

Análise espacial de dados de contagem com excesso de zeros aplicado ao estudo da incidência de dengue em Campinas, São Paulo, Brasil

Spatial analysis of counting data with excess zeros applied to the study of dengue incidence in Campinas, São Paulo State, Brazil

Análisis espacial de datos de conteo con exceso de ceros, aplicado al estudio de la incidencia de dengue en Campinas, São Paulo, Brasil

José Vilton Costa ¹
Líciana Vaz de Arruda Silveira ²
Maria Rita Donalísio ³

Resumo

A incidência de dengue ocorre predominantemente em áreas urbanas das cidades. Identificar o padrão de distribuição espacial da doença no nível local contribui na formulação de estratégias de controle e prevenção da doença. A análise espacial de dados de contagem para pequenas áreas comumente viola as suposições dos modelos tradicionais de Poisson, devido à quantidade excessiva de zeros. Neste estudo, comparou-se o desempenho de quatro modelos de contagem utilizados no mapeamento de doenças: Poisson, Binomial negativa, Poisson inflacionado de zeros e Binomial negativa inflacionado de zeros. Os métodos foram comparados em um estudo de simulação. Os modelos analisados no estudo de simulação foram aplicados em um estudo ecológico espacial, aos dados de dengue agregados por setores censitários, do Município de Campinas, São Paulo, Brasil, em 2007. A análise espacial foi conduzida por modelos hierárquicos bayesianos. O modelo de Poisson inflacionado de zeros apresentou melhor desempenho para estimar o risco relativo de incidência de dengue nos setores censitários.

Análise Espacial; Dengue; Controle de Doenças Transmissíveis

¹ Programa de Pós-graduação em Demografia, Universidade Federal do Rio Grande do Norte, Natal, Brasil.

² Instituto de Biociências, Universidade Estadual Paulista Julio de Mesquita Filho, Botucatu, Brasil.

³ Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, Brasil.

Correspondência

J. V. Costa
Programa de Pós-graduação em Demografia, Universidade Federal do Rio Grande do Norte.
Av. Salgado Filho 3000, Campus Universitário, Natal, RN 59078-970, Brasil.
josevilton@gmail.com

Introdução

A incidência de dengue ocorre predominantemente em áreas urbanas das cidades ¹, cuja transmissão é influenciada por um conjunto complexo de fatores ², incluindo aspectos ambientais, clima, comportamento humano, a forma de combate ao mosquito vetor e a imunidade da população humana para os diferentes sorotipos do vírus ^{3,4,5,6}. A precariedade das condições de saneamento básico de algumas cidades, particularmente o abastecimento irregular de água e a coleta inadequada de lixo, têm favorecido o aumento de criadouros potenciais do principal mosquito vetor e contribuído para a incidência da doença ⁷.

Compreender a associação entre fatores ambientais e dengue é essencial para o desenvolvimento de ações de prevenção e controle da doença ⁷. A incidência de inúmeras doenças infecciosas, particularmente aquelas transmitidas por vetores, geralmente exibem padrões espaciais ⁸, os quais podem ser resultantes das interações entre patógeno/hospedeiro/ambiente ⁹, em locais e em períodos específicos ¹⁰. A natureza, intensidade, direção e o resultado dessas interações específicas dependem da forma como as populações envolvidas ocupam e se apropriam do espaço ⁴.

Assim, considera-se apropriado analisar a distribuição espacial dos casos de dengue e sua associação com fatores socioambientais no processo de desenvolvimento da doença. A utilização de ferramentas de epidemiologia espacial ¹¹ tem contribuído para estimar e representar a incidência de dengue em diferentes escalas geográficas, bem como identificar fatores de risco associados à sua distribuição no espaço geográfico ^{12,13,14,15,16,17}.

Em relação à análise espacial e mapeamento de doenças com dados agregados, quanto menor o nível de agregação das unidades espaciais, maior é a capacidade para captação das interações que ocorrem no território, favorecendo a identificação de microáreas prioritárias para as ações de vigilância epidemiológica. A opção por polígonos espaciais com tamanhos mínimos pode resultar em um excesso de áreas sem notificação do evento de interesse, ou seja, com contagens da doença igual a zero. Considerando-se que o modelo de regressão de Poisson é comumente empregado no mapeamento de doenças ¹¹, alguns autores têm chamado a atenção para os problemas (por exemplo, superdispersão) e limitações (imprecisão dos estimadores) que o excesso de zeros traz no processo de modelagem estatística, indicando a necessidade de se empregar modelos que considerem estas características dos dados ^{18,19,20,21,22,23}.

O presente trabalho avaliou, por meio de simulações, o desempenho do modelo de regressão de Poisson em relação aos modelos Binomial negativa (BN), Poisson inflacionado de zeros e Binomial negativa inflacionada de zeros, na análise de dados espaciais agregados, com a presença de excesso de zeros. Com base em registros de dengue, analisou-se como exemplo empírico a distribuição espacial dos casos e a relação entre incidência da doença e condições socioambientais, no ano de 2007, no Município de Campinas, Estado de São Paulo.

Métodos

Desenho e área de estudo

Realizou-se um estudo de simulação para ilustrar o potencial dos modelos inflacionados de zeros (ZI, do inglês *zero-inflated*) em modelar dados caracterizados por excesso de zeros, na presença de diferentes proporções de zeros, comparando-se ao desempenho dos modelos de Poisson e BN. Conduziu-se um estudo ecológico sobre a incidência de dengue em setores censitários em Campinas, notificados na epidemia de 2007. Baseando-se em casos registrados e georreferenciados foram comparados modelos de análise mais adequados para se enfrentar a frequente ocorrência de áreas sem registro de casos.

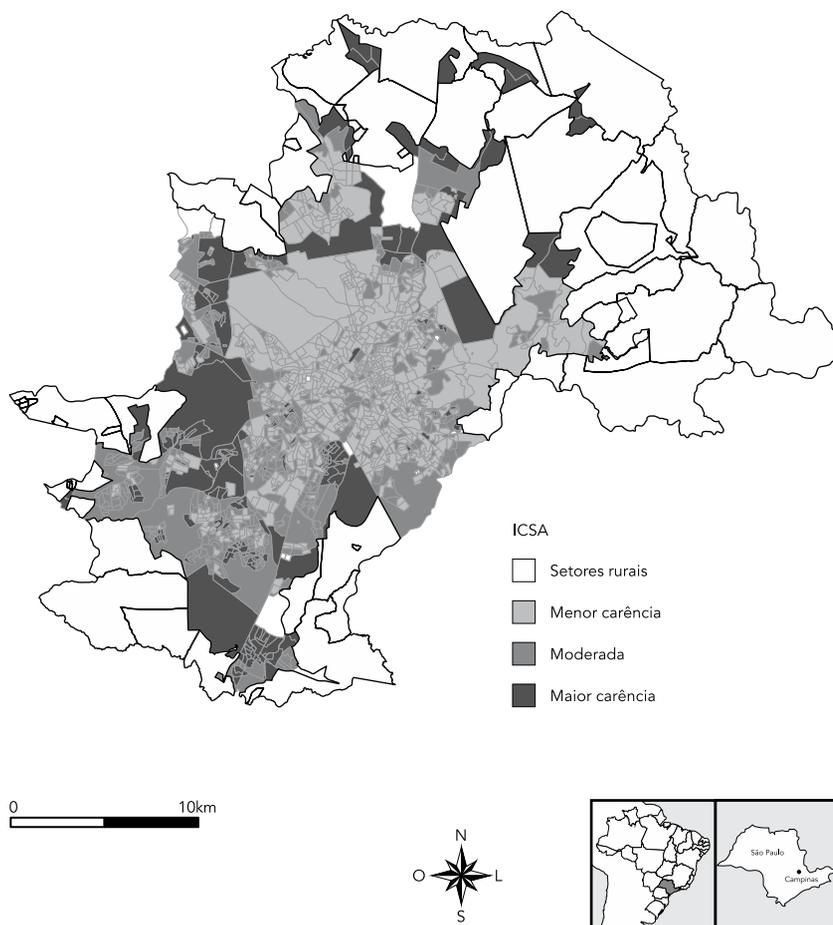
No estudo de simulação e aplicação aos dados empíricos, considerou-se o espaço urbano do Município de Campinas, adotando-se por unidade de análise os setores censitários (Figura 1). A escolha do setor censitário como unidade espacial de análise permite representar os grupos populacionais e socioeconômicos (variáveis de exposição) no nível mais desagregado de dados, coletados de forma sistemática de âmbito nacional. Segundo dados do *Censo Demográfico* de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE; <http://www.ibge.gov.br>), o município era subdividido em 1.749 setores, 1.695 (97%) destes localizados em áreas urbanas, cuja população era de 1.061.540 habitantes. A área mediana dos setores censitários correspondia a 0,080km² (intervalo interquartil: 0,041-0,144km²), com densidade populacional mediana nos setores de 7.797 habitantes/km² (intervalo interquartil: 4.390-13.450 habitantes/km²). A população majoritariamente (98,3%) urbana correspondia a 1.061.540 habitantes.

Mapeamento de doenças

A incidência Y_i de uma dada doença, distribuída na região i ($i = 1, 2, \dots, N$), comumente é modelada

Figura 1

Região de estudo, segundo a distribuição do índice de carência socioambiental (ICSA).



por um modelo de Poisson, com $Y_i | \theta_i \sim \text{Poisson}(\lambda_i = E_i \theta_i)$ ^{11,24}. Considera-se que E_i é o número de casos esperados, e θ_i o risco relativo desconhecido da incidência da doença na i -ésima área de estudo. O risco relativo é estimado assumindo-se a função de ligação logarítmica para o preditor linear:

$$\eta_i = \log(\lambda_i) = \log(E_i) + \alpha + (X_i)\beta + \psi_i + \Phi_i$$

Em que η_i é um preditor linear para estimar o risco relativo θ_i , α é o intercepto comum a todas as áreas, X é uma matriz de covariáveis, β é o vetor dos parâmetros associado à covariável, ψ_i descreve a componente aleatória espacialmente estruturada (autocorrelação espacial local) do risco relativo e Φ_i a componente aleatória não estruturada espacialmente ²⁴.

O estimador de máxima verossimilhança de λ_i , denotado por $\lambda = SIR = Y_i / E_i$, coincide com

a taxa de incidência padronizada (SIR, do inglês *standardized incidence ratio*), uma taxa bruta que apresenta instabilidade frente a pequenos valores observados ou para populações pequenas ¹¹.

O modelo de Poisson considera a hipótese de igualdade nos valores da média e variância ²⁵. Entretanto, quando se estuda a distribuição dos casos de uma doença em pequenas áreas geográficas, é comum identificar áreas sem nenhuma ocorrência de casos. O excesso de zeros frequentemente resulta em superdispersão dos dados (a variância é maior que a média), comprometendo a aplicabilidade do modelo de Poisson ¹⁹.

Observando-se a presença de superdispersão nos dados, é importante incluir no modelo componentes que ajustem esta superdispersão. O modelo BN $Y_i \sim BN(\lambda_i, \tau)$, em que τ é um parâmetro de escala e representa a superdispersão

dos dados, é a primeira opção para modelar os dados na presença de superdispersão. Contudo, o modelo BN não ajusta adequadamente os dados quando a variabilidade está relacionada ao excesso de zeros na distribuição destes dados¹⁹.

Os modelos ZI, definidos com base na mistura de uma distribuição de Bernoulli e uma distribuição discreta (por exemplo, Poisson ou BN), representam uma abordagem flexível para contornar este problema de excesso de zeros, quando a contagem de zeros observada nos dados excede a contagem predita por um modelo de contagem.

Modelos inflacionados de zeros

Dois modelos comumente utilizados em dados com excesso de zeros são o Poisson inflacionado de zeros (ZIP) e o Binomial negativo inflacionado de zeros (ZINB)^{18,19,20,21,22,23}. O excesso de zeros é ajustado assumindo-se que os zeros surgem baseando-se em dois distintos processos.

O primeiro, ocorre segundo um processo de Bernoulli com probabilidade p_i , e produz apenas verdadeiros zeros (ou zeros estruturais). O segundo, denominado falsos zeros (ou zeros amostrais), com probabilidade $(1 - p_i)$ é produzido segundo uma distribuição de Poisson ou BN, com média λ_i ^{18,19,20,21,22,23}. A probabilidade total de zeros é a combinação das probabilidades dos dois processos, ponderada pela probabilidade do respectivo processo, ou seja, $Pr(Y_i = 0) = p_i + (1 - p_i) Pr(Z_i = 0)$, em que $Pr(Z_i = 0)$ denota a probabilidade de Poisson ou BN para a ocorrência de um valor zero. A probabilidade de contagens positivas (maior que zero) é definida por $Pr(Y_i = y_i) = (1 - p_i)Pr(Z_i = y_i)$, em que $Pr(Z_i = y_i)$ representa uma probabilidade de Poisson ou BN para a realização de contagens positivas, em que $0 < p_i < 1$.

Segundo a descrição anterior, o modelo ZIP apresenta a seguinte formulação:

$$Y_i \sim ZIP(\lambda_i, p_i) ; Y_i \sim \begin{cases} 0 & , \text{com probabilidade } p_i \\ Pois(\lambda_i) & , \text{com probabilidade } (1-p_i) \end{cases}$$

Dessa forma, com probabilidade p_i a variável resposta assume o valor zero e com probabilidade $(1 - p_i)$ assume o valor de uma variável aleatória com distribuição de Poisson com média λ_i , conforme descrito anteriormente. A média de Y_i é $E(Y_i) = (1 - p_i)\lambda_i$ e variância igual a $Var(Y_i) = (1 - p_i)\lambda_i + p_i$. Observa-se que a variância da mistura é maior que a média da distribuição. Quanto maior a probabilidade do excesso de zeros, maior a variância da variável. Na medida em que p_i se aproxima de zero, a variância se aproxima de λ_i , ou seja, voltamos a lidar somente com uma distribuição Poisson padrão.

O modelo ZINB, por sua vez, é expresso por:

$$Y_i \sim ZINB(\lambda_i, \tau, p_i) \text{ e } Y_i \sim \begin{cases} 0 & , \text{com probabilidade } p_i \\ BN(\lambda_i, \tau) & , \text{com probabilidade } (1 - p_i) \end{cases}$$

A média e a variância de Y_i são: $E(Y_i) = (1 - p_i)\lambda_i$ e $Var(Y_i) = (1 - p_i)\lambda_i + p(1 - p_i)\lambda_i^2$. A distribuição ZINB aproxima-se da ZIP quando $\tau \rightarrow 0$ e aproxima-se da binomial negativa quando $p_i \rightarrow 0$. Se ambos $1/\tau$ e p_i convergem para zero, então a distribuição ZINB é reduzida a uma distribuição de Poisson padrão.

Em ambos os modelos ZIP e ZINB, o risco relativo é estimado por um preditor linear médio (η_i), conforme descrito previamente para o modelo de Poisson.

Modelagem bayesiana

Os modelos de contagem apresentados previamente foram analisados segundo a metodologia de modelo hierárquico inteiramente bayesiano²⁴. Sob a abordagem bayesiana, prioris são especificadas para os parâmetros do modelo. O efeito espacial ψ_i foi modelado por um processo de Campos Aleatórios Markovianos Gaussianos (CAMG)²⁶, assumindo priori com estrutura autorregressiva condicional (CAR). A relação espacial entre os setores foi modelada por uma matriz de vizinhança (adjacência) binária, adotando-se o valor 1 para cada elemento quando os setores são vizinhos, e 0 em caso contrário. Para os parâmetros do modelo (α e β), atribuiu-se prioris com distribuição normal com média zero e variância grande ($N[0, 10.000]$). Para a precisão dos efeitos aleatórios (Φ_i e ψ_i), seguiu-se a recomendação da literatura²⁷, adotando-se *hyperprioris* de uma distribuição Gama (0,5; 0,0005).

A análise bayesiana para estimativa dos parâmetros foi conduzida pelo método INLA (*Integrated Nested Laplace Approximation*). A abordagem INLA destaca-se por empregar aproximações de Laplace para aproximar as distribuições a posteriori via métodos computacionais determinísticos, sem a necessidade de utilizar simulações. As estimativas obtidas são eficientes e computacionalmente mais rápidas em relação aos métodos tradicionais de Monte Carlo via cadeias de Markov (MCMC)²⁶.

Utilizou-se o critério DIC (*Deviance Information Criterion*)²⁸ e *Deviance* baseada na Ordenada Preditiva Condicional (CPO – *Conditional Predictive Ordinate*)²⁹ para identificar o modelo com o melhor ajuste aos dados. Em relação ao DIC, quanto menor o valor estimado, melhor é o ajuste do modelo. Considerou-se que uma diferença menor que cinco entre os valores do DIC de dois modelos ($D = DIC1 - DIC2$) não é significativa²⁸.

O critério CPO avalia o poder preditivo do modelo e consiste em densidades de validação cruzada, isto é, distribuições preditivas condicionadas sobre o conjunto de dados observados, exceto a k -ésima informação. O CPO proporciona uma medida de ajuste para cada observação individualmente, e comparações entre os modelos são realizadas usando-se a média do logaritmo do CPO (-média(log(CPO)) ou LCPO) ²⁹.

As análises foram realizadas com o software R, versão 3.2.0 (The R Foundation for Statistical Computing, Viena, Áustria; <http://www.r-project.org>), utilizando-se o pacote R-INLA.

Estudo de simulação

Adotando-se a malha urbana dos setores censitários do Município de Campinas em 2010, simulou-se 100 conjuntos de dados, para cinco diferentes cenários em relação à proporção (p) de setores censitários com contagem zero de casos de dengue: $p = 0,1$; $p = 0,2$; $p = 0,4$; $p = 0,6$ e $p = 0,8$. Os setores com contagem igual a zero foram selecionados aleatoriamente, de acordo com a proporção (p) de regiões com contagens zero da doença. Com base em uma distribuição uniforme, gerou-se um número aleatório $r_i \sim U(0,1)$ para todas os setores. Selecionamos os setores com contagem zero se $r_i < p$ e para os setores com $r_i \geq p$, as contagens observadas (O_i) foram geradas por uma distribuição multinomial, conforme descrito por Richardson et al. ³⁰:

$$O_i \sim \text{Multinomial} \left(N_i; \frac{E_i \theta_i}{\sum_j E_j \theta_j} \right)$$

Em que N_i é o número total de casos esperados. Esse procedimento assegura que em cada conjunto de dados simulado a soma dos casos observados permanece igual à soma dos casos esperados. Em relação ao risco, adotou-se um modelo básico, com a hipótese de risco constante e igual a um ($\theta_i = \theta = 1$) sobre a região de estudo.

Dados de dengue em Campinas

Analisou-se os casos de dengue autóctones, ocorridos nos setores censitários urbanos da cidade de Campinas, registrados entre 1º de janeiro e 31 de dezembro de 2007. Adotou-se esse período de estudo por ser aquele com maior incidência da doença com dados disponíveis e georreferenciados na base cartográfica da cidade, historicamente é o período com a segunda maior incidência (11.442 casos), superado pelo ano de 2014 com 42.109 casos. O banco de dados com os casos de dengue foi obtido do Sistema de Informação de Agravos de Notificação (SINAN), corri-

gido e fornecida nova versão pela Coordenadoria de Vigilância e Saúde da Secretária Municipal de Saúde do município (Covisa/SMS-Campinas). Os dados populacionais utilizados para a construção das taxas de incidência foram obtidos do *Censo Demográfico* de 2010.

Na análise empírica dos dados, construiu-se uma variável sintética, denominada índice de carência socioambiental (ICSA), segundo três classes de carência socioambiental: maior; moderada e menor. Essa variável foi obtida baseando-se nos seguintes indicadores provenientes do *Censo Demográfico* de 2010: uma variável relacionada à renda domiciliar (proporção de domicílios com renda *per capita* até meio salário mínimo), três relacionadas à inadequação de serviços urbanos nos domicílios (proporção de domicílios com inadequação de: esgotamento sanitário, destinação do lixo e forma de abastecimento de água) e três variáveis relacionadas à infraestrutura urbana no entorno dos domicílios (proporção de domicílios com: ausência de pavimentação, esgoto a céu aberto e lixo acumulado no entorno).

Esses indicadores foram submetidos à técnica de componentes principais (ACP). O ICSA é a média aritmética dos indicadores primários, após cada um ter sido padronizado em um intervalo de valores entre 0 (zero) e 1 (um), de acordo com os valores mínimos e máximos observados para cada setor. Posteriormente, a técnica de análise de agrupamento (*cluster analysis*), via o método de k -médias (*k-means*), foi utilizada para a obtenção dos três estratos de carência socioambiental dos setores censitários urbanos da cidade de Campinas (Figura 1).

Adicionalmente, foi estimada e mapeada a probabilidade a posteriori do risco relativo (RR) de dengue exceder 1,2 ($Pr[RR > 1,2 | Y]$), obtendo-se assim, uma medida de evidência (estatística) de excesso de risco em cada setor censitário. O valor de $RR = 1,2$ reflete o risco médio de incidência de dengue, observado na região urbana de Campinas. Estimativas de probabilidades a posteriori superiores a 0,8 representam evidências de excesso de risco nos setores ³⁰, permitindo identificar possíveis efeitos de agregação espacial.

Resultados

Estudo de simulação

Os resultados deste estudo estão resumidos na Tabela 1, em que é apresentada a comparação do desempenho do modelo padrão de Poisson em relação aos modelos BN, ZIP e ZINB, nos diferentes cenários de p . Nossos resultados mostram que, segundo o valor médio do DIC, o modelo

Tabela 1

Modelos bayesianos analisados e critério de ajuste *Deviance Information Criterion* (DIC) e *deviance* baseada na ordenada preditiva condicional (LCPO), aplicado ao estudo do dengue em Campinas, São Paulo, Brasil, 2007.

Distribuição	Modelos	DIC	LCPO
ZIP	1. ψ_i	8.025,6	2,74
	2. Φ_i	8.003,6	2,75
	3. $\psi_i + \Phi_i$	7.966,8	2,65
	4. ψ_i + covariável (ICSA)	8.034,2	2,69
	5. $\psi_i + \Phi_i$ + covariável (ICSA)	7.964,3	2,66
ZINB	6. ψ_i	8.864,7	4,45
	7. Φ_i	8.827,2	4,52
	8. $\psi_i + \Phi_i$	8.833,1	4,15
	9. ψ_i + covariável (ICSA)	8.753,3	4,22
	10. $\psi_i + \Phi_i$ + covariável (ICSA)	8.827,4	4,16

ICSA: indicador de condições socioambientais; ZINB: modelo binomial negativo inflacionado de zeros; ZIP: modelo de Poisson inflacionado de zeros; ψ_i : efeito espacial estruturado; Φ_i : efeito espacial não estruturado.

ZIP foi aquele que apresentou o menor valor deste critério para todos os cenários de p , sendo que o modelo ZINB foi o segundo com melhor ajuste. Diferenças substanciais foram observadas na comparação dos valores do DIC do modelo Poisson em relação aos modelos ZIP e ZINB. Conforme se aumentou os valores da proporção p de zeros, maiores foram as diferenças entre os critérios desses modelos.

Em relação ao LCPO, os resultados obtidos seguiram na mesma direção dos obtidos com o DIC, apontando melhor capacidade preditiva para o modelo ZIP.

Aplicação aos dados empíricos

Um total de 11.442 casos de dengue foi confirmado e notificado ao SINAN, na cidade de Campinas, em 2007. Após revisão, retirada de duplicações, análise de consistência do banco e correção de ortografia de endereços registrados no SINAN³¹, 8.956 (77%) foram georreferenciados sobre a malha digital dos setores censitários do município, sendo 8.912 (99,5%) registrados nos setores urbanos. Entretanto, no processo de modelagem, 31 setores foram excluídos por inexistência dos dados populacionais do *Censo Demográfico* de 2010, ocasionando a exclusão de 5 casos nestes setores. O banco de dados final para modelagem foi constituído de 8.907 casos de dengue.

Na Figura 2, observa-se a distribuição de frequência das contagens dos casos de dengue. O histograma mostra alta variação na distribuição de frequência, caracterizada por uma assimetria positiva e com excesso de áreas sem registros

de casos da doença. Verificou-se que em 25,8% dos setores censitários não foram registrados nenhum caso de dengue. A média e o desvio padrão (DP) do número de casos da doença e da SIR, por setor censitário, foram iguais a 5,3 (DP = 10,2) casos e 1,03 (DP = 2,03), respectivamente. Essas estatísticas descritivas, em que a variância das contagens dos casos é maior que a sua média, juntamente com a distribuição de frequência observada no histograma da Figura 2, indicam a presença de superdispersão na distribuição das contagens de casos de dengue.

A análise da autocorrelação espacial, segundo o índice de Moran bayesiano empírico³², foi igual a 0,19 (valor de $p < 0,001$), indicando autocorrelação espacial positiva, ou seja, as taxas de incidência de dengue nos setores censitários não são independentes das taxas apresentadas em setores vizinhos.

Considerando-se as características dos dados, com elevada proporção de setores censitários com contagens zero da doença, e os resultados do estudo de simulação, foram ajustadas estimativas do risco relativo de dengue, adotando-se os modelos ZIP e ZINB. Estimaram-se cinco modelos para cada uma das distribuições, ajustando-se a presença de efeitos aleatórios espacialmente estruturado (ψ_i) e não estruturado (Φ_i), e o impacto da covariável ICSA.

A Tabela 2 apresenta uma análise comparativa das medidas de ajuste dos modelos analisados (ZIP e ZINB), com seus respectivos valores do critério DIC e LCPO. Os cinco modelos ZIP apresentaram os menores valores do DIC, ou seja, a distribuição ZIP ajustou-se melhor aos

Figura 2

Distribuição de frequência dos casos de dengue por setor censitário.

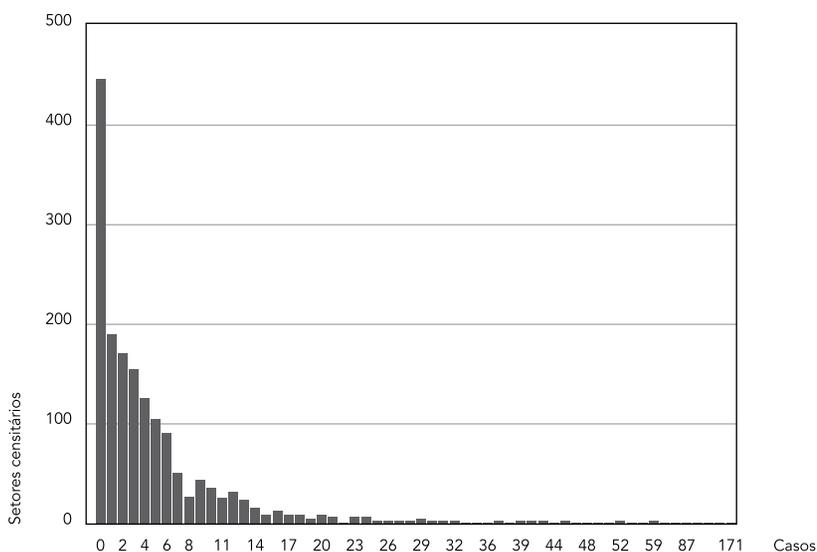


Tabela 2

Medidas de ajuste dos modelos *Deviance Information Criterion* (DIC) e *deviance* baseada na ordenada preditiva condicional (LCPO), para diferentes proporções (p) de setores com contagem igual a zero, aplicado ao estudo de simulação.

p	Modelo	DIC		LCPO	
		Média	DP	Média	DP
0,1	Poisson	9.034,7	114,9	2,92	0,04
	Binomial negativa	9.081,5	105,2	2,73	0,03
	ZIP	8.605,3	93,9	2,59	0,04
	ZINB	8.959,1	99,7	2,72	0,04
0,2	Poisson	10.543,0	187,5	2,75	0,06
	Binomial negativa	9.854,3	91,7	2,76	0,05
	ZIP	8.409,5	66,7	2,63	0,02
	ZINB	9.715,1	74,5	2,69	0,05
0,4	Poisson	14.381,1	337,7	4,32	0,10
	Binomial negativa	9.832,4	95,7	2,92	0,09
	ZIP	8.213,5	61,2	2,47	0,08
	ZINB	9.514,9	68,4	2,62	0,08
0,6	Poisson	20.368,0	455,7	6,13	0,14
	Binomial negativa	8.285,8	130,0	2,63	0,13
	ZIP	7.852,3	72,6	2,27	0,06
	ZINB	8.045,5	85,5	2,45	0,15
0,8	Poisson	30.996,6	973,6	9,33	0,29
	Binomial negativa	9.932,6	200,2	1,98	0,08
	ZIP	6.649,4	89,0	1,48	0,04
	ZINB	7.588,9	135,7	1,53	0,09

DP: desvio padrão; ZINB: modelo binomial negativo inflacionado de zeros; ZIP: modelo de Poisson inflacionado de zeros.

dados. O modelo 5, ajustado por efeitos aleatórios espacialmente estruturado (ψ_i) e não estruturado (ϕ_i), adicionado ao efeito da covariável ICSPA, apresentou o menor valor do DIC (7964,30) entre todos os modelos. Entretanto, ao comparar a diferença entre o DIC do modelo 5 com o respectivo DIC do modelo 3 (7966,81), essa diferença foi inferior a cinco. Segundo o critério adotado neste trabalho e considerando-se a complexidade dos modelos, essa diferença nos valores do DIC é insuficiente para considerá-la significativa em favor do modelo 5. O ajuste segundo o critério LCPO, foi semelhante ao observado com o DIC. Dessa forma, considerou-se o modelo 3, modelado por uma distribuição ZIP e ajustado para os efeitos aleatórios estruturado espacialmente e não estruturado, aquele com melhor ajuste entre os dez modelos. Desse ponto em diante, a análise do risco relativo refere-se ao modelo 3.

A Figura 3a apresenta o mapa com a SIR, e na Figura 3b o mapa ilustra as estimativas *a posteriori* do RR, as quais variaram de 0,22 a 36,52, com valor médio de 1,27 e DP de 1,55. Contrastando-se a distribuição espacial da SIR com a distribuição do RR, nota-se que há uma redução de áreas com $RR < 0,5$, e em função do efeito de suavização do modelo espacial, observa-se um estreitamento das estimativas do risco em direção ao risco médio.

A Figura 3c ilustra as probabilidades do RR ajustado ser maior do que 1,2 ($RR > 1,2$), ou seja, dos setores censitários apresentarem RR de dengue 20% maior que a média dos 1.664 setores urbanos. Observou-se que 566 (43,6%) setores censitários apresentaram excesso de risco, com probabilidade acima de 0,8.

Discussão

Neste estudo, no contexto de análise espacial de dados agregados, foram avaliados quatro modelos de regressão para dados de contagem: Poisson, BN, ZIP e ZINB, na presença de superdispersão nos dados. A superdispersão nos dados não apresenta parâmetro de ajuste no modelo de Poisson, é modelada por um parâmetro adicional no modelo BN e por uma proporção extra de zeros, utilizando-se modelos inflacionados de zeros (ZIP e ZINB). Adicionalmente, efeitos aleatórios espaciais foram incluídos nos modelos para ajustar heterogeneidade não observada, permitindo correlações espaciais entre as observações. Os dados foram analisados sob o enfoque bayesiano.

Esses modelos foram aplicados inicialmente em um estudo de simulação, na malha de seto-

res censitários do Município de Campinas, em diferentes cenários para a proporção p de zeros nos dados. Obtendo-se, segundo os critérios DIC e LCPO, melhor desempenho para o modelo ZIP. O modelo de Poisson apresentou desempenho bem inferior aos modelos inflacionados de zero, principalmente com proporção de zeros igual ou superior a 40% ($p = 0,4$). Esse resultado é reflexo da falta de um parâmetro adicional no modelo de Poisson para correção de superdispersão nos dados. O modelo BN também apresentou desempenho inferior em relação aos modelos inflacionados de zero. Isso porque o parâmetro adicional da distribuição BN considera que a sobredispersão decorre de uma heterogeneidade não observada²⁰, sendo que nos nossos dados a principal fonte da superdispersão é o excesso de regiões com contagens iguais a zero.

Os resultados desse estudo de simulação evidenciaram a importância de se empregar métodos adequados que ajustem o excesso de zeros presentes na distribuição dos dados.

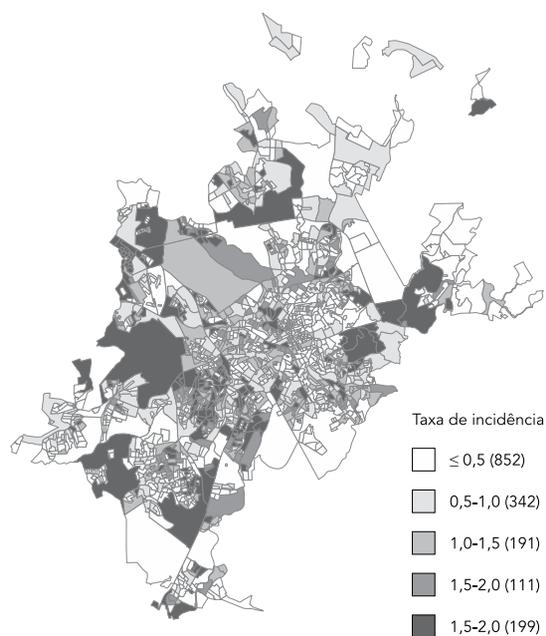
Com base nos resultados desse estudo de simulação foram aplicados os modelos inflacionados de zero (ZIP e ZINB) a um conjunto de dados reais com contagens de casos de dengue do Município de Campinas, em 2007. Nessa análise, segundo os critérios DIC e LCPO, o modelo ZIP apresentou melhor ajuste, estimando de forma consistente o risco relativo da incidência de dengue nos setores censitários urbanos, corrigindo os efeitos de correlação espacial, flutuações aleatórias, superdispersão e excesso de zeros. Foi possível identificar áreas onde ocorreu maior incidência em 2007, assim prioritárias para ações de controle naquele ano. A distribuição espacial dos casos de dengue não se mostrou associada às condições socioambientais dos setores. Ressalta-se que embora áreas periféricas apresentem menor cobertura de saneamento e intermitência de oferta de água encanada com pior situação socioeconômica, outros fatores interferem na circulação viral. A circulação prévia do sorotipo do vírus, resultando na imunidade adquirida pela população, medidas de controle específicas por parte do poder público, mudança no comportamento da população após a notificação de casos, são fatores que interferem nos caminhos da transmissão^{12,14,16}.

Em relação ao nível de agregação dos dados, os setores censitários são as menores unidades geográficas disponíveis para os dados socioambientais, permitindo a visualização espacial mais precisa do território onde se concentram os maiores níveis de transmissão de dengue. O aumento da escala (áreas menores) também permite obter informações contextuais com maior homogeneidade interna, diferentemente do que

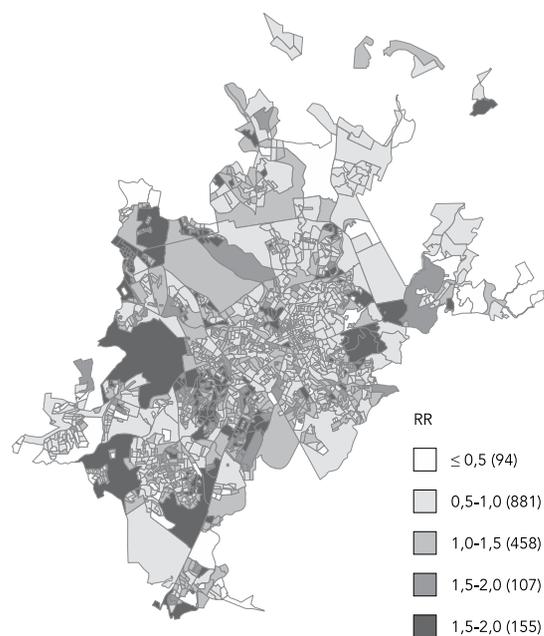
Figura 3

Mapas com a distribuição do risco de dengue em Campinas, São Paulo, Brasil: (3a) razão de incidência padronizada (SIR); (3b) estimativas do risco relativo (RR) obtidas com o modelo de Poisson inflacionado de zeros (ZIP); e (3c) probabilidade *a posteriori* do RR > 1,2.

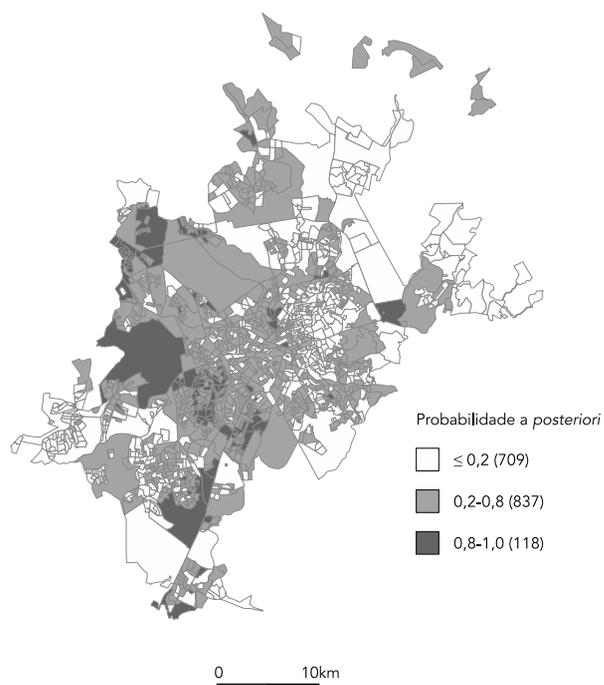
3a)



3b)



3c)



se observa com dados agregados em escala menor (áreas maiores).

Um método alternativo foi proposto para permitir a agregação dos dados censitários em uma grade regular com cerca de 0,0625km² ou 250x250m na Amazônia brasileira ³³. A disponibilidade dos dados em grade regular apresenta algumas vantagens em relação aos setores censitários: permite obter estabilidade temporal destas unidades, sendo possível trabalhar com regiões ainda menores, do mesmo tamanho e também se obter maior homogeneidade interna nos dados populacionais. Dessa forma, a aplicação do modelo de Poisson inflacionado de zeros viabiliza a análise de dados que se apresentam na forma de contagens para pequenas áreas, na presença de contagens nulas, sem a necessidade de se agregar um conjunto de áreas para a obtenção de áreas maiores.

A inclusão do efeito aleatório espacial ajustou a heterogeneidade espacial não observada e corrigiu a presença de correlação espacial presente entre as observações. A correlação espacial foi modelada segundo uma matriz de vizinhança binária, definida baseando-se na relação de vizinhança entre os setores censitários. É importante considerar que diferentes especificações na matriz de vizinhança produzem diferentes estimativas de vizinhança. Esses critérios também podem ser determinados com base no tamanho da fronteira, com ou sem a presença de barreiras naturais, ou outra característica geográfica que possa interferir na ligação entre as áreas ¹¹.

A utilização da abordagem INLA para estimativa dos parâmetros mostrou-se atrativa e conveniente para a realização de inferências. O processo de estimação dos parâmetros não apresentou problemas de convergência, estimando os parâmetros de forma bem rápida (aproximadamente 13 segundos), conforme já demonstrado em outros estudos ²⁹.

A identificação dos parâmetros é uma das principais dificuldades encontradas no contexto em que a inferência bayesiana é realizada via o método de MCMC. Para superar esse problema, o processo de estimação pela INLA impõe uma restrição de soma zero ao efeito aleatório estruturado espacialmente ³⁴.

Os resultados obtidos demonstraram a presença de variação espacial na distribuição geográfica da incidência de dengue em Campinas. Embora as estimativas do RR não tenham sido significantes, não há como desconsiderar totalmente a importância de fatores socioambientais na dinâmica do vetor transmissor. Um estudo identificou que nesse período de análise a transmissão originou-se na região Norte da cidade, uma área carente de infraestrutura urbana, com

moradores apresentando precárias condições socioeconômicas e marcada pela presença de importantes rodovias que favorecem o intenso fluxo de pessoas e veículos ³¹.

Estudos demonstram que o espaço e o tempo são duas dimensões importantes a serem consideradas na descrição da dinâmica de epidemias e na distribuição de risco da doença, sendo necessário o acompanhamento sistemático das áreas mais vulneráveis à doença. Wen et al. ³⁵ desenvolveram um modelo espaço-temporal de risco com base em indicadores de frequência, duração e intensidade para o mapeamento e acompanhamento da dinâmica da doença, identificação de áreas de risco e comparação dos diferentes padrões de risco. Estudos com o modelo espaço-temporal realizados nos municípios de São José do Rio Preto ³⁶ e Tupã ³⁷, em São Paulo, e Rio de Janeiro ¹⁴ identificaram diferentes áreas de risco nos períodos analisados, demonstrando-se que as áreas de risco podem variar entre alças epidêmicas. A persistência de ondas epidêmicas na cidade denotam que estas variações em escala temporal e espacial podem estar mais relacionadas à redução de indivíduos suscetíveis ao sorotipo do vírus circulante do que a medidas de controle ³⁸.

Alguns autores encontraram evidência da associação entre risco de dengue e condições socioambientais após a reintrodução de novo sorotipo do vírus, sendo que depois da dispersão do vírus as incidências ocorrem de forma geral em todos os extratos socioeconômicos ³⁸. Nesse sentido, verifica-se que, em anos anteriores, a incidência da doença em Campinas foi mais concentrada em áreas periféricas, em especial nas regiões Sul e Norte da cidade ^{39,40}, e em 2007, distribuiu-se de forma generalizada pelo município ³¹. Destaca-se que essa epidemia foi caracterizada pela circulação do sorotipo DENV-3. Esse sorotipo já havia sido isolado em 2006, na região Norte da cidade e em municípios vizinhos ³¹ no ano de 2003 e na epidemia de 2001/2002, juntamente com o DENV-1 ⁴¹, mas sem grande disseminação. O histórico sobre os sorotipos circulantes na região ainda revela que houve isolamento dos sorotipos DENV-1 nos primeiros casos detectados em 1995 e 1996, e epidemia de 1998, o DENV-2 ³⁹.

Estudos sobre a associação entre fatores socioambientais e incidência de dengue apresentam resultados contraditórios ⁴². Almeida et al. ¹⁴ na cidade do Rio de Janeiro, Pessanha et al. ⁴³ em Belo Horizonte (Minas Gerais), Resendes et al. ⁴⁴ em Niterói (Rio de Janeiro) e Costa et al. ¹⁷ em Campinas encontraram relação positiva entre condições socioambientais e incidência de dengue. Distintamente, Mondini & Chiaraval-

loti-Neto ⁴⁵ em São José do Rio Preto, Machado et al. ⁴⁶ em Nova e Iguaçu (Rio de Janeiro) e Teixeira et al. ⁴⁷ no Rio de Janeiro encontraram resultados semelhantes aos registrados no presente estudo.

Os resultados conflitantes encontrados nos diferentes estudos demonstram a complexa relação entre fatores socioambientais e o risco de infecção da doença, cuja variação espacial também pode estar associada a outros fatores não mensurados. Estudos anteriores demonstraram que a dinâmica do dengue também é afetada por fatores ambientais e climáticos, interações hospedeiro/vetor e imunidade da população ^{4,48,49,50}. As variações nas condições climáticas (aumento da temperatura, pluviosidade e umidade do ar) afetam a sobrevivência, a reprodução do vetor, as mudanças na sua distribuição e a densidade.

Diferentes estratégias metodológicas adotadas para avaliar a relação entre dengue e condições socioambientais também podem influenciar os resultados obtidos. O tipo de dado utilizado (dados primários ou secundários), diferentes variáveis adotadas para a caracterização socioambiental e o tipo de agregação usada (setor censitário, bairro, distritos e/ou municípios) podem gerar distorções nos diferentes estudos ³⁸.

Diante da complexa teia de fatores envolvidos na incidência de dengue e da indisponibilidade de uma vacina eficaz e segura, os esforços devem ser concentrados para o controle do vetor transmissor ^{3,4}, direcionados às realidades locais com base na intersectorialidade, integralidade e participação social no enfrentamento da doença ^{43,44}.

Destacam-se algumas limitações observadas nos estudos de análise espacial que trabalham com dados secundários. Quanto ao tipo de agregação adotada, os diferentes resultados obtidos segundo diferentes recortes espaciais são chamados de problema da unidade de área modificável ⁴⁵. Nesse caso, mudanças na escala e nas dimensões da unidade de análise podem levar a conclusões diferentes. Um estudo de revisão sobre a relação entre dengue e indicadores socioeconômicos e ambientais, ao comparar diferentes estudos com dados secundários, concluiu que resultados semelhantes foram observados principalmente nos estudos com a mesma escala de agregação.

A possibilidade de subnotificação de casos e da variabilidade da qualidade dos dados registrados no SINAN constitui outra limitação do estudo. São registrados principalmente aqueles casos de evolução mais grave, havendo subnotificação das formas mais leves da doença. Essas subnotificações ocorrem frequentemente por falhas do próprio sistema de informações em saúde e treinamento dos profissionais de saúde sobre o diagnóstico, vigilância e assistência à doença ⁴⁶. Adicionalmente, distorções no conhecimento da distribuição da circulação dos vírus do dengue podem ocorrer devido aos dados registrados no SINAN retratarem mais fielmente a busca à assistência médica nos serviços públicos, não incluindo parte dos indivíduos que buscam atendimento nos serviços privados, embora se observe aumento das notificações nestes serviços.

Por fim, os estudos ecológicos por não utilizarem dados sobre indivíduos, apenas agregados sobre grupos de pessoas, estão sujeitos ao “viés ecológico”, que se refere à impossibilidade das associações obtidas no nível ecológico (agregado) refletirem associações no nível individual ¹⁴. Uma alternativa para reduzir esse tipo de viés seria a combinação de dados de área e uma amostra de dados individuais ⁴⁹. Trabalhos futuros precisam considerar o ajuste para outros fatores de exposição no nível agregado e no nível individual, além de considerar a variação espaço-temporal da incidência de dengue.

Portanto, conclui-se que na análise espacial de dados agregados é fundamental avaliar de forma adequada a distribuição discreta que melhor ajusta o conjunto de dados em análise, não se restringindo somente à distribuição de Poisson. Os modelos inflacionados de zero permitem analisar os dados de área em níveis mais desagregados, sem a necessidade de juntar áreas com baixa frequência do evento de interesse com áreas em que não houve a ocorrência do evento. Com base na abordagem bayesiana, a aplicação dos modelos inflacionados de zeros com a inclusão de efeitos aleatórios, permite obter estimativas robustas da distribuição espacial do risco relativo. Somando-se a outras variáveis obtidas junto aos serviços de vigilância epidemiológica, virológica e vetorial, contribuem para refinar o conhecimento sobre a transmissão em pequenas áreas.

Colaboradores

J. V. Costa participou no desenho do estudo, análise dos dados, discussão dos resultados, redação e revisão final do artigo. L. V. A. Silveira contribuiu no desenho do estudo, na análise dos dados e revisão do artigo. M. R. Donalísio colaborou no desenho do estudo, redação e revisão do artigo.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela bolsa de doutorado concedida parcialmente a J. V. Costa. À Heloísa Girardi Malavasi por geocodificar os casos de dengue.

Referências

- Gubler DJ. Dengue/dengue haemorrhagic fever: history and current status. *Novartis Found Symp* 2006; 277:3-16.
- Gubler DJ. Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev* 1998; 11:480-96.
- Martina BEE, Koraka P, Osterhaus ADME. Dengue virus pathogenesis: an integrated view. *Clin Microbiol Rev* 2009; 22:564-81.
- Rodenhuis-Zybert IA, Wilschut J, Smit JM. Dengue virus life cycle: viral and host factors modulating infectivity. *Cell Mol Life Sci* 2010; 67:2773-86.
- Gubler DJ. Dengue, urbanization and globalization: the unholy trinity of the 21st century. *Trop Med Health* 2011; 39(4 Suppl):S3-11.
- Lambrechts L, Paaijmans KP, Fansiri T, Carrington LB, Kramer LD, Thomas MB, et al. Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proc Natl Acad Sci U S A* 2011; 108:7460-5.
- Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. *Lancet Infect Dis* 2011; 11:131-41.
- Liebman KA, Stoddard ST, Morrison AC, Rocha C, Minnick S, Sihuíncha M, et al. Spatial dimensions of dengue virus transmission across interepidemic and epidemic periods in Iquitos, Peru (1999-2003). *PLoS Negl Trop Dis* 2012; 6:e1472.
- Halstead SB. Dengue virus-mosquito interactions. *Annu Rev Entomol* 2008; 53:273-91.
- Eisen L, Lozano-Fuentes S. Use of mapping and spatial and space-time modeling approaches in operational control of *Aedes aegypti* and dengue. *PLoS Negl Trop Dis* 2009; 3:e411.
- Elliott P, Wakefield J, Best N, Briggs D. *Spatial epidemiology: methods and applications*. London: Oxford University Press; 2001.
- Siqueira JB, Martelli CMT, Maciel IJ, Oliveira RM, Ribeiro MG, Amorim FP, et al. Household survey of dengue infection in central Brazil: spatial point pattern analysis and risk factors assessment. *Am J Trop Med Hyg* 2004; 71:646-51.
- Mondini A, Chiaravalloti Neto F, Gallo Y, Sanches M, Lopes JCC. Análise espacial da transmissão de dengue em cidade de porte médio do interior paulista. *Rev Saúde Pública* 2005; 39:444-51.
- Almeida AS, Medronho RA, Valencia LIO. Análise espacial da dengue e o contexto socioeconômico no Município do Rio de Janeiro, RJ. *Rev Saúde Pública* 2009; 43:666-73.
- Flauzino RF, Souza-Santos R, Barcellos C, Gracie R, Magalhães MAFM, Oliveira RM. Heterogeneidade espacial da dengue em estudos locais, Niterói, RJ. *Rev Saúde Pública* 2009; 43:1035-43.
- Cordeiro R, Donalísio MR, Andrade VR, Mafra AC, Nucci LB, Brown JC, et al. Spatial distribution of the risk of dengue fever in southeast Brazil, 2006-2007. *BMC Public Health* 2011; 11:355.
- Costa JV, Donalísio MR, Silveira LVA. Spatial distribution of dengue incidence and socio-environmental conditions in Campinas, São Paulo State, Brazil, 2007. *Cad Saúde Pública* 2013; 29:1522-32.
- Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 1992; 34:1-14.

19. Ridout M, Malling W, Me K. Models for count data with many zeros. *North* 1998; 19:179-92.
20. Hall DB. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 2000; 56:1030-9.
21. Ghosh S, Mukhopadhyay P, Lu J. Bayesian analysis of zero-inflated regression models. *J Stat Plan Inference* 2006; 136:1360-75.
22. Gschlößl S, Czado C. Modelling count data with overdispersion and spatial effects. *Stat Pap (Berl)* 2008; 49:531-52.
23. Haining R, Law J, Griffith D. Modelling small area counts in the presence of overdispersion and spatial autocorrelation. *Comput Stat Data Anal* 2009; 53:2923-37.
24. Besag J, York J, Molli A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math* 1991; 43:1-20.
25. Cameron AC, Trivedi PK. Regression analysis of count data. Cambridge: Cambridge University Press; 1998. (Econometric Society Monograph, 30).
26. Rue H, Martino S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *J Stat Plan Inference* 2007; 137:3177-92.
27. Kelsall J, Wakefield J. Discussion of "Bayesian models for spatially correlated disease and exposure data". In: Bernardo J, Berger J, Dawid A, Smith A, editors. *Bayesian statistics 6*. Oxford: Oxford University Press; 1999. p. 131-56.
28. Spiegelhalter D, Thomas A, Best N, Way R. WinBUGS user manual. Cambridge: MRC Biostat Unit; 2003.
29. Held L, Schrödle B, Rue H. Posterior and cross-validated predictive checks: a comparison of MCMC and INLA. In: Kneib T, Tutz G, editors. *Statistical modelling and regression structures*. Heidelberg: Physica-Verlag HD; 2010. p. 91-110.
30. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* 2004; 112:1016-25.
31. Malavasi H. Análise espacial da epidemia de dengue em Campinas/SP no ano de 2007 [Dissertação de Mestrado]. Campinas: Faculdade de Ciências Médicas, Universidade Estadual de Campinas; 2011.
32. Assunção RM, Reis EA. A new proposal to adjust Moran's I for population density. *Stat Med* 1999; 18:2147-62.
33. D'Antona AO, Bueno MCD, Dagnino RS. Estimativa da população em unidades de conservação na Amazônia Legal brasileira: uma aplicação de grades regulares a partir da Contagem 2007. *Rev Bras Estud Popul* 2013; 30:401-28.
34. Schrödle B, Held L. A primer on disease mapping and ecological regression using INLA. *Comput Stat* 2011; 26:241-58.
35. Wen T-H, Lin NH, Chao D-Y, Hwang K-P, Kan C-C, Lin KC-M, et al. Spatial-temporal patterns of dengue in areas at risk of dengue hemorrhagic fever in Kaohsiung, Taiwan, 2002. *Int J Infect Dis* 2010; 14:e334-43.
36. Galli B, Chiaravalloti Neto F. Modelo de risco tempo-espacial para identificação de áreas de risco para ocorrência de dengue. *Rev Saúde Pública* 2008; 42:656-63.
37. Barbosa GL, Lourenço RW. Análise da distribuição espaço-temporal de dengue e da infestação larvária no Município de Tupã, Estado de São Paulo. *Rev Soc Bras Med Trop* 2010; 43:145-51.
38. Almeida MCM, Assunção RM, Proietti FA, Caiáffa WT. Dinâmica intra-urbana das epidemias de dengue em Belo Horizonte, Minas Gerais, Brasil, 1996-2002. *Cad Saúde Pública* 2008; 24:2385-95.
39. Kemp B. Os casos que não se confirmaram como dengue durante a epidemia de dengue no Município de Campinas/SP, 2002 [Tese de Doutorado]. Campinas: Faculdade de Ciências Médicas, Universidade Estadual de Campinas; 2005.
40. Andrade VR. Distribuição espacial do risco de dengue em região do Município de Campinas [Tese de Doutorado]. Campinas: Faculdade de Ciências Médicas, Universidade Estadual de Campinas; 2009.
41. Lima VLC, Rangel O, Andrade VR, Silveira NYJ, Oliveira SS, Figueiredo LTM. Dengue: inquérito populacional para pesquisa de anticorpos e vigilância virológica no Município de Campinas, São Paulo, Brasil. *Cad Saúde Pública* 2007; 23:669-80.
42. Flauzino RF, Souza-Santos R, Oliveira RM. Dengue, geoprocessamento e indicadores socioeconômicos e ambientais: um estudo de revisão. *Rev Panam Salud Pública* 2009; 25:456-61.
43. Pessanha JEM, Caiáffa WT, Kroon EG, Proietti FA. Dengue em três distritos sanitários de Belo Horizonte, Brasil: inquérito soropidemiológico de base populacional, 2006 a 2007. *Rev Panam Salud Pública* 2010; 27:252-8.
44. Resendes APC, Silveira NAPR, Sabroza PC, Souza-Santos R. Determinação de áreas prioritárias para ações de controle da dengue. *Rev Saúde Pública* 2010; 44:274-82.
45. Mondini A, Chiaravalloti Neto F. Socioeconomic variables and dengue transmission. *Rev Saúde Pública* 2007; 41:923-30.
46. Machado JP, Oliveira RM, Souza-Santos R. Análise espacial da ocorrência de dengue e condições de vida na cidade de Nova Iguaçu, Estado do Rio de Janeiro, Brasil. *Cad Saúde Pública* 2009; 25:1025-34.
47. Teixeira TRA, Cruz OG. Spatial modeling of dengue and socio-environmental indicators in the city of Rio de Janeiro, Brazil. *Cad Saúde Pública* 2011; 27:591-602.
48. Barclay E. Is climate change affecting dengue in the Americas? *Lancet* 2008; 371:973-4.
49. Kyle JL, Harris E. Global spread and persistence of dengue. *Annu Rev Microbiol* 2008; 62:71-92.
50. Sessions OM, Barrows NJ, Souza-Neto JA, Robinson TJ, Hershey CL, Rodgers MA, et al. Discovery of insect and human dengue virus host factors. *Nature* 2009; 458:1047-50.

Abstract

Dengue incidence occurs predominantly within city limits. Identifying spatial distribution of the disease at the local level helps formulate strategies to control and prevent the disease. Spatial analysis of counting data for small areas commonly violates the assumptions of traditional Poisson models due to the excessive amount of zeros. This study compared the performance of four counting models used in mapping diseases: Poisson, negative binomial, zero-inflated Poisson, and zero-inflated negative binomial. The methods were compared in a simulation study. The models analyzed in the simulation were applied to a spatial ecological study of dengue data aggregated by census tracts in the city of Campinas, São Paulo State, Brazil, 2007. Spatial analysis was conducted with Bayesian hierarchical models. The zero-inflated Poisson model showed the best performance for estimating relative risk of dengue incidence in the census tracts.

Spatial Analysis; Dengue; Communicable Disease Control

Resumen

La incidencia de dengue se produce, predominantemente, en las áreas urbanas de las ciudades. Identificar el patrón de distribución espacial de la enfermedad a nivel local contribuye a la formulación de estrategias de control y prevención de la enfermedad. El análisis espacial de datos de conteo para pequeñas áreas comúnmente transgrede las suposiciones de los modelos tradicionales de Poisson, debido a la cantidad excesiva de ceros. En este estudio, se comparó el desempeño de cuatro modelos de conteo utilizados en el mapeo de enfermedades: Poisson, binomial negativo, Poisson con exceso de ceros y binomial negativo con exceso de ceros. Los métodos fueron comparados en un estudio de simulación. Los modelos analizados en el estudio de simulación fueron aplicados en un estudio ecológico espacial, a los datos de dengue agregados por sectores censales, del Municipio de Campinas, São Paulo, Brasil, 2007. El análisis espacial fue realizado con modelos jerárquicos bayesianos. El modelo de Poisson con exceso de ceros presentó un mejor desempeño para estimar el riesgo relativo de incidencia de dengue en los sectores censales.

Análisis Espacial; Dengue; Control de Enfermedades Transmisibles

Recebido em 09/Mar/2015
Versão final rerepresentada em 18/Fev/2016
Aprovado em 06/Abr/2016