. Slum health: Diseases of neglected populations. *BMC International Health Humans and Rights*, 7:1–6, 2007.
- [66] L. Roberts, U. Confalonieri, and J. Aaron. *Ecosystem Change and Public Health: A Global Perspective*. Johns Hopkins University Press, Baltimore, 2001.
- [67] E. Romero, C. Bernardo, and P. Yasuda. Human leptospirosis: a twenty-nine-year serological study in sao paulo, brazil 45, 245–248. *Revista do Instituto de Medicina Tropical de São Paulo*, 45:245–248, 2003.
- [68] P. Rosenberg, H. Katki, C. Swanson, L. Brown, S. Wacholder, and R. Hoover. Quantifying epidemiologic risk factors using non-parametric regression: Model selection remains the greatest challenge. *Statistics in Medicine*, 22:3369–3381, 2003.
- [69] B. Rowlingson, P. Diggle, adapted, packaged for R by Roger Bivand, pcp functions by Giovanni Petris, and goodness of fit by Stephen Eglen. *splanacs: Spatial and Space-Time Point Pattern Analysis*, 2008. R package version 2.01-24.
- [70] S. Santos and C. Barcellos. Space approaches in the public health. health department, osvaldo cruz foundation, brazil (in portuguese). available at: <http://portal.saude.gov.br/saude/> (accessed 9 april 2007).
- [71] U. Sarkar, S. Nascimento, R. Barbosa, R. Martins, H. Nuevo, I. Kalafanos, I. Grunstein, B. Flannery, J. Dias, L. Riley, M. Reis, and A. Ko. Population-based case-control investigation of risk factors for leptospirosis during na urban epidemic. *American Journal of Tropical Medicine and Hygiene*, 66:605–6100, 2002.
- [72] D. Sasaki, L. Pang, and H. Minette. Active surveillance and risk factors for

- leptospirosis in hawaii. *American Journal of Tropical Medicine and Hygiene*, 48:35–43, 1993.
- [73] E. Sclar, P. G. P, and G. Carolini. The 21st century health challenge of slums and cities. *The Lancet*, 365:901–903, 2005.
- [74] SECOMP. Mapamento da pobreza em areas urbanas do estado da bahia. Cd-rom, Secretaria de Combate a Pobreza e as Desigualdades Sociais, 2005.
- [75] H. Silva, J. Tavares-Neto, J. Bina, and R. Meyer. Leptospirose-infecção e forma subclínica em crianças de salvador, bahia. *Revista da Sociedade Brasileira de Medicina Tropical*, 36(2), 2003.
- [76] D. Skaba, M. Carvalho, C. Barcellos, P. Martins, and S. Terron. Geoprocessamento dos dados da saúde:o tratamento dos endereços. *Cadernos de Saúde Pública*, 20(6):1753–1756, 2004.
- [77] L. Smythe. Leptospirosis worldwide. *Weekly Epidemiological*, 74:237–242, 1999.
- [78] G. Snedecor and W. Cochran. *Statistical methods*. 7th edition, 1982.
- [79] T. Snijders and R. Bosker. *Multilevel Analysis*. 1999.
- [80] M. Susser. The logic in ecological: I. the logic of analysis. *American Journal of Public Health*, 84(5):825–829, 1994.
- [81] M. Susser. The logic in ecological: Ii. the logic of design. *American Journal of Public Health*, 84(5):830–835, 1994.
- [82] W. Tangkanakul, H. Smits, S. Jatanasen, and D. Ashford. Leptospirosis: an emerging health problem in thailand. *Southeast Asian Journal of Tropical Medicine & Public Health*, 36:281–288, 2005.
- [83] W. Tassinari, D. Pellegrini, M. Carvalho, and P. Sabroza. Distribuição espacial da leptospirose no município do rio de janeiro ao longo dos anos de 1996-1999. *Cadernos de Saúde Pública*, 20(6):1721–1729, 2004.

- [84] W. Tassinari, D. Pellegrini, C. Sa, R. Reis, A. Ko, and M. Carvalho. Detection and modelling of case clusters for urban leptospirosis. *Tropical Medicine and International Health*, 13(4):1–10, 2008.
- [85] P. Tauil. Aspectos críticos do controle do dengue no brasil. *Cadernos de Saúde Pública*, 18(3):867–871, 2002.
- [86] R. D. C. Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.
- [87] R. H. W. Terpstra. Leptospirosis in wild animals. *The Veterinary Quarterly*, 18:149–150, 1996.
- [88] UN-HABITAT. *Slums of the world: The face of urban poverty in the new millennium*. (United Nations Human Settlements Programme), Nairobi, 2003.
- [89] Vado-Solís, M. Cárdenas-Marrufo, B. Jiménez-Delgadillo, A. Alzina-López, H. Laviada-Molina, V. Suarez-Solís, and J. Zavala-Velazquez. Clinical-epidemiological study of leptospirosis in humans and reservoirs in yucatán, méxico. *Revista do Instituto de Medicina Tropical de São Paulo*, 44(6):335–340, 2002.
- [90] J. Vinetz. Leptospirosis. *Current Opinion in Infectious Diseases*, 14:527–538, 2001.
- [91] J. Watson, G. Michelle, and M. C. MA. Epidemics after natural disasters. emerging infectious diseases. *Emerging Infectious Diseases*, 13:001–005, 2007.
- [92] WHO. *Human leptospirosis: guidance for diagnosis surveillance and control*. World Health Organization, Geneva, 2003.
- [93] S. Wood. *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC Texts in Statistical Science Series, 2006.

- [94] V. Wuthiekanun, N. Sirisukkarn, P. Daengsupa, and et al. Clinical diagnosis and geographic distribution of leptospirosis, thailand. *Emerging Infectious Disease*, 13:124–126, 2007.
- [95] C. Yersin. Human leptospirosis in 1994 in the seychelles. *Medical and Dental Journal*, 4:8–9, 1996.
- [96] C. Yersin, P. Bovet, T. Merien, J. Wong, Panowsky, and Perolat. Human leptospirosis in the seychelles (indian ocean): a population- based study. *Am J Trop Med Hyg*, 59(6):933–940, 1998.