

Inundações e leptospirose nos municípios brasileiros no período de 2003 a 2013: utilização de técnicas de mineração de dados

Floods and leptospirosis in Brazilian municipalities from 2003 to 2013: use of data mining techniques

Inundaciones y leptospirosis en municipios brasileños durante el período de 2003 a 2013: utilización de técnicas de mineración de datos

Renata Gracie ¹
Diego Ricardo Xavier ¹
Roberto Medronho ^{2,3}

doi: 10.1590/0102-311X00100119

Resumo

A leptospirose se relaciona a problemas de saneamento ambiental, com incremento de casos em períodos de inundações. Levando-se em consideração as questões relacionadas a mudanças climáticas, as inundações tendem a um aumento. As inundações não atingem as populações de maneira homogênea, em geral os menos favorecidos em termos socioeconômicos são os mais acometidos. Para saber se o número de inundações aumentaria a incidência de leptospirose e sua relação com as variáveis contextuais, utilizou-se dados socioeconômicos, ambientais e de ocorrência da doença no nível municipal. Os municípios que tinham problemas no esgotamento sanitário apresentaram maior risco para a ocorrência da leptospirose. O total de inundações adquirida a partir da decretação pela autoridade municipal constituiu um importante marcador de risco para a ocorrência de leptospirose. A modelagem de árvore de regressão mostrou-se útil para estimar a ocorrência de leptospirose no Brasil.

Leptospirose; Inundações; Fossas Secas; Árvores de Decisões; Mudança Climática

Correspondência

R. Gracie
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz.
Av. Brasil 4365, Rio de Janeiro, RJ 21045-900, Brasil.
gracie.renata@gmail.com

¹ Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

² Faculdade de Medicina, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.

³ Instituto de Estudos de Saúde Coletiva, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil.



Introdução

A leptospirose tem ocorrência em quase toda a superfície terrestre, com exceção das calotas polares. Em países com desenvolvimento consolidado, a ocorrência desse agravo acontece devido à exposição durante o exercício e nas atividades de trabalho ^{1,2}. Nos países não desenvolvidos, a ocorrência da leptospirose está vinculada às relações de trabalho e, principalmente, a problemas de saneamento ambiental ³. O aumento da ocorrência de chuva e inundações favorece a possibilidade de transmissão em quase todos os países ⁴.

As mudanças climáticas são um fato que diversos pesquisadores já afirmam sua existência ^{5,6,7} e a partir deste fenômeno em muitas regiões do mundo identifica-se um aumento da ocorrência de inundações. Contudo, os eventos extremos de ocorrência de inundações não atingem as populações de maneira homogênea ^{8,9}.

As populações com condições socioeconômicas menos favorecidas no tocante ao acesso a saneamento ambiental, nível de instrução e de renda inferiores tendem a sofrer mais os efeitos negativos dessas mudanças e uma maior exposição aos riscos à saúde ¹⁰.

Em alguns lugares, esses eventos poderiam ser minimizados se a construção dos espaços urbanos ocorresse de maneira mais planejada. Assim, devemos manter o monitoramento e vigilância nos eventos extremos, bem como o diagnóstico das situações ambientais que geram maiores riscos para inundações ¹¹.

Em relação à gerência e organização de informações para a tomada de decisão, de uma maneira geral nos municípios mais populosos identifica-se um nível de complexidade e capacidade técnica maior do que nos municípios menores. Isso ocorre devido aos diferentes arranjos e concentrações populacionais dos municípios brasileiros ¹², e reflete-se na gestão, captação e sistematização das informações de diferentes instituições públicas, de gerência e da tomada de decisão tanto na área ambiental quanto na saúde pública. Isso acontece por causa do gasto financeiro para estruturação e sistematização de dados de interesse para saúde ¹³. Esse diferencial pode ter impactos nas decretações de estado de calamidade em períodos de eventos extremos, como no total de inundações.

Diante do contexto de diferentes informações com diferentes origens analisadas em territórios estruturados de diversas maneiras, podemos ressaltar que os usos da combinação de técnicas de geoprocessamento com métodos estatísticos podem ser úteis para a identificação das situações de risco, apontando de forma mais adequada a locação de recursos públicos para a melhora das condições de vida e de saúde das populações ¹⁴. As técnicas de mineração de dados podem auxiliar na busca por relações entre as variáveis contextuais e de desfecho, apontando quais os limiares socioeconômicos apresentam aumento de risco ¹⁵.

O objetivo deste trabalho foi analisar a ocorrência de leptospirose e sua associação com a ocorrência de inundações, variáveis socioeconômicas e ambientais no nível municipal.

Método

Estudo ecológico nos municípios do Brasil. Estruturou-se o banco de dados com base nessas unidades de análises com informações sobre a ocorrência de leptospirose, total de inundações e dados socioeconômicos e ambientais no período de 2003 a 2013.

Fonte de dados

Utilizou-se os casos notificados de leptospirose do banco do Sistema de Informação de Agravos de Notificação (SINAN), no período de 2003 a 2013, sistematizados pelo Departamento de Informática do SUS (DATASUS) para a construção de taxas de incidência acumulada da doença.

Estima-se que haja subnotificação da leptospirose devido a problemas no sistema de vigilância epidemiológica e ao fato do aumento da ocorrência da doença, em sua grande maioria, coincidir com o aumento de dengue, Zika e chikungunya, que apresentam sintomas semelhantes aos da leptospirose ¹⁶. Para tentar diminuir a flutuação da taxa de incidência acumulada da leptospirose, utilizou-se o método de suavização bayesiano local calculado no software Terra View 4.2.2 (<http://www.dpi.inpe.br/>)

terraview). Após a aplicação do método de suavização da taxa optou-se por retirar os municípios que mesmo após a suavização mantiveram valor zero, com o objetivo de controlar a superdispersão dos dados na variável contextual.

Os totais de inundações foram adquiridos a partir dos decretos de situação de emergência e estados de calamidade pública obtidos no Sistema Nacional de Proteção e Defesa Civil (SINPDEC), que organiza os dados do Formulário de Avaliação de Danos (AVADAN) e do Formulário de Informações do Desastre (FIDE). Esses documentos são enviados ao Ministério do Desenvolvimento Regional, que reconhece o desastre ¹⁷.

Os dados socioeconômicos e ambientais, tais como: renda, escolaridade, esgotamento sanitário, abastecimento de água, coleta de lixo, altitude da sede do município, população, proporção da população residente em área de favela, índice de Gini e PIB *per capita* foram adquiridos do censo de 2010 do Instituto Brasileiro de Geografia e Estatísticas (IBGE) e do IPEADATA, base de dados do Instituto de Pesquisa Econômica Aplicada (IPEA). Os dados de cobertura da população atendida pela Estratégia Saúde da Família (ESF) foram obtidos no DATASUS e o total de chefes de família que recebem benefício do Programa Bolsa Família no Ministério da Cidadania. Para a confecção dos mapas foram utilizadas as bases gráficas dos municípios do Brasil do IBGE do ano de 2013 (*datum SIRGAS 2000*).

Devido à heterogeneidade das unidades de análise, optou-se pela estratificação dos municípios segundo o tamanho da população: 0-5.000, 5.001-10.000, 10.001-50.000, 50.001-100.000, ≥ 100.001 ¹². Do ponto de vista metodológico, a identificação dos arranjos e das concentrações de população tem como objetivo sistematizar as informações estatísticas, provenientes das bases de dados de diferentes instituições, como o IBGE, e identificar as diversas estruturas territoriais que apontam para o reconhecimento de vulnerabilidades e riscos diferenciados ¹².

Análise de dados

• Teste de correlação não paramétrico

Devido aos dados socioeconômicos não apresentarem distribuição normal, usou-se o teste de correlação não paramétrico de Spearman, no software SPSS 17.0 (<https://www.ibm.com/>), para identificar a correlação entre a taxa de incidência acumulada de leptospirose ajustada pelo método bayesiano local no período de 2003 a 2013 e as seguintes variáveis socioeconômicas e ambientais: total de inundações por município no período de 2003 a 2013; categoria de cidades pelo tamanho da população; altitude da sede do município; proporção de domicílios com: abastecimento de água rede geral, abastecimento de água poço, abastecimento de água poço fora do domicílio, abastecimento de água captada rio e lago, abastecimento de água poço na aldeia, abastecimento de água poço fora da aldeia, abastecimento de água carro pipa, abastecimento de água por chuva da cisterna, abastecimento de água chuva e abastecimento de água outra forma; valor esperado de precipitação dezembro a fevereiro; valor esperado de precipitação junho a agosto; valor esperado de precipitação março a maio; valor esperado de precipitação setembro a novembro; proporção de domicílios com: esgotamento rede geral, esgotamento fossa, esgotamento fossa seca, esgotamento em vala, esgotamento em rio, esgotamento outro, esgotamento não definido; índice de Gini; proporção de domicílios com: coleta direta de lixo, coleta de lixo caçamba, lixo queimado, lixo enterrado, lixo jogado no rio, lixo jogado em terreno baldio ou em logradouro, coleta de lixo com outro fim; proporção de população atendida pela ESF; proporção de chefes de família com pelo menos 10 anos de estudos; proporção de chefes de família com benefício do Bolsa Família.

Essas variáveis contextuais foram selecionadas de acordo com levantamentos de trabalhos anteriores e disponibilidade do dado em âmbito nacional e escala municipal. As variáveis contextuais com correlação significativa no teste de Spearman foram analisadas utilizando-se técnicas de mineração de dados no software R (<http://www.r-project.org>) denominada árvore de regressão, que é formada por um conjunto de nós de decisão ¹⁸. A análise de regressão múltipla permitida pela árvore de regressão consiste em prever o valor de uma variável de desfecho em função das variáveis contextuais utilizadas.

- **Árvore de regressão com inferência condicional**

As árvores de regressão apresentam três níveis primários de resultados consistindo em um nó raiz, nós internos e nós terminais (folhas). Cada árvore começa com um “nó raiz” que contém a amostra de dados com base na qual a árvore será cultivada. Os dados então são divididos em dois “nós filhos” baseando-se no valor da variável contextual que melhor atende a algum critério de particionamento. Os nós secundários resultantes contêm um subconjunto dos dados originais. Cada nó filho pode ser ainda particionado, novamente com base no valor de uma variável contextual. Esse processo continua até que não exista mais partições ou algum conjunto de critérios de particionamento já não seja mais atendido, resultando em nós terminais. O particionamento recursivo é uma abordagem de regressão não paramétrica na forma de agrupamento hierárquico em que os dados são divididos sequencialmente em grupos dicotômicos, de modo que cada grupo resultante apresenta como resposta valores ajustados próximos da variável de desfecho, configurando o valor estimado^{19,20}. Essa abordagem é eficaz para identificar interações complexas entre as variáveis contextuais sem especificação *a priori* de termos de interação. Além disso, comportam um elevado número de variáveis contextuais em relação ao número de observações.

A árvore de regressão com inferência condicional utiliza o critério de particionamento baseado na significância estatística e avalia as relações condicionais entre as variáveis contextuais em busca de grupos homogêneos que descrevam o comportamento da variável de desfecho. O primeiro passo do algoritmo considera a hipótese nula global de independência entre as variáveis contextuais e a variável de desfecho, se esta hipótese não puder ser rejeitada, o particionamento para. Se a hipótese global nula for rejeitada, então a variável contextual mais significativa no modelo condicional em relação a outras variáveis contextuais é selecionada. Quando a variável selecionada é dicotômica, a escolha da melhor divisão binária é trivial; para as variáveis não dicotômicas, o algoritmo identifica a melhor divisão binária de todas as divisões possíveis²¹. As árvores condicionais são baseadas na inferência estatística, a poda não é necessária, como ocorre na abordagem usada nos algoritmos de partição recursiva proposta por Breiman et al.²².

A estatística para o particionamento é baseada na teoria assintótica das estatísticas de permutação proposta por Strasser & Weber²³. Inicialmente, são criadas as partições possíveis com base em todas as variáveis contextuais e classificadas as unidades de análises, depois é aplicado o teste estatístico para a seleção dos grupos que irão compor a árvore. Foi utilizado o teste de Bonferroni para a avaliação de independência e a comparação do valor de p e seleção dos nós e $\alpha \leq 0,05$ (nível de significância adotado para usar como referência para o valor de p) para a inclusão do particionamento das variáveis contextuais. Com base nesses resultados foram criados os grupos e encontrados os valores preditos resultantes dos nós terminais. Neste estudo, foi utilizada a função *ctree* do pacote (ou biblioteca) do R denominada *partykit*, versão 3.4.0, disponível em <http://cran.r-project.org/web/packages/party/index.html>²⁴.

Baseado no método descrito foram construídas duas árvores de regressão com inferência condicional, a primeira foi estruturada com caráter exploratório considerando como variável de desfecho a taxa de incidência acumulada de leptospirose e suavizada pelo método bayesiano local, e as variáveis contextuais o total de inundações por município no período de 2003 a 2013 e a classificação dos municípios brasileiros de acordo com o total de população segundo o IBGE (categoria de municípios). Essa análise buscou avaliar a heterogeneidade da distribuição da variável de desfecho segundo a variável categoria de municípios e o comportamento da variável total de inundações.

A segunda árvore buscou classificar a variável de desfecho segundo os diferentes níveis de resposta das variáveis contextuais selecionadas pelo teste de Spearman, e com isto verificar o efeito destas variáveis contextuais para a ocorrência de leptospirose. Essa regressão com inferência condicional permitiu produzir um valor estimado para a ocorrência de leptospirose. Ao final, calculou-se o erro quadrático médio do modelo. O resultado preditivo junto com a taxa de incidência acumulada bruta e a taxa de incidência acumulada de leptospirose ajustada pelo método bayesiano local foi processado em ambiente do Sistema de Informações Geográficas (SIG) para a construção dos mapas finais.

Resultados

A Tabela 1 mostra a distribuição de casos de leptospirose e de inundações segundo tamanho populacional dos municípios.

No primeiro grupo dos municípios com até 5.000 habitantes, 38,9% destes apresentaram inundações e a taxa de incidência de leptospirose nos municípios com inundações foi 77% maior do que nos municípios sem inundações (RT = 1,77). No segundo grupo com população entre 5.001 e 10.000, 47,1% dos municípios apresentaram inundações e a taxa de incidência de leptospirose foi 36% maior nestes municípios em relação aos que não apresentaram inundações (RT = 1,36). O terceiro grupo com população entre 10.001 e 50.000 concentra quase a metade dos municípios brasileiros com um total de 2.459. Nesse grupo, observou-se que a taxa de incidência de leptospirose foi 80% maior nos municípios com inundações em relação aos sem inundações (RT = 1,80). No grupo com população entre 50.001 e 100.000, a taxa de incidência de leptospirose nos municípios com inundações foi mais de duas vezes maior que nos municípios sem inundações (RT = 2,14). O grupo que tem ≥ 100.001 , embora apresente o menor número de unidades de análise, concentra a maior parte da população. Também observa-se nesse grupo que os municípios com inundações apresentam taxa de incidência da doença 72% maior que os sem inundações (RT = 1,72). Observa-se também que a proporção de municípios com inundações aumenta com o aumento do tamanho da população, atingindo um platô em torno de 62% naqueles com população maior que 50.000 habitantes.

Na Figura 1, verifica-se que na medida em que aumenta o número do total de inundações nos municípios com inundações, as taxas de incidência de leptospirose suavizadas também aumentam.

A Figura 2 mostra a primeira árvore de regressão, que apresenta 6 partições (em que se lê taxa $p < 0.001$) e 7 nós terminais (linhas que apontam para os gráficos da parte inferior da figura). Essa árvore foi construída para a análise exploratória da dispersão da taxa de incidência suavizada em função do número de inundações representadas pelos gráficos de *boxplot* e das categorias de cidades, segundo o tamanho populacional, representadas pelos gráficos de barras que estão numeradas de 1 a 5 em que cada número representa as diferentes categorias de cidade (1 – cidades com população até 5.000 habitantes, 2 – cidades com população entre 5.001 e 10.000, 3 – cidades entre 10.001 e 50.000, 4 – cidades entre 50.001 e 100.000 e 5 – cidades com mais de 100.000). No gráfico de barras verifica-se que a distribuição do agravo ocorre de forma heterogênea segundo a categoria de cidades representada pelos números de 1 a 5. É possível observar nos gráficos de *boxplot* abaixo da árvore de regressão que o número de inundações apresenta relação direta com o comportamento da taxa de leptospirose.

Tabela 1

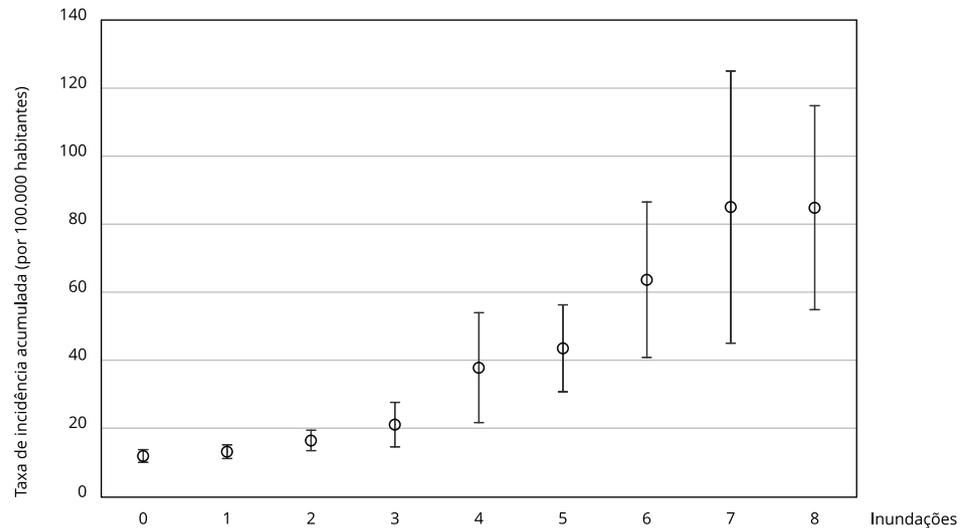
Distribuição de casos de leptospirose, de inundações e taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local atribuída às inundações segundo tamanho populacional dos municípios brasileiros, no período entre 2003 e 2013.

Indicadores	Tipos de municípios segundo tamanho da população				
	0-5.000	5.001-10.000	10.001-50.000	50.001-100.000	≥ 100.001
Total de municípios	1.247	1.226	2.459	340	298
Total de municípios sem inundações	761	648	1.119	127	113
Total de municípios com inundações	486	578	1.340	213	185
Municípios com inundações (%)	38,9	47,1	54,4	62,6	62,0
Taxa média de incidência acumulada de leptospirose ajustada pelo método bayesiana local com inundações	31,31	19,87	21,02	17,59	25,18
Taxa média de incidência acumulada de leptospirose ajustada pelo método bayesiana local sem inundações	17,65	14,65	11,68	8,23	14,64
Razão de taxas de incidência *	1,77	1,36	1,80	2,14	1,72

* Taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local com inundações/taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local sem inundações.

Figura 1

Taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local, segundo o número do total de inundações no período de 2003 a 2013.



A primeira partição considerou os municípios com taxa de incidência maior que 33,69 casos por 100.000 habitantes e taxa igual ou menor que 33,69 casos por 100.000 habitantes. O nó que está à extrema direita representa um grupo com 534 municípios com taxa de incidência superior a 33,69 casos por 100.000 habitantes, e 40% destes estão na categoria de município com população de 10.000 a 50.000 habitantes.

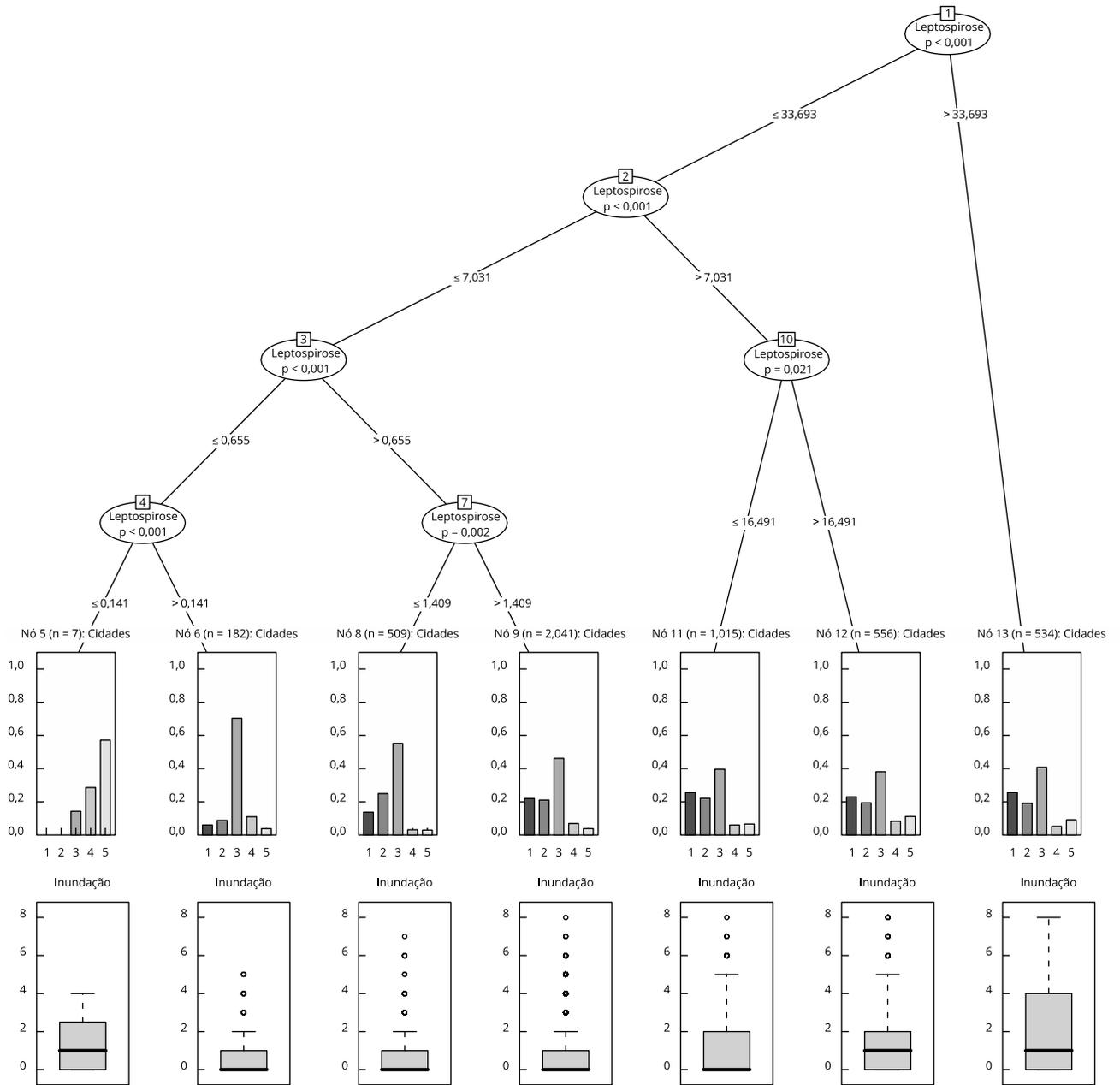
De maneira geral, o aumento do número de inundações expressou o aumento nas taxas de incidência suavizada pelo método bayesiano local. O comportamento da taxa em relação à categoria populacional e o número de inundações mostram que não há relação direta entre o porte populacional de categoria da cidade, total de inundações e incremento da taxa. As menores taxas de incidência foram observadas no nó da extrema esquerda (total de 7 municípios) que é o grupo em que estão os municípios com as maiores populações e as menores taxas ($\leq 0,141$), com mediana de pelo menos uma inundações.

Os dois maiores grupos classificados em termos de números de municípios têm 1.657 (nó 10) e 1.179 (nó 11) municípios, com taxa suavizada, respectivamente, acima de 5.048 e menor ou igual a 5.048. Com relação ao número de ocorrências de inundações no nó 10, observa-se o máximo de 7 ocorrências e no nó 11, 8 ocorrências.

Para a segunda árvore de regressão com inferência condicional (Figura 3) utilizou-se as seguintes variáveis contextuais que apresentaram correlação significativa com a taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local pelo teste de correlação de Spearman: proporção de domicílios abastecidos de água por chuva da cisterna, valor esperado de precipitação média de junho a agosto, PIB *per capita*, proporção de domicílios com esgotamento por fossa, proporção de domicílios com coleta de lixo jogado em terreno baldio ou em logradouro, proporção de chefes de família com benefício do Programa Bolsa Família, proporção de população atendida pela ESF, proporção de chefes de família com pelo menos dez anos de estudos, número de inundações, categoria de cidade (0-5.000, 5.001-10.000, 10.001-50.000, 50.001-100.000 e ≥ 100.001), proporção de população residente em área de favela em 2010.

Figura 2

Árvore de regressão com a taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local e as variáveis contextuais do total de inundações e categoria de cidade.



A árvore apresentou 16 partições e 17 nós terminais. A principal variável contextual dentre as variáveis de entrada no modelo foi a proporção de domicílios com esgotamento por fossa. O grupo de 125 municípios (nó 33) foi o conjunto de unidades de análise que apresentou a maior taxa estimada de leptospirose, 93,33 (IC95%: 81,08-105,58) casos por 100.000 habitantes, e ocorreram em municípios com população acima de 31,09% das habitações com fossa e mais de três eventos de inundações. A segunda variável contextual com maior associação à variável de desfecho foi a de total de inundações. O grupo de 52 municípios contidos no nó 30 e que apresentou proporção de domicílios com esgotamento por fossa menor ou igual a 31,09% e cinco ou mais eventos de inundações apresentou a segunda maior taxa estimada de leptospirose, com 62,5 (IC95%: 48,9-73,0) por 100.000 habitantes. O terceiro grupo com a maior taxa predita estava no nó 32 com 488 municípios classificados com taxa de 44,0 (IC95%: 38,5-49,5) casos por 100.000 habitantes. Esses municípios apresentaram população com proporção de fossa inferior a 31,09% e total de inundações igual ou menor que três.

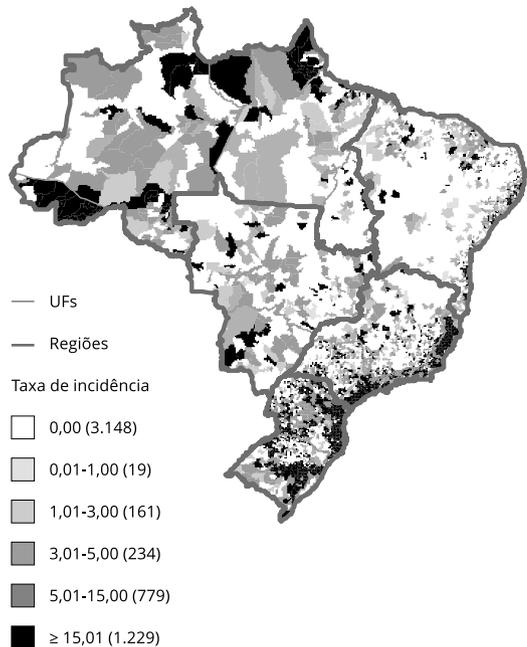
A variável total de inundações ainda apresenta influência nas partições 4 e 2, a variável fossa apresenta influência no ramo 6 e 14. A variável categoria de cidades segundo tamanho populacional apresenta associação com a de desfecho no ramo 5 quando ocorre uma partição à esquerda que classifica municípios de pequeno porte (até 5.000 habitantes) e de grande porte (com mais de 100.000 habitantes), e à direita as demais categorias de municípios. À esquerda, no ramo 6, a variável proporção de domicílios com esgotamento por fossa volta a particionar a variável de desfecho em grupos de municípios com proporção de fossa menor ou igual a 9,12 e maior que 9,12. No nó terminal 7 estão classificados 638 municípios com taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local por 100.000 habitantes de 11,6 (IC95%: 10,6-12,5). No nó terminal 8 estão 30 municípios classificados com taxa de 28,6 (IC95%: 19,52-37,7) casos por 100.000 habitantes. O modelo final apresentou erro médio de 53,10343 casos por 100.000 habitantes.

A Figura 4 mostra a distribuição das taxas de incidência de leptospirose segundo diferentes técnicas. A Figura 4a representa a taxa acumulada bruta, em que nota-se um maior número de áreas com taxa entre 0 e 1 caso por 100.000 habitantes. A Figura 4b representa a taxa suavizada e nota-se que as áreas com taxas entre 0 e 1 estão em menor número, já que a taxa suavizada procura corrigir a possível subnotificação da doença. Destaca-se também que a distribuição espacial das taxas representadas nas Figuras 4a e 4b são muito semelhantes, incluindo as áreas com maior incidência da doença. Nota-se nessas figuras que existe uma extensa área de maior incidência da doença, que vai desde a região central do extremo sul e prolonga-se por todo o litoral das regiões Sul e Sudeste do país. No interior da Região Nordeste do país encontra-se uma grande área com as menores incidências da doença. Essa região é a mais seca do país, é quase coincidente com o que se denomina polígono da seca²⁵. Na Região Norte, destacam-se algumas áreas com a maior incidência da doença: no entorno do Rio Amazonas e seus afluentes, bem como nos estados do Amapá e do Acre. A Figura 4c mostra que a taxa estimada pela árvore de regressão apresenta semelhança em relação à distribuição espacial com as Figuras 4a e 4b. Observam-se áreas de concentração de casos nas regiões central e sudoeste do Estado do Rio Grande do Sul, uma mancha quase única e contínua de todo o litoral a partir de Santa Catarina até o Espírito Santo e nos estados do Nordeste, em alguns municípios no litoral. Na Região Norte, o Amapá e o Acre são os estados que apresentam uma grande área de incidência do agravo, além de alguns municípios do Pará e do Amazonas no entorno do Rio Amazonas e seus afluentes. Na Região Centro-oeste, destaque para alguns municípios das regiões central e sudoeste de Mato Grosso do Sul. Nota-se também que as taxas preditas superestimaram as taxas brutas e ajustadas.

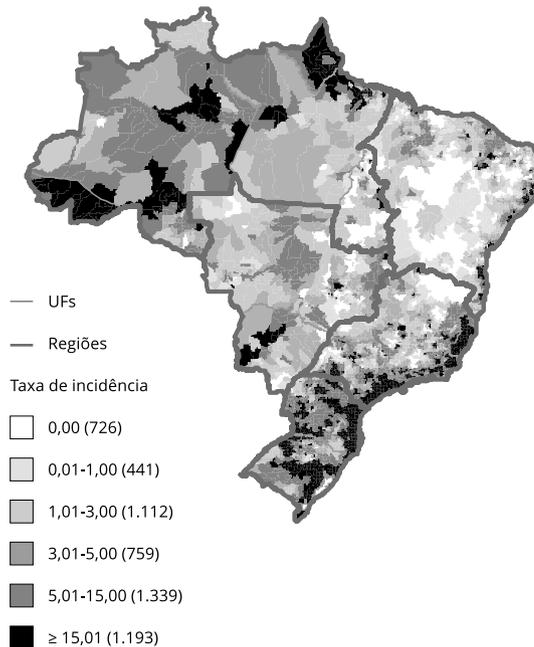
Figura 4

Distribuição espacial da taxa bruta acumulada, da taxa bayesiana local, da taxa estimada pelo modelo de árvore de regressão com inferência condicional.

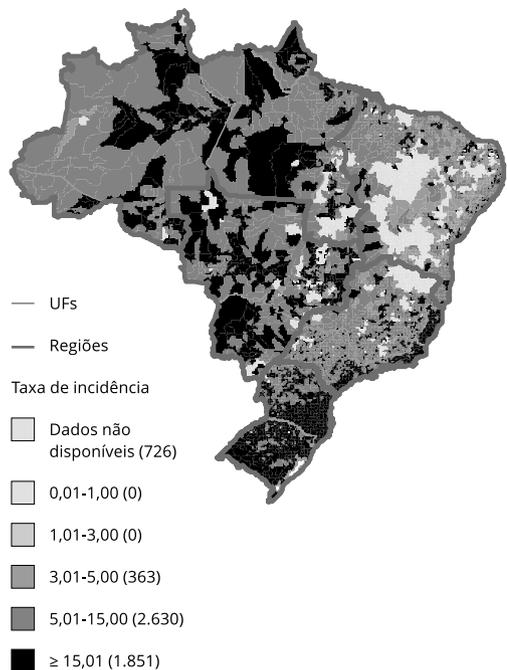
4a) Taxa de incidência acumulada bruta



4b) Taxa de incidência suavizada pelo método bayesiano local



4c) Taxa estimada pelo modelo de árvore de regressão com inferência condicional



UFs: Unidades da Federação.

Discussão

Principais achados

Com base na análise desses dados percebeu-se que o risco de ocorrência de leptospirose foi maior nos municípios que decretaram inundações em relação aos que não decretaram, independentemente do tamanho da população. Além disso, a incidência de leptospirose aumentou na medida em que cresceu o total de inundações nos municípios. Esses achados mostram a importância desse fenômeno natural para a ocorrência da doença e alertam para a possibilidade do aumento da leptospirose fruto das mudanças climáticas. Nos países em desenvolvimento, que têm dificuldades na governança para gerir desastres, a magnitude do problema pode ser agravada ²⁶.

O processo desordenado de ocupação urbana também contribuiu muito para a ocorrência de leptospirose. Este trabalho mostrou que os municípios com a maior proporção de domicílios com esgotamento por fossa e que tiveram três inundações ou mais no período analisado foram aqueles com as maiores taxas de incidência de leptospirose. Ressalta-se que, no Brasil, 94% dos domicílios contam com abastecimento de água por rede geral e somente 56% com esgotamento por rede, ou seja, a população recebe água limpa, embora em muitos municípios exista problema de sua regularidade e quase a metade devolve esta água de maneira inadequada ao ambiente ²⁷.

Com relação à distribuição espacial da doença, observam-se áreas no extremo sul do país com incidência elevada, como no Estado do Rio Grande do Sul já descrita em trabalhos anteriores ²⁸. A Região Norte mostra áreas com taxas elevadas nos estados do Acre e do Amapá que não apresentavam esta distribuição anteriormente ^{29,30}. O Estado do Espírito Santo também demonstra incremento na ocorrência do agravo, bem como os municípios do litoral das regiões Sudeste e Sul ³¹. Entretanto, cada uma dessas áreas pode apresentar uma dinâmica diferente de ocorrência do agravo. Por exemplo, no Rio Grande do Sul é conhecida uma relação com a produção agrícola ²⁸; em áreas da Região Norte, como na Amazônia, até mesmo a natureza do evento extremo é diferenciada porque as inundações são provocadas pela cheia de rios que invadem as casas ³². Em grandes centros urbanos, como os do litoral nordestino ou os dos estados do São Paulo e Rio de Janeiro, a ocupação desordenada, a falta de drenagem e coleta inadequada de lixo são responsáveis pelo aumento do número de casos ^{2,3,14}. Entretanto, independentemente do contexto no qual está inserido, este trabalho demonstrou que o aumento no total de inundações está associado a um aumento significativo do número de casos de leptospirose.

Com relação ao modelo da árvore de regressão com inferência condicional, nem todas as variáveis selecionadas no teste de correlação e inseridas no desenho da árvore estão presentes no modelo final (Figura 3). Esse fato ocorre devido a essa técnica realizar testes de hipótese por meio de permutações que buscam selecionar tanto as variáveis com maior capacidade de predição quanto seus particionamentos binários de resposta ²¹. A taxa predita pelo modelo final mostrou-se aderente com os valores observados e destacou as áreas de maior incidência da doença e as áreas de baixíssima incidência, já nas áreas de incidências média e baixa o modelo não apresentou ajuste eficiente, classificando em geral estas áreas como de alta incidência. Contudo, deve-se verificar que o erro médio do modelo pode ser considerado elevado quando levadas em conta as observações individuais. Entretanto, devido à amplitude da taxa de incidência acumulada de leptospirose suavizada pelo método bayesiano local, que variou de 0,04 a 1.667,8, pode-se considerar que na avaliação geral do modelo o erro é baixo. Isso pode acontecer devido às questões de dificuldades no diagnóstico diferencial da doença, pois leptospirose ocorre justamente no mesmo período da incidência de dengue, Zika e chikungunya, que são agravos que mantêm uma vigilância mais atenta ¹⁷ e, portanto, tende-se a notificar casos anictéricos de leptospirose como sendo uma destas arboviroses. Além disso, possivelmente esse comportamento pode estar relacionado à flutuação aleatória da taxa bruta, que mesmo suavizada pelo método bayesiano local não seria suficiente para ajustá-la ¹³.

Ressalta-se a importância da atenção às áreas detectadas com taxas estimadas elevadas neste estudo, para um aprimoramento da vigilância de leptospirose, pois são áreas que têm características ecológicas sugestivas de maior risco, mas talvez o serviço de vigilância não esteja sendo tão eficiente na detecção destes casos. Devido provavelmente a esses fatores, observa-se no Brasil uma incidência de leptospirose baixa e uma letalidade alta ³³.

Este estudo permitiu um melhor conhecimento da ecologia de transmissão da leptospirose. Além da importância das inundações e dos problemas no esgotamento sanitário discutidos anteriormente, destacam-se também as condições socioeconômicas baixas representadas pela proporção de chefes de família com pelo menos dez anos de estudos, a proporção da população residente em área de favela e problemas relacionados ao saneamento ambiental³³.

O Brasil tem uma extensa experiência e já trabalha com diversas técnicas de análise estatística espacial na área de Saúde e Ambiente, a qualidade dos dados de saúde vem melhorando muito com o crescente desenvolvimento dos sistemas de informação³⁴, mas a sistematização e disponibilização de dados de emergência pela defesa civil é mais recente¹⁷. Acredita-se que com o uso combinado dessas bases de dados contribuiremos para o melhor entendimento da ocorrência da doença.

A utilização dos dados do total de inundações parece sintetizar o efeito da chuva na produção do agravo e possibilitar análises mais simples do que as que usam os dados de pluviosidade. Possivelmente, essa questão ocorra porque nem sempre o volume pluviométrico elevado configura necessariamente que a intensidade e o intervalo de tempo foram suficientes para promover inundações. Na maioria dos artigos utiliza-se para a análise da ocorrência da leptospirose as variáveis de precipitação, período seco e úmido. O uso de informações do total de inundações em uma escala nacional ainda não é comum no Brasil^{35,36}. São necessários mais estudos ao longo do tempo para identificar se as taxas de incidência aumentam nos mesmos períodos das inundações.

Os dados utilizados na escala nacional sobre a taxa de incidência de leptospirose em trabalhos anteriores apresentaram maior correlação com a presença de favela, e na escala mais detalhada o indicador que apresentou maior importância foram as áreas sujeitas a inundações. Contudo, esses dados de inundações na escala nacional não existiam até há pouco tempo e utilizava-se uma variável menos sensível, como a altitude da sede do município³.

Este trabalho envolvendo todos os municípios do Brasil, apesar de apresentar complexidade na análise, pode ser bastante útil pois em geral a escala de análise da leptospirose é quase sempre local. Dessa forma, pretende-se contribuir com os gestores nos apontamentos para políticas públicas nacionais, a fim de aprimorar a vigilância da leptospirose e alertar para a necessidade de investimentos em saneamento ambiental no Brasil. Ressalta-se que o investimento em monitoramento, obras de saneamento e a consequente diminuição na ocorrência de inundações pode acarretar uma redução no número de casos de leptospirose com a diminuição dos gastos de recurso no setor saúde e a redução de óbitos³⁷.

A decretação de inundações pela autoridade municipal constituiu um importante marcador de risco para a ocorrência de leptospirose. Os municípios que tinham como solução para o esgotamento sanitário o uso de fossas apresentaram maior risco para a ocorrência da doença. Com as condições ambientais resultantes das mudanças climáticas, urge que as cidades melhorem sua infraestrutura urbana para tentar mitigar a ocorrência de leptospirose e se adaptar às mudanças climáticas. Destaca-se também que as iniciativas para melhorar a infraestrutura devem ser avaliadas dentro de um processo de adaptação e mitigação.

Colaboradores

R. Gracie participou da concepção do estudo, estruturação do artigo, coleta de dados, análise, estruturação do mapa e dos resultados, escrita, discussão e revisão do artigo. D. R. Xavier participou da concepção do estudo, análise dos dados, escrita e revisão do texto. R. Medronho participou da concepção do estudo e revisão do texto.

Informações adicionais

ORCID: Renata Gracie (0000-0003-0225-3696); Diego Ricardo Xavier (0000-0001-5259-7732); Roberto Medronho (0000-0002-4617-6904).

Referências

1. Vinetz JM. The paradox of the chicken and the egg: lack of actionable diagnostics prevents global burden of disease assessment and deployment of public health preventive measures. *Anuradhapura Medical Journal* 2015; 9:1-4.
2. Felzemburgh RDM, Ribeiro GS, Costa F, Reis RB, Hagan JE, Melendez AXTO, et al. Prospective study of leptospirosis transmission in an urban slum community: role of poor environment in repeated exposures to the *Leptospira* agent. *PLoS Negl Trop Dis* 2014; 8:e2927.
3. Gracie R, Barcellos C, Magalhães M, Souza-Santos R, Barrocas P. Geographical scale effects on the analysis of leptospirosis determinants. *Int J Environ Res Public Health* 2014; 11:10366-83.
4. Pappas G, Papadimitriou P, Siozopoulou V, Christou L, Akritidis N. The globalization of leptospirosis: worldwide incidence trends. *Int J Infect Dis* 2008; 12:351-7.
5. McMichael AJ, Woodruff RE, Hales S. Climate change and human health: present and future risks. *Lancet* 2006; 367:859-69.
6. Intergovernmental Panel on Climate Change. *Climate Change 2014: impacts, adaptation and vulnerability. Part B: regional aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. New York: Cambridge University Press; 2014.
7. Patz JA, Campbell-Lendrum A, Holloway D, Foley T. Impact of regional climate change on human health. *Nature* 2005; 438:310-7.
8. Barcellos C, Monteiro AMV, Corvalan C, Gurgel HC, Carvalho MS, Artaxo P, et al. Mudanças climáticas e ambientais e as doenças infecciosas: cenários e incertezas para o Brasil. *Epidemiol Serv Saúde* 2009; 18:285-304.
9. Lowe R, Bailey TC, Stephenson DB, Graham RJ, Coelho CAS, Carvalho MS, et al. Spatio-temporal modelling of climate-sensitive disease risk: towards an early warning system for dengue in Brazil. *Comput Geosci* 2011; 37:371-81.
10. Leichenko R, Silva JA. Climate change and poverty: vulnerability, impacts, and alleviation strategies. *WIREs Climate Change* 2014; 5:539-56.
11. Engle NL, Bremond A, Moss EL, Moss RH. Towards a resilience indicator framework for making climate-change adaptation decisions. *Mitig Adapt Strateg Glob Chang* 2014; 19:1295-312.
12. Instituto Brasileiro de Geografia e Estatística. Arranjos populacionais e concentrações urbanas do Brasil. http://www.ibge.gov.br/apps/arranjos_populacionais/2015/pdf/publicacao.pdf (acessado em 24/Jan/2017).
13. Carvalho MS, Souza-Santos R. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cad Saúde Pública* 2005; 21:361-78.

14. Gracie R. O efeito da escala geográfica na análise dos determinantes da leptospirose [Dissertação de Mestrado]. Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz; 2008.
15. Chiavegatto F, Porto AD. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiol Serv Saúde* 2015; 24:325-32.
16. Sumi A, Telan EFO, Chagan-Yasutan H, Piolo MB, Hattori T, Kobayashi N. Effect of temperature, relative humidity and rainfall on dengue fever and leptospirosis infections in Manila, the Philippines. *Epidemiol Infect* 2016; 145:78-86.
17. Xavier DR, Barcellos C, Barros HS, Magalhães MAFM, Matos VP, Pedroso MM. Organização, disponibilização e possibilidades de análise de dados sobre desastres de origem climática e seus impactos sobre a saúde no Brasil. *Ciênc Saúde Colet* 2014; 19:3657-68.
18. Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol* 2008; 26:1011-3.
19. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods* 2009; 14:323-48.
20. Breiman L. Random forests. *Mach Learn* 2001; 45:5-32.
21. Zhang Z. Decision tree modeling using R. *Ann Transl Med* 2016; 4:268-75.
22. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and regression trees*. Monterey: Wadsworth & Brooks; 1984.
23. Strasser H, Weber C. On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics* 1999; 8:220-50.
24. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Stat* 2006; 15:651-74.
25. Ab'Saber NA. Sertões e sertanejos: uma geografia humana sofrida. *Estud Av* 1999; 13:7-59.
26. Freitas CM, Carvalho ML, Ximenes EF, Arraes EF, Gomes JO. Vulnerabilidade socioambiental, redução de riscos de desastres e construção da resiliência: lições do terremoto no Haiti e das chuvas fortes na Região Serrana, Brasil. *Ciênc Saúde Colet* 2012; 17:1577-86.
27. Heller L. Visão geral dos serviços de água e esgotamento sanitário no Brasil. São Paulo: Departamento Intersindical de Estatística e Estudos Socioeconômicos; 2016.
28. Rezende RS, Barcelos Y, Laurindo TVF, Dias CAGM, Oliveira E, Utzig ILA, et al. Epidemiology of leptospirosis: cases from 2011 to 2015 in the Brazilian northern region in the Brazilian northern region. *Biota Amazônia* 2016; 6:94-9.
29. Pereira FCS, Santos EF. Mortalidade por leptospirose no Estado do Amapá (2005-2014). *Revista Eletrônica Estácio Saúde* 2016; 5(2). <http://revistaadmmade.estacio.br/index.php/saudesantacatarina/article/viewFile/2891/1297>.
30. Ministério da Saúde. Leptospirose: análise dos dados epidemiológicos de 2010 a 2014. *Informe Epidemiológico* 2016. <http://portalarquivos2.saude.gov.br/images/pdf/2016/maio/20/Informe-epidemiol-gico-leptospirose.pdf> (acessado em 01/Mai/2018).
31. Barcellos C, Lammerhirt CB, Almeida MAB, Santos E. Distribuição espacial da leptospirose no Rio Grande do Sul, Brasil: recuperando a ecologia dos estudos ecológicos. *Cad Saúde Pública* 2003; 19:1283-92.
32. Freitas CM, Silva DRX, Sena ARM, Silva EL, Sales LBF, Carvalho ML, et al. Desastres naturais e saúde: uma análise da situação do Brasil. *Ciênc Saúde Colet* 2014; 19:3645-56.
33. Pereira CAR, Barata MML, Trigo AGM. Social cost of leptospirosis cases attributed to the 2011 disaster striking Nova Friburgo, Brazil. *Int J Environ Res Public Health* 2014; 11:4140-57.
34. Lima CRA. Qualidade dos sistemas de saúde no Brasil. *Cad Saúde Colet (Rio J.)* 2010; XVIII:5.
35. Guimarães RM, Cruz OGC, Pereira VGP, Mazoto ML, Vieira JD, Asmus CIRF. Análise temporal da relação entre leptospirose e ocorrência de inundações por chuvas no Município do Rio de Janeiro, Brasil, 2007-2012. *Ciênc Saúde Colet* 2014; 19:3683-92.
36. Soares TSM, Latorre MRDO, Laporta GZL, Buzzar MR. Análise espacial e sazonal da leptospirose no Município de São Paulo – SP, 1998 a 2006. *Rev Saúde Pública* 2010; 44:283-91.
37. Heller L. Relação entre saúde e saneamento na perspectiva do desenvolvimento. *Ciênc Saúde Colet* 1998; 3:73-84.

Abstract

Leptospirosis is related to problems with environmental sanitation, and the incidence tends to increase during flood periods. Considering issues related to climate change, floods can be expected to increase. Floods do not affect populations homogeneously, and communities with worse socioeconomic conditions tend to be impacted more heavily. In order to determine whether the number of floods increases the incidence of leptospirosis and its relationship to contextual variables, the study used socioeconomic, environmental, and disease occurrence data at the municipal (county) level. Municipalities suffering problems with sewage disposal showed a higher risk of leptospirosis incidence. Total flooding since the municipality's declaration of flood emergency was an important risk marker for leptospirosis incidence. Regression tree modeling proved useful for estimating leptospirosis incidence in Brazil.

Leptospirosis; Floods; Dry-Pit; Decision Trees; Climate Change

Resumen

La leptospirosis se relaciona con problemas de saneamiento ambiental, así como con el incremento de casos en períodos de inundaciones. Teniendo en consideración las cuestiones relacionadas con el cambio climático, las inundaciones tienden a aumentar. Las inundaciones no afectan a las poblaciones de manera homogénea, en general, los menos favorecidos en términos socioeconómicos son los más afectados. Para saber si el número de inundaciones aumentaría la incidencia de leptospirosis, y su relación con variables contextuales, se utilizaron datos socioeconómicos, ambientales y de ocurrencia de la enfermedad en el nivel municipal. Los municipios que poseían problemas en el alcantarillado sanitario presentaron un mayor riesgo para la ocurrencia de leptospirosis. El total de inundaciones sufridas a partir de su reconocimiento oficial por parte de la autoridad municipal constituyó un importante marcador de riesgo para la ocurrencia de leptospirosis. El modelo de árbol de regresión se mostró útil para estimar la ocurrencia de leptospirosis en Brasil.

Leptospirosis; Inundaciones; Fosas Secas; Árboles de Decisión; Cambio Climático

Recebido em 07/Nov/2019
Versão final rerepresentada em 11/Jul/2020
Aprovado em 16/Set/2020