



Trabalho de conclusão de curso

Modelo hierárquico bayesiano para mapeamento da aids no Rio Grande do Sul entre os anos 2001 e 2018

Renan Baiocco Pereira

11 de maio de 2021

Renan Baiocco Pereira

**Modelo hierárquico bayesiano para mapeamento da aids no
Rio Grande do Sul entre os anos 2001 e 2018**

Trabalho de conclusão apresentado à comissão de graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Rodrigo Citton Padilha dos Reis
Coorientadora: Prof^a. Dr^a. Márcia Helena Barbian

Porto Alegre
Maio de 2021

Renan Baiocco Pereira

Modelo hierárquico bayesiano para mapeamento da aids no
Rio Grande do Sul entre os anos 2001 e 2018

Este trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pelo orientador, pela coorientadora e pela banca examinadora.

Orientador: _____

Prof. Dr. Rodrigo Citton Padilha dos Reis, UFMG
Doutor pela Universidade Federal de Minas Gerais,
Belo Horizonte, MG

Coorientadora: _____

Prof^ª. Dr^ª. Márcia Helena Barbian, UFMG
Doutora pela Universidade Federal de Minas Gerais,
Belo Horizonte, MG

Banca examinadora:

Prof^ª. Dr^ª. Silvana Schneider, UFMG
Doutora pela Universidade Federal de Minas Gerais – Belo Horizonte, MG

Porto Alegre
Maio de 2021

“Ciência não é uma escolha. É uma necessidade.” (Marcelo Gleiser)

Agradecimentos

À minha mãe e ao meu pai, que, além de sempre me incentivarem a focar nos estudos, nunca mediram esforços para que eu o pudesse priorizar. Se não fosse por todo o apoio e amparo deles, eu não teria chegado até aqui.

À minha irmã e ao meu irmão, por toda a ajuda nesses anos de graduação.

Às minhas amigas e aos meus amigos, por todo o companheirismo e auxílio que me foram dados em forma de alegrias, risadas, conversas, partilhas de tristezas e inclusive desesperos com os estudos, durante esses últimos anos.

A todos os professores e professoras do Departamento de Estatística que fizeram parte da minha jornada acadêmica, por todo o conhecimento a mim transmitido. Em especial ao meu orientador e à minha coorientadora, por toda a ajuda durante a construção deste trabalho.

Resumo

A síndrome da imunodeficiência adquirida (aids) é uma possível decorrência do vírus da imunodeficiência humana. Alguns autores alegam que o vírus pode ter esporadicamente infectado humanos ao longo do século XX, entretanto, foi só na década de 1980 que esse recebeu atenção mundial. No Brasil, entre os anos 1980 e 2019, quase um milhão de casos foram notificados. Com o objetivo de verificar a distribuição do risco relativo dos casos de aids pelo estado do Rio Grande do Sul ao longo dos anos, foi ajustado um modelo hierárquico bayesiano. Utilizando dados de órgãos públicos e o modelo espaço-temporal proposto por Bernardinelli, foram estimados os riscos relativos dos municípios do estado do Rio Grande do Sul, entre os anos de 2001 e 2018. A distribuição *a posteriori* do modelo foi calculada através da aproximação de Laplace aninhada integrada, com o auxílio do pacote R-INLA. Foi observado que o aumento no índice de desenvolvimento humano municipal está associado com o aumento no risco relativo de casos de aids nos municípios do Rio Grande do Sul. Com o passar dos anos, nota-se uma decrescência/estabilização no risco relativo de casos de aids.

Palavras-chave: Síndrome da imunodeficiência adquirida, Espaço-temporal, Estatística bayesiana, Risco relativo, R-INLA.

Abstract

The acquired immunodeficiency syndrome (aids) is a possible consequence of the human immunodeficiency virus. Some authors say that the virus may have sporadically infected humans throughout the 20th century, however, it was only in the 1980's that it received worldwide attention. In Brazil, between the years of 1980 and 2019, almost a million cases were notified. Aiming to verify the distribution of the relative risk of aids cases in the state of Rio Grande do Sul over the years, a bayesian hierarchical model was fitted. Using data from governmental sources and a spatiotemporal model proposed by Bernardinelli, relative risks from municipalities of Rio Grande do Sul between the years of 2001 and 2018 were estimated. The model's posterior distribution was calculated via integrated nested Laplace approximation, assisted by the R-INLA package. It was noticed that an increase in the municipal human development index is associated with an increase in the relative risk of aids cases in the municipalities of Rio Grande do Sul. Over the years, a decrease/stabilization in the relative risk of aids cases was observed.

Keywords: Acquired immunodeficiency syndrome, Spatiotemporal, Bayesian statistics, Relative risk, R-INLA.

ARTIGO CIENTÍFICO

ARTIGO CIENTÍFICO

Modelo hierárquico bayesiano para mapeamento da aids no Rio Grande do Sul entre os anos 2001 e 2018.

Bayesian hierarchical model for the mapping of aids in Rio Grande do Sul between the years of 2001 and 2018.

**Renan Baiocco Pereira, bacharelado em Estatística pela UFRGS.
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS).**

A ser enviado para:
Revista Brasileira de Biometria

MODELO HIERÁRQUICO BAYESIANO PARA MAPEAMENTO DA AIDS NO RIO GRANDE DO SUL ENTRE OS ANOS 2001 E 2018

Renan Baiocco PEREIRA¹

- RESUMO: A síndrome da imunodeficiência adquirida (aids) é uma possível decorrência do vírus da imunodeficiência humana. Alguns autores alegam que o vírus pode ter esporadicamente infectado humanos ao longo do século XX, entretanto, foi só na década de 1980 que esse recebeu atenção mundial. No Brasil, entre os anos 1980 e 2019, quase um milhão de casos foram notificados. Com o objetivo de verificar a distribuição do risco relativo dos casos de aids pelo estado do Rio Grande do Sul ao longo dos anos, foi ajustado um modelo hierárquico bayesiano. Utilizando dados de órgãos públicos e o modelo espaço-temporal proposto por Bernardinelli, foram estimados os riscos relativos dos municípios do estado do Rio Grande do Sul, entre os anos de 2001 e 2018. A distribuição *a posteriori* do modelo foi calculada através da aproximação de Laplace aninhada integrada, com o auxílio do pacote R-INLA. Foi observado que o aumento no índice de desenvolvimento humano municipal está associado com o aumento no risco relativo de casos de aids nos municípios do Rio Grande do Sul. Com o passar dos anos, nota-se uma decrescência/estabilização no risco relativo de casos de aids.
- PALAVRAS-CHAVE: Síndrome da imunodeficiência adquirida; Espaço-temporal; Estatística bayesiana; Risco relativo; R-INLA

¹Universidade Federal do Rio Grande do Sul - UFRGS, Instituto de Matemática e Estatística, Departamento de Estatística, CEP: 90650-001, Porto Alegre, RS, Brasil. E-mail: renanbaiocco2@gmail.com

1 Introdução

O vírus da imunodeficiência humana (HIV, sigla em inglês para *human immunodeficiency virus*) é um vírus que, além de atacar o sistema imunológico, responsável pela defesa de doenças no organismo, também consegue mudar o ácido desoxirribonucleico (DNA, do inglês *deoxyribonucleic acid*) das células e fazer cópias de si mesmo. Pessoas que vivem com o HIV podem, ou não, desenvolver a aids (sigla em inglês para síndrome da imunodeficiência adquirida), mas isso não interfere na transmissão do vírus, que é feita por relações sexuais sem preservativo, compartilhamento de seringas, por transmissão vertical (de mãe para filho durante a gravidez) e na amamentação por pessoas infectadas. Ou seja, uma pessoa infectada pelo HIV pode transmitir o vírus mesmo sem ter desenvolvido a aids.

Deeks *et al.* (2015) relatam que o vírus pode ter, de forma esporádica, infectado humanos ao decorrer dos anos 1900, mas que foi apenas na década dos anos 1980 que o vírus recebeu uma atenção mundial. Os autores alegam que a descoberta do HIV como causa da aids aconteceu somente em 1983, dois anos depois da notificação do primeiro caso de aids. Até o final de 2019 estimava-se que o HIV havia infectado cerca de 75 milhões de pessoas em todo o mundo e que 38 milhões de pessoas viviam com o vírus, sendo que a maior parte desses casos (aproximadamente 20 milhões) se concentrava nas partes oriental e austral da África (UNAIDS, 2020). Apesar da maior parte dos casos se concentrar no continente africano, Naghavi *et al.* (2015) afirmam que a infecção pelo HIV é causa de mortalidade e morbidade em todo o mundo.

De acordo com o boletim epidemiológico de HIV/aids de 2019, divulgado pelo Ministério da Saúde, foram identificados quase um milhão de casos de aids no Brasil entre 1980 e junho de 2019. Nesse mesmo período, a região Sudeste teve uma maior concentração de casos, com 51,3%. A região Sul assume a segunda posição, com 19,9%, Nordeste, Norte e Centro-Oeste possuem, respectivamente, 16,1%, 6,6% e 6,1%. No ano de 2013, quase 43 mil casos foram notificados no Brasil, já em 2018, foram 37.161 (MINISTÉRIO DA SAÚDE, 2019). Num informativo sobre a situação do HIV no Brasil, foi divulgado que se estima que mais de 130 mil brasileiros vivem com o vírus e não sabem. O mesmo boletim também mostrou que 2,5 mil mortes por aids foram evitadas no período entre 2014 e 2018 e que o número de mortes teve uma queda de 22,8% no mesmo período (AQUINO, 2019).

O mapeamento de casos de aids é importante para que seja possível a visualização das estimativas locais, como taxas de incidência/prevalência, e para avaliar uma possível heterogeneidade espaço-temporal. Essa forma de apresentar os dados facilita a visualização de fatores de risco, padrões espaciais em áreas que apresentam maiores taxas de casos e outras características que podem auxiliar gestores em saúde a tomarem decisões mais precisas no combate à aids. Notificações de casos que são divulgadas a nível de municípios podem ser utilizadas para a construção de mapas de regiões de risco elevado, que seriam monitoradas pelos órgãos de vigilância das macrorregião ou mesorregião específicas. Fazendo com que as políticas públicas municipais, estaduais e federais de prevenção e atendimento

sejam direcionadas à essas áreas.

A avaliação da heterogeneidade espacial pode ser feita através de estimativa de taxas ou da razão de incidência padronizada (SIR, do inglês *standardized incidence ratio*), uma medida que estima o risco relativo de doenças em áreas. A SIR é obtida por meio de uma razão entre o número de casos observados e o número de casos esperados nas populações, mas não é o mais recomendado quando estamos lidando com áreas que apresentam uma população pequena ou poucos casos registrados. Nesses cenários, preferencialmente, são utilizados modelos espaciais em que a estrutura de vizinhança é levada em consideração na análise. Isso é, modelos estatísticos que levem em consideração que locais mais próximos de uma certa área possuem uma influência maior sobre ela do que locais mais distantes (MORAGA, 2019). Quando as notificações são dispostas ao longo do tempo, modelos espaço-temporais podem ser ajustados e, dentre as metodologias espaciais e espaço-temporais, destacam-se os modelos hierárquicos bayesianos.

O objetivo deste artigo é estimar o risco relativo referente aos casos de aids nos municípios do estado do Rio Grande do Sul, entre os anos de 2001 e 2018, por meio da razão de incidência padronizada e por meio do modelo de Bernardinelli (BERNARDINELLI *et al.*, 1995).

O restante do artigo está estruturado da seguinte forma: na [Seção 2](#) são apresentados os métodos, onde são descritas algumas informações sobre os dados e sobre as variáveis de estudo, além de apresentar a modelagem espaço-temporal de dados de área. Na [Seção 3](#) são apresentados os resultados obtidos com o estudo e a discussão desses. Por fim, são apresentadas as [Conclusões](#). No [Apêndice](#) estão os códigos, em R, utilizados para a elaboração do artigo.

2 Métodos

2.1 Descrição sobre os dados e variáveis do estudo

Foram utilizados todos os casos de aids diagnosticados em pessoas com mais de 15 anos nos municípios do estado do Rio Grande do Sul, entre os anos 2001 e 2018, disponibilizados pelo Sistema de Informação de Agravos de Notificação (SINAN) para o Ministério da Saúde (SINAN, 2019). A população residente nos municípios do RS foi obtida a partir do Censo, para 2010 (IBGE, 2012), dados de contagem para o ano de 2007 (SIDRA, 2020a) e estimativas de população para os intervalos 2001-2006, 2008-2009 e 2011-2018 (SIDRA, 2020b).

O índice de desenvolvimento humano municipal (IDHM), que reúne informações sobre educação, renda e saúde de cada município, foi utilizado como covariável, a fim de verificar o impacto do fator socioeconômico no número de casos de aids dos municípios. O IDHM dos municípios do estado do RS foram consultados no Atlas do Desenvolvimento Humano no Brasil (ATLAS BRASIL, 2020). Este índice é disponibilizado para os anos de realização do censo demográfico. Assim, para as análises deste artigo, o IDHM referente ao ano de 2000 foi repetido para anos de 2001 a 2009 e, da mesma forma, o IDHM referente ao ano de 2010 foi

repetido para os anos de 2010 a 2018.

Dos 497 municípios do RS, apenas Pinto Bandeira não foi avaliado no estudo. Por ter sido fundado em 2013, possui informação do número de população para seis dos dezoito anos de interesse, além de não ter informações sobre o seu IDHM.

2.2 Mapeamento de doenças

A maioria das abordagens de mapeamento de doenças estimam os riscos de doenças em unidades de área, que formam uma partição da região de estudo (WALLER *et al.*, 2004). Frequentemente, o risco de doença na área $i = 1, \dots, N$ é estimado pela razão de incidência padronizada (SIR). Neste estudo as áreas são representadas pelos municípios e a região de estudo é o estado do Rio Grande do Sul. A SIR é obtida como a razão do número de casos da doença observado e o número de casos esperados (também chamado de contagem esperada): $SIR_i = Y_i/E_i$. As contagens esperadas em cada uma das áreas são calculadas com base em suas populações. Especificamente,

$$E_i = r \times n_i,$$

em que r é a taxa de casos na população padrão (o número de casos da região em estudo dividido pela população da região em estudo) e n_i é a população da área i . As contagens esperadas representam o número total de eventos que se esperaria se i , a população observada, se comportasse da mesma maneira que a população padrão (ou população de referência) se comportou.

Em áreas com populações pequenas, as SIRs podem ser muito extremas e não são confiáveis para relatar a distribuição espacial de determinado agravo. Os modelos bayesianos de mapeamento de doenças são uma alternativa para obter estimativas de risco, uma vez que permitem o “empréstimo” de informações de áreas vizinhas e também podem incorporar informações de covariáveis para melhorar as estimativas locais (GELFAND *et al.*, 2010).

Modelos de mapeamento de doenças espaço-temporais são usados quando o interesse é entender os padrões de doenças espaciais e temporais (KNORR-HELD, 2000). Neste artigo foi adotado o modelo de Bernardinelli *et al.* (1995). Este modelo assume que o número de casos Y_{ij} observado no município i e no ano j segue uma distribuição *Poisson* com parâmetro $E_{ij}\theta_{ij}$, em que θ_{ij} é o risco e E_{ij} é o número de casos esperados no município i e no ano j :

$$Y_{ij} \sim \text{Poisson}(E_{ij}\theta_{ij}). \quad (1)$$

Bernardinelli *et al.* (1995) propuseram que tendências temporais paramétricas podem ser expressas em relação ao logaritmo dos riscos relativos como

$$\log(\theta_{ij}) = \alpha + u_i + v_i + (\beta + \gamma_i) \times t_j, \quad (2)$$

em que α denota o intercepto, $u_i + v_i$ é um efeito aleatório de área (u_i é um componente espacial estruturado, enquanto v_i é um componente espacial não

estruturado), β é um efeito de tendência linear global γ_i é um termo de interação entre espaço e tempo, que representa a diferença entre a tendência global e a tendência específica de área. Para levar em consideração o IDHM como um fator de risco, a seguinte estrutura do risco foi assumida:

$$\log(\theta_{ij}) = \beta_0 + \beta_1 \times IDHM_i + u_i + v_i + (\beta_2 + \gamma_i) \times t_j, \quad (3)$$

em que β_0 é o termo de intercepto e β_1 o coeficiente associado à covariável IDHM.

Os efeitos u_i e γ_i foram modelados conforme o modelo Besag-York-Mollie (BYM) (BESAG *et al.*, 1991), que utiliza uma distribuição autorregressiva condicional (CAR). A distribuição CAR suaviza os dados de acordo com uma certa estrutura de vizinhança binária, com uma matriz de vizinhança que especifica que duas áreas são vizinhas se compartilham um limite comum. A distribuição CAR é expressa como:

$$u_i | \mathbf{u}_{-i} \sim N \left(\bar{u}_{\delta_i}, \frac{\sigma_u^2}{n_{\delta_i}} \right),$$

em que $\bar{u}_{\delta_i} = \sum_{j \in \delta_i} u_j / n_{\delta_i}$, δ_i representa o conjunto de vizinhos da área i , e n_{δ_i} é o número de vizinhos da área i .

Os efeitos v_i foram modelados como variáveis normais independentes e identicamente distribuídas, dados por:

$$v_i \sim N(0, \sigma_v^2).$$

Este modelo permite que cada uma das áreas tenha seu próprio intercepto $\beta_0 + \beta_1 \times IDHM_i + u_i + v_i$, e sua própria tendência linear dada por $\beta_2 + \gamma_i$. O risco relativo θ_{ij} quantifica se o risco de doença no município i e no ano j é maior ($\theta_{ij} > 1$) ou menor ($\theta_{ij} < 1$) do que o risco médio no Rio Grande do Sul durante o período de estudo.

A inferência bayesiana pode ser realizada usando a aproximação de Laplace aninhada integrada (INLA, do inglês *integrated nested Laplace approximation*) (RUE *et al.*, 2007). O método INLA foi utilizado para aproximar a distribuição *a posteriori* do modelo descrito acima. No INLA, que deve ser utilizado em modelos gaussianos latentes (classe ampla e flexível de modelos, que engloba tanto o modelo linear generalizado misto como os modelos espaciais e espaço-temporais) (LINDGREN *et al.*, 2015), assume-se que a variável resposta Y pertence a uma distribuição da família exponencial onde a média μ_i é relacionada a um preditor aditivo estruturado η_i por meio de uma função de ligação $g(\cdot)$, de modo que $g(\mu_i) = \eta_i$ (RUE *et al.*, 2009). Este método, computacionalmente eficiente para obter a distribuição *a posteriori*, está implementado em R no pacote R-INLA (BLANGIARDO *et al.*, 2013).

Todas as análises foram realizadas no software R, versão 3.6.1 (R CORE TEAM, 2019). Os códigos em R utilizados para a implementação computacional são apresentados no [Apêndice](#).

3 Resultados

Os 496 municípios avaliados somaram 74.678 casos entre os anos 2001 e 2018, tendo 3.348 como total no primeiro ano avaliado e 3.121 no último, sendo esse, o ano com o menor número. Na Tabela 1 é possível verificar a soma e a taxa de casos por 10 mil habitantes, a cada ano. O ano com o maior número total de casos foi 2007, com 4.908, e a maior taxa de casos foi no mesmo ano, com 0,0046 casos por 10 mil habitantes.

Tabela 1 – Total e taxa de casos de aids no RS entre os anos 2001 e 2018.

Ano	Total de casos	Taxa de casos por 10 mil	Ano	Total de casos	Taxa de casos por 10 mil
2001	3.348	0,0032	2010	4.513	0,0042
2002	4.101	0,0039	2011	4.626	0,0043
2003	4.274	0,0041	2012	4.670	0,0043
2004	3.883	0,0036	2013	4.656	0,0042
2005	3.845	0,0035	2014	4.342	0,0039
2006	4.155	0,0038	2015	3.937	0,0035
2007	4.908	0,0046	2016	3.611	0,0032
2008	4.761	0,0044	2017	3.367	0,0030
2009	4.560	0,0042	2018	3.121	0,0028

Conforme descrito anteriormente, uma forma de verificar uma possível heterogeneidade espacial é por meio da SIR. Na Figura 1 são apresentados os SIR para os 496 municípios do RS aqui estudados, ao passar dos dezoito anos avaliados. O município de Pinhal ocupa metade das posições das dez maiores taxas, sendo 6,015, em 2017, a maior taxa do estado. A segunda e a terceira posições, também ocupadas por Pinhal, são taxas de 5,218 em 2010 e 5,026 em 2015. Aparecem também entre as dez maiores taxas os municípios de Fagundes Varela, Mato Queimado, São Valentim do Sul, Pouso Novo e Tramandaí. Pode-se verificar com base na Figura 1 que, ao passar dos anos, as taxas nos municípios do estado diminuem levemente, apesar de alguns municípios pontuais apresentarem uma taxa relativamente alta em comparação com os outros, como é o caso nos anos 2007, 2011 e 2013, por exemplo.

Na Figura 2 é apresentado um gráfico da SIR ao longo do tempo de estudo. É dado destaque às cidades com maiores valores de SIR, sendo elas: Pinhal, Fagundes Varela, Mato Queimado, São Valentim do Sul, Pouso Novo, Tramandaí e, além dessas, a capital do estado, Porto Alegre. É possível verificar que a taxa na capital gaúcha, após aumentar em 2005, segue uma tendência de queda, enquanto os outros municípios destacados, com exceção de Pinhal, apresentam pontos altos apenas em anos isolados. Os outros municípios avaliados, representados pelas linhas em cinza na Figura 2, apresentam, em sua maioria, taxas menores em todo o período avaliado.



Figura 1 – SIR dos casos de aids do RS entre os anos 2001 e 2018.

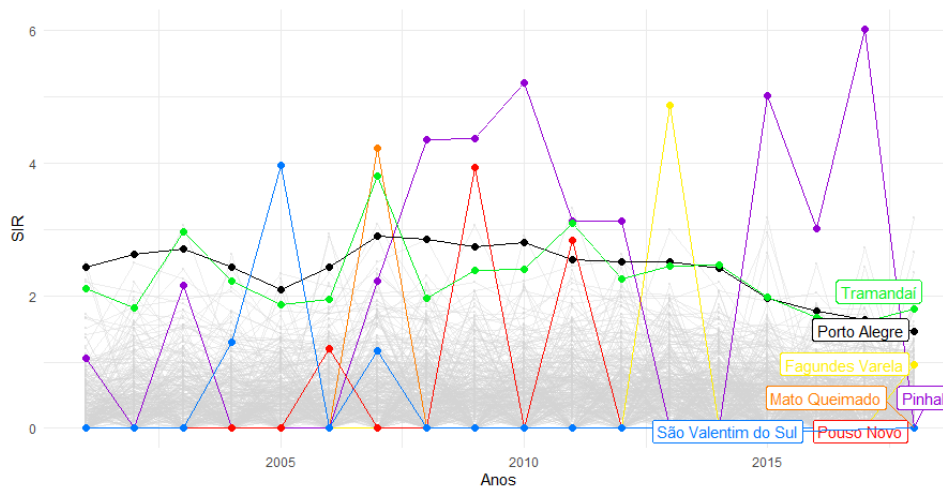


Figura 2 – *Timeplot* dos SIR dos casos de aids do RS entre os anos 2001 e 2018.

Na Tabela 2 são apresentados os resultados obtidos com o modelo espaço-temporal de Bernardinelli. As estimativas para os efeitos fixos do modelo, baseadas na média, o desvio padrão, os quartis e o valor da moda da distribuição *a posteriori*. Pode-se verificar que a média *a posteriori* para o intercepto é $-2,345$, com um intervalo de credibilidade de $(-2,560; -2,130)$. Com uma média *a posteriori* de $1,455$ para o coeficiente associado à covariável IDHM, podemos concluir que quanto maior for o IDHM de um município do RS, maior é o risco relativo de casos de aids nesse. Além disso, com um intervalo de credibilidade de $(1,105; 1,805)$, temos a variável IDHM como significativa no modelo. Essa relação entre o IDHM e o risco relativo foi encontrada também por Prado e Castilho (2009), onde, em um estudo realizado no estado de São Paulo, os autores encontraram que municípios com maiores IDHM apresentavam também maior risco relativo de casos de aids.

Tabela 2 – Estimativas para os efeitos fixos do modelo.

	Média	Desvio padrão	0,025	0,5	0,975	Moda
Intercepto	-2,345	0,110	-2,560	-2,345	-2,130	-2,345
IDHM	1,455	0,178	1,105	1,455	1,805	1,455
Tempo	-0,003	0,003	-0,008	-0,003	0,002	-0,003

Para os hiperparâmetros, além do modelo retornar a média, o desvio padrão e os quartis, ele também retorna o *mod*. Esses valores estão descritos na Tabela 3, onde o componente independente e identicamente distribuído “idarea” possui uma média *a posteriori* para a precisão estimada em $3,83$ e o componente espacial “idarea”, de $1,49$. É possível verificar que a variância (inverso do parâmetro de precisão,

apresentado na coluna "Média") para o componente da interação entre área e tempo ("idarea1") é a menor.

Tabela 3 – Estimativas para os hiperparâmetros do modelo.

Precisão para...	Média	Desvio padrão	0,025	0,5	0,975	Mod
idarea ¹	3,83	0,64	2,73	3,78	5,25	3,67
idarea ²	1,49	0,35	0,91	1,45	2,30	1,38
idarea1 ³	1782,59	276,18	1296,01	1763,54	2381,11	1727,85

A Figura 3 apresenta o mapa dos riscos relativos estimados pela média *a posteriori* do modelo espaço-temporal de Bernardinelli para os municípios do RS, no período de estudo. Pode-se verificar que risco relativo parece diminuir ao passar dos anos, apesar de algumas áreas ainda possuírem um risco relativo visivelmente maior do que outras, como é o caso da região metropolitana de Porto Alegre e da região Sul do estado, por exemplo.

Apesar dessa possível interpretação com base nos gráficos, a média *a posteriori* para o tempo (Tabela 2) não indica uma interferência direta no risco, no sentido que, com base no tempo, o risco relativo dos casos de aids no RS permanece estabilizado. Assim como na Figura 1, aqui também é possível verificar casos pontuais de municípios que apresentam um risco relativo alto, como por exemplo nos anos 2010 e 2016.

Na Figura 4 são apresentados os riscos relativos dos casos de aids nos municípios avaliados ao longo do tempo de estudo, estimados com base na média *a posteriori*. Os municípios destacados são os mesmos da Figura 2. É possível verificar que os riscos estimados estão suavizados, com valores máximos em torno de 3. Pinhal e Tramandaí, dois dos municípios com maiores valores de SIR, continuam com valores altos no risco relativo, diferente dos municípios de Pouso Novo, Fagundes Varela, São Valentim do Sul e Mato Queimado. Porto Alegre, a capital do estado, apresentava o maior risco relativo de casos de aids entre os anos 2001 e 2012, mas foi substituída por Pinhal, que apresenta um comportamento crescente e o maior valor no estado a partir do ano de 2013.

Conclusões

Neste artigo foi apresentado um modelo hierárquico bayesiano, cuja aproximação da distribuição *a posteriori* foi realizada pelo método INLA, por meio do R-INLA, a fim de verificar a heterogeneidade do risco relativo dos casos de aids nos municípios do estado do Rio Grande do Sul. Com auxílio de mapas, foram apresentados SIRs e riscos relativos dos casos de aids no estado estimados

¹Componente independente e identicamente distribuído

²Componente espacial

³Interação entre área e tempo



Figura 3 – Risco relativo dos casos de aids do RS entre os anos 2001 e 2018.

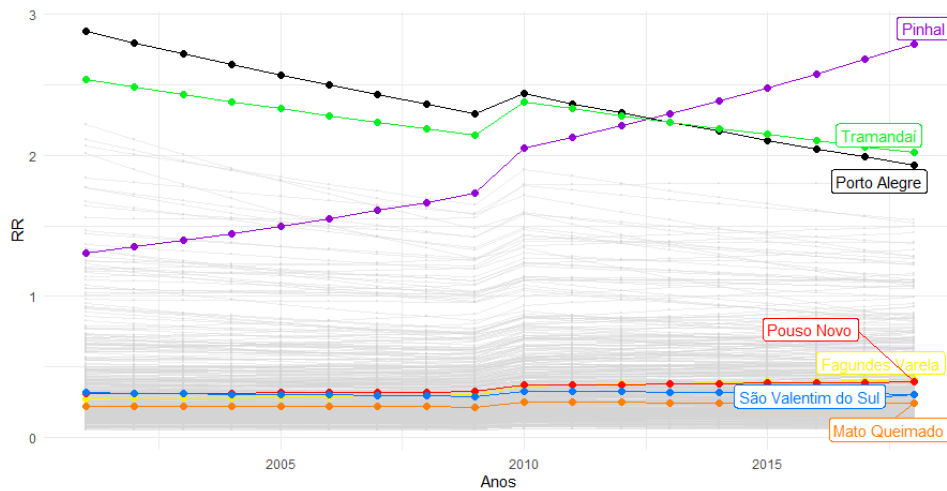


Figura 4 – *Timeplot* do risco relativo dos casos de aids no RS entre os anos 2001 e 2018.

através de um modelo hierárquico bayesiano espaço-temporal. Foi possível verificar que ambos apresentam uma estabilização/diminuição ao passar dos anos avaliados. Essa visualização espacial permite o reconhecimento dos municípios com maior risco relativo, que é de extrema importância para que ações ao combate da aids possam ser direcionadas para áreas que mais necessitam. Após o ajuste do modelo, foi possível observar que quanto maior o IDHM de um município, maior é o risco relativo de casos de aids. Para trabalhos futuros seria interessante avaliar não apenas o estado do Rio Grande do Sul, mas a distribuição dos riscos relativos dos casos de aids em todo o Brasil. Além disso, a adição de novas covariáveis ao modelo deixariam esse risco relativo estimado mais acurado.

PEREIRA, R. B. Bayesian hierarchical model for the mapping of aids in Rio Grande do Sul between the years of 2001 and 2018. *Rev. Bras. Biom.*, Lavras, v.xx, n.x, p.xx-xx, 20xx.

- **ABSTRACT:** *The acquired immunodeficiency syndrome (aids) is a possible consequence of the human immunodeficiency virus. Some authors say that the virus may have sporadically infected humans throughout the 20th century, however, it was only in the 1980's that it received worldwide attention. In Brazil, between the years of 1980 and 2019, almost a million cases were notified. Aiming to verify the distribution of the relative risk of aids cases in the state of Rio Grande do Sul over the years, a bayesian hierarchical model was fitted. Using data from governmental sources and a spatiotemporal model proposed by Bernardinelli, relative risks from municipalities of Rio Grande do Sul between the years of 2001 and 2018 were estimated. The model's posterior distribution was calculated via integrated nested Laplace approximation, assisted by the R-INLA package. It was noticed that an increase in the municipal human development index is associated with an increase in the relative risk of aids cases in the municipalities of Rio Grande do Sul. Over the years, a decrease/stabilization in the relative risk of aids cases was observed.*
- **KEYWORDS:** *Acquired immunodeficiency syndrome; Spatiotemporal; Bayesian statistics; Relative risk; R-INLA*

Referências

- ALLENBY, G. M.; ROSSI, P. E.; MCCULLOCH, R. E. Hierarchical Bayes Models: A Practitioners Guide. *SSRN Electronic Journal*, 2015.
- AQUINO, V. (2019). *135 mil brasileiros vivem com HIV e não sabem*. Disponível em: <http://www.aids.gov.br/pt-br/noticias/135-mil-brasileiros-vivem-com-hiv-e-nao-sabem> (Acesso em: 9 de maio de 2021).
- ATLAS DO DESENVOLVIMENTO HUMANO NO BRASIL - Atlas Brasil (2020). *Consulta em Tabela*. Disponível em: <http://www.atlasbrasil.org.br/consulta/planilha>. (Acesso em: 10 de maio de 2021).
- BERNARDINELLI, L.; CLAYTON, D.; PASCUTTO, C.; *et al.* Bayesian analysis of space—time variation in disease risk. *Statist. Med.*, v.14, p.2433-2443, 1995.
- BESAG, J.; YORK, J.; MOLLIE, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, v.43, p.1–20, 1991.
- BIVAND, S. R.; PEBESMA, E.; GOMEZ-RUBIO, V. *Applied spatial data analysis with R*. 2.ed. New York, Heidelberg, Dordrecht, London: Springer, 2013.
- BLANGIARDO, M.; CAMELETTI, M.; BAIO, G.; *et al.* Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-temporal Epidemiology*, v.4, p.33–49, 2013.
- DEEKS, S. G.; OVERBAUGH, J.; PHILLIPS, A.; *et al.* HIV infection. *Nature Reviews Disease Primers*, v.1, 2015.

GELFAND, A. E.; DIGGLE, P.; GUTTORP, P.; *et al.* *Handbook of spatial statistics*. 1.ed. Boca Raton: CRC Press, 2010.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. *Censo Brasileiro de 2010*. Rio de Janeiro: IBGE, 2012.

KIM; Y. A.; WAKEFIELD, J. SpatialEpi: Methods and Data for Spatial Epidemiology. R package version 1.2.3. 2018.

KNORR-HELD, L. Bayesian modelling of inseparable space-time variation in disease risk. *Statist. Med.*, v.19, p.2555-2567, 2000.

LINDGREN, F.; RUE, H. Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, v.63(19), p.1-25, 2015.

MINISTÉRIO DA SAÚDE (2019). *Boletim Epidemiológico de HIV/Aids 2019*. Disponível em: <http://www.aids.gov.br/pt-br/pub/2019/boletim-epidemiologico-de-hiv-aids-2019> (Acesso em: 25 de maio de 2020).

MORAGA, P. *Geospatial health data: modeling and visualization with R-INLA and Shiny*. 1.ed. London: Chapman and Hall, 2019.

NAGHAVI, M.; WANG, H.; LOZANO, R.; *et al.* Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, v.385, p.117-171, 2015.

PRADO, R. R. DO; CASTILHO, E. A. DE. A epidemia de aids no Estado de São Paulo: uma aplicação do modelo espaço-temporal bayesiano completo. *Revista da Sociedade Brasileira de Medicina Tropical*, v.42, n.5, p.537–542, 2009.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. 2019.

RUE, H.; MARTINO, S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of Statistical Planning and Inference*, v.137, n.10, p.3177–3192, 2007.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v.71, n.2, p.319–392, 2009.

SISTEMA IBGE DE RECUPERAÇÃO AUTOMÁTICA - SIDRA (2020a). *Contagem da População*. Disponível em: <https://sidra.ibge.gov.br/tabela/793>. (Acesso em: 10 de maio de 2021).

SISTEMA IBGE DE RECUPERAÇÃO AUTOMÁTICA - SIDRA (2020b). *Estimativas de População*. Disponível em: <https://sidra.ibge.gov.br/tabela/6579>. (Acesso em: 10 de maio de 2021).

SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO - SINAN (2019). *AIDS Adulto*. Disponível em: <http://www.portalsinan.saude.gov.br/aids-adulto>. (Acesso em: 10 de maio de 2021).

UNAIDS (2020). *Global HIV & AIDS statistics — 2020 fact sheet*. Disponível em: <https://www.unaids.org/en/resources/fact-sheet>. (Acesso em: 15 de março de 2021).

WALLER, L. A.; GOTWAY, C. A. *Applied spatial statistics for public health data*. 1.ed. Hoboken, N.J: John Wiley & Sons, 2004.

WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.

WICKHAM, H.; FRANÇOIS, R.; HENRY L; *et al.* *dplyr: A Grammar of Data Manipulation*. R package version 1.0.5. 2021.

YUTANI, H. *gghighlight: Highlight Lines and Points in 'ggplot2'*. R package version 0.3.1. 2020.

Recebido em 0x.0x.20xx.

Aprovado após revisão em 0x.0x.20xx.

Apêndice

O código em R utilizado para a implementação computacional das análises geradas nesse artigo é apresentado a seguir. Os seguintes pacotes foram utilizados: *dplyr* (WICKHAM, 2021), *SpatialEpi* (KIM *et al.*, 2018), *ggplot2* (WICKHAM, 2016), *gghighlight* (YUTANI, 2020), *INLA* (RUE *et al.*, 2009), e *spdep* (BIVAND *et al.*, 2013).

```
#carregando os pacotes necessários para rodar o código
library(dplyr)
library(SpatialEpi)
library(ggplot2)
library(gghighlight)
library(INLA)
library(spdep)

#juntando as bases de dados com casos e população (no formato
"long") e salvando em "df"
df = left_join(casos1, pop1, by = c("mun", "ano"))

#colocando na base "df" as informações do idhm
df2 = left_join(df, idhm, by = c("mun"))

#removendo o município "pinto bandeira" da base e mantendo apenas as
variáveis de interesse
df3 = subset(df2, mun != "Pinto Bandeira", c("cod", "mun", "ano",
"casos", "pop", "idhm"))

#salvando o shapefile em "map"
map = st_read(dsn = file.path("RS_Municipios_2020.shp"), quiet =
TRUE)

#ordenar o banco antes de rodar o expected
df4 = df4[order(df4$cod, df4$ano),]

#calculando o valor esperado de casos nos municípios do estudo
(expected)
E = expected(population = df4$pop, cases = df4$casos, n.strata = 1)

#criando um banco de dados com municípios, anos e E:
nanos = length(unique(df4$ano)); munE = rep(unique(df4$mun), each =
nanos)
nmun = length(unique(df4$mun)); anosE = rep(unique(df4$ano), times =
nmun)
dE = data.frame(mun = munE, ano = anosE, E = E)
```



```

#juntando o último banco dos dados criado com o que possui casos,
população e idhm
d = merge(df4, dE, by = c("mun", "ano"))

#calculando SIR para fazer descritivas
d$SIR = (d$casos/d$E)

#transpondo a base de dados para juntar com o shapefile
dw = reshape(d, timevar = "ano", idvar = "mun", direction = "wide")

#alterações no shapefile para poder fazer o merge
map2 = rename(map, cod = CD_MUN, mun = NM_MUN)
map2$SIGLA_UF = map2$AREA_KM2 = NULL

#merge entre base de casos/população/sir/casos esperados + shapefile
map = merge(map2, dw, by.x = "mun", by.y = "mun")

##### descritivas
#criando base para plotar os gráficos descritivos
map_sf = merge(map, d, by.x = "mun", by.y = c("mun"))

#mapeando o SIR
ggplot(map_sf) +
  geom_sf(aes(fill = SIR), size = NA) +
  facet_wrap(~ ano, dir = "h", ncol = 4) +
  theme_minimal() +
  theme(axis.text.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks = element_blank()) +
  scale_fill_gradient(low = "#D0C4CF", high = "#9400D3")

#criando o timeplot
ggplot(d, aes(x = ano, y = SIR, group = mun, color = mun)) +
  geom_line() +
  geom_point(size = 2) +
  xlab("Anos") +
  theme_minimal() +
  theme(legend.position = "none") +
  scale_color_manual(values = c(rep("#FFEF00", 161), rep("#FF7F00",
252-161), rep("#9400D3", 313-252), rep("#000000", 325-313),
rep("#FF0900", 330-325), rep("#0079FF", 415-330), rep("#00F11D",
496-415))) +

```

```

gghighlight(mun == "Pinhal" | mun == "Fagundes Varela" | mun ==
"Mato Queimado" | mun == "São Valentim do Sul" | mun == "Pouso
Novo" | mun == "Porto Alegre" | mun == "Tramandaí",
  label_key = mun,
  keep_scales = TRUE,
  unhighlighted_params = list(size = 0.5, colour =
alpha("lightgrey", 0.4)))

##### modelagem
#criando uma lista com os vizinhos de cada um dos municípios
nb = poly2nb(map)

#salvando a lista no diretório de trabalho
nb2INLA("maprs", nb)

#salvando em "g" a lista que está no diretório de trabalho
g = inla.read.graph(filename = "maprs")

#criando variáveis necessárias para rodar o modelo
d$mun = as.factor(d$mun)
d$ano = as.numeric(d$ano)

#índices para as áreas (1, 2, ..., I)
d$idarea = d$idareal = as.numeric(d$mun)

#índice para o tempo (1, 2, ..., J)
d$idtime = 1 + d$ano - min(d$ano)

#modelo:
formula = casos ~ idhm + f(idarea, model = "bym", graph = g) +
  f(idareal, idtime, model = "iid") + idtime

ajuste = inla(formula, family = "poisson", data = d, E = E,
  control.predictor = list(compute = TRUE))

#mapeando o risco relativo
d$RR = ajuste$summary.fitted.values[, "mean"]

map_sf = merge(map, d, by.x = c("mun"), by.y = c("mun"))

ggplot(map_sf) +
  geom_sf(aes(fill = RR), size = NA) +
  facet_wrap(~ ano, dir = "h", ncol = 4) +

```

```

theme_minimal() +
theme(axis.text.x = element_blank(),
      axis.text.y = element_blank(),
      axis.ticks = element_blank()) +
scale_fill_gradient(low = "#D0C4CF", high = "#9400D3")

#timeplot do risco relativo
ggplot(map_sf, aes(x = ano, y = RR, group = mun, color = mun)) +
  geom_line() +
  geom_point(size = 2) +
  xlab("Anos") +
  theme_minimal() +
  theme(legend.position = "none") +
  scale_color_manual(values = c(rep("#FFEF00", 161), rep("#FF7F00",
252-161), rep("#9400D3", 313-252), rep("#000000", 325-313),
rep("#FF0900", 330-325), rep("#0079FF", 415-330), rep("#00F11D",
496-415))) +
  gghighlight(mun == "Pinhal" | mun == "Fagundes Varela" | mun ==
"Mato Queimado" | mun == "São Valentim do Sul" | mun == "Pouso
Novo" | mun == "Porto Alegre" | mun == "Tramandaí",
              label_key = mun,
              keep_scales = TRUE,
              unhighlighted_params = list(size = 0.5, colour =
alpha("lightgrey", 0.4)))

```