

Análise espacial da Tuberculose no Rio de Janeiro no período de 2005 a 2008 e fatores socioeconômicos associados utilizando microdado e modelos de regressão espaciais globais

Spatial analysis of Tuberculosis in Rio de Janeiro in the period from 2005 to 2008 and associated socioeconomic factors using micro data and global spatial regression models

Monica de Avelar Figueiredo Mafra Magalhães¹
Roberto de Andrade Medronho²

Abstract *The present study analyses the spatial pattern of tuberculosis (TB) from 2005 to 2008 by identifying relevant socioeconomic variables for the occurrence of the disease through spatial statistical models. This ecological study was performed in Rio de Janeiro using new cases. The census sector was used as the unit of analysis. Incidence rates were calculated, and the Local Empirical Bayesian method was used. The spatial autocorrelation was verified with Moran's Index and local indicators of spatial association (LISA). Using Spearman's test, variables with significant correlation at 5% were used in the models. In the classic multivariate regression model, the variables that fitted better to the model were proportion of head of family with an income between 1 and 2 minimum wages, proportion of illiterate people, proportion of households with people living alone and mean income of the head of family. These variables were inserted in the Spatial Lag and Spatial Error models, and the results were compared. The former exhibited the best parameters: $R^2 = 0.3215$, Log-Likelihood = -9228, Akaike Information Criterion (AIC) = 18,468 and Schwarz Bayesian Criterion (SBC) = 18,512. The statistical methods were effective in the identification of spatial patterns and in the definition of determinants of the disease providing a view of the heterogeneity in space, allowing actions aimed more at specific populations.*

Key words *Tuberculosis, Spatial analysis, Global spatial regression model*

Resumo *Este trabalho analisa o padrão espacial da tuberculose no período de 2005 a 2008 identificando variáveis socioeconômicas relevantes para a ocorrência da doença através de modelos estatísticos espaciais. É um estudo ecológico realizado no Rio de Janeiro com casos novos. Utilizou-se o setor censitário como unidade de análise. Foram calculadas as taxas de incidência e usado o método Bayesiano Empírico Local. Foi constatada a autocorrelação espacial com Índice de Moran e LISA. Usando teste de Spearman, as variáveis com correlação estatisticamente significativas a 5% foram utilizadas nos modelos. No modelo de regressão multivariado clássico as variáveis Proporção de responsável com renda entre 1 e 2 salários-mínimos, Proporção de analfabetos, Proporção de domicílios com pessoas que moram sozinhas e Renda média do responsável se ajustaram melhor. Essas variáveis foram inseridas nos modelos Spatial Lag e Spatial Error e os resultados comparados. O primeiro apresentou os melhores parâmetros: $R^2 = 0,3215$, Log da Verossimilhança = -9228, AIC = 18468 e SBC = 18512. Os métodos estatísticos apresentaram-se eficientes na identificação de padrões espaciais e definição de determinantes da doença dando uma visão da heterogeneidade no espaço, possibilitando uma atuação mais direcionada a populações específicas.*

Palavras-chave *Tuberculose, Análise espacial, Modelo de regressão espacial global*

¹ Instituto de Comunicação e Informação Científica e Tecnologia em Saúde, Fiocruz. Av. Brasil 4365, Manguinhos. 21040-360 Rio de Janeiro RJ Brasil. monica.avelar.magalhaes@gmail.com

² Instituto de Estudos em Saúde Coletiva, UFRJ. Rio de Janeiro RJ Brasil.

Introdução

A tuberculose (TB) é causada pelo *Mycobacterium tuberculosis* (bacilo de Koch). A doença é um grave problema de saúde pública no mundo sendo endêmica em vários países e mata cerca de 1,5 milhões de pessoas a cada ano¹.

A tuberculose é uma doença diretamente ligada a condições precárias de vida. A probabilidade de um indivíduo ser infectado e desenvolver a doença depende de vários fatores, dentre eles as condições socioeconômicas e as médicas a que esse indivíduo está submetido². Em centros urbanos de países em desenvolvimento, determinantes sociais como pobreza, baixa escolaridade, adensamento populacional, moradias insalubres e abuso de drogas formam um grupo de indivíduos vulneráveis à doença^{3,4}.

Desigualdade social, advento da AIDS, envelhecimento da população e grandes movimentos migratórios são alguns dos fatores apontados por Ruffino-Neto⁵ como sendo as principais causas para a gravidade da situação atual da tuberculose no mundo. Em 2010, havia uma estimativa de cerca de 8,5-9,2 milhões de casos e 1,2-1,5 milhões de mortes (incluindo as mortes por tuberculose entre pessoas HIV positivas). É a segunda principal causa de morte por doenças infecciosas no mundo perdendo apenas para o HIV em 2008¹.

A tuberculose no Brasil ainda é um grave problema de saúde pública. A maior concentração dos casos acontece na Região Sudeste com o Rio de Janeiro apresentando a maior taxa de incidência. No ano de 2010, o Estado do Rio de Janeiro notificou 14.206 casos de tuberculose e o município do Rio de Janeiro era o que se posicionava com o maior número de notificações, 7.664 casos. Embora no município do Rio de Janeiro tenha havido uma queda na taxa de incidência de 13% nos últimos 9 anos, em 2010 possuía uma taxa de 95/100 mil habitantes. A tuberculose pulmonar é a forma clínica da doença predominante no município^{4,6}.

A principal fonte de dados para a tuberculose é o Sinan e a notificação se baseia na definição de caso confirmado na investigação e o acompanhamento dos casos⁷.

No município do Rio de Janeiro, a tuberculose não acontece igualmente distribuída no espaço geográfico. Conhecer a distribuição espacial da tuberculose e seus determinantes sociais nas diferentes áreas do município permite identificar populações mais vulneráveis e planejar ações governamentais mais focadas às necessidades dos

diferentes territórios. Assim, a incorporação da dimensão espacial nas análises da doença pode extrair significados adicionais às convencionais, auxiliando na compreensão da dinâmica deste agravo. Sendo assim, a utilização de técnicas de Geoprocessamento, principalmente de Sistemas de Informações Geográficas (SIG) juntamente com Estatística Espacial permitem incorporar diversas variáveis como localização, tempo, características socioeconômicas e características ambientais nos estudos em saúde⁸. Elas permitem gerar hipóteses etiológicas referentes à origem das doenças em diferentes populações⁹. Com isso, buscam um modelo inferencial que inclua as relações espaciais constituintes do fenômeno estudado¹⁰.

Um aspecto fundamental na aplicação dessas técnicas é a caracterização da dependência espacial, mostrando como os valores das variáveis de interesse estão correlacionados no espaço. Os modelos de regressão espaciais globais foram utilizados em alguns estudos de associação da tuberculose e outras doenças transmissíveis e determinantes socioeconômicos, porém utilizando unidades espaciais maiores como bairros e regiões administrativas¹¹⁻¹³ poucos utilizaram o setor censitário para análise.

Este estudo buscou analisar o padrão espacial da tuberculose identificando variáveis socioeconômicas relevantes para a ocorrência de tuberculose comparando o modelo estatístico de regressão linear clássico com modelos com efeitos espaciais globais.

Metodologia

Trata-se de um estudo ecológico realizado no município do Rio de Janeiro. Com uma população de 6.323.037, o município é composto por 160 bairros e 763 favelas, onde vivem 22 % da população¹⁴.

A análise espacial usou a estratégia de análise de dados de área. A unidade espacial escolhida foi o setor censitário. Os dados de tuberculose estudados foram os casos novos notificados à Vigilância Epidemiológica do município do Rio de Janeiro através do Sistema de Informação de Agravos de Notificação (Sinan) entre os anos de 2005 e 2008, disponibilizados pela Secretaria Municipal de Saúde e Defesa Civil do Rio de Janeiro (SMSDC-RJ). Foram utilizados os registros cuja informação sobre o município de residência e sobre o atendimento foram realizados no Rio de Janeiro. Os dados socioeconômicos foram retira-

dos do Censo Demográfico de 2010 e serviram de base para construir os indicadores utilizados nas análises dos dados por setor censitário: Número médio de pessoas por domicílios (POP_DOM), Renda familiar média (RM_DOM); Proporção de responsável com renda maior que 1 salário-mínimo e menor que 2 salários-mínimos (P_R_2SM); Proporção de responsável com renda mensal menor que 1 salário-mínimo (P_R_1SM); Proporção de analfabetos (P_ANALF); Proporção de domicílio com abastecimento de água da rede geral (P_DOM_AG); Proporção de domicílio com banheiro de uso exclusivo dos moradores ou sanitário e esgotamento sanitário via rede geral de esgoto ou pluvial (P_DOM_SA); Proporção de domicílio com coleta de lixo por serviço de limpeza (P_DOM_CL); Proporção de domicílios com banheiro de uso exclusivo dos moradores ou sanitário (P_DOM_BA); Proporção de cor de pele declarada branca (P_P_BRA); Proporção de cor de pele declarada preta (P_P_PRE); Proporção de cor de pele declarada amarela (P_P_AM); Proporção de cor de pele declarada parda (P_P_PAR); Proporção de cor de pele declarada indígena (P_P_IN); Proporção de domicílios com pessoas que moram sozinhas (P_DOM_MS); Renda média do responsável (RM_RESP). Os indicadores foram construídos com objetivo de contemplar as dimensões de renda, escolaridade e condições de habitação fazendo um *proxy* da condição social do paciente.

Os casos de tuberculose foram georreferenciados por endereço de residência utilizando a metodologia descrita em Magalhães et al.¹⁵. Foram excluídos do estudo os casos em que a residência se referia a presídio e hospitais. Inicialmente foi feita uma análise exploratória a partir da taxa de incidência do período para cada 1.000 habitantes calculada por setor censitário.

Foram excluídos do estudo 314 setores censitários que não tinham população (em sua maioria áreas ocupadas por mato, lagoas, áreas verdes etc.) ou que, por terem poucos domicílios particulares permanentes, tiveram seus dados omitidos pelo IBGE a fim de manter o sigilo dos dados.

Visando minimizar a instabilidade das taxas brutas e eliminar a flutuação aleatória foi aplicado o alisamento das taxas de incidência usando o método Bayesiano Empírico Local¹⁶⁻¹⁸. Através de um histograma foi verificado que a distribuição dos dados não era normal e para isso foi empregada a transformação do tipo logaritmo neperiano (Ln) na variável dependente para aproximá-la a uma distribuição normal.

Para verificar a presença de aglomerados, isto é, áreas com dinâmica espacial própria e que merecem análise detalhada foram utilizados o índice de Moran Global.

Como o estudo trabalhou com número grande de áreas foi utilizado o Indicador Local de Associação Espacial (LISA – *Local Indicator of Spatial Association*) baseado na matriz de vizinhança gerada com os vizinhos de primeira ordem. Esse indicador possibilita a identificação de padrões significativos de associação espacial e permite representar uma decomposição do índice global¹⁹.

O LISA classificou os setores censitários em função do nível de significância dos valores de seus índices locais em: alta/alta, baixa/baixa, significa que possuem associação positiva, isto é a localização possui vizinhos com valores próximos, e alta/baixa e baixa/alta, indicam associação negativa, ou seja, que a localização possui vizinhos com valores distintos.

Constatada a autocorrelação espacial, foi construída a matriz de correlação de *Spearman* e sob um olhar estatístico e epidemiológico, as variáveis independentes com correlação estatisticamente significativas a 5% com a variável dependente e as variáveis não colineares, isto é, com correlação < 0,7 foram utilizadas na análise das regressões clássica e espacial.

Foi aplicada a regressão linear multivariada (*Ordinary Least Squares Estimation - OLS*). Utilizando o método *backward* e critério epidemiológico, buscou-se por variáveis que descrevessem melhor a ocorrência da doença com correlação estatisticamente significativa a 5%.

Com o objetivo de incorporar os efeitos espaciais, aplicou-se os chamados modelos com efeitos espaciais globais que tratam a estrutura espacial de forma global, isto é, supõe que é possível capturar a estrutura de correlação espacial num único parâmetro, que é adicionado ao modelo de regressão tradicional²⁰. Com o intuito de verificar qual melhor modelo se encaixaria nas variáveis estudadas foram aplicadas duas alternativas.

O primeiro modelo utilizado foi o modelo espacial autoregressivo misto (*Spatial Lag Model*) que atribui à variável resposta Y a autocorrelação espacial ignorada. Dado que se considera a dependência espacial através da adição ao modelo de regressão de um novo termo na forma de uma relação espacial para a variável dependente.

$$Y = X\beta + \rho WY + \varepsilon$$

onde: W é a matriz de proximidade espacial; WY expressa a dependência espacial em Y; e ρ

é o coeficiente espacial autoregressivo¹⁶. Neste modelo a autocorrelação espacial é incorporada como componente do modelo.

O outro modelo foi o de erro espacial (*Spatial Error Model*), que considera os efeitos espaciais como um ruído, isto é, como um fator a ser removido. Esse modelo parte do princípio de que não é possível modelar todas as características de uma unidade geográfica que podem influenciar as regiões vizinhas. Os efeitos da autocorrelação espacial são associados ao termo de erro ε e o modelo é descrito por:

$$Y = X\beta + \varepsilon, \quad \varepsilon = \lambda W\varepsilon + \xi$$

onde: $W\varepsilon$ é a componente do erro com efeitos espaciais; λ é o coeficiente autoregressivo; ξ é a componente do erro com variância constante e não correlacionada¹⁹.

Na avaliação de qual modelo melhor se ajustaria às variáveis disponíveis foi escolhido aquele que apresenta maior valor do Log da Verossimilhança e menores valores do critério de informação de *Akaike* (AIC) e do critério bayesiano *Schwarz*^{10,20,21}.

Os resíduos nos modelos *Spatial Lag* e *Spatial Error* foram analisados através do índice de Moran para verificar quantitativamente se a autocorrelação espacial foi eliminada com a aplicação do modelo.

O programa utilizado para gerar os modelos e cálculo do Índice de Moran e o LISA foi o GeoDa.

Resultados

Houve uma perda de 11% na localização dos casos.

Na análise por área, pode-se analisar, através da Figura 1, a variação da taxa de incidência após o alisamento bayesiano ao longo do município do Rio de Janeiro. Com as maiores taxas aparecendo em áreas dos bairros do Centro e adjacências, se estendendo pelos bairros de Benfica, Manguinhos, Maré, Penha, Vila da Penha e parte da zona oeste.

A existência de autocorrelação espacial da variável dependente no nível de setor censitário pôde ser observada através do índice de Moran Global ($I = 0,402$, $p = 0,001$). Calculando-se o LISA foi possível classificar os setores censitários em função do nível de significância dos valores de seus índices locais. A Figura 2 mostra que existem, para a variável estudada, áreas com índices estatisticamente significativos. Na figura 2A,

identifica-se pelo menos quatro grandes aglomerados de setores. Um primeiro (aglomerado 1) formado por parte bairros do Centro, São Cristóvão, Tijuca, Catumbi, Glória; outro (aglomerado 2) formado por parte dos bairros da Penha e Penha Circular, este é quase uma extensão do primeiro. Um terceiro (aglomerado 3) formado Barra da Tijuca, Recreio; e outro (aglomerado 4) formado por setores dos bairros de Pedra de Guaratiba e Barra de Guaratiba. A Figura 2B mostra que os aglomerados 1, 2 e 4 são compostos por setores com alta taxa de incidência de TB. E o aglomerado 3 é um agrupamento de setores com baixa taxa de incidência.

Com o resultado da matriz de correlação de *Spearman*, observou-se que dentre todas as variáveis independentes estudadas, apenas Proporção de pessoas por domicílio (POP_DOM) não apresentou correlação significativa com a variável dependente. Entretanto, algumas variáveis apresentaram correlação muito forte entre si: RM_DOM, P_P_BRA; P_P_PRE, P_P_PAR. Avaliando as correlações com uma conotação epidemiológica, nota-se que as variáveis: Proporção de analfabetos (P_ANALF) e Proporção de domicílio com coleta de lixo por serviço de limpeza (P_DOM_CL) apresentaram correlação positiva e por isso sentido contrário.

A partir dessa avaliação optou-se por retirar do modelo OLS as variáveis que não apresentaram associação significativa com a variável dependente e as que apresentaram colinearidade. A partir desse primeiro resultado, buscou-se por modelos que descrevessem melhor a relação entre essas variáveis e optou-se pelo modelo final apresentado na Tabela 1.

Com esse modelo o coeficiente de determinação R^2 foi de 0,044, o valor do Log da Verossimilhança foi de -10598,8; o critério de informação de *Akaike* (AIC) foi de 21207,5; e o critério bayesiano *Schwarz* (SBC) foi de 21243,7.

A existência de autocorrelação espacial pode ser observada a partir dos resíduos da regressão clássica. O índice de Moran dos resíduos foi de 0,3609 ($p < 0,01$). Embora, tenham distribuição normal, não estão distribuídos aleatoriamente pelo município como observado na Figura 3.

Em função da presença de autocorrelação espacial, foi utilizado o modelo *Spatial Lag* com as mesmas variáveis utilizadas no OLS. Nesse modelo obteve-se o coeficiente de determinação R^2 de 0,3215, o valor do Log da Verossimilhança (valor do *logaritmo da função de verossimilhança* calculado para os valores estimados dos coeficientes) foi de -9228,39; o critério de informação

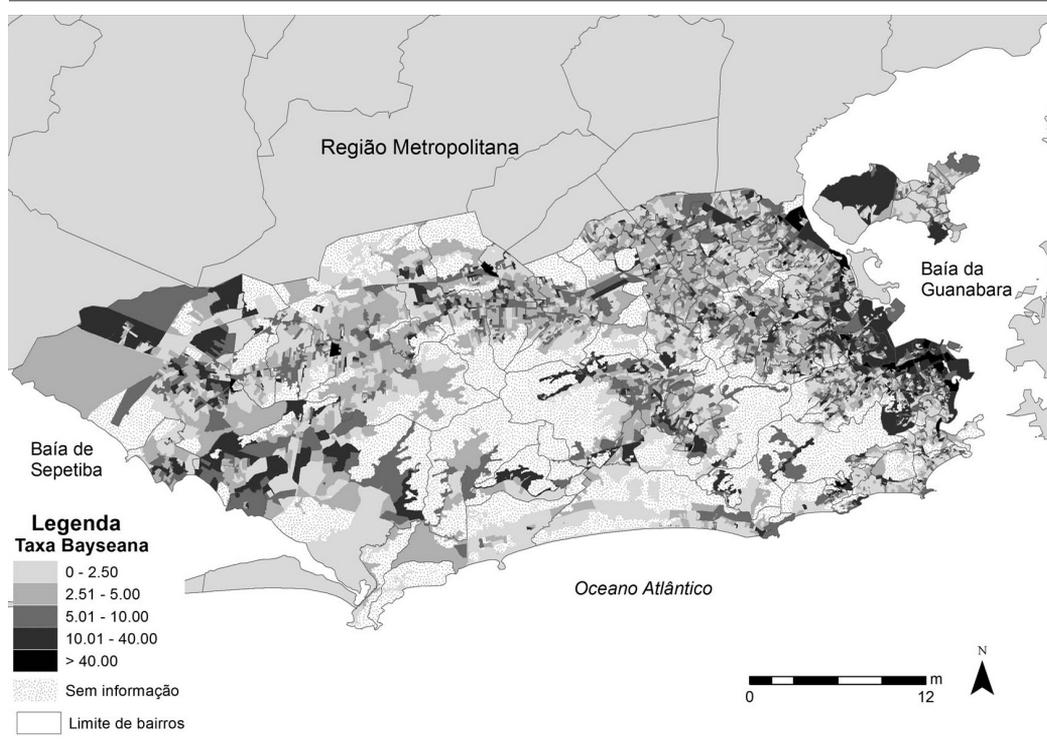


Figura 1. Mapa da taxa de incidência de tuberculose após o alisamento bayesiano por setor censitário.

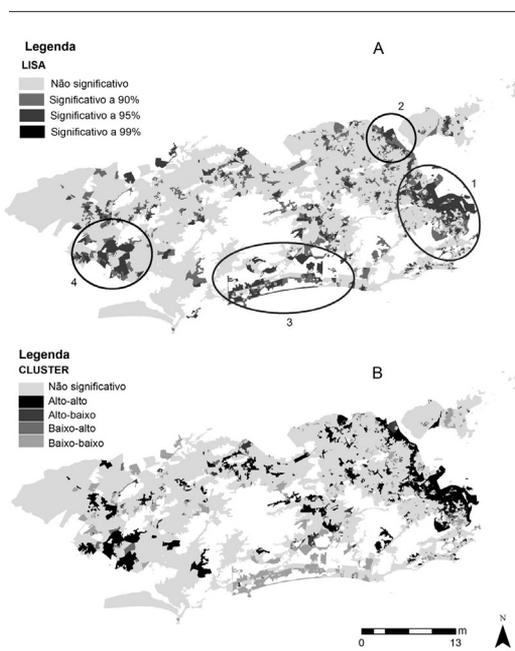


Figura 2. Mapa do Indicador local de autocorrelação espacial (LISA) para a variável dependente. (A) Áreas com valores significativos; (B) Espalhamento do LISA.

de Akaike (AIC) foi de 18468; e critério bayesiano Schwarz (SBC) foi de 18512,2.

Os resíduos desse modelo têm distribuição normal e o índice global de Moran foi de $-0,0183$ ($p < 0,001$). Esse valor baixo do índice de Moran indica que a inclusão da componente espacial no modelo praticamente eliminou a autocorrelação espacial.

Para fins de comparação, foi utilizado a seguir o modelo *Spatial Error* com as mesmas variáveis do OLS. Obteve-se o coeficiente de determinação R^2 que foi de 0,319, o valor do Log da Verossimilhança foi de $-9270,72$; o critério de informação de Akaike (AIC) foi de 18551,4; e critério bayesiano Schwarz (SBC) foi de 18587,6.

Os resíduos desse modelo apresentam distribuição normal e o índice global de Moran foi de $-0,0195$ ($p < 0,001$). Esse valor baixo do índice de Moran indica que a inclusão da componente espacial nesse modelo também eliminou a autocorrelação espacial.

A Tabela 2 apresenta o resumo dos índices que permitem avaliar a qualidade dos modelos. Observa-se que quando essa autocorrelação espacial foi introduzida nos modelos, através do *Spatial Lag* e do *Spatial Error*, há uma melhora nos resultados. Porém, dentre os métodos de re-

gressão espacial considerados, os resultados fornecidos pelo *Spatial Lag* indicaram que este foi o modelo que proporcionou o melhor ajuste das variáveis estudadas com maior valor do Log da Verossimilhança e menores valores do critério de informação de *Akaike* e do critério bayesiano *Schwarz* (SBC).

Considerações

A opção pelo uso da suavização das taxas de incidência de tuberculose pelo método bayesiano local foi uma tentativa de se minimizar possíveis distorções que a variabilidade fornecida pelo cálculo de taxas brutas em função do tamanho da população sob risco nos setores censitários podem acarretar. Entretanto uma desvantagem desse método é o possível potencial na indução de autocorrelação espacial. Segundo Morais Neto et al.²², este efeito pode ocasionar superestimativa no cálculo dos coeficientes de autocorrelação global e local nas áreas com poucos casos de TB, onde o alisamento das taxas do setor censitário em direção à média de seus vizinhos é mais acentuada. Embora reconhecendo o possível viés, a opção por usar esse recurso para correção das taxas de incidência se deu por dois motivos. O primeiro, pelo fato da tuberculose ser uma doença de transmissão respiratória com marcada in-

fluência das condições socioeconômicas. Assim, tratando-se de um estudo em microáreas, a ocorrência de um setor censitário com taxa de incidência baixa, cujos vizinhos a têm elevada, acre-

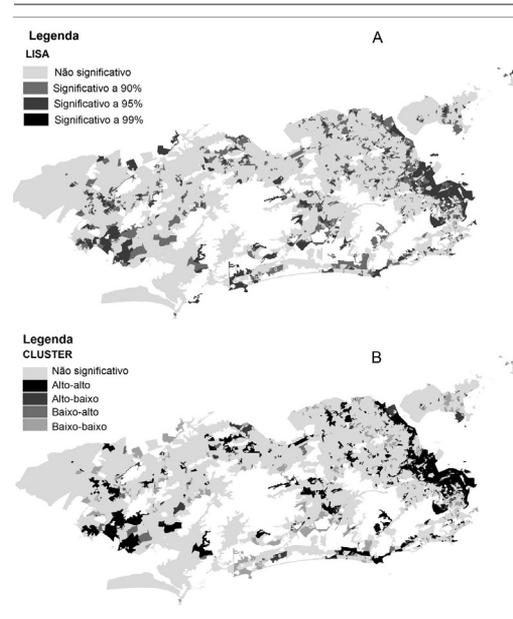


Figura 3. Mapa do Indicador local de autocorrelação espacial (LISA) para os resíduos do modelo OLS. (A) Áreas com valores significativos; (B) Espalhamento do LISA.

Tabela 1. Modelo OLS para o log da taxa de incidência após alisamento bayesiano.

Variáveis independentes	Coefficiente	Erro Padrão	t	p
P_R_2SM ^a	0,002958	0,000742	3,99	<0,001
P_ANALF ^b	-0,017731	0,001211	-14,64	<0,001
P_DOM_MS ^c	0,004201	0,000855	4,91	<0,001
RM_RESP ^d	-0,000062	0,000004	-14,28	<0,001

^a Proporção de responsável com renda maior que 1 salário-mínimo e menor que 2 salários-mínimos; ^b Proporção de analfabetos;

^c Proporção de domicílios com pessoas que moram sozinhas; ^d Renda média do responsável.

Tabela 2. Índices R², log da verossimilhança, critério de informação de *Akaike* e do critério bayesiano *Schwarz* (SBC) para os três modelos estudados.

Variável	OLS	Spatial Lag	Error Lag
R ²	0,044	0,3215	0,319
Log da Verossimilhança	-10598,8	-9228,39	-9270,72
AIC	21207,5	18468	18551,4
SBC	21243,7	18512,2	18587,6
Moran (I) - Resíduo	0,364	-0,018	-0,019

dita-se que o ajuste da taxa deste setor seja mais coerente do ponto de vista epidemiológico. O outro motivo foi pelo fato da grande quantidade de zeros quando se utilizou a taxa de incidência de tuberculose bruta. Quase metade dos setores censitários não possuía caso e, portanto, tinham o valor da taxa bruta igual a zero. Resalta-se que o índice de Moran também foi calculado com a taxa bruta e o valor não diferiu muito do valor calculado utilizando a taxa após o alisamento pelo método Bayesiano Local.

As variáveis P_ANALF e P_DOM_CL apresentaram correlação no sentido contrário ao esperado, porém optou-se por deixá-las nas análises para avaliar como se comportariam nos modelos de regressão uma vez que, nas primeiras análises exploratórias eram variáveis importantes no poder explicativo do modelo. E no modelo final a P_DOM_CL foi excluída, pois perdeu a significância estatística, e a P_ANALF se manteve. A informação sobre anos de estudo seria extremamente importante neste estudo, porém essa informação não foi coletada no Censo 2010. A informação sobre escolaridade se restringiu a alfabetizados/analfabetos e emprobecu as análises a partir dos dados censitários, uma vez que essa variável juntamente com a renda era um bom estimador de condição de vida da população.

O baixo valor para o coeficiente de determinação no OLS pode indicar que outras variáveis podem estar relacionadas à incidência da doença. Comparando-se os três modelos, nota-se que aqueles que levaram em conta a dependência espacial (*Spatial Lag* e *Spatial Error*) tiveram um desempenho melhor que o modelo clássico (OLS). Esse fato pode ser explicado pela dependência espacial da variável dependente comprovada pelo valor observado do índice Global de Moran e pelos mapas de espalhamento do índice Local de Moran. Outro parâmetro importante na avaliação na melhoria dos modelos espaciais em comparação ao clássico pode ser visto pelo ganho significativo no poder de explicação das variáveis através do valor de R^2 .

Entre os dois modelos de regressão espacial, o *Spatial Lag* apresentou melhores valores para *log da função de verossimilhança*, critério de informação de Akaike (AIC) e critério bayesiano Schwarz (SBC).

A análise dos valores dos índices de Moran dos resíduos dos modelos espaciais capturou a dependência espacial existente, uma vez que os valores saíram de 0,3609 para o modelo clássico para -0,0183 e -0,0195 para os modelos *Spatial Lag* e *Spatial Error*, respectivamente.

O modelo melhor ajustado ficou com duas variáveis que se referem a renda: P_R_2SM e RM_RESP, porém considerou-se que a primeira tem uma relação indireta, por se tratar de uma proporção e de representar uma parte da população bem definida. Considerando que a variável P_R_1SM também foi inserida nas análises iniciais, mas não permaneceu no modelo final por se tornar estatisticamente não significativa, pode-se indicar que a variável P_R_2SM seja um ponto de corte importante no estudo da tuberculose. Alguns estudos^{11,23} indicaram uma possível associação entre a infecção por tuberculose e fazer parte de uma classe social média/baixa, principalmente quando observamos a infecção da tuberculose associada ao vírus HIV e às formas resistentes¹¹. Um estudo desenvolvido em Olinda²³ mostrou que a relação entre a taxa de incidência de tuberculose não é diretamente relacionada à população de baixa renda, principalmente quando já existam condições que favorecem a propagação da doença mesmo em locais onde a população possui maior renda.

Corroborar essa ideia os resultados desse estudo que mostrou que os bairros do Centro e adjacências, se estendendo pelos bairros de Benfica, Manguinhos, Maré, Penha, Vila da Penha e parte da zona oeste (Figura 1) apresentaram as maiores taxas de incidência após o alisamento bayesiano e são áreas onde a renda média dos responsáveis está em torno de 1.600 reais.

Os três aglomerados formados por: 1) parte bairros do Centro, São Cristóvão, Tijuca, Catumbi, Glória; 2) parte dos bairros da Penha e Penha Circular; 3) por setores dos bairros de Pedra de Guaratiba e Barra de Guaratiba (Figura 2) são compostos por setores com alta taxa de incidência de TB e podem ser consideradas como áreas de risco para a transmissão da doença.

A variável P_P_MS permaneceu no modelo final. Pela literatura, parece incoerente quando se trata da tuberculose, porém, a pessoa que mora sozinha pode ter maior risco em função da não aderência ao tratamento¹¹. Gonçalves et al.²⁴ em um estudo realizado em Pelotas identificaram que a participação familiar é importante na aderência e permanência do paciente no tratamento da tuberculose.

A utilização de setores censitários como unidade de análise ao mesmo tempo em que traz o problema da instabilidade estatística das taxas de incidência por conta das pequenas populações, traz também a possibilidade de se trabalhar com populações mais homogêneas que os bairros, por exemplo. No estudo da tuberculose no município

do Rio de Janeiro, a utilização dos setores censitários como unidade de análise se mostrou bastante importante. As análises dos aglomerados formados pela variável log da taxa de incidência após o alisamento bayesiano permitem estudar a ocorrência da doença com maior detalhe. Entretanto uma limitação nesse estudo se refere a perda do georreferenciamento a partir do endereço descritos em Magalhães et al.¹⁵. Não foi possível determinar o percentual dos casos não localizados em áreas carentes, mas a experiência tem mostrado que a grande maioria dos casos não localizados possivelmente ocorre em favelas. Muitas vezes, o endereço fornecido pelo paciente é o da entrada da favela ou algum ponto de referência dentro da comunidade. Isso faz com que alguns setores censitários recebem pontos além ou aquém do esperado.

A perda no georreferenciamento pode ter retirado do estudo casos referentes à população mais carente descritos em Magalhães et al.¹⁵, no caso da tuberculose em particular é de se esperar que o impacto na perda seja maior nesse grupo. Além disto, um estudo sobre óbito atribuídos à tuberculose no Estado do Rio de Janeiro, realizado em alguns hospitais por Selig et al.²⁵, mostrou que a hospitalização no momento da notificação

ocorreu em 3495 (21%), o que mostra problemas no acesso, e que apenas 41,4% dos óbitos ocorreram em pacientes notificados, mostra que a tuberculose no Rio de Janeiro é subnotificada e mal detectada. Piller⁴, num outro estudo desenvolvido no Rio de Janeiro avaliou as fontes notificadoras e verificou que 26% dos casos ainda são notificados em hospitais, quando esses casos deveriam ter sido detectados e tratados precocemente pela atenção básica. Este achado reforça a evidência de que a população mais afetada pela doença é formada por indivíduos com dificuldades no acesso aos serviços de saúde e, portanto, mais carentes. Esta população é justamente aquela que mora em lugares de difícil localização e consequentemente com maior probabilidade de ser perdida no presente estudo. Portanto é necessário olhar o resultado do modelo com a ressalva de que algumas variáveis podem não ter tido associação com a taxa de tuberculose simplesmente porque a população com piores condições de vida pode ter sido involuntariamente excluídas das análises. Entretanto, de forma geral, os métodos estatísticos aplicados neste estudo apresentaram-se eficientes na identificação de padrões espaciais da tuberculose e na definição de alguns determinantes para ocorrência da doença.

Colaboradores

MAFM Magalhães trabalhou na concepção, delineamento, análise, interpretação dos dados, redação do artigo e revisão crítica. RA Medronho trabalhou na concepção, delineamento, análise e na revisão crítica.

Referências

- World Health Organization. Global tuberculosis control. WHO report 2013. Geneva: WHO; 2013. [acessado 2014 mar 6]. Disponível em: http://www.who.int/tb/publications/global_report/2013/en/
- Brasil. Ministério da Saúde (MS). *Guia de vigilância epidemiológica*. 6ª ed. Brasília: MS; 2005.
- Buss PM, Pellegrini Filho A. A saúde e seus determinantes sociais. *Physis* 2007; 17(1):77-93.
- Piller RVB. Epidemiologia da tuberculose. *Pulmão RJ* 2012; 21(1):4-9.
- Ruffino-Netto A. Tuberculose: a calamidade negligenciada. *Revista da Sociedade Brasileira de Medicina Tropical* 2002; 35(1):51-58.
- Secretária de Estado de Saúde do Rio de Janeiro, Subsecretária de Vigilância em Saúde, Sistema de Informação de Agravos de Notificação (SINAN). [acessado 2015 nov 5]. Disponível em: <http://sistemas.saude.rj.gov.br/tabnet/deftohtm.exe?sinan/tuberculose.def>
- Pinheiro RS, Oliveira GP, Oliveira PB, Coeli CM. Melhoria da qualidade do sistema de informação para a tuberculose: uma revisão da literatura sobre o uso do linkage entre bases de dados. In: Brasil. Ministério da Saúde (MS). *Saúde Brasil 2010: uma análise da situação de saúde e de evidências selecionadas de impacto de ações de vigilância em saúde*. Brasília: MS; 2010.
- Barcellos C, Bastos FI. Geoprocessamento, ambiente e saúde: uma união possível? *Cad Saude Publica* 1996; 12(3):389-397.
- Medronho RA, Werneck GL. Análise de dados espaciais em saúde. In: Medronho RA, Carvalho DM, Bloch KV, Luiz RR, Werneck GL. *Epidemiologia*. Rio de Janeiro: Atheneu; 2009. p. 493-511.
- Prado FA, Berveglieri A, Tachibana VM, Imai NN. Aplicação e análise de modelos de regressão clássica e espacial para os distritos da cidade de São Paulo. *III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação*; 27-30 jul. 2010; Recife - PE, p. 1-8.
- Pereira AGL. *Distribuição espacial da tuberculose e sua correlação com variáveis socioeconômicas no município do Rio de Janeiro nos anos de 2004 a 2006* [dissertação]. Rio de Janeiro: UFRJ; 2010.
- Sobral de Almeida A, de Andrade Medronho R, Ortiz Valencia LI. Análise espacial da dengue e o contexto socioeconômico no município do Rio de Janeiro, RJ. *Rev Saude Publica* 2009; 43(4):666-673.
- Acosta LMW. *O mapa de Porto Alegre e a tuberculose: distribuição espacial e determinantes sociais* [dissertação]. Porto Alegre: UFRGS; 2008.
- Instituto Brasileiro de Geografia e Estatística (IBGE). Censo demográfico, 2010. [acessado 2012 jul 2]. Disponível em <http://www.ibge.gov.br>
- Magalhães MAFM, Pascoal VM, Medronho RA. Avaliação do dado sobre endereço no Sistema de Informação de Agravos de Notificação utilizando georreferenciamento em nível local de casos de tuberculose por dois métodos no município do Rio de Janeiro. *Cad Saúde Coletiva* 2014; 22(2):192-199.
- Bailey TC, Gatrell AC. *Interactive spatial data analysis*. Essex: Longman Scientific and Technical; 1995.
- Assunção RM, Barreto SM, Guerra HL, Sakurai E. Figuras de taxas epidemiológicas: uma abordagem Bayesiana. *Cad Saude Publica* 1998; 14(4):713-723.
- Santos AE, Rodrigues AL, Lopes DL. Aplicações de estimadores bayesianos empíricos para análise espacial de taxas de mortalidade. VII Simpósio Brasileiro de Geoinformática – GEOINFO; 20-23 nov. 2005; Campos do Jordão - SP. [acessado 2011 jul 31]. Disponível em: <http://www.geoinfo.info/geoinfo2005/papers/P63.PDF>.
- Câmara G, Carvalho MS, Cruz OG, Correa V. Análise espacial de áreas. In: Druck S, Carvalho MS, Câmara G, Monteiro AMV, organizadores. *Análise espacial de dados geográficos*. Brasília: Embrapa; 2004. p. 107-151.
- Anselin L. Exploring Spatial Data with GeoDaTM: A Workbook. Centre for Spatially Integrated Social Science. 2005. [cited 2011 Jul 31]. Available from: <http://geodacenter.asu.edu/software/downloads>.
- Cunha MD. *Estatística espacial na investigação epidemiológica de fatores associados à detecção de casos de hanseníase no Rio de Janeiro* [tese]. Rio de Janeiro: Fiocruz; 2012.
- Morais Neto OL, Barros MBA, Martelli CMT, Silva AS, Cavenaghi SM, Siqueira Júnior JB. Diferenças no padrão de ocorrência da mortalidade neonatal e pós-neonatal no Município de Goiânia, Brasil, 1992-1996: análise espacial para identificação das áreas de risco. *Cad Saude Publica* 2001; 17(5):1241-1250.
- Souza WV, Ximenes R, Albuquerque MFM, Lapa TM, Portugal JL, Lima MLC, Martelli CMT. The use of socioeconomic factors in mapping tuberculosis risk areas in a city of northeastern Brazil. *Rev Panam Salud Publica* 2000; 8(6):403-410.
- Gonçalves H, Costa JSD, Menezes AMB, Knauth D, Leal OF. Adesão à terapêutica da tuberculose em Pelotas, Rio Grande do Sul: na perspectiva do paciente. *Cad Saude Publica* 1999; 15(4):777-787.
- Selig L, Belo M, Cunha AJLA, Teixeira EG, Brito R, Luna AL, Trajman A. Óbitos atribuídos à tuberculose no Estado do Rio de Janeiro. *Jornal Brasileiro de Pneumologia* 2004; 30(4):327-334.

Artigo apresentado em 19/05/2015

Aprovado em 16/03/2016

Versão final apresentada em 18/03/2016

