

Utilização da modelagem inteiramente bayesiana na detecção de padrões de variação de risco relativo de mortalidade infantil no Rio Grande do Sul, Brasil

Utilization of fully Bayesian modeling to detect patterns in relative risk variation for infant mortality in Rio Grande do Sul State, Brazil

Sérgio Kakuta Kato ^{1,2,3}
Diego de Matos Vieira ²
Jandyra Maria Guimarães Fachel ^{2,4}

Abstract

The infant mortality rate is one of the key indicators used to measure a population's quality of life. The State of Rio Grande do Sul has a social and economic indicator called the Socioeconomic Development Index (IDESE). Most studies analyze the infant mortality rate in relation to risk factors, visually aided by maps. This study presents the methodology and an application of a Spatial Epidemiology method called the ecological correlation, using hierarchical Bayesian procedures. The article discusses the main problems found in ecological correlations, such as spatial autocorrelation and the estimator's instability for small areas. To overcome these difficulties, the relative risk estimate obtained by spatial regression analysis using fully Bayesian estimation is presented. The infant mortality rate is analyzed in all 496 municipalities of Rio Grande do Sul for the years 2001 to 2004. Several models with spatial component and different variables from the IDESE/2003 were compared. The model using spatial structure along with the variable "education" was considered the best choice. With this methodology, it was possible to obtain a more interpretable pattern of infant mortality risk in Rio Grande do Sul.

Infant Mortality; Statistical Analysis; Risk Assessment

Introdução

A taxa de mortalidade infantil é um dos indicadores mais usados para medir a qualidade de vida da população ¹, além de compor o Índice de Desenvolvimento Humano (IDH), divulgado pela Organização das Nações Unidas (ONU), e outros indicadores sócio-econômicos. O Rio Grande do Sul, Brasil, possui um indicador sócio-econômico próprio, denominado Índice de Desenvolvimento Sócio-econômico (IDESE), da Fundação de Economia e Estatística (FEE), que também utiliza a taxa de mortalidade infantil para compor um índice referente à saúde. Nos estados do Brasil, essa taxa é obtida de dados do SINASC (Sistema de Informações sobre Nascidos Vivos) e SIM (Sistema de Informações sobre Mortalidade), porém é importante salientar que, em alguns estados, ela é estimada por meio de métodos demográficos indiretos, em virtude da baixa cobertura dos sistemas SIM e SINASC.

No Brasil, existe uma grande variação na taxa de mortalidade. Em geral, os municípios da região Norte e Nordeste do Brasil são os que apresentam as piores taxas, e, de acordo com dados disponíveis no Departamento de Informática do SUS (DATASUS) ², em 2004, o estado com maior taxa de mortalidade foi Alagoas, com 47,09 mortes até um ano de idade para cada mil nascidos vivos. Por outro lado, Santa Catarina foi o destaque positivo, com apenas 13,62 mortes para cada mil nascidos vivos. No Rio Grande do Sul, entre

¹ Faculdade de Matemática, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brasil.

² Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.

³ Laboratório de Pesquisa em HIV/AIDS, Universidade de Caxias do Sul, Caxias do Sul, Brasil.

⁴ Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil.

Correspondência

S. K. Kato

Departamento de Estatística, Faculdade de Matemática, Pontifícia Universidade Católica do Rio Grande do Sul.

Av. Ipiranga 6681, Prédio 30, Bloco C, sala 104, Porto Alegre, RS

90619-900, Brasil.

klmsergio@terra.com.br

os anos de 1997 e 2004, observou-se uma estabilidade em relação à taxa de mortalidade, que variou entre 15,10 e 17,30. No entanto, é possível observar grande variabilidade entre municípios.

A demanda por análises que envolvam informações geográficas é muito grande. A Organização Pan-Americana da Saúde (OPAS) estima que cerca de 80% das necessidades de informações dos dirigentes políticos está relacionada com a localização geográfica.

Existem estudos propondo ações ou intervenções para a redução da mortalidade infantil, apontando que sua ocorrência é determinada, em última instância, por fatores sociais, econômicos e culturais – como renda, educação e posse de terra³.

Os estudos que têm como foco a comparação de grupos ao invés de pessoas são denominados ecológicos. Os estudos ecológicos são ideais para dados de área, como taxas por área geográfica, além de serem úteis para levantar hipóteses. As informações utilizadas nos estudos ecológicos relacionados à área da saúde, como a mortalidade infantil, estão, em grande parte, disponíveis no DATASUS², além de em outras bases oficiais, como as do Instituto Brasileiro de Geografia e Estatística (IBGE) e da FEE. Os estudos que utilizam modelos para avaliar a distribuição geográfica de taxas de mortalidade ou de incidência ou alguma outra medida epidemiológica de risco se enquadram na área da Epidemiologia Espacial.

A utilização da Epidemiologia Espacial por parte dos pesquisadores está em ampla expansão, principalmente devido aos desenvolvimentos recentes do Sistema de Informação Geográfica (SIG), da maior disponibilidade de informações de saúde e dos avanços dos métodos estatísticos, destacando-se a inferência bayesiana, ainda pouco empregada na área da saúde⁴. Esses avanços possibilitam a confecção de mapas apresentando estimativas mais precisas dos indicadores de saúde, bem como a obtenção de modelos capazes de mensurar o efeito de covariáveis, que podem contribuir para formular hipóteses a respeito da distribuição espacial desses indicadores, além de avaliar sua relação com indicadores sócio-econômicos⁵.

Em grande parte dos estudos ecológicos, a análise da associação entre as covariáveis e a taxa de mortalidade infantil é realizada apenas descritivamente ou utilizando técnicas de modelagem muitas vezes inadequadas.

A avaliação descritiva, apenas visual, é bastante fácil de realizar e consiste em comparar mapas que apresentam a taxa de mortalidade infantil por município ou por bairro com mapas que apresentam a distribuição espacial das covariáveis. A grande dificuldade neste tipo de

análise é interpretar o mapa que apresenta a taxa de mortalidade infantil, pois esse indicador geralmente tem alta variabilidade (pouca precisão) em regiões com população reduzida, apresentando, em muitos casos, valores extremamente superiores ou inferiores à taxa ou risco médio da área como um todo, sem que essas regiões realmente se caracterizem como regiões de alto ou baixo risco. Sendo assim, esses mapas podem não representar o verdadeiro processo subjacente a esses indicadores^{6,7}.

Outras vezes são utilizados modelos de regressão linear para relacionar as covariáveis à taxa de mortalidade, ou seja, consideram-se as áreas como indivíduos. O grande problema neste tipo de análise, que também é bastante acessível e por isso muito utilizado, reside no fato de as unidades amostrais (municípios ou bairros) geralmente não poderem ser consideradas independentes, por isso os resíduos do modelo continuam apresentando a autocorrelação espacial presente nos dados. Nos dados espaciais, áreas vizinhas são mais similares do que áreas mais distantes. A interdependência entre as unidades amostrais altera o poder explicativo do modelo, podendo produzir associações errôneas^{8,9}.

Para superar as dificuldades de instabilidade das estimativas e autocorrelação espacial na variável-resposta, várias modelagens têm sido propostas, mas as que apresentam melhores resultados são os métodos de estimação inteiramente bayesianos¹⁰. Nessa linha, para entender melhor a distribuição espacial da mortalidade infantil, alguns estudos já foram realizados em Minas Gerais¹¹ e no Rio de Janeiro¹². No Rio Grande do Sul, esses métodos foram utilizados em uma estimação espaço-temporal¹³, porém sem o uso de covariáveis.

O objetivo deste artigo é demonstrar a utilização de modelos de regressão espacial inteiramente bayesianos na detecção de padrões de variação, por meio do mapeamento do risco de mortalidade infantil nos municípios do Rio Grande do Sul (dados acumulados para os anos de 2001 a 2004), utilizando como covariáveis subescalas do IDESE de 2003 da FEE.

Metodologia

Dados utilizados

As informações sobre o número de nascimentos e óbitos em recém-nascidos nos 496 municípios do Rio Grande do Sul entre os anos de 2001 e 2004 foram obtidas no site do DATASUS².

O IDESE produz um índice sintético composto por 12 indicadores divididos em quatro blocos

temáticos: educação, renda, saneamento e domicílio, saúde. Esses indicadores são agregados e transformados em índices, um para cada bloco. O índice resulta da média ponderada dos indicadores que o compõem. Devido a reparametrização, é um índice que varia entre 0 e 1: quanto mais próximo de 0, menor o desenvolvimento; quanto mais próximo de 1, maior o desenvolvimento. Utilizaremos como covariáveis os índices relativos aos três primeiros blocos temáticos. Não foi utilizado o bloco saúde, visto que a taxa de mortalidade infantil compõe este bloco.

Análise estatística

Considerando o risco de mortalidade infantil como variável-resposta no modelo de regressão, neste artigo serão avaliados os nove modelos a seguir: (1) sem efeito espacial, controlando por educação, renda e condições de saneamento e domicílio; (2) com efeito espacial, sem covariáveis; (3) com efeito espacial, controlando por educação; (4) com efeito espacial, controlando por saneamento e domicílio; (5) com efeito espacial, controlando por renda; (6) com efeito espacial, controlando por educação e renda; (7) com efeito espacial, controlando por educação e saneamento; (8) com efeito espacial, controlando por renda e saneamento; (9) com efeito espacial, controlando por educação, renda e saneamento.

A análise de regressão utilizando modelos inteiramente bayesianos é própria para a modelagem dos diferentes tipos de efeitos espaciais, temporais e espaço-temporais. Nessa modelagem, supõe-se que o número de eventos observados em cada área possua distribuição binomial. Como a maioria dos dados epidemiológicos mapeados é rara ou com grande variação das taxas/riscos entre diferentes áreas, o modelo binomial pode ser aproximado pela distribuição Poisson ¹⁴, com número de eventos observados (Y_i) e com valor esperado (μ_i):

$$Y_i \sim \text{Poisson}(\mu_i)$$

onde $\mu_i = E_i \theta_i$, sendo θ_i o risco relativo de óbito infantil na i -ésima área e E_i a quantidade esperada de óbitos infantis na i -ésima área sob a hipótese de que o risco seja constante em todas as áreas e igual ao risco geral da região. O risco relativo θ_i é uma quantidade que pode assumir valores reais entre 0 e $+\infty$, sendo igual a 1 nas áreas onde o risco possui exatamente o mesmo risco médio para toda a região.

O primeiro nível hierárquico do modelo é dado por:

$$Y_i | E_i \sim \text{Poisson}(E_i \theta_i)$$

A especificação dos componentes utilizados na estimação do logaritmo do risco θ_i é dada por:

$$\log(\theta_i) = \beta_0 + \beta_p + X_{ip} + \mu_i + v_i$$

onde β_0 representa o logaritmo do risco médio global; $\beta_p = \{\beta_1, \dots, \beta_k\}$ é um vetor que representa os efeitos de cada uma das k covariáveis no $\log(\theta_i)$; X_{ip} é um vetor de covariáveis e μ_i e v_i são vetores de efeitos aleatórios.

O modelo é construído de modo a relacionar os componentes com o logaritmo de θ_i e não diretamente com o risco θ_i , pelas inúmeras vantagens matemáticas e computacionais que a transformação logarítmica pode proporcionar no presente caso, pois a distribuição Poisson é uma distribuição de probabilidade enquadrada na família de distribuições exponenciais. Depois de obtidas as estimativas de $\log(\theta_i)$, aplica-se a função exponencial para a obtenção das estimativas do risco θ_i .

No segundo nível hierárquico do modelo, estão as distribuições *a priori* para cada um dos parâmetros do modelo: β_0, β_p, μ_i e v_i . Neste artigo, serão testados nove diferentes modelos, desde aqueles sem covariáveis até os com os três blocos do IDESE. Para todos os modelos para β_0 será atribuída uma distribuição *a priori* uniforme ($-\infty; +\infty$), e a cada um dos elementos do vetor β_p , uma distribuição normal com média igual a zero e parâmetro de precisão τ_β o menor possível, a fim de que estas *prioris* sejam pouco informativas.

Quanto aos componentes aleatórios, o componente μ_i representa um efeito aleatório espacial não estruturado, que pode ser visto como decorrente de particularidades de cada área, ou seja, efeitos de pequena escala que não ultrapassam as fronteiras das áreas. A cada um dos componentes μ_i será atribuída independentemente uma distribuição *a priori* normal com média zero e parâmetro de precisão τ_μ , que em geral é desconhecido e denominado hiperparâmetro, já que é um parâmetro de uma *a priori*.

No terceiro nível hierárquico do modelo, são apresentadas as *hiperprioris*. Para τ_μ será atribuído *a priori* uma distribuição *gamma* com parâmetros 0,5 e 0,0005 (esperança igual a 1.000; variância igual a 2.000.000), atribuídos de forma que a hiperpriori seja pouco informativa, ou seja, uma distribuição com grande variabilidade.

Já o componente v_i incorpora a estrutura espacial, captando a influência das áreas vizinhas referentes aos efeitos de larga escala, através de uma *a priori* espacialmente estruturada mediante a definição de uma matriz de vizinhança entre as áreas. Neste estudo, será utilizada apenas a matriz de vizinhança binária, na qual cada elemento assumirá valor 1 quando as áreas são vizinhas e 0, em caso contrário. Ao componente v_i , atribui-se uma distribuição *a priori* denominada CAR (Condicional Auto-Regressiva) normal. Como a matriz de vizinhança utilizada é binária, o

efeito espacial médio da i -ésima área v_i é dado pela média aritmética dos efeitos dos seus vizinhos, e a variância é inversamente proporcional à quantidade de áreas vizinhas. Assim, quanto maior o número de vizinhos, maior é a precisão da estimativa de v_i .

Além disso, a especificação completa da distribuição CAR normal depende de um único parâmetro τ_v (o inverso da variância de v_i), ao qual também será atribuída uma distribuição *a priori gamma*, analogamente à distribuição *a priori* para v_i .

As estimativas de risco são obtidas através da distribuição *a posteriori* de θ , que é o resultado da combinação dos diversos níveis do modelo e de suas *prioris*. Em virtude da complexidade do modelo, a distribuição *a posteriori* não pode ser encontrada analiticamente. Dessa forma, é necessário utilizar métodos de simulação estocástica chamados MCMC (*Markov Chain Monte Carlo*). Um programa que atualmente está sendo utilizado para modelagem de dados com enfoque bayesiano, valendo-se de métodos MCMC, é o WinBUGS (*Win Bayesian inference Using Gibbs Sampling*)¹⁵, um programa livre, que conta com uma biblioteca para análise de dados espaciais e que será utilizado na aplicação dos dados neste trabalho. Em adição, será utilizado o programa TerraView (Instituto Nacional de Pesquisas Espaciais; <http://www.dpi.inpe.br/terraview>), igualmente um programa livre, para a elaboração dos mapas.

Esta distribuição conjunta *a posteriori* é uma distribuição de probabilidade de todos os parâmetros do modelo conjuntamente e de onde são obtidas as estimativas de cada um dos parâmetros. Serão consideradas como estimativas dos parâmetros (que serão apresentadas nos mapas) as médias obtidas desta distribuição, e para cada parâmetro podem-se obter informações adicionais como intervalos de credibilidade, que não costumam ser apresentados em mapas, mas que são úteis na avaliação da associação entre as covariáveis e a taxa de mortalidade infantil.

Utilizando-se a metodologia bayesiana em estudos de correlação ecológica, deve-se fazer a análise dos parâmetros do modelo para encontrar fatores realmente explicativos para a variável em estudo. Neste contexto, avaliaremos as distribuições *a posteriori* para os parâmetros de regressão do segundo nível do modelo, o uso de critérios para escolha do modelo e a análise dos riscos. Um critério utilizado é o *Deviance Information Criterion* (DIC), uma forma semelhante ao *Akaike's Information Criterion* (AIC), utilizado em modelos de regressão clássica. O DIC pode ser usado quando se modelam os dados através de MCMC, além de ser amplamente utilizado

para comparar modelos com diferentes níveis de complexidade¹⁶. O modelo que apresentar o menor DIC é considerado aquele que melhor se ajusta aos dados.

Resultados

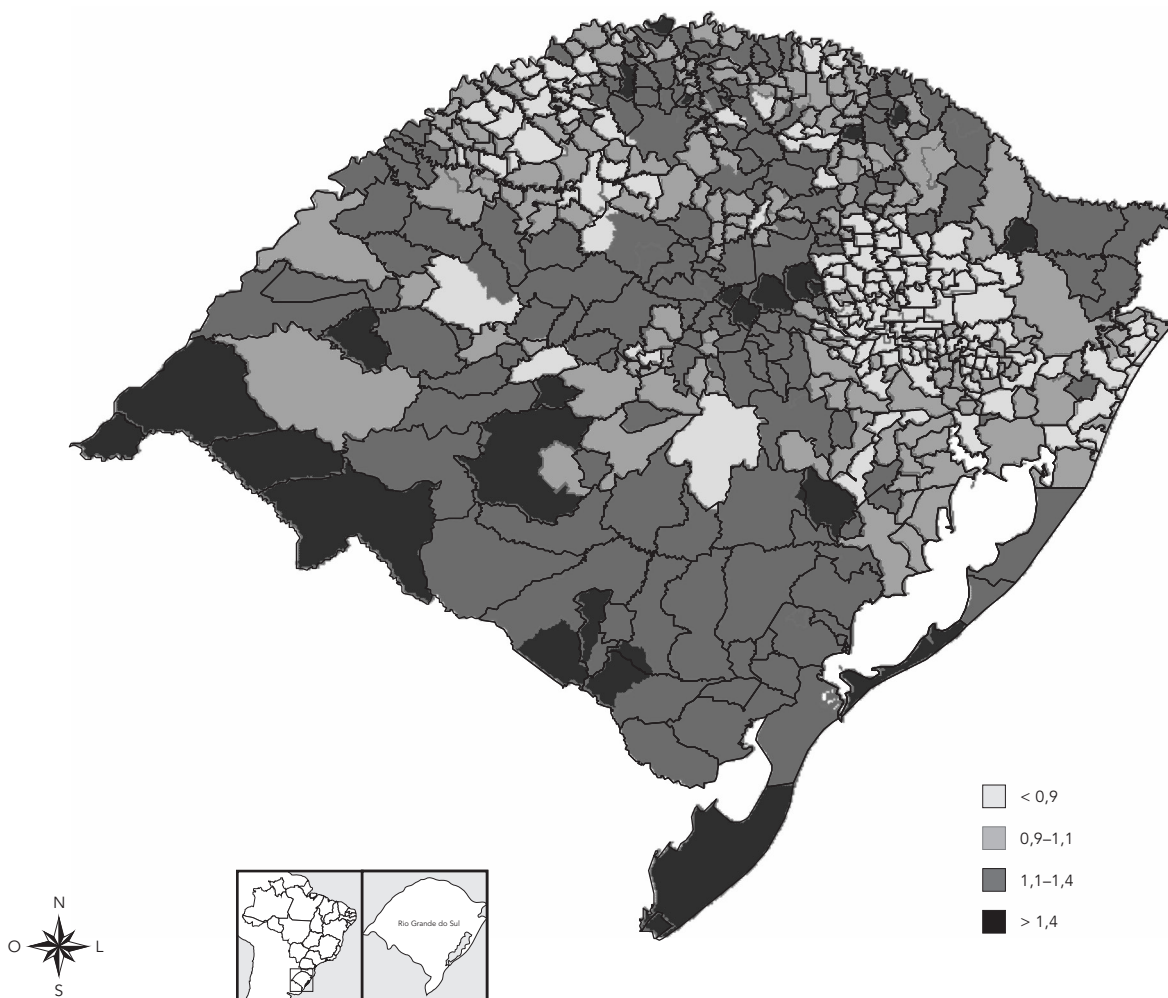
O Rio Grande do Sul, no período entre 2001 e 2004, tinha em sua composição 496 municípios com características muito diferentes em relação ao risco de mortalidade infantil e em relação a indicadores sócio-econômicos. Com o objetivo de buscar estimativas mais robustas, foi utilizada a soma de nascimentos e óbitos no período estudado para o cálculo dos riscos. O número de nascidos vivos por município no Rio Grande do Sul, totalizando os anos de 2001 a 2004, variou de 35 a 79.672. Além do risco relativo estimado pelo modelo de regressão utilizando modelagem inteiramente bayesiana, foi ainda considerada a SMR (*Standardised Mortality Ratio*), estimação de máxima verossimilhança para o risco, também denominada estimativa clássica do risco ou risco relativo bruto, que é a razão entre a taxa da região i e a taxa geral para o conjunto de regiões estudadas. No caso do Rio Grande do Sul entre os anos de 2001 a 2004, a taxa de mortalidade geral é de 15,63 óbitos para cada mil nascidos vivos. Na modelagem bayesiana, as covariáveis utilizadas, como descrito anteriormente, foram os blocos do IDESE de 2003: educação, renda, e condições de saneamento e domicílio.

Na Figura 1 é apresentado o mapa com as SMRs de cada município e na Figura 2, a dispersão entre a SMR e o número de nascidos vivos acumulado (2001 a 2004). O Município de Porto Alegre não foi considerado neste diagrama por causa da grande quantidade de nascidos vivos, o que poderia distorcer os resultados. O que se pode perceber é que a SMR apresenta flutuações aleatórias muito grandes, influenciadas pelo tamanho da população em risco, particularmente quando o número de nascimentos é inferior a 200. Além disso, não considera as informações dos municípios vizinhos na estimativa do risco.

Para superar as dificuldades de instabilidade das estimativas e incluir no modelo a autocorrelação espacial da variável-resposta, todas as regressões foram realizadas utilizando-se métodos de estimação inteiramente bayesianos, e o critério DIC foi empregado para a escolha do modelo que melhor se ajustasse aos dados. Em todos os modelos, utilizaram-se 150 mil simulações, descartando-se as cinco mil primeiras iterações (*burn-in*), com espaçamento de 50 entre os pontos amostrados (*thin*); além disso, verificou-se a convergência através de inspeção visual das

Figura 1

Mapa do risco relativo da mortalidade infantil no Rio Grande do Sul, Brasil (dados acumulados de 2001 a 2004), obtido pelo método de máxima verossimilhança (*Standardised Mortality Ratio*).



trajetórias das cadeias, utilizando-se diferentes valores iniciais, a fim de retirar possíveis autocorrelações na simulação, restando uma amostra de tamanho 2.900.

A Tabela 1 apresenta os modelos avaliados com seus respectivos valores de DIC. Pode-se perceber que o modelo sem efeito espacial, controlando para educação, renda e condições de saneamento e domicílio, é o que apresenta o maior valor de DIC, ou seja, é o que apresenta menor ajuste. Por outro lado, os demais modelos que incorporam o efeito espacial apresentam valores

muito próximos, com destaque para o modelo controlando para educação do IDESE.

Segundo o critério DIC (Tabela 1), o modelo escolhido foi o com efeito espacial que utiliza como covariável apenas o bloco educação do IDESE. Também por intermédio do critério DIC, verificou-se que a inclusão das covariáveis renda e saneamento não melhorou o ajuste, uma vez que todas as três covariáveis são correlacionadas.

A partir de agora, sempre que nos referirmos ao risco relativo obtido através da análise de regressão utilizando modelagem inteiramente

Figura 2

Relação entre *Standardised Mortality Ratio* (SMR) e número de nascidos vivos.

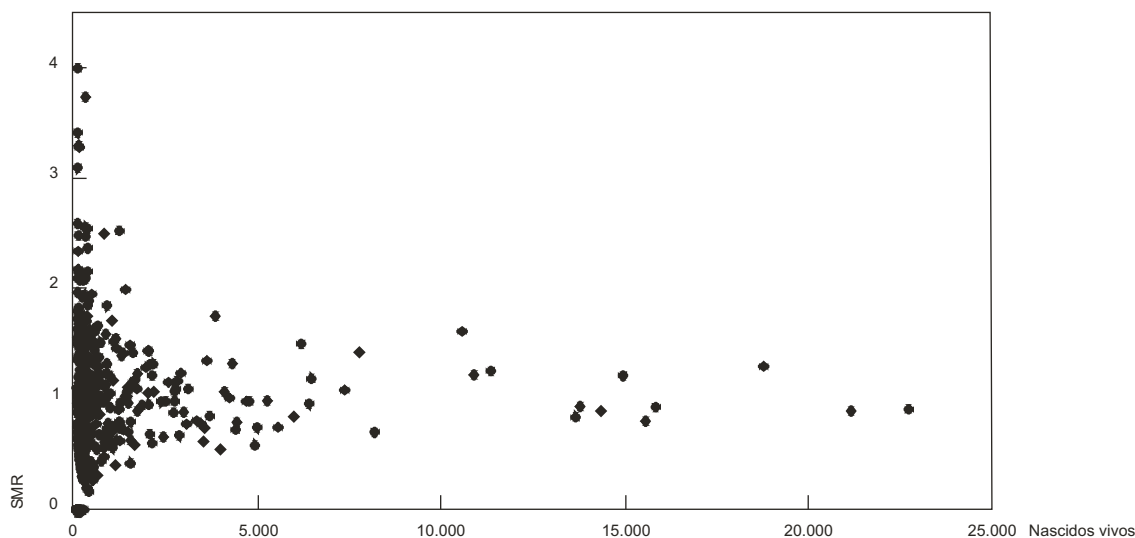


Tabela 1

Modelos avaliados e respectivos valores de *Deviance Information Criterion* (DIC), obtidos pelo *Win Bayesian inference Using Gibbs Sampling* (WinBUGS), baseado em 150 mil simulações.

Modelo	DIC
Sem efeito espacial, controlando por educação, renda e saneamento	2622,96
Com efeito espacial, sem covariável	2332,99
Com efeito espacial, controlando por educação	2306,17
Com efeito espacial, controlando por renda	2334,04
Com efeito espacial, controlando por saneamento	2334,34
Com efeito espacial, controlando por educação e renda	2307,25
Com efeito espacial, controlando por educação e saneamento	2310,89
Com efeito espacial, controlando por renda e saneamento	2335,71
Com efeito espacial, controlando por educação, renda e saneamento	2307,28

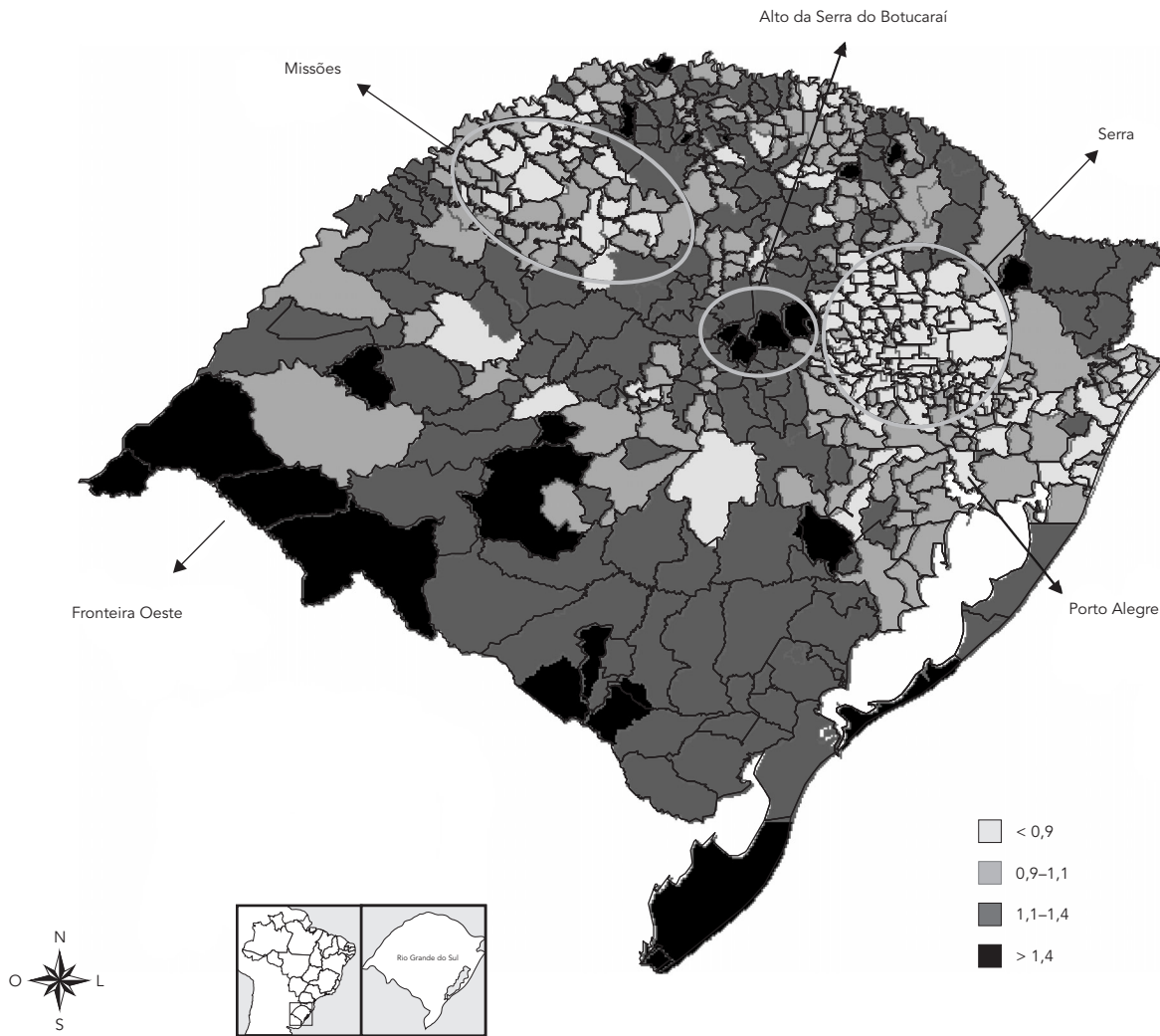
bayesiana, estaremos utilizando apenas o melhor modelo segundo o critério DIC, aquele com efeito espacial que utiliza como covariável o bloco educação do IDESE. A interpretação é similar à das SMRs.

O efeito visual das estimativas obtidas pela modelagem inteiramente bayesiana pode ser verificado na Figura 3. O mapa apresenta padrões de risco com estimativas suavizadas e melhor interpretáveis, além de menores flutuações aleatórias. De acordo com a Figura 4, não existem mais

municípios com risco de mortalidades nula ou extremamente superiores ao risco médio estadual. As estimativas inteiramente bayesianas do risco relativo apresentam bem menos variabilidade, ou seja, são mais concentradas do que as fornecidas pela SMR (Figura 2). Riscos altos obtidos na modelagem clássica foram reduzidos, enquanto valores próximos de 0 ficaram próximos da média da distribuição. Na modelagem bayesiana, aqueles municípios com grande quantidade de nascimentos apresentam pequena variação

Figura 3

Mapa do risco relativo da mortalidade infantil no Rio Grande do Sul, Brasil (dados acumulados de 2001 a 2004), obtido pela modelagem inteiramente bayesiana.



entre as novas estimativas e as SMR. Por outro lado, as estimativas dos municípios com pequena população em risco são mais afetadas pelas informações dos municípios vizinhos, além de serem influenciadas pela covariável educação, sendo assim observadas diferenças maiores entre o SMR e o risco relativo estimado.

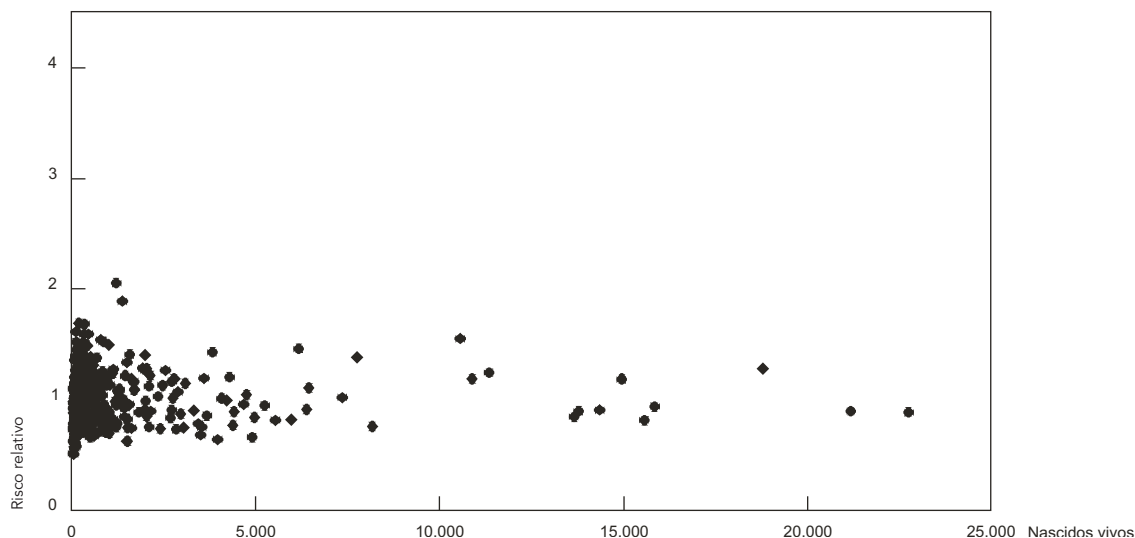
O efeito do tamanho da população sob os riscos relativos de SMR e de estimativa inteiramente bayesiana pode ser exemplificado com o Município Alto Alegre. Nesta cidade, ocorreram 64

nascimentos entre os anos de 2001 a 2004, sendo a SMR aproximadamente quatro vezes (3,998) o risco do estado. Usando a estimativa bayesiana, o risco foi praticamente o mesmo encontrado para o estado (0,968).

Além de o mapa utilizando modelagem inteiramente bayesiana apresentar padrões de risco mais bem definidos, as estimativas municipais podem ser analisadas com maior segurança, pois apresentam maior estabilidade. Assim, os valores extremos podem ser analisados desconsideran-

Figura 4

Relação entre *Standardised Mortality Ratio* e número de nascidos vivos.



do-se a hipótese de que sejam flutuações aleatórias causadas por pequena população em risco.

Na modelagem clássica do risco (SMR), 100 municípios apresentaram risco 40% superior ao risco médio do Estado, ao passo que, no melhor modelo inteiramente bayesiano, foram apenas 25, identificados nos mapas na tonalidade mais escura.

Entre 2001 e 2004, contam-se 50 municípios sem óbito infantil e com poucos nascimentos em geral. A estimativa clássica (SMR) para o risco relativo desses municípios seria 0, porém essa é uma estimativa teoricamente incorreta, pois, dessa forma, não haveria risco de ocorrência de óbito infantil nessas localidades, quando, na realidade, o risco existe, ainda que não manifestado na prática. Por meio da modelagem inteiramente bayesiana, os municípios passaram a ter pelo menos a metade do risco de mortalidade infantil do Estado, chegando a atingir, como é o caso do Município de Herveiras, 1,231, ou seja, um risco 23% maior do que a média estadual. Nessa cidade, o índice educação do IDESE foi o menor entre os municípios que não apresentaram mortalidade infantil no período.

Ainda avaliando os padrões de risco, é possível destacar positivamente a região da serra gaúcha, com baixas taxas de risco e muito homogênea. Outras regiões com agrupamentos de baixo risco de mortalidade infantil são: Região Metro-

politana de Porto Alegre, Litoral Norte, Missões e Fronteira Nordeste. Por outro lado, algumas regiões apresentam alto risco de mortalidade infantil, principalmente nos municípios localizados nas regiões da Fronteira Oeste, Sul e Alto da Serra do Botucará.

Conclusões e considerações finais

Neste artigo, utilizamos modelos de regressão espacial inteiramente bayesiano para estimar o risco de mortalidade infantil nos municípios do Rio Grande do Sul (dados acumulados para os anos de 2001 a 2004), utilizando como co-variáveis subescalas do IDESE de 2003. O método baseia-se em simulações MCMC para estimar as distribuições posteriores dos riscos relativos, incorporando efeitos de origem aleatória na ocorrência do evento em cada município e testando diferentes funções de regressão para o risco, além de possibilitar a estimação dos parâmetros. Como resultado, temos mapas com padrões de risco mais precisos, com estimativas suavizadas e com menos flutuações aleatórias.

Na modelagem inteiramente bayesiana, foram avaliadas as possíveis influências das covariáveis, além do efeito espacial. Observou-se que os modelos que incorporam a estrutura espacial além de covariáveis apresentaram melhor de-

sempenho, padrão também detectado em estudo que utilizou a abordagem inteiramente bayesiana para avaliar a distribuição de homicídios na cidade de Curitiba, Paraná, Brasil ¹⁷.

Comparando as SMR com o risco relativo obtido pela modelagem inteiramente bayesiana, foi possível observar um ganho substancial na interpretação e na detecção de padrões de variação no risco de mortalidade infantil nos municípios do Rio Grande do Sul. Na modelagem inteiramente bayesiana, as estimativas municipais puderam ser analisadas com maior segurança, pois apresentaram maior estabilidade. Desse modo, os valores extremos puderam ser analisados desconsiderando-se a hipótese de flutuações aleatórias causadas por pequena população em risco, ou seja, os municípios com altos riscos relativos de mortalidade infantil devem sofrer ações prioritárias.

Além das vantagens já apontadas da modelagem inteiramente bayesiana, o método permite a incorporação da modelagem espaço-temporal, a fim de aumentar a precisão das estimativas do risco de mortalidade infantil. Comparando os achados com uma análise espaço-temporal ¹³ que avalie também a mortalidade infantil no Rio Grande do Sul, mas sem o uso de covariáveis, é possível observar semelhança nos padrões de risco.

Na epidemiologia espacial, existe uma grande flexibilidade para explorar melhor os métodos de mapeamento de doenças e de correlação es-

pacial, a partir de diferentes estruturas de vizinhança. Embora existam várias formas de definir a matriz de vizinhança, neste artigo utilizamos apenas a matriz binária, mas poderia ser considerado o tamanho da fronteira, o tamanho das fronteiras com ou sem a presença de barreiras naturais, ou outra característica geográfica que possa interferir na ligação entre áreas ¹⁸. Diferentes especificações na matriz de vizinhança geram diferentes estimativas. Na literatura, a comparação de estimativas utilizando diferentes matrizes de vizinhança ainda é pouco explorada.

Atualmente também estão sendo conduzidos estudos de simulação ¹⁴ com o objetivo de comparar os modelos bayesianos com modelos semiparamétricos.

Com base nos resultados, acredita-se que os governantes e as secretarias de saúde municipais e estaduais podem ter uma melhor visão do risco de mortalidade infantil subjacente para os municípios, sem as flutuações aleatórias inerentes aos dados brutos. Com isso, possibilita-se a geração de novas políticas públicas para a mortalidade infantil no Rio Grande do Sul.

Várias áreas do conhecimento podem se beneficiar dos métodos aqui expostos, no entanto a obtenção das estimativas a partir da modelagem inteiramente bayesiana ainda é um processo complexo e exige tempo na execução das interações, além de ainda não existir um programa que combine a modelagem com a criação de mapas.

Resumo

Neste artigo são analisados os fatores possivelmente associados à mortalidade infantil nos 496 municípios do Rio Grande do Sul, Brasil, com base em dados acumuladas entre os anos de 2001 a 2004, obtidos pela análise de regressão utilizando modelagem inteiramente bayesiana como alternativa para superar a autocorrelação espacial e a instabilidade dos estimadores clássicos, como a taxa bruta e a SMR (Standardised Mortality Ratio). Foram comparadas diferentes especificações de componente espacial e covariáveis, provenientes dos blocos do Índice de Desenvolvimento Sócio-econômico da Fundação de Economia e Estatística (IDESE/FEE-2003). Verificou-se que o modelo que utiliza a estrutura espacial além da covariável educa-

ção apresenta melhor desempenho, quando comparado pelo critério DIC (Deviance Information Criterion). Comparando as estimativas das SMR com os riscos relativos obtidos pela modelagem inteiramente bayesiana, foi possível observar um ganho substancial na interpretação e na detecção de padrões de variação do risco de mortalidade infantil nos municípios do Rio Grande do Sul ao utilizar essa modelagem. A região da Serra Gaúcha destacou-se com baixo risco relativo e estimativas muito homogêneas.

Mortalidade Infantil; Análise Estatística; Medição de Risco

Colaboradores

S. K. Kato participou da concepção do estudo, da análise dos dados, da discussão dos resultados, da redação e da revisão do manuscrito. J. M. G. Fachel participou da concepção do estudo, da redação e da revisão do manuscrito. D. M. Vieira contribuiu na concepção do estudo e participou da análise dos dados e revisão do manuscrito.

Referências

1. Maia SF, Sousa TRV. Uma investigação dos determinantes da redução da taxa de mortalidade infantil nos estados da Região Nordeste do Brasil. In: I Congresso da Associação Latino-Americana de População. Caxambu: Associação Latino-Americana de População; 2004. p. 1-17.
2. Departamento de Informática do SUS. Informações de saúde. Estatísticas vitais – mortalidade e nascidos vivos. <http://www.datasus.gov.br> (acessado em 12/Mar/2007).
3. Victora CG. Intervenções para reduzir a mortalidade infantil pré-escolar e materna no Brasil. *Rev Bras Epidemiol* 2001; 4:3-69.
4. Carvalho MS, Santos RS. Análise de dados espaciais em Saúde Pública: métodos, problemas, perspectivas. *Cad Saúde Pública* 2005; 21:361-78.
5. Rojas LI, Barcellos C, Peiter P. Utilização de mapas no campo da epidemiologia no Brasil. *Inf Epidemiol SUS* 1999; 8:25-35.
6. Lawson A. *Statistical methods in spatial epidemiology*. Sussex: John Wiley & Sons; 2001.
7. Elliott P, Wakefield J, Best N, Briggs D. *Spatial epidemiology: methods and applications*. London: Oxford University Press; 2001.
8. Bailey TC, Gatrell AC. *Interactive spatial data analysis*. London: Longman; 1995.
9. Cressie N. *Statistics for spatial data*. New York: Wiley; 1991.
10. Assunção RM. *Estatística espacial com aplicações em epidemiologia, economia e sociologia*. <http://www.est.ufmg.br/leste/publicacoes.htm> (acessado em 08/Mai/2006).
11. Assunção RM, Barreto SM, Guerra HL, Sakura E. Mapas de taxas epidemiológicas: uma abordagem bayesiana. *Cad Saúde Pública* 1998; 14:713-23.
12. Santos SM, Noronha CP. Padrões espaciais de mortalidade e diferenciais sócio-econômicos na cidade do Rio de Janeiro. *Cad Saúde Pública* 2001; 17:1099-110.
13. Vieira DM. *Análise espaço-temporal na estimação de taxas de incidência/mortalidade [Dissertação de Mestrado]*. Porto Alegre: Faculdade de Medicina, Universidade Federal do Rio Grande do Sul; 2006.
14. Richardson S, Thomson A, Best NG, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* 2004; 112:1016-25.
15. Thomas A, Spiegelhalter DJ, Gilks WR. BUGS: a program to perform Bayesian inference using Gibbs Sampling. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian statistics*. v. 4. Oxford: Clarendon Press; 1992. p. 837-42.
16. Spiegelhalter DJ, Best NG, Carlin BP, Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Series B Stat Methodol* 2002; 64:583-639.
17. Ehlers RS, Silva SA, Melo LLM. Fully Bayesian spatial analysis of homicide rates. *Estadística* 2006; 58:43-59.
18. Mollié A. Bayesian mapping of disease. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. *Monte Carlo Markov Chain in practice*. London: Chapman and Hall; 1996.

Recebido em 11/Mar/2008

Versão final reapresentada em 02/Out/2008

Aprovado em 02/Mar/2009