

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**



**DISSERTAÇÃO DE MESTRADO**

**ANÁLISE ESPAÇO-TEMPORAL NA ESTIMAÇÃO DE TAXAS  
DE INCIDÊNCIA / MORTALIDADE**

Diego de Matos Vieira

Orientador: Profa. Dra. Jandyra Maria Guimarães Fachel

Porto Alegre, Dezembro de 2006

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA**



**DISSERTAÇÃO DE MESTRADO  
ANÁLISE ESPAÇO-TEMPORAL NA ESTIMAÇÃO DE TAXAS  
DE INCIDÊNCIA / MORTALIDADE**

Diego de Matos Vieira

**Orientador: Profa. Dra. Jandyra Maria Guimarães Fachel**

A apresentação desta dissertação é exigência do Programa de Pós-graduação em Medicina: Epidemiologia, Universidade Federal do Rio Grande do Sul, para obtenção do título de Mestre.

Porto Alegre, Brasil.  
2006

## **BANCA EXAMINADORA**

Profa. Dra. Maria Inês Azambuja, Programa de Pós Graduação em Epidemiologia, UFRGS.

Prof. Dr. Sérgio Luiz Bassanesi, Programa de Pós Graduação em Epidemiologia, UFRGS.

Profa. Dra. Patrícia Klarmann Ziegelmann, Instituto de Matemática, Depto. de Estatística, UFRGS.

Profa. Dra. Jandyra Maria Guimarães Fachel, Programa de Pós Graduação em Epidemiologia, UFRGS (Orientadora).

## **AGRADECIMENTOS**

Agradeço primeiramente a minha orientadora, Profa. Jandyra Maria Guimarães Fachel pelo apoio quando da decisão de entrar para o Programa de Pós-Graduação pela orientação na formulação desta dissertação e pela paciência que teve comigo durante estes dois anos.

Aos membros da banca, professores Sérgio Luiz Bassanesi, Maria Inês Azambuja e Patrícia Klarmann Ziegelmann pelas importantes sugestões realizadas para a qualificação do trabalho.

Aos colegas do curso que sempre pude contar quando foi necessário.

E finalmente aos meus familiares, por ter me dado a condição de iniciar e continuar meus estudos, e pelo incentivo e compreensão demonstrados durante este período.

## SUMÁRIO

1. INTRODUÇÃO .....	1
1.1 Estatística Espacial .....	1
1.2 Mapeamento de Doenças.....	2
1.3 Análise Espaço-Temporal .....	4
2. REVISÃO DE LITERATURA .....	7
2.1 Modelo Clássico.....	7
2.2 Inferência Bayesiana .....	12
2.3 Modelos Hierárquicos Bayesianos .....	13
2.3.1 Definição das Distribuições <i>a Priori</i> .....	14
2.3.2 Definição das Hiperprioris .....	17
2.3.3 Distribuições a Posteriori .....	19
2.4 Modelos Espaço-Temporais .....	21
3. OBJETIVOS .....	26
3.1 Objetivo Geral.....	26
3.2 Objetivos Específicos .....	26
4. REFERÊNCIAS BIBLIOGRÁFICAS .....	27
5. ARTIGO.....	32
ANEXO 1 – PROJETO DE PESQUISA .....	56
1. INTRODUÇÃO .....	57
1.1 Epidemiologia Espacial .....	57
1.2 Modelos Espaço-Temporais .....	58
2. OBJETIVOS .....	60
2.1 Objetivo Geral:.....	60
2.1 Objetivos Específicos: .....	60
3. METODOLOGIA .....	61
4. CRONOGRAMA .....	61
5. REFERÊNCIAS .....	62
ANEXO 2 – LINHAS DE COMANDO .....	63

## RESUMO

A análise de dados espaciais, aplicada ao contexto epidemiológico, também conhecida como Epidemiologia Espacial, vem apresentando crescente utilização por parte dos pesquisadores da área da saúde como uma das ferramentas de estudos de mapeamento e comportamento de doenças e de construção de indicadores de saúde. O presente trabalho apresenta uma revisão bibliográfica sobre a análise espacial de dados agregados por área, aplicada ao Mapeamento de Doenças, relatando os principais problemas da utilização das taxas de mortalidade brutas na construção dos mapas. Como as taxas brutas são estimativas de pequena precisão em regiões onde a população em risco tem tamanho reduzido, e o número de casos é raro, os modelos Bayesianos Hierárquicos são apresentados como alternativas na estimação destas taxas, através da suavização das estimativas produzidas pela incorporação de um componente espacialmente estruturado ao modelo. Nos modelos apresentados pode-se incorporar o fator temporal ao modelo espacial, através de diferentes formas de especificação. No artigo são analisadas as taxas de mortalidade infantil nos 496 municípios do Rio Grande do Sul no período de 2001 a 2004. Foram comparados três modelos com diferentes especificações dos componentes espaciais e temporais. Verificou-se que o modelo que utiliza a configuração espacial dos municípios na estimação das taxas obteve maior adequação aos dados quando comparado ao modelo onde as taxas são estimadas sem a utilização do componente espacialmente estruturado. Pode-se observar uma melhora sensível na interpretação dos mapas das taxas de mortalidade estimadas, através dos modelos apresentados, em relação aos mapas que utilizam as taxas brutas de mortalidade, podendo-se observar diferentes padrões de evolução temporal em diferentes regiões do Estado.

## ABSTRACT

Spatial data analysis applied to the context of Epidemiology, also known as Spatial Epidemiology, has been used by health researchers as one of the tools for understand the behavior of the disease and for disease mapping studies. This paper presents a bibliographic revision about spatial analysis of data aggregated by area applied to Disease Mapping. The major problems about using gross mortality rates in the map construction are related here, as those rates are poor-precision estimates on areas with reduced population at risk. In this paper, Hierarchical Bayesian model are presented as an alternative in the estimation of those rates, through estimate smoothing produced by inserting a spatially structured component to the model, in addition to the temporal factor, using a different model specifications. Infant mortality rates from 2001 to 2004 in all 496 cities in Rio Grande do Sul were analyzed and three models with different specifications of spatial and temporal components were compared. Noticeably, the model that uses the spatial configuration of the cities for estimating the rates was the best, in comparison to the model where rates are estimated without the spatially-structured component. Considerable improvement is noted in the interpretation of mortality rate maps in relation to maps that use gross mortality rates, allowing the observation of different infant mortality patterns in different areas of the State.

## **LISTA DE TABELAS**

Tabela 1 – Descritivas Taxa de Mortalidade Infantil 2001-2004.....	45
Tabela 2 – Descritivas Taxas Bayesianas 2001-2004.....	51



## LISTA DE FIGURAS

### Revisão Bibliográfica

Figura 1 – Distribuição Amostral da Taxas de Incidência N=50 .....	10
Figura 2 – Distribuição Amostral da Taxas de Incidência, N=500 .....	11
Figura 3 - Critério da adjacência .....	16

### Artigo

Figura 1 – Taxas brutas para a Mortalidade Infantil em 2001 .....	46
Figura 2 – Taxas brutas para a Mortalidade Infantil em 2002 .....	46
Figura 3 – Taxas brutas para a Mortalidade Infantil em 2003 .....	47
Figura 4 – Taxas brutas para a Mortalidade Infantil em 2004 .....	47
Figura 5 – Diagrama de dispersão dos municípios segundo as taxas brutas de mortalidade infantil e o número de nascidos vivos, Rio Grande do Sul, 2004....	48
Figura 6 – Taxas Bayesianas para a Mortalidade Infantil em 2001 .....	49
Figura 7 – Taxas Bayesianas para a Mortalidade Infantil em 2002 .....	49
Figura 8 – Taxas Bayesianas para a Mortalidade Infantil em 2003 .....	50
Figura 9 – Taxas Bayesianas para a Mortalidade Infantil em 2004 .....	50

## APRESENTAÇÃO

Este trabalho consiste na dissertação de mestrado intitulada “Análise Espaço Temporal na Estimção de Taxas de Incidência / Mortalidade” apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, em 21 de dezembro de 2006. O trabalho é apresentado em duas partes, na ordem que segue:

1. Introdução, Revisão da Literatura e Objetivos
2. Artigo(s)

Documentos de apoio, incluindo o Projeto de Pesquisa, estão apresentados nos anexos.

# 1. INTRODUÇÃO

## 1.1 Estatística Espacial

“A análise de dados espaciais ocorre quando os dados são espacialmente localizados e se considera explicitamente a possível importância de seu arranjo espacial na análise ou interpretação dos resultados”(Bailey e Gatrell, 1995).

Neste contexto pressupõe-se que a localização espacial apresenta-se como um fator relevante na explicação da ocorrência de determinado fenômeno. Como exemplo, pode-se citar doenças contagiosas onde unidades mais próximas no espaço possuem maiores chances de contágio e, a medida que esta distância aumenta, o risco de contágio tende a diminuir.

A análise de dados espaciais, aplicada ao contexto epidemiológico, também conhecida como Epidemiologia Espacial, vem apresentando crescente utilização por parte dos pesquisadores da área da saúde como uma das ferramentas de desenvolvimento de informação quanto ao comportamento de doenças e construção de indicadores de saúde (Carvalho e Souza-Santos, 2005).

Este recente crescimento deve-se em grande parte ao enorme desenvolvimento da computação nos últimos anos, possibilitando um maior acesso aos sistemas geográficos de informação (GIS), bem como aos inúmeros avanços metodológicos na área da estatística, destacando-se os modelos inferenciais Bayesianos.

A Epidemiologia Espacial pode ser dividida em três grandes áreas (Lawson, 2001):

1. Mapeamento de doenças: que tem como objetivo avaliar a variação espacial de taxas de mortalidade ou incidência de doenças, visando identificar diferenciais de risco, realizar previsões de epidemias, orientar ações em saúde e levantar hipóteses.
2. Estudos ecológicos: tem por objetivo estudar a relação entre potenciais fatores etiológicos e a incidência de doenças ou mortalidade. Essencialmente constroem modelos onde se busca a explicação das variações das taxas de incidência através de indicadores sócio-econômicos e comportamentais além da sua localização espacial.
3. Detecção de cluster: o objetivo principal desta análise é estabelecer a significância de um sobre-risco em um determinado espaço, que em geral pode ser pensado com um foco da doença.

Este trabalho está inserido no contexto do Mapeamento de doenças e, portanto não serão realizados maiores comentários sobre os Estudos ecológicos e Detecção de cluster.

## **1.2 Mapeamento de Doenças**

O mapeamento de doenças tem como objetivo avaliar a variação espacial de medidas resumo de eventos em saúde, como riscos relativos, taxas de mortalidade ou incidência de doenças entre outros, através da utilização de mapas, visando

identificar diferenciais de risco, realizar predições de epidemias, orientar ações em saúde e levantar hipóteses (Elliott et al., 2001).

A construção de mapas de doenças é utilizada como uma ferramenta básica na análise de eventos de saúde há muito tempo, apresentando um considerável avanço nos últimos anos (Lawson, 2001). Um dos primeiros exemplos de aplicação do mapeamento de doenças foi realizado por John Snow em 1855 que plotou em um mapa da cidade de Londres os locais de residência das pessoas infectadas com cólera, onde ficou evidenciado o relacionamento dos casos com a localização da bomba de um poço que continha água contaminada (Snow, 1855).

Estes mapas podem ser realizados através da utilização da análise de dados pontuais, ou seja, quando se tem acesso a localização exata do evento, ou através da análise de dados agregados por área, ou dados de contagem, que utilizam apenas a informação do número de eventos e do tamanho da população em risco nas áreas analisadas. Estas áreas em geral são definidas pelos limites de abrangência de bairros, setores censitários, municípios, entre outros.

A maior parte destes mapas são realizados com dados agregados por área, onde áreas com diferentes níveis de taxa apresentam cores diferentes e áreas com taxas semelhantes, cores iguais. Dados agregados são os mais comumente utilizados no mapeamento de doenças, devido a uma maior disponibilidade destes em relação aos dados pontuais.

Em uma análise de dados pontuais necessita-se do exato local de ocorrência do evento que por muitas vezes não pode ser divulgado, pois poderá ferir o princípio da privacidade. Outra grande limitação é o fato de que atualmente apenas algumas cidades possuem grande parte dos locais mapeados.

A análise descritiva dos dados agregados por área, ou seja, a construção de mapas das taxas observadas nas regiões é uma tarefa bastante simples, porém em muitos casos resultam em mapas de difícil compreensão ou não-informativos, devido ao fato de que as taxas observadas em áreas com população reduzida geralmente apresentam alta variabilidade, apresentando muito facilmente valores extremamente superiores (superestimação do risco), ou muito inferiores à taxa média (subestimação do risco), sem que estas regiões realmente se caracterizem como regiões de alto ou baixo risco, ou seja, estes mapas podem não representar o verdadeiro processo, ou risco subjacente responsável por gerar estas taxas (Assunção 2001).

Atualmente os grandes desafios no mapeamento de doenças são o desenvolvimento e a utilização de métodos estatísticos alternativos na estimação das taxas, de modo a diminuir a variabilidade aleatória das estimativas para regiões com pequena população em risco e a incorporação de covariáveis na análise de modo a testar hipóteses relacionadas à distribuição espacial das taxas (Lawson, 2001).

O método mais frequentemente utilizado, pelos pesquisadores da área, é o método de modelagem que utiliza os modelos Bayesianos Hierárquicos (Besag et al., 1991 e Mollié, 1996) e mais recentemente, os modelos dinâmicos (West e Harrison, 1997), que também utilizam a inferência bayesiana.

### **1.3 Análise Espaço-Temporal**

Com o desenvolvimento da epidemiologia espacial, naturalmente existe um interesse da comunidade científica em aprofundar seus conhecimentos através da construção de modelos espaciais mais sofisticados, entre estes se destacam os

modelos que incorporam o fator tempo à análise espacial, que devido a sua complexidade, só obteve desenvolvimento a partir da última década assim como os métodos Bayesianos de análise.

A análise temporal pode ser utilizada na identificação de padrões sazonais, endêmicos ou epidêmicos e na construção de previsões futuras sobre o comportamento do fenômeno. Já a análise espacial pode ser utilizada na identificação de fatores demográficos, ambientais ou sócio-culturais que podem ser identificados como fatores de risco no desenvolvimento de um determinado evento.

Devido a este desenvolvimento ainda ser bastante recente, por diversas vezes tem-se acesso aos dados referentes à localização espacial e temporal de eventos relacionados à saúde, porém estas informações não são utilizadas na análise de forma conjunta, ou seja, analisam-se os eventos relacionando-os apenas à dimensão tempo, caracterizando assim uma análise puramente temporal, ou relacionam-se os eventos apenas à dimensão espacial, caracterizando assim uma análise puramente espacial(Werneck e Struchiner, 1997).

Entretanto estes modelos de análise, apenas espacial ou apenas temporal, justificam-se somente quando o fenômeno pesquisado tem comportamento pouco variável em uma das dimensões, ou seja, a incorporação de efeitos temporais num modelo espacial não altera as estimativas de risco nas regiões analisadas e a incorporação de efeitos espaciais num modelo temporal não altera as estimativas de risco no tempo, o que, em muitos fenômenos de saúde, não se pode aceitar como uma verdade.

Por exemplo: Suponha duas áreas A e B com mesma população em risco, analisadas durante dois períodos de tempo  $t$  e  $t+1$ . No período  $t$  foram observados 10

casos na área A e 5 na área B enquanto no período  $t + 1$  foram observados 5 casos na área A e 10 na área B.

Analisando os dados apenas através do enfoque espacial, ou seja, desconsiderando o tempo, não seria observada uma associação entre o espaço (as áreas) e o número de casos, pois em todo o período foram registrados 15 casos em cada uma das duas áreas.

Sob um enfoque apenas temporal também não seriam observadas diferenças entre os períodos (efeito temporal), já que, nos dois períodos foram observados 15 casos.

Ou seja, através destas duas análises se concluiria que a ocorrência dos casos independe da área e independe do tempo, o que não é correto, pois claramente existe interação espaço-temporal, ou seja, as cidades apresentam diferentes padrões de ocorrência de casos no período.

Este exemplo é bastante simplificado, porém pode ser estendido a análises de um número maior de áreas e de períodos.

O foco desta dissertação será a apresentação de metodologias de análise espaço-temporal no mapeamento de doenças, através da apresentação de modelos utilizados na análise espacial bem como a incorporação do fator temporal à análise, utilizando estimadores Bayesianos. Também mostraremos a aplicação desta metodologia na análise das taxas de mortalidade infantil no Rio Grande do Sul nos anos de 2001 a 2004.



## 2. REVISÃO DE LITERATURA

### 2.1 Modelo Clássico

Suponha uma região R dividida em n áreas. No modelo clássico as taxas de incidência/ mortalidade para cada uma das n áreas são estimadas através da taxa observada (também denominada taxa bruta):

$$t_i = y_i / N_i ; i = \{1, \dots, n\}$$

onde  $y_i$  é o número observado de casos na área i em um determinado período especificado;  $N_i$  é a população em risco no período e  $t_i$  é a taxa observada para a i-ésima área, sendo comum sua multiplicação por 100, 1.000 ou outros valores, de modo a facilitar a interpretação (Assunção, 2001).

Apesar de  $t_i$  depender apenas de quantidades populacionais  $y_i$  e  $N_i$ , supõe-se que exista um processo subjacente à distribuição destas taxas, ou seja, que as taxas observadas são apenas uma manifestação (resultado amostral) deste processo onde estão definidas as taxas de incidência/ mortalidade verdadeiras.

Assim sendo supõe-se que o número de casos na i-ésima área ( $y_i$ ) é uma variável aleatória, que geralmente supõe-se que possua distribuição Binomial. Quando a ocorrência dos eventos é rara, o modelo Binomial pode ser aproximado pela distribuição Poisson (Richardson et. al., 2004).

$$y_i \sim P(\lambda_i) ; i = \{1, \dots, n\} \quad (1)$$

Neste caso  $\lambda_i$  representa a quantidade esperada de casos na  $i$ -ésima área e o número de casos em cada área  $y_1, \dots, y_n$  são modelados independentemente.

Notas-se que o parâmetro  $\lambda_i$ , representa apenas o número esperado de casos, ou seja, não fornece resultados que possam ser interpretados como uma medida de risco de ocorrência de casos.

A estimação das taxas é realizada através da estimação do risco relativo do evento na área  $i$ . Este risco é estimado através da estimação do parâmetro  $\lambda_i$ , que é decomposto em duas partes, onde uma se refere ao número esperado de casos ( $e_i$ ) na área  $i$ , caso esta apresentasse o risco médio global, e a outra,  $\theta_i$  representa o risco relativo (a partir deste ponto será utilizado apenas a palavra risco para representar o risco relativo) verdadeiro de ocorrência de casos na área  $i$ , e deste modo a distribuição em (1) pode ser descrita como (Elliott et al., 2001):

$$y_i/e_i \sim P(e_i \theta_i) ; i = \{1, \dots, n\}$$

Pode-se demonstrar que o estimador de máxima verossimilhança de  $\theta_i$  é dado por  $r_i$ ,

$$r_i = y_i / e_i ; i = \{1, \dots, n\}$$

que também pode ser denominado  $SMR_i$  (Razão de Mortalidade/ Morbidade Padronizada).

A relação entre a estimativa da taxa de incidência  $t_i$  e o risco estimado  $r_i$  é dado pela multiplicação da taxa de incidência global com o risco estimado na  $i$ -ésima área:

$$t_i = \frac{\sum y_i}{\sum N_i} r_i ; i = \{1, \dots, n\}$$

Em relação ao estimador  $r_i$ , pode-se provar que sua variância é inversamente proporcional ao número esperado de casos  $e_i$ , que por sua vez é proporcional a  $N_i$ .

$$\text{Var}(r_i) = y_i / e_i^2 ; i = \{1, \dots, n\}$$

Ou seja, quanto menor for o tamanho da população em risco, maior a incerteza na estimativa do risco e conseqüentemente maior incerteza na estimação das taxas de incidência/ mortalidade. Esta instabilidade nas estimativas é o fator que causa maior dificuldade na interpretação dos mapas, pois principalmente em doenças raras, poderão ocorrer estimativas extremas em algumas regiões, que se destacarão no mapa, sem que estas regiões se configurem em reais áreas de alto ou baixo risco (Assunção, 2001).

Como um exemplo da ocorrência destas instabilidades, suponha uma área onde a taxa de incidência verdadeira de uma determinada doença é de 20 por 1000. Num determinado período, o tamanho da população em risco é 50, e, portanto, o número de casos desta doença, nesta área e neste período, possui distribuição Poisson com número esperado de casos igual a 1.

Utilizando o modelo probabilístico Poisson pode-se obter a distribuição amostral das taxas de incidência nesta área, onde se observa que a probabilidade de ocorrência de apenas 1 caso nesta área é de 0,368, obtendo-se assim uma estimativa exatamente igual à taxa verdadeira de 20 por mil. Porém se ocorresse um caso a menos (0,368 de probabilidade) ou a mais (0,184 de probabilidade), as taxas estimadas seriam 0 e 40 por 1000 respectivamente, ou seja, pela diferença de apenas 1 caso, esta área seria considerada sem risco ou com o dobro do risco estimado. Além disso, pode-se observar que existe aproximadamente 0,08 de probabilidade de ocorrência de mais de 3 casos nesta área, ou seja, existe 8% de probabilidade de que a taxa de incidência seja estimada em 60 ou mais casos por 1000, quando na realidade este indicador é de 20 por 1000.

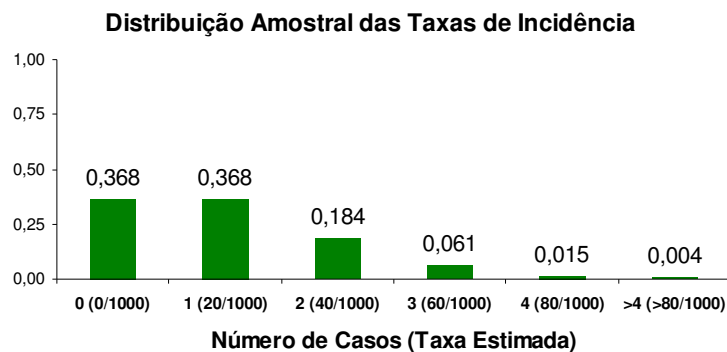


Figura 1 – Distribuição Amostral da Taxas de Incidência N=50

À medida que o tamanho da população em risco aumenta, a variabilidade das estimativas da taxa de incidência diminui, fazendo com que as estimativas tenham maior precisão e apresentando uma maior concentração em torno da verdadeira taxa.

Numa área onde a população em risco tem tamanho 500, a probabilidade de obter uma estimativa entre 10 e 30 por 1000 é de aproximadamente 0,922, enquanto

numa área onde  $N=50$ , esta probabilidade é de apenas 0,368. Outro exemplo que demonstra as diferenças de qualidade das estimativas é o seguinte: numa área com  $N=50$  a probabilidade de obter uma estimativa igual ou superior a  $60/1000$  é de 0,08, enquanto para  $N=500$  esta probabilidade é aproximadamente nula, igual a  $3 \times 10^{-7}$ .

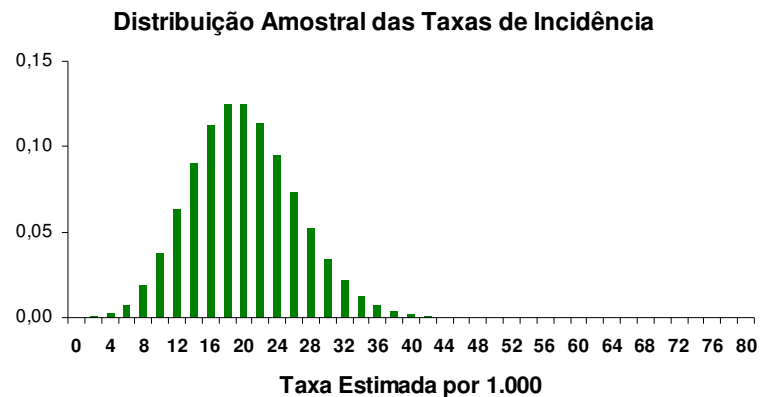


Figura 2 – Distribuição Amostral da Taxas de Incidência,  $N=500$

Existem diversos métodos para diminuir o efeito da aleatoriedade na estimação destas taxas, tais como mapas de probabilidade e a utilização de métodos Bayesianos. Os mapas de probabilidade estão sendo menos utilizados dado o advento dos métodos Bayesianos, pois nos modelos Bayesianos, diferentemente dos mapas de probabilidade utiliza-se diretamente a configuração espacial das  $n$  regiões na estimação das taxas (Lawson, 2001).

## 2.2 Inferência Bayesiana

Diferentemente da Inferência Clássica, na Inferência Bayesiana, nem toda a informação sobre o parâmetro é proveniente da amostra. Parte dela está contida na amostra que é representada pela função de verossimilhança e outra parte é devida ao conhecimento que o pesquisador possui antes de observar os dados através de uma distribuição de probabilidade para o parâmetro que é conhecida como distribuição *a priori* do parâmetro.

A contribuição de cada uma destas informações (da amostra e *a priori*) na estimação do parâmetro é uma função do tamanho da amostra e da precisão (variância atribuída à distribuição do parâmetro) da distribuição *a priori*. Quanto maior a amostra, maior a importância da informação proveniente da amostra, e quanto menor a variância atribuída à distribuição do parâmetro, mais precisa é a distribuição *a priori* (mais informativa), e por consequência maior a importância da informação da distribuição *a priori*.

A estimação do parâmetro é realizada com a utilização destas duas informações conjuntamente, através do teorema de Bayes, e tem como resultado a distribuição *a posteriori* do parâmetro, ou seja, é a distribuição final encontrada para o parâmetro e que será utilizada nas inferências a serem realizadas (Carlin e Louis, 1996).

Em grande parte dos problemas, não é possível a obtenção analítica das distribuições *a posteriori* dos parâmetros de interesse, e neste caso as inferências são realizadas através de amostras destas distribuições, obtidas por rotinas intensivas de simulação computacional como Simulação de Monte Carlo Via Cadeias de Markov

(método MCMC) e mais especificamente através do Amostrador de Gibbs (Gelfand e Smith, 1990).

Atualmente, os modelos mais utilizados no mapeamento de doenças são os modelos Bayesianos Hierárquicos que serão apresentados na próxima seção (Best et al., 2005).

### 2.3 Modelos Hierárquicos Bayesianos

Seja uma região geográfica  $R$  subdividida em  $n$  áreas. O objetivo da construção do modelo é estimar o risco de ocorrência do evento em cada uma das  $n$  áreas analisadas, ou seja, o vetor de riscos relativos  $\theta = (\theta_1, \dots, \theta_n)$ .

Devido às dificuldades numéricas, constrói-se um modelo para estimar os logaritmos dos riscos, ao invés de se estimar os riscos diretamente.

A definição da forma funcional do modelo dá-se no segundo nível de hierarquia do modelo, onde o  $\log(\theta_i)$  é decomposto em dois componentes (Besag, 1991), sendo também conhecida como priori de convolução:

$$\log(\theta_i) = \alpha + b_i + u_i \quad (2)$$

onde  $\alpha$  é uma constante que representa a média global,  $b_i$  representa o efeito aleatório devido à correlação espacial entre as áreas, sendo modelados através de uma priori espacialmente estruturada e  $u_i$  é um efeito aleatório devido à diferença entre os riscos nas diferentes áreas, que são modelados através de uma priori sem qualquer estrutura de dependência entre os coeficientes.

No terceiro nível hierárquico do modelo são definidas as distribuições *a priori* dos componentes do modelo  $\alpha$ ,  $\mathbf{b}_i$  e  $\mathbf{u}_i$ .

Note que este modelo não possui covariáveis, porém pode-se incorporar ao modelo covariáveis.

$$\log(\theta_i) = \alpha + \beta X + \mathbf{b}_i + \mathbf{u}_i$$

### 2.3.1 Definição das Distribuições *a Priori*

Em geral os efeitos aleatórios sem estrutura espacial  $\mathbf{u}_i$  possuem uma distribuição *a priori*  $P(\mathbf{u}_i/\sigma^2)$  i.i.d com distribuição normal com média 0 e parâmetro de precisão constante  $\tau$ , definido como o inverso da variância  $\sigma^2$ . Ou seja,

$$\mathbf{u} = (u_1, \dots, u_n) \sim \text{Normal} (0; \sigma^2 \mathbf{I}_{n \times n})$$

onde, o vetor de efeitos aleatórios  $\mathbf{u}$  possui distribuição normal multivariada com média 0 e dependente apenas do parâmetro desconhecido  $\sigma^2$ . Este parâmetro é denominado hiperparâmetro.

Os efeitos aleatórios espacialmente estruturados  $\mathbf{b}_i$  geralmente seguem um processo de Campos Aleatórios de Markov (CAM) (Mollié, 1996).

Num processo CAM a distribuição de probabilidade de uma variável aleatória em qualquer estado depende apenas das distribuições de probabilidade da variável nos estados vizinhos.

O exemplo simplificado de um processo CAM pode ser descrito num passeio aleatório na reta, onde a distribuição de uma variável aleatória  $X_{i+1}$  depende apenas



da distribuição de  $X_i$ , ou seja, a distribuição de probabilidade do próximo estado depende apenas da distribuição de probabilidade do estado atual.

A utilização de processos CAM no modelo (2) é o fator responsável pela suavização e melhoria das estimativas dos riscos, principalmente em áreas de pequena população em risco, pois produz estimativas dos efeitos  $b_i$  e, conseqüentemente, de  $\theta_i$ , utilizando as estimativas das regiões vizinhas à área  $i$ , ou seja, as estimativas  $b_i$  são localmente suavizadas.

Dentre as distribuições de probabilidade que podem ser modeladas segundo um processo CAM, a mais amplamente utilizada no mapeamento de doenças é a distribuição Normal sendo denominada por distribuição Condicional Autorregressiva Normal ou CAR Normal (Besag e Kooperberg, 1995).

Seja  $\delta_i$  o conjunto de regiões, que segundo um critério de vizinhança, são consideradas vizinhas a  $i$ -ésima área. A distribuição *a priori*  $P(b_i/\sigma_b^2)$  para o efeito espacial  $\mathbf{b}_i$  da  $i$ -ésima área está condicionada às estimativas  $b_{j \neq i} \in \delta_i$  e é dada por:

$$b_i | b_{j \neq i} \sim \text{Normal} \left( \frac{\sum_{j \in \delta_i} W_{ij} b_j}{\sum_{j \in \delta_i} W_{ij}} ; \frac{\sigma_b^2}{\sum_{j \in \delta_i} W_{ij}} \right), i = \{1, \dots, n\}$$

onde  $W_{ij}$  é o peso atribuído à área  $j$  na estimativa de  $\mathbf{b}_i$ .

Existem inúmeros critérios para a definição das regiões vizinhas a uma área  $i$ , e seus respectivos pesos.

Os critérios mais utilizados na definição da matriz de vizinhanças são:

- ✓ Critério da adjacência: são considerados vizinhos de uma área  $i$  apenas as regiões que fazem fronteira com a área  $i$  (Regiões em

verde na Figura 3). Existem variações deste critério, onde não apenas as regiões que fazem fronteira com a área  $i$  são consideradas, mas também as regiões que fazem fronteira com áreas vizinhas a área  $i$  (Regiões em cor azul na Figura 3).

- ✓ Critério de número de vizinhos: é definido um número fixo ( $n$ ) de vizinhos para cada uma das  $k$  regiões, sendo estes vizinhos definidos por proximidade. Calcula-se a distância entre os centróides da área de origem (a área que se quer definir os vizinhos) e as outras  $k-1$  áreas. São consideradas vizinhas a área  $k$  as  $n$  áreas de menor distância entre os centróides.
- ✓ Critério da distância entre os centróides: define-se como regiões vizinhas à área  $i$  aquelas cuja distância entre o centróide e o centróide da área  $i$  não ultrapasse um determinado valor. A diferença entre este critério e o anterior é o fato de este critério não limitar o número de vizinhos.

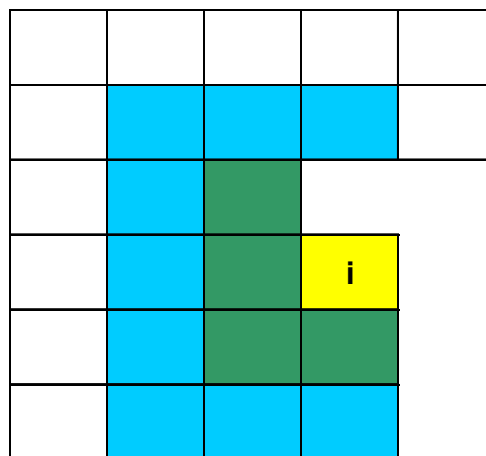


Figura 3 - Critério da adjacência

Os critérios mais utilizados na definição dos pesos (matriz  $W_{ij}$ ) das regiões vizinhas na estimação do efeito espacial na área  $i$  são:

- ✓ Uniforme: todas as regiões vizinhas à área  $i$  possuem o mesmo peso na estimação do efeito em  $i$ , ou seja  $W_{ij} = 1$  caso  $i$  e  $j$  sejam regiões vizinhas e  $W_{ij} = 0$  caso  $i$  e  $j$  não sejam regiões vizinhas.
- ✓ Tamanho da fronteira: os pesos  $W_{ij}$  são definidos pelo tamanho da fronteira entre  $i$  e  $j$ , onde regiões que possuem uma fronteira maior com a área  $i$  possuem peso superior na estimação de  $b_i$  em relação às regiões de menor fronteira.
- ✓ Distância entre os centróides: os pesos  $W_{ij}$  são definidos pela distância entre os centróides de  $i$  e  $j$ , onde, quando maior a distância, menor o peso  $W_{ij}$ .

Uma discussão mais aprofundada sobre a definição das matrizes de vizinhanças e pesos pode ser encontrada em Bernardinelli et al., 1995b, Best et al, 1999, Kelsal e Wakefield, 2001 e Schmidt e Ferreira, 2006.

### 2.3.2 Definição das Hiperprioris

Após a definição do modelo e das distribuições *a priori* para os efeitos, deve-se ainda definir as distribuições *a priori* (hiperprioris) para os parâmetros (hiperparâmetros) das distribuições *a priori* dos efeitos.

No presente caso, o modelo possui dois efeitos aleatórios,  $u_i$  e  $b_i$ , cada um com uma determinada distribuição *a priori*, sendo:

$$u_i \sim \text{Normal}(0; \sigma^2), i = \{1, \dots, n\} \text{ e,}$$

$$b_i | b_{j \neq i} \sim \text{Normal} \left( \frac{\sum_{j \in \tilde{\delta}_i} W_{ij} b_j}{\sum_{j \in \tilde{\delta}_i} W_{ij}} ; \frac{\sigma_b^2}{\sum_{j \in \tilde{\delta}_i} W_{ij}} \right), i = \{1, \dots, n\}$$

onde pode-se observar que apenas  $\sigma^2$  e  $\sigma_b^2$  são quantidades ainda desconhecidas.

Em geral, quando se trabalha num contexto Bayesiano realizam-se as transformações,  $\tau = 1 / \sigma^2$  e  $\tau_b = 1 / \sigma_b^2$  com o objetivo de facilitar os cálculos necessários para a obtenção da distribuição *a posteriori*.

Estes parâmetros,  $\tau$  e  $\tau_b$  são responsáveis pelo grau de precisão das estimativas dos efeitos  $u_i$  e  $b_i$ . Como  $\tau$  e  $\tau_b$  são definidos como o inverso da variância, quanto menor o valor de  $\tau$  e  $\tau_b$ , maior a variabilidade das estimativas dos efeitos e por conseqüência menor precisão.

Em geral a distribuição *a priori* é definida subjetivamente, sem a utilização dos dados observados, ao contrário da função de verossimilhança que reflete os dados observados e quando não se tem informações *a priori* deve-se utilizar distribuições *a priori* pouco informativas (com pequena precisão obtida através de valores próximos a zero para  $\tau$  e  $\tau_b$ ) de modo que a informação proveniente dos dados tenha maior importância na obtenção da distribuição *a posteriori* dos riscos, em relação à informação subjetiva proveniente das distribuições *a priori*.

Como  $\tau$  e  $\tau_b$  são hiperparâmetros desconhecidos, deve-se definir uma distribuição *a priori* para estes hiperparâmetros. Em geral as prioris atribuídas para  $\tau$  e  $\tau_b$ ,  $P(\tau)$  e  $P(\tau_b)$  são distribuições conjugadas Gamma independentes, denotadas por  $\Gamma(\alpha_\tau; \beta_\tau)$  e  $\Gamma(\alpha_{\tau_b}; \beta_{\tau_b})$ .

Numa distribuição Gamma  $\Gamma(a;b)$ , a esperança é dada pela razão entre  $a$  e  $b$  e a variância é dada pela razão entre  $a$  e  $b^2$ . No presente contexto, por razões já colocadas anteriormente, é indicado que se determinem valores de  $a$  e  $b$  que produzam variância grande. As quantidades mais utilizadas para  $a$  e  $b$  são 0,001 para os dois parâmetros. Uma maior discussão relativa à escolha das hiperprioris pode ser encontrada em (Bernardinelli et al., 1995b) e (Banerjee et al., 2004).

O grau de suavização das estimativas dos riscos é controlado através destes parâmetros  $\tau$  e  $\tau_b$ . A utilização de diferentes hiperprioris leva a diferentes estimativas para o risco não havendo um consenso na literatura sobre qual grau de suavização deve ser obtido e quanto aos valores de  $a$  e  $b$ .

Deste modo, é sempre adequado realizar a análise atribuindo diferentes valores para  $a$  e  $b$  sendo escolhido para o modelo final aquele que melhor se ajusta aos dados na opinião do pesquisador, ou através da utilização de algum critério de escolha de modelos.

### 2.3.3 Distribuições a Posteriori

Por fim é obtida a distribuição conjunta a posteriori  $P(\alpha, b_i, u_i, \tau, \tau_b / y)$  de todos os parâmetros envolvidos no modelo  $(\alpha, b_i, u_i, \tau, \tau_b)$ . Esta distribuição contém toda a informação probabilística acerca destes parâmetros e é obtida através da utilização do teorema de Bayes, que combina as distribuições a priori  $P(b_i/\tau_b)$ ,  $P(u_i/\tau)$ , as hiperprioris  $P(\alpha)$ ,  $P(\tau)$  e  $P(\tau_b)$  e a função de verossimilhança  $P(y/\alpha, b_i, u_i, \tau, \tau_b)$ . Na prática a distribuição a posteriori é proporcional ao produto das distribuições dos três níveis hierárquicos.

$$P(\alpha, b_i, u_i/ y) \propto P(y/\alpha, b_i, u_i, \tau, \tau_b) P(b_i/\tau_b) P(u_i/\tau) P(\alpha) P(\tau)P(\tau_b)$$

Em muitos casos, em geral quando o número de parâmetros envolvidos no modelo é grande, a distribuição a posteriori conjunta, bem como as distribuições a posteriori marginais não podem ser obtidas analiticamente, sendo assim se recorre a métodos de que aproximam estas distribuições através de simulações. Estas simulações são realizadas através do método de Simulação de Monte Carlo Via Cadeias de Markov (método MCMC) e mais especificamente através do Amostrador de Gibbs (Gelfand e Smith, 1990).

Através da distribuição conjunta a posteriori podem ser obtidas as distribuições marginais de todos os parâmetros envolvidos no modelo bem como as distribuições a posteriori marginais para  $\log(\theta_i)$ , para os riscos relativos  $\theta_i$  além das taxas de incidência/ mortalidade possibilitando assim a realização de inferências acerca dos parâmetros, como estimativas pontuais, intervalos de credibilidade entre outros. Em geral as estimativas pontuais são dadas pela esperança quando as distribuições marginais são simétricas e pela mediana quando as distribuições marginais são fortemente assimétricas.

Assim como na inferência clássica a precisão das estimativas pontuais depende do tamanho amostral, porém, sob o enfoque bayesiano e mais especificamente através dos métodos de simulação, o tamanho amostral é dado pelo número de iterações utilizadas na estimação, ou seja, pela quantidade de valores simulados para a obtenção da distribuição a posteriori. Usualmente o conjunto de valores simulados é denominado de cadeia.

A qualidade das distribuições a posteriori bem como das estimativas pontuais para os parâmetros deve ser avaliada através da avaliação da convergência da cadeia para a distribuição de interesse, ou distribuição estacionária, e através da verificação da independência, ou ausência de autocorrelação, entre os valores simulados.

A condição de convergência da cadeia é obtida quando o aumento do tamanho amostral não produz mais modificações circunstanciais na distribuição a posteriori. A convergência pode ser avaliada através de critérios estatísticos ou de inspeção visual.

É desejável também que os valores simulados não possuam forte autocorrelação, ou seja, sejam aproximadamente independentes. A autocorrelação existe quando, por exemplo, o segundo valor simulado é dependente do primeiro, o terceiro do segundo (e por vezes até do primeiro, com um menor grau de dependência) e assim por diante.

Quando se utiliza o método de simulação MCMC os valores da cadeia não são gerados independentemente e o grau de relacionamento entre os valores obtidos nas iterações é dependente da complexidade do modelo a ser estimado e do número de parâmetros envolvidos. Uma das formas de diminuir a autocorrelação dos valores simulados é utilizar apenas uma simulação a cada conjunto de  $k$  simulações.

## **2.4 Modelos Espaço-Temporais**

Em diversas ocasiões, é de interesse do pesquisador a análise da evolução do risco de uma determinada doença, durante um certo período, dentro de uma determinada região (Werneck e Struchiner, 1997). Através deste tipo de análise

pode-se observar e analisar as variações deste risco, podendo-se determinar se há um aumento, ou decréscimo no risco, e até mesmo, se existe um deslocamento deste risco ao longo da região em estudo.

Os modelos espaço-temporais que utilizaremos neste trabalho são próprios para captação dos diferentes tipos de efeitos determinados por estas variações. Através destes modelos é possível realizar a estimação dos riscos, e conseqüentemente, das taxas de incidência, em cada uma das regiões estudadas, para cada unidade de tempo observada, sendo que cada diferente área pode apresentar um diferente padrão de variação das taxas ao longo do tempo. Em geral são utilizados modelos que pertencem à classe dos Modelos Lineares Generalizados Mistos (GLM) (Mc Cullagh e Nelder, 1989).

O número de casos observados na  $i$ -ésima área, na  $k$ -ésima unidade de tempo, ( $y_{ik}$ ) é modelado através de uma distribuição Poisson com média  $e_{ik}\theta_{ik}$ , onde  $e_{ik}$  é definido como o número esperado de casos na  $i$ -ésima área, na  $k$ -ésima unidade de tempo, caso não existam diferenças de risco entre as regiões e as unidades de tempo, ou seja, caso o risco seja constante em toda a região analisada e constante ao longo do tempo.

O risco  $\theta_{ik}$  é uma quantidade que pode assumir valores reais entre 0 e  $+\infty$ , sendo igual a 1, quando o risco possui exatamente a mesma magnitude do risco médio para toda a área e todo o tempo. Valores superiores a 1 expressam sobre-risco, isto é, risco aumentado, e inferiores a 1, baixo risco, sempre em relação à média global.



A estimativas de  $\theta_{ik}$  são obtidas através da sua decomposição em efeitos espaciais, temporais e espaço-temporais através da utilização da função de ligação logarítmica entre o número de casos observados e os componentes do modelo.

$$y_{ik} \sim \text{Poisson}(e_{ik}\theta_{ik}), \text{ onde } \theta_{ik} \text{ é estimado por } r_{ik} = y_{ik}/e_{ik}$$

Utilizando a função de ligação logarítmica para estimar  $\theta_{it}$ ,

$$\log(\theta_{ik}) = c + f(s_i) + f(t_k) + f(s_i, t_k)$$

onde  $c$  é uma constante,  $f(s_i)$  é uma função que estima o efeito puramente espacial,  $f(t_k)$  uma função que estima o efeito puramente temporal e  $f(s_i, t_k)$  uma função que estima o efeito da interação entre as regiões e as unidades de tempo analisadas, também denominado efeito espaço-temporal.

As estimativas de  $\theta_{ik}$  são obtidas através da aplicação da função exponencial,

$$\theta_{it} = \exp\{c + f(s_i) + f(t_k) + f(s_i, t_k)\},$$

A forma funcional de cada uma das funções,  $f(s_i)$ ,  $f(t_k)$  e  $f(s_i, t_k)$  é determinada através do objetivo específico que se postula ao construir o modelo, sendo que em alguns casos, a própria natureza do fenômeno indica a forma funcional adequada aos componentes do modelo, porém na maioria dos casos a decisão é baseada em conhecimento prévio (Wickle, 2003).

Conforme (Lawson, 2001), uma das primeiras aplicações deste tipo de modelagem foi realizada por Bernardinelli et al., 1995 onde,

$$\log(\theta_{ik}) = \alpha + b_i + u_i + \beta t + \delta_{it},$$

onde  $\alpha$  é uma constante,  $b_i$  é um efeito aleatório espacialmente estruturado CAR Normal, onde se assume que regiões vizinhas possuem perfis de risco similares,  $u_i$  é um efeito aleatório não estruturado, que capta as variações das taxas entre as regiões não captadas por  $b_i$ ,  $\beta$  é o efeito fixo de tendência linear no tempo e  $\delta_i$  é estruturado assim como  $b_i$ , e representa o efeito de interação espaço-temporal, onde cada uma das regiões analisadas pode apresentar diferentes efeitos temporais.

Note que, neste modelo, não existem coeficientes temporais aleatórios, ou seja, o tempo é modelado sempre como uma função linear, que pode variar entre as regiões.

Xia e Carlin (1998) propuseram um modelo com a seguinte forma funcional:

$$\log(\theta_{ik}) = c + \beta t + b_i^{(k)}$$

onde  $b_i^k$  representa o efeito espacial em cada um dos  $m$  períodos analisados, sendo modelados independentemente através de uma distribuição CAR Normal.

$$k = 0 ; b_i^{(0)}/b_j^{(0)} \in \delta \sim \text{CAR Normal } (\sigma^2_{i0})$$

$$k = 1 ; b_i^{(1)}/b_j^{(1)} \in \delta \sim \text{CAR Normal } (\sigma^2_{i1})$$

$$k = m ; b_i^{(m)}/b_j^{(m)} \in \delta \sim \text{CAR Normal } (\sigma^2_{im})$$

Neste modelo pode-se observar que  $\beta$  é o efeito de tendência linear temporal e  $b_i^k$  é o efeito de interação espaço-temporal.

Waller et al (1997), propuseram um modelo onde o log do risco no tempo  $k$  é modelado pela adição de dois componentes espaciais aleatórios independentes,  $u_i^{(k)}$  sem estrutura espacial e  $b_i^{(k)}$  espacialmente estruturado.

$$\log(\theta_{ik}) = \text{covariáveis} + b_i^{(k)} + u_i^{(k)}$$

$$k = 0 ; u_i^{(0)} \sim \text{Normal}(\sigma_{u0}^2) ; b_i^{(0)}/b_j^{(0)} \in \delta \sim \text{CAR Normal}(\sigma_{i0}^2)$$

$$k = 1 ; u_i^{(1)} \sim \text{Normal}(\sigma_{u1}^2) ; b_i^{(1)}/b_j^{(1)} \in \delta \sim \text{CAR Normal}(\sigma_{i1}^2)$$

$$k = m ; u_i^{(m)} \sim \text{Normal}(\sigma_{um}^2) ; b_i^{(m)}/b_j^{(m)} \in \delta \sim \text{CAR Normal}(\sigma_{im}^2)$$

Pode-se notar que os componentes  $u_i^{(k)}$  e  $b_i^{(k)}$  possuem distribuições independentes em relação ao tempo, ou seja, as estimativas em cada um dos diferentes períodos não dependem de outros períodos, sendo assim, a interação espaço-temporal é modelada como variações nas correlações espaciais ao longo do tempo.

Lagazio et al. (2001) utilizou a distribuição CAR Normal para modelar o efeito aleatório temporal, onde o efeito temporal no k-ésimo período depende apenas dos períodos k-1 (imediatamente anterior) e k+1 (imediatamente posterior). Este método de modelagem no efeito temporal pode ser realizada quando existe o objetivo de suavizar as estimativas dos efeitos temporais, e não apenas dos efeitos espaciais.

Mais recentemente estão sendo desenvolvidos modelos espaços-temporais da classe dos modelos lineares dinâmicos (West e Harrison, 1997) onde o efeito espacial é modelado dinamicamente por:

$$\log(\theta_{ik}) = f(s_i) + \beta_{ik}$$

onde,

$$\beta_{ik} = \beta_{i(k-1)} + e_{ik}, e_{ik} \sim \text{Normal}(\sigma^2) \text{ ou } \text{CAR Normal}(\sigma_i^2)$$

e,

$$\beta_{i0} \sim \text{Normal}(\sigma^2) \text{ ou } \text{CAR Normal}(\sigma_i^2)$$

Sendo assim a estimativa temporal para o k-ésimo período é dependente da estimativa obtida para o período k-1.

### **3. OBJETIVOS**

#### **3.1 Objetivo Geral**

O objetivo geral deste trabalho é apresentar o método de estimação inteiramente Bayesiana na estimação de taxas de incidência ou taxas de mortalidade para dados agregados em área e em unidades de tempo, analisando de maneira adequada sua distribuição espaço-temporal, através da construção de modelos estatísticos, além da utilização de mapas para a representação desta distribuição.

#### **3.2 Objetivos Específicos**

Apresentar modelos que suavizam as estimativas de incidência ou de mortalidade em regiões de pequena população em risco através da utilização da informação espacial e temporal dos casos.

Apresentar uma aplicação dos modelos espaço-temporais na análise das taxas de mortalidade infantil nos municípios do Rio Grande do Sul no período de 2001 a 2004.

#### 4. REFERÊNCIAS BIBLIOGRÁFICAS

Assunção RM.(2001). *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*, São Carlos-SP, Associação Brasileira de Estatística. Disponível em <http://www.est.ufmg.br/leste/publicacoes.htm>. Data de acesso: 04/06/2005

Bailey TC & Gatrell AC. (1995). *Interactive Spatial Data Analysis*. Essex: Longman.

Bailey TC. (2001). Spatial statistical methods in health. *Cadernos de Saúde Pública*, Rio de Janeiro, 17(5):1083-1098.

Banerjee S, Carlin BP, and Gelfand AE. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton: Chapman and Hall/CRC,.

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. (1995a). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* ; 14: 2433-43

Bernardinelli L, Clayton D, Montomoli C. (1995b). Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine*, 14, 2411–2431.

Besag J, Kooperberg C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733-746.

Besag J, York J, Mollié A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1–59.

Best N, Richardson S, Thomson A.(2005). A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res*, 14:35-59.

Best N, Arnold RA, Thomas A, Waller LA, Conlon EM. (1999). Bayesian models for spatially correlated disease and exposure data. In: J. M. Bernardo, J. O. Dawid A. P. Berger and A. F. M. Smith(eds),*Bayesian Statistics 6* (, pp. 131–156. Oxford University Press.

Carlin B, Louis TA.(1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London:. Chapman & Hall, , 399 p.

Carvalho MS, Souza-Santos R. (2005). Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cad. Saúde Pública*, vol.21, no.2, p.361-378.

Elliott P, Wakefield J, Best N, Briggs D. (2001). *Spatial epidemiology: methods and applications*. London: Oxford University Press;.

Gelfand AE, Smith AFM. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.

Kelsal J, Wakefield J. (2001). Modeling spatial variation in disease risk: A geostatistical approach. *Journal of the American Statistical Association*, 97, 692–701.

Knorr-Held L, Besag J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, 17, 2045-2060.

Lagazio C, Dreassi E, Biggeri A. (2001). A hierarchical Bayesian model for space-time variation of disease risk. *Statistical Modelling*, 1: 17–29.

Lawson A. (2001). *Statistical methods in spatial epidemiology*. Sussex: John Wiley & Sons.

McCullagh P, Nelder JA. (1989). *Generalized Linear Models*. 2nd ed. Londres: Chapman and Hall.

Mollié A. (1996). Bayesian Mapping of Disease. In: W.R. Gilks, S. Richardson and D.J. Spiegelhalter (editors). *Monte Carlo Markov Chain in Practice*. .

Richardson S, Thomson A, Best NG, Elliott P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*; 112: 1016-1025.

Schmidt AM, Ferreira GS (2006). Spatial modelling of the relative risk of dengue fever in Rio de Janeiro for the epidemic period between 2001 and 2002. *Brazilian Journal of Probability and Statistics*. Vol 20. Number 1.

Snow J. (1855). *Mode of communication of cholera*, 2.ed., London: John Churchill.

Sun D, Tsutakawa RK, Kim H, He Z. (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine*.; 19(15):2015–2036.

Waller LA, Carlin BP, Xia H, Gelfand AE. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92, 607-617.

Werneck GL, Struchiner CJ. (1997). Estudos de Agregados de Doença No Espaço-Tempo: Conceitos, Técnicas e Desafios. *Cadernos de Saúde Pública*, v. 13, n. 8, p. 435-448.

West M, Harrison PJ. (1997). *Bayesian Forecasting and Dynamic Models* (2nd.ed.). New York: Springer.



Wikle CK, Berliner LM, Cressie N. (1998). Hierarchical Bayesian space-time models. *J. Envir. Ecol. Stat.*, 5, 117–154.

Wikle CK. (2003). Hierarchical models in environmental science. *International Statistical Review*, 71, 181-199.

Xia H, Carlin BP.(1998). Spatio-Temporal models with errors in covariates: Mapping Ohio lung cancer mortality. *Statistics in Medicine*, V.17, p.2025-2043.

## 5. ARTIGO

### **Padrões Espaço-Temporais da Mortalidade Infantil no Rio Grande do Sul entre 2001 e 2004.**

### **Spatio-Temporal Patterns of the Infant Mortality Rates in Rio Grande do Sul from 2001 to 2004.**

Diego de Matos Vieira, Programa de Pós-Graduação em Epidemiologia UFRGS;  
Jandyra Maria Guimarães Fachel, Pós-Graduação em Epidemiologia da UFRGS e  
Departamento de Estatística, Instituto de Matemática UFRGS

*A ser enviado ao Cadernos de Saúde Pública.*

**Padrões Espaço-Temporais da Mortalidade Infantil no Rio Grande do Sul entre 2001 e 2004.**

**Spatio-Temporal Patterns of the Infant Mortality Rates in Rio Grande do Sul from 2001 to 2004.**

Diego de Matos Vieira<sup>1</sup>

Jandyra Maria Guimarães Fachel<sup>12</sup>

1 – Programa de Pós-Graduação em Epidemiologia da Faculdade de Medicina da Universidade Federal do Rio Grande do Sul.

Rua Ramiro Barcelos, 2600, sala 419, Porto Alegre, RS, Brasil, CEP:90035-003

[dmatosv@yahoo.com.br](mailto:dmatosv@yahoo.com.br)

2 – Departamento de Estatística, Instituto de Matemática, UFRGS

## **Resumo**

Neste artigo são analisadas as taxas de mortalidade infantil nos 496 municípios do Rio Grande do Sul no período de 2001 a 2004. A utilização de taxas brutas de mortalidade na construção de mapas dificulta a interpretação da distribuição espacial e da evolução temporal das taxas devido a alta variabilidade destas taxas em regiões onde o número de nascidos vivos é reduzida. Devido a este problema foram utilizados os modelos Bayesianos Hierárquicos na estimação destas taxas. Foram comparados três modelos com diferentes especificações dos componentes espaciais e temporais, onde se verificou que o modelo que utiliza a configuração espacial dos municípios na estimação das taxas obteve maior adequação aos dados quando comparado ao modelo onde as taxas são estimadas sem a utilização do componente espacialmente estruturado. Pode-se observar uma melhora sensível na interpretabilidade dos mapas das taxas de mortalidade em relação aos mapas que utilizam as taxas brutas de mortalidade, podendo-se observar diferentes padrões de evolução temporal em diferentes regiões do Estado.

**Palavras chave:** Epidemiologia Espacial; Modelos Espaço-Temporais; Modelos Bayesianos Hierárquicos; Taxa de Mortalidade Infantil; Mapeamento de doenças.

## **Abstract**

In this article, infant mortality rates from 2001 to 2004 of all 496 cities in Rio Grande do Sul are analyzed. The use of gross mortality rates in map construction impairs the interpretation spatial distribution of the rates and temporal evolution due to high variability of such rates in areas with reduced number of infants born alive. Thus, Hierarchical Bayesian models were used for estimating those rates. Three models with different specifications of the spatial and temporal components were compared. Noticeably, the model that uses the spatial configuration of the cities for estimating the rates obtained the best fit to the data, in comparison to the model where rates are estimated without the spatially structured component. Considerable improvement is noted in the interpretation of mortality rate maps in relation to maps that use gross mortality rates, allowing the observation of different infant mortality patterns in different areas of the State, along the years.

**Key Words:** Spatial Epidemiology; Spatio-Temporal Models; Hierarchical Bayesian Models; Infant Mortality Rate; Disease mapping.

## Introdução

As análises espaciais são reconhecidas há muito tempo na literatura epidemiológica como uma ferramenta básica na descrição e análise de eventos em saúde<sup>1</sup>.

Uma das primeiras análises espaciais aplicada à área da saúde foi realizado por John Snow em 1855. Em Londres ocorria uma epidemia de cólera e pouco se sabia sobre os mecanismos causais da doença. Uma das hipóteses relacionava os casos aos miasmas enquanto outra relacionava a ingestão de água contaminada. No mapa da cidade foram localizados os casos de cólera e as bombas de água que abasteciam a cidade, onde pode-se visualizar claramente uma destas, localizada em Broad Street, como o epicentro da epidemia<sup>2</sup>. Estudos posteriores confirmaram a hipótese de que a ingestão de água contaminada era um dos mecanismos de transmissão da doença.

Desde então, trabalhos com informações espaciais em saúde, área que atualmente é denominada de Epidemiologia Espacial<sup>3</sup>, vêm apresentando crescente utilização por parte dos pesquisadores, impulsionados pela grande disponibilidade de informações em sistemas de saúde, evolução dos sistemas geográficos de informação (SIG) e dos métodos estatísticos, destacando-se a inferência Bayesiana, que ainda é pouco empregada em aplicações na área da saúde<sup>4</sup>.

A Epidemiologia Espacial pode ser dividida em três grandes áreas<sup>3</sup>:

- Mapeamento de doenças: que tem como objetivo avaliar a variação espacial de taxas de mortalidade ou incidência de doenças, visando identificar diferenciais de risco, realizar predições de epidemias, orientar ações em saúde e levantar hipóteses.

- Estudos ecológicos: tem por objetivo estudar a relação entre potenciais fatores etiológicos e a incidência de doenças ou mortalidade. Essencialmente constroem-se modelos onde se busca a explicação das variações das taxas de incidência por área através de indicadores sócio-econômicos, comportamentais e ambientais, incluindo-se na modelagem informações sobre sua localização espacial.
- Detecção de cluster: cujo objetivo principal é identificar e estabelecer a significância de um sobre-risco, não-aleatório, em um determinado espaço, que em geral pode ser pensado como um foco da doença.

Este trabalho tem como foco o Mapeamento de Doenças onde as análises espaciais são realizadas através da construção de mapas que podem ser realizados através da utilização da análise de dados pontuais, ou seja, quando se tem acesso a localização exata do evento, ou através da análise de dados agregados, também denominados como dados de contagem, que utiliza apenas a informação do número de eventos e do tamanho da população em risco nas áreas analisadas. Estas áreas em geral são definidas pelos limites de abrangência de bairros, setores censitários, municípios entre outros<sup>5</sup>. Devido a maior disponibilidade de informações agregadas em relação às informações pontuais, a maior parte destes mapas é realizado com dados agregados por área, onde áreas com diferentes níveis de taxa apresentam cores diferentes e áreas com taxas semelhantes cores iguais.

A construção de mapas utilizando dados agregados (taxas de incidência ou risco por área) é uma tarefa bastante simples. Porém os mapas resultantes são, em muitos casos, de difícil compreensão ou não-informativos. Isto deve ao fato de que indicadores provenientes de regiões com população reduzida, geralmente apresentam

alta variabilidade, mostrando valores extremamente superiores à taxa média da área como um todo (superestimação do risco), ou muito inferiores (subestimação do risco), sem que estas regiões realmente se caracterizem como regiões de alto ou baixo risco, ou seja, estes mapas podem não representar o verdadeiro processo subjacente a estes indicadores <sup>3,6</sup>.

Existem inúmeras proposições de métodos que podem ser utilizados na correção, ou suavização destas estimativas. Atualmente um dos métodos mais adequados e ainda pouco utilizado na Epidemiologia são os modelos Bayesianos Hierárquicos <sup>4</sup>.

A metodologia Bayesiana é bastante flexível em relação aos pressupostos básicos nos modelos fazendo com que estes modelos possam ser generalizados mais facilmente do que poderia ser feito sob o enfoque da inferência clássica <sup>7,8</sup>.

O avanço teórico, a disponibilidade de informações e a evolução das técnicas de estimação (principalmente os estimadores Bayesianos) e predição têm aumentado a preferência por estudos que levam em conta a interação espaço-tempo na identificação de cofatores associados e na predição de tendências.

Através do desenvolvimento dos estimadores Bayesianos, os modelos espaciais puderam ser generalizados incorporando o fator tempo (modelos espaço-temporais), sendo atualmente os modelos mais adequados na análise das variações das taxas de incidência, risco e mortalidade <sup>9,10,11</sup>.

O objetivo deste artigo é apresentar um modelo de análise espaço-temporal utilizando dados de mortalidade infantil através da utilização de estimadores Bayesianos, apresentando e comparando modelos espaço-temporais com diferentes especificações.



## **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 DATASUS <sup>12</sup>. Os óbitos e nascimentos foram atribuídos ao município de residência da mãe.

Segundo o SINASC no período de 2001 a 2004, no Rio Grande do Sul, observou-se 618.031 nascidos vivos com 9.661 óbitos infantis (nascidos vivos com menos de 1 ano de idade), ou seja, globalmente o Estado durante este período obtém-se uma taxa de incidência de 15,63 óbitos por 1.000 nascidos vivos.

O DATASUS disponibiliza os dados referentes à mortalidade infantil por município a partir do ano de 1994, porém devido a criação de novos municípios, divisões e agrupamentos de municípios ao longo dos últimos anos preferiu-se não utilizar todos os dados disponíveis.

### **Análise Estatística**

Os modelos espaço-temporais utilizados são próprios para a modelagem dos diferentes tipos de efeitos determinados por estas variações (espaciais e temporais). Através destes modelos será possível realizar a estimação dos riscos e, conseqüentemente das taxas de incidência, em cada uma das regiões estudadas, para cada unidade de tempo observada, podendo apresentar diferentes variações nas taxas

ao longo do tempo nas diferentes regiões, ou seja, cada área pode apresentar um diferente padrão de variação das taxas ao longo do tempo.

O número de casos observados na  $i$ -ésima área, na  $k$ -ésima unidade de tempo ( $y_{ik}$ ) é modelado através de uma distribuição Poisson com média  $e_{ik}\theta_{ik}$ , onde  $e_{ik}$  é o número esperado de casos na  $i$ -ésima área na  $k$ -ésima unidade de tempo, caso não existam diferenças de risco entre as regiões e as unidades de tempo, e  $\theta_{ik}$ , o risco relativo de ocorrência de um caso na  $i$ -ésima área, na  $k$ -ésima unidade de tempo em relação a média global.

O risco relativo  $\theta_{ik}$  é uma quantidade que pode assumir valores reais entre 0 e  $+\infty$ , sendo igual a 1, nas áreas onde o risco possui exatamente a mesma magnitude do risco médio para toda a região durante todo o período. Valores superiores a 1 expressam risco aumentado (sobre- risco) e inferiores a 1, baixo risco, sempre em relação à média global.

As estimativas de  $\theta_{ik}$  são obtidas através da decomposição do risco em efeitos espaciais, efeitos temporais e efeitos espaço-temporais através da utilização da função de ligação logarítmica entre o risco relativo  $\theta_{ik}$  e os componentes do modelo.

$$y_{ik} \sim \text{Poisson}(e_{ik}\theta_{ik}),$$

Utilizando a função de ligação logarítmica para estimar  $\theta_{ik}$ ,

$$\log(\theta_{ik}) = c + f(s_i) + f(t_k) + f(s_i, t_k)$$

As estimativas de  $\theta_{ik}$  são obtidas através da aplicação da função exponencial,

$$\theta_{ik} = \exp\{c + f(s_i) + f(t_k) + f(s_i, t_k)\}$$

onde  $c$  é uma constante que representa a média do logaritmo do risco global, ou seja em toda a região ao longo de todo o período de tempo,  $f(s_i)$  é uma função que estima o efeito puramente espacial,  $f(t_k)$  uma função que estima o efeito puramente temporal e  $f(s_i, t_k)$  uma função que estima o efeito da interação entre as regiões e as unidades de tempo analisadas, também denominado efeito espaço-temporal.

Em modelos com enfoque Bayesiano deve-se determinar as distribuições *a priori* para cada um dos componentes do modelo, a constante  $c$ ,  $f(s_i)$ ,  $f(t_k)$  e  $f(s_i, t_k)$ . A utilização de diferentes prioris leva a diferentes distribuições *a posteriori* e conseqüentemente a estimativas diferentes.

A forma funcional dos modelos, bem como a determinação das distribuições *a priori* dos componentes do modelo devem ser definidas de acordo com o processo que se deseja modelar. Em geral se deve testar diferentes modelos quanto a forma funcional e especificação das distribuições *a priori* de modo a utilizar o modelo que melhor representa os dados observados. Neste artigo serão testados três modelos com diferentes especificações para as prioris.

As distribuições *a posteriori* para cada um dos componentes do modelo foram realizadas através da utilização de Simulação de Monte Carlo via Cadeias de Markov (MCMC) através do Amostrador de Gibbs<sup>13</sup> implementadas no software WinBugs<sup>14</sup> versão 1.4.1. As estimativas das taxas de incidência de cada um dos municípios em cada um dos anos foram calculadas através da esperança das respectivas distribuições *a posteriori*.

A verificação da convergência foi realizada através de inspeção visual das trajetórias de duas cadeias de Markov utilizando diferentes valores iniciais. Com o objetivo de diminuir a autocorrelação nos valores amostrados, será utilizado apenas um valor a cada 5 valores amostrados.

Será utilizado o critério de comparação de modelos DIC<sup>15</sup> (Deviance Information Criterion) para a escolha do modelo. O modelo que apresentar o menor valor para o DIC é considerado como o que tem melhor capacidade de representar os dados observados.

#### Modelo 1

O primeiro modelo testado é dado por:

$$\log(\theta_{ik}) = \alpha + v_i + (\beta + \beta_i)t$$

onde  $t$  é a covariável tempo codificada como 1 para o ano de 2001, igual a 2 para o ano de 2002, igual a 3 para o ano de 2003 e igual a 4 para o ano de 2004.

Quanto aos coeficientes do modelo define-se  $\alpha$  como uma constante, com distribuição a priori Uniforme entre  $-\infty$  e  $+\infty$ ,  $v_i$  representa o efeito espacial para cada um dos 496 municípios, modelados independentemente através de priors distribuídas normalmente.

$$v_i \sim \text{Normal}(0, \tau_v), \tau_v = 1/\sigma_v^2$$

Note que  $v_i$  é um efeito espacial, porém não utiliza explicitamente a informação espacial na estimação dos riscos.

O fator temporal é modelado através de uma função linear onde  $\beta$  possui distribuição *a priori* Uniforme entre  $-\infty$  e  $+\infty$  e representa o efeito temporal e  $\beta_i$  é modelado através de distribuições *a priori* Normal independentes, representando a interação espaço-tempo, já que este coeficiente permite que cada um dos 496 municípios possuam seu próprio perfil evolutivo.

$$\beta_i \sim \text{Normal}(0, \tau_\beta), \tau_\beta = 1/\sigma_\beta^2$$

Aos hiperparâmetros foram atribuídas distribuições *a priori* Gamma independentes,

$$\tau_v \sim \text{Gamma}(0,001;0,001)$$

$$\tau_\beta \sim \text{Gamma}(0,001;0,001)$$

Estes hiperparâmetros são os responsáveis pelo grau de suavização das estimativas.

Quanto menor os valores de  $\tau_v$  e  $\tau_\beta$  menor a suavização das estimativas.

## Modelo 2

O segundo modelo testado é dado por<sup>16</sup>:

$$\log(\theta_{ik}) = \alpha + v_i + (\beta + \beta_i)t$$

que apesar de ter forma funcional idêntica ao Modelo 1, se difere deste pelas distribuições *a priori* para os componentes espaciais  $v_i$  e  $\beta_i$ .

Estes componentes são modelados através de prioris espacialmente estruturadas com distribuição Condicional Autorregressiva (CAR) Normal, que permitem que municípios vizinhos possuam estimativas semelhantes.

$$v_i \sim \text{CAR Normal}(\tau_v), \tau_v = 1/\sigma_v^2$$

$$\beta_i \sim \text{CAR Normal}(\tau_\beta), \tau_\beta = 1/\sigma_\beta^2$$

A matriz de vizinhanças foi construída através do critério de adjacências, ou seja, apenas os municípios que fazem fronteiras são considerados vizinhos. A matriz de pesos  $W_{ij}$  foi definida como 1 para os municípios vizinhos e 0 para os municípios não vizinhos.

Aos hiperparâmetros foram atribuídas distribuições *a priori* Gamma independentes,

$$\tau_v \sim \text{Gamma}(0,001;0,001)$$

$$\tau_\beta \sim \text{Gamma}(0,001;0,001)$$

A covariável  $t$  que representa o tempo é definida como no Modelo 1, bem como os coeficientes  $\alpha$  e  $\beta$  que possuem distribuição *a priori* como definida no Modelo 1.

### Modelo 3

O terceiro modelo testado é dado por:

$$\log(\theta_{ik}) = v_i + \beta_i t_k$$

onde todos os coeficientes espaciais do modelo  $v_i$  e  $\beta_i$  possuem distribuições *a priori* definidas como no Modelo 2.

O fator temporal  $t_k$  tem uma estrutura dinâmica<sup>17</sup>.

$$t_1 \sim \text{Normal}(0, \tau_t)$$

para  $k = \{2, 3, 4\}$

$$t_k \sim \text{Normal}(t_{k-1}, \tau_t)$$

Aos hiperparâmetros foram atribuídas distribuições *a priori* Gamma,

$$\tau_v \sim \text{Gamma}(0,001;0,001)$$

$$\tau_t \sim \text{Gamma}(0,001;0,001)$$

## Resultados

Através do critério DIC e da análise da convergência da cadeia, optou-se pela utilização do Modelo 2 (DIC = 6.094,57), pois este apresentou DIC inferior ao Modelo 1 (DIC = 6.120,06) enquanto que para o Modelo 3 não foi obtida convergência.

As estimativas dos parâmetros do modelo foram obtidas através do Amostrador de Gibbs, utilizando-se 5.000 iterações de Burn-in, e 30.000 iterações sendo utilizadas apenas 1 valor a cada 5 valores amostrados, ou seja 6.000 valores foram utilizados na geração das distribuições a posteriori dos parâmetros e estimação dos mesmos através da esperança da posteriori.

Durante o período analisado observou-se pequenas variações na taxa de mortalidade infantil para o Estado do Rio Grande do Sul e variações maiores entre os municípios.

Tabela 1 – Descritivas das Taxas Brutas de Mortalidade Infantil 2001-2004

Ano	Taxa	Mínimo	1º Quartil	2º Quartil	3º Quartil	Máximo
2001	15,75	0,00	0,00	12,20	21,62	100,00
2002	15,64	0,00	0,00	13,33	24,69	93,75
2003	15,97	0,00	0,00	12,20	21,74	166,67
2004	15,16	0,00	0,00	12,09	23,36	96,77

Através das Figuras 1, 2, 3 e 4 e da Tabela 1, pode-se observar que em todos os anos analisados, observa-se a grande quantidade de municípios onde não ocorreram óbitos infantis sendo a taxa bruta igual a 0. Além disso observa-se também que existem municípios com taxas brutas de mortalidade altíssimas quando comparadas a taxa estadual. A maior taxa observada foi em 2003, onde a estimativa foi de 166,67 óbitos a cada 1.000 nascimentos, aproximadamente 10 vezes superior a taxa estadual de 15,97. Pode-se observar que esta taxa é proveniente de um município onde em 18

nascimentos, ocorreram 3 óbitos, sendo que no ano anterior não ocorreram óbitos e portanto a taxa observada foi igual a 0.

Em relação ao padrão temporal, também não se consegue observar um padrão claro. É possível verificar que em muitos municípios a taxa em um ano é muito elevada em relação a média estadual enquanto no ano posterior a taxa é muito baixa ou igual a 0.

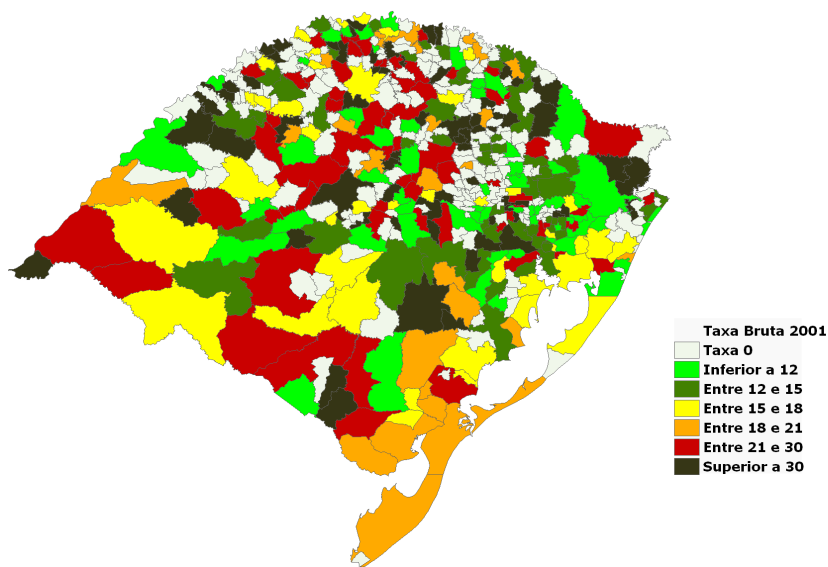


Figura 1 – Taxas brutas para a Mortalidade Infantil em 2001

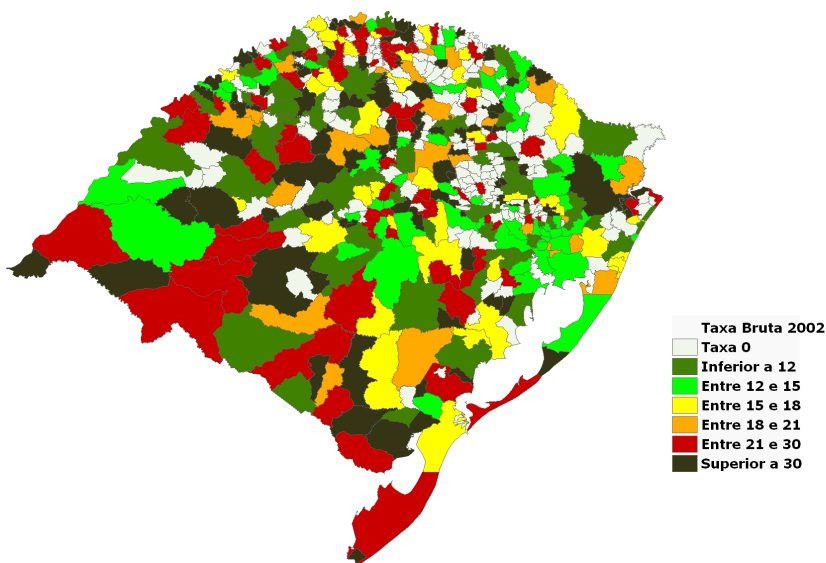


Figura 2 – Taxas brutas para a Mortalidade Infantil em 2002



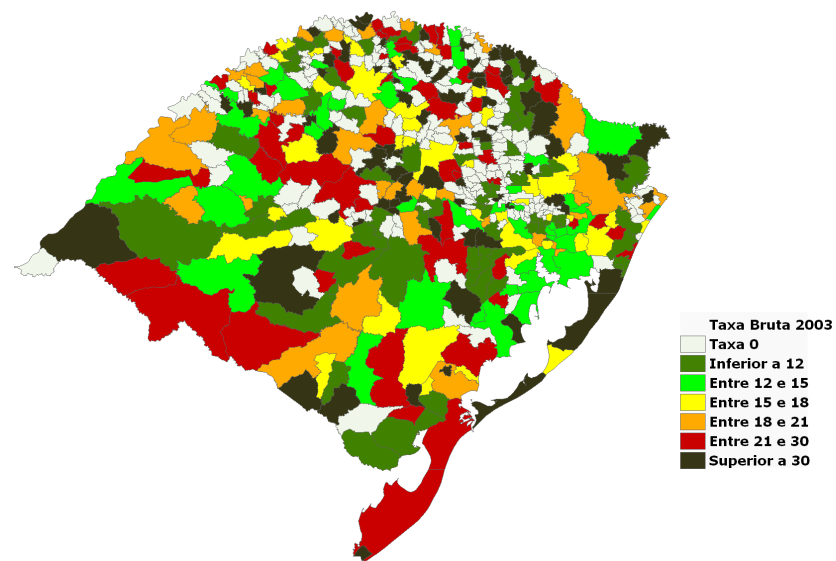


Figura 3 – Taxas brutas para a Mortalidade Infantil em 2003

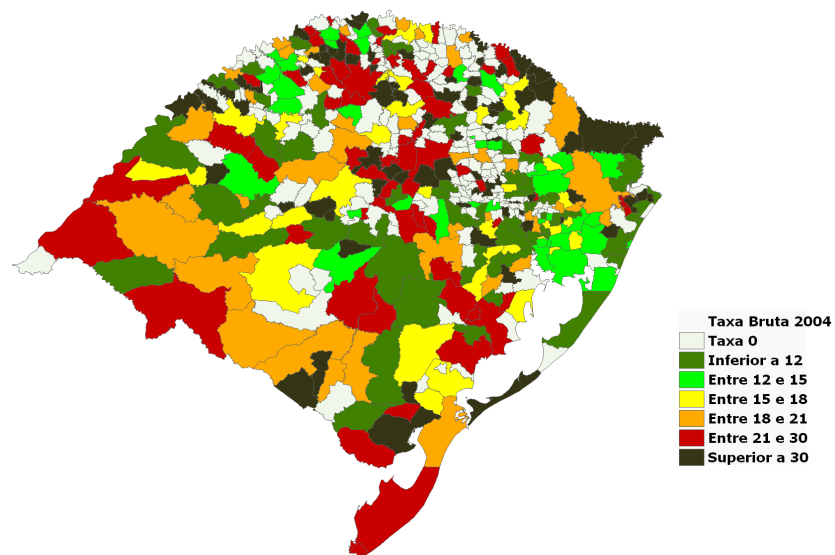
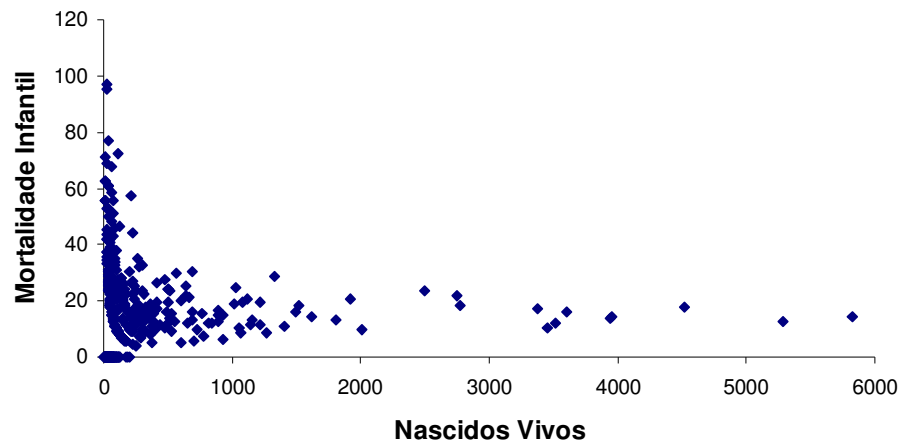


Figura 4 – Taxas brutas para a Mortalidade Infantil em 2004

### Taxa Bruta vs Nascidos Vivos



\* por problemas de escala visual do gráfico, o município de Porto Alegre não foi considerado por obter alto número de nascidos vivos

Figura 5 – Diagrama de dispersão dos municípios segundo as taxas brutas de mortalidade infantil e o número de nascidos vivos, Rio Grande do Sul, 2004

Pode-se observar através do Gráfico 1, que as taxas brutas muito diferenciadas da média global (taxa bruta igual a 0 ou superior a 40 óbitos por 1.000 nascidos vivos) são obtidas apenas em municípios onde o número de nascidos vivos (denominador da taxa de mortalidade) é pequeno. Estas estimativas possuem grande margem e erro, ou seja, as taxas brutas podem não representar a taxa de mortalidade verdadeira (no contexto, taxa de mortalidade verdadeira refere-se à taxa subjacente que gerou a taxa observada. Não significa que as taxas observadas não representem a verdade por problemas de sub-registros). Além disso, também se pode observar que estes mapas são de difícil interpretação no sentido de observar se existe um incremento ou decréscimo nas taxas ao longo dos anos.

Após a suavização das taxas através dos estimadores Bayesianos pode-se observar claramente um padrão na dispersão espacial da mortalidade infantil no Estado do Rio Grande do Sul, e mais discretamente a evolução das taxas no período analisado.

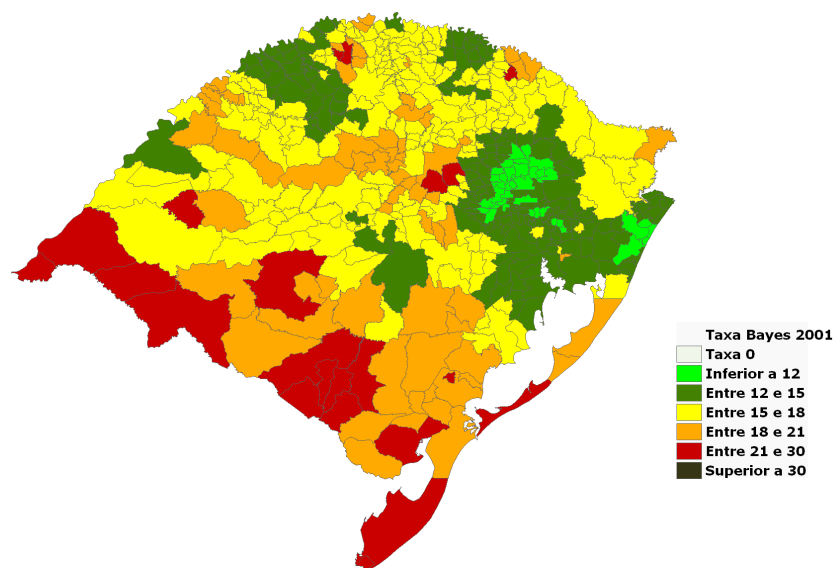


Figura 6 – Taxas Bayesianas para a Mortalidade Infantil em 2001

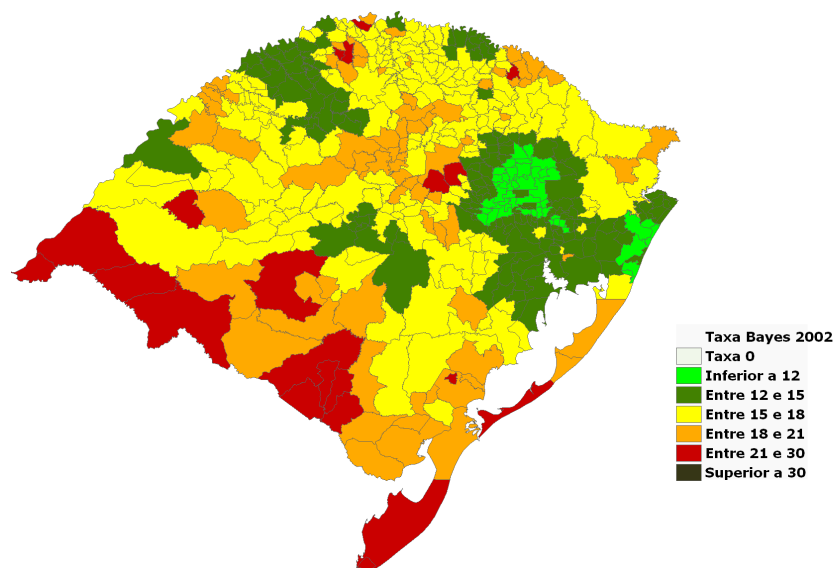


Figura 7 – Taxas Bayesianas para a Mortalidade Infantil em 2002

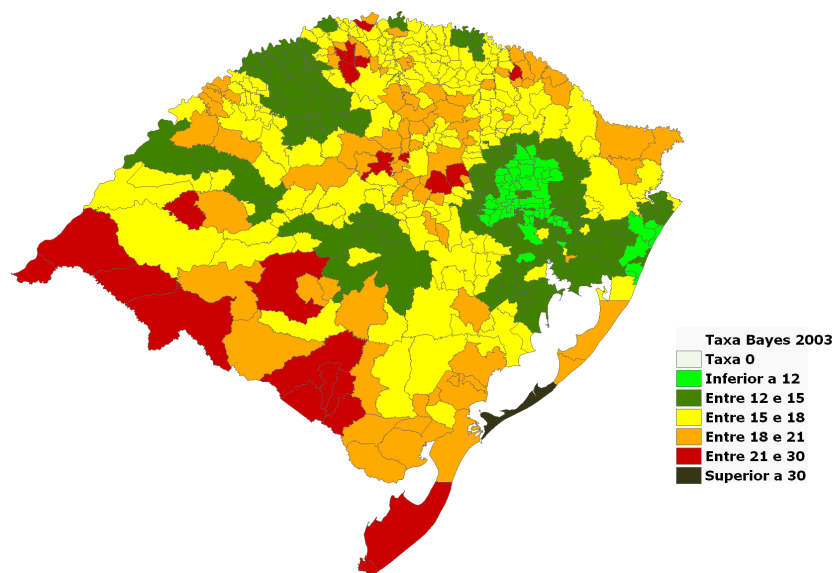


Figura 8 – Taxas Bayesianas para a Mortalidade Infantil em 2003

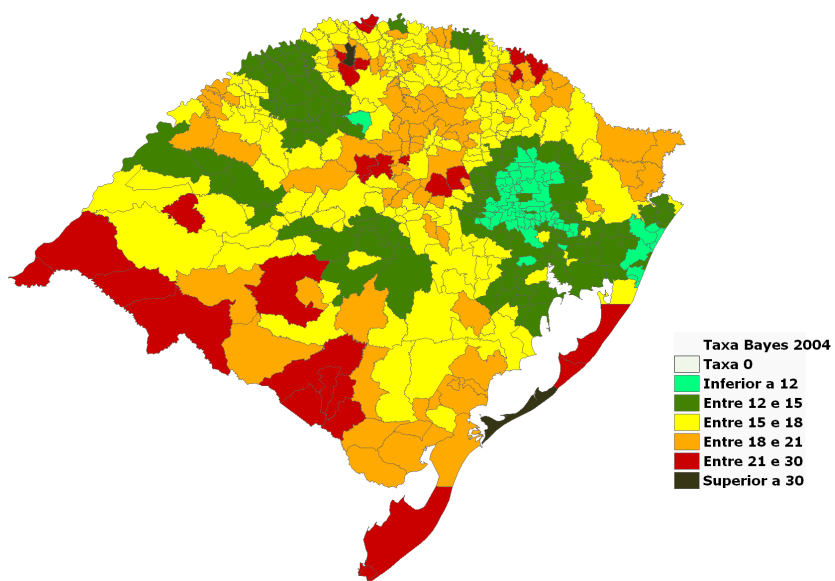


Figura 9 – Taxas Bayesianas para a Mortalidade Infantil em 2004

Tabela 2 – Descritivas das Taxas Bayesianas de Mortalidade Infantil 2001-2004

Ano	Taxa	Mínimo	1º Quartil	2º Quartil	3º Quartil	Máximo
2001	15,85	10,12	13,77	15,60	17,46	25,93
2002	15,72	9,86	13,44	15,55	17,58	28,05
2003	15,59	9,62	13,61	15,63	17,73	31,32
2004	15,48	9,41	13,50	15,77	17,94	35,27

Pode-se observar que as estimativas Bayesianas são suavizadas, ou seja, não existem mais municípios com taxas de mortalidade nulas ou extremamente superiores a taxa estadual. Pode-se destacar que a taxa mais alta foi observada em 2004 sendo aproximadamente 128% superior a taxa estadual, estimada em 15,48.

Em relação ao padrão espacial, destaca-se a região da Serra, que apresenta as menores taxas do Estado, com taxas inferiores a 12 óbitos por 1.000 nascidos vivos, em grande parte dos municípios, enquanto a região situada na Metade-Sul, mais especificamente a região da Campanha, fronteira com o Uruguai, apresenta as maiores taxas, variando, em grande parte dos municípios, entre 18 e 30 óbitos por 1.000 nascidos vivos.

Em relação à evolução das taxas no período, pode-se destacar um aumento na desigualdade onde algumas regiões apresentam tendência de decréscimo na taxa de mortalidade, enquanto em outras as taxas apresentam tendência de incremento. Sob este aspecto destaca-se novamente a região da Serra, onde as taxas apresentam uma tendência de decréscimo, podendo ser visualizado pelo aumento da região onde a taxa de mortalidade é inferior a 12. Esta tendência também pode ser visualizado ao Norte/Noroeste do Estado mais especificamente nas regiões próximas a Santo Ângelo, Santa Rosa e Ijuí e na região localizada ao sul entre Pelotas e Santa Maria. Por outro lado, algumas regiões apresentam uma tendência de incremento nas taxas, podendo ser observadas principalmente nos municípios localizados no centro/norte

do Estado (regiões de Passo Fundo e Carazinho) e nos municípios próximos à fronteira com o Estado de Santa Catarina (regiões de Sananduva e Vacaria).

Ainda em relação a evolução das taxas, verificou-se diferenças nos padrões temporais nas variações das taxas entre os municípios. Em alguns municípios a taxa estimada para o ano de 2004 é até 1,63 óbitos por 1.000 nascidos vivos inferior a taxa de 2001, enquanto em alguns municípios a taxa de 2004 é até de 9,91 óbitos por 1.000 nascidos vivos superior a taxa de 2001.

### **Conclusões e Considerações Finais**

Os modelos utilizados foram construídos de maneira a estimar o risco relativo dos municípios em cada um dos anos analisados em relação a taxa global do Estado no período. Através destas estimativas, foram obtidas as estimativas das taxas de mortalidade infantil por 1.000 nascidos vivos, já que este é o indicador mais utilizado quando se analisa a mortalidade infantil.

Através dos mapas das taxas Bayesianas foi possível observar um ganho substancial na interpretação e na detecção de um padrão de variação na taxas de mortalidade infantil nos municípios do Rio Grande do Sul, em relação aos mapas que utilizam apenas as taxas brutas de mortalidade infantil.

Além disso, o modelo utilizado foi capaz de fornecer estimativas evidenciando o padrão de evolução das taxas de mortalidade entre 2001 e 2004, onde cada um dos 496 municípios apresentou diferentes intensidades na variação das taxas, podendo-se assim observar as variações espaço-temporais das taxas de mortalidade infantil.

Este método de modelagem pode ser generalizado através da análise de um período maior e tempo, podendo assim, ser utilizado na previsão das taxas de mortalidade e até mesmo do número de óbitos para os períodos futuros.

Além disso, podem ser incorporadas ao modelo covariáveis, de modo a aumentar a precisão das estimativas, além de ser, por si só, útil na avaliação do grau de associação entre estas covariáveis e as taxas de mortalidade infantil.

Uma das grandes discussões sobre os modelos Bayesianos aplicados a análise espacial é o fato destes apresentarem estimativas muito suavizadas o que conduz a um problema na detecção de áreas de risco elevado. Atualmente estão sendo conduzidos estudos de simulação<sup>1</sup> com o objetivo de comparar os modelos Bayesianos com modelos Não-paramétricos Bayesianos, aonde se verifica uma melhora na detecção de áreas de alto risco com a utilização de modelos Não-paramétricos Bayesianos.

## Referências

1. Richardson S, Best NG, Thomson A, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives* 2004; 112: 1016-1025.
2. Snow J. *Mode of communication of cholera*, 2.ed., London: John Churchill; 1855.
3. Lawson A. *Statistical methods in spatial epidemiology*. Sussex: John Wiley & Sons; 2001.
4. Carvalho MS & Souza-Santos R. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cadernos de Saúde Pública* 2005, vol.21, no.2, p.361-378.
5. Assunção RM. *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*, Associação Brasileira de Estatística 2001. Disponível em <http://www.est.ufmg.br/leste/publicacoes.htm>. Data de acesso: 04/06/2005.
6. Elliott P, Wakefield J, Best N, Briggs D. *Spatial epidemiology: methods and applications*. London: Oxford University Press; 2001.
7. Besag J, York J and Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 1991, 43, 1–59.
8. Mollié A. Bayesian Mapping of Disease. In W.R. Gilks, S. Richardson and D.J. Spiegelhalter (editors). *Monte Carlo Markov Chain in Practice*.;1996.



9. Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of spacetime variation in disease risk. *Statistics in Medicine* 1995; 14: 2433-43.
10. Lagazio C, Dreassi E, Biggeri A. A hierarchical Bayesian model for space-time variation of disease risk. *Statistical Modelling* 2001; 1: 17–29.
11. Wikle CK. Hierarchical models in environmental science. *International Statistical Review* 2003; 71, 181-199.
12. Ministério da Saúde. Informações de Saúde. Estatísticas Vitais – Mortalidade e Nascidos Vivos. Disponível em: <http://www.datasus.gov.br>. Data de acesso: 05/08/2006.
13. Gamerman D. (1997). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. Chapman and Hall.
14. WinBUGS with DoodleBUGS. Version 1.4.1. Imperial College & MRC, UK. Disponível em <http://www.mrcbsu.cam.ac.uk/bugs/winbugs/contents.shtml>
15. Spiegelhalter DJ, Best NG, Carlin BP e Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* 2002; B, 64, 583-639.
16. Castro MS, Vieira VA, Assunção RM. Space-time analysis of lung cancer mortality in Southern Brazil. *Revista Brasileira de Epidemiologia*. 2004; 7(2): 131-143.
17. Ferreira GS. *Análise Espaço-Temporal da Distribuição dos Casos de Dengue na Cidade do Rio de Janeiro no Período de 1986 a 2002*. 2004; 170 f. Dissertação (Mestrado em Estatística) - Universidade Federal do Rio de Janeiro.

**ANEXO 1 – PROJETO DE PESQUISA**

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE MEDICINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

**PROJETO DE PESQUISA**

**ANÁLISE ESPAÇO-TEMPORAL NA ESTIMAÇÃO DE TAXAS DE  
INCIDÊNCIA/ MORTALIDADE**

AUTOR: DIEGO DE MATOS VIEIRA

ORIENTADOR: JANDYRA GUIMARÃES FACHEL

Porto Alegre, 2005

## **1. INTRODUÇÃO**

### **1.1 Epidemiologia Espacial**

“A análise de dados espaciais se dá quando os dados são espacialmente localizados e se considera explicitamente a possível importância de seu arranjo espacial na análise ou interpretação dos resultados”(1).

Neste contexto pressupõe-se que a localização espacial se apresenta como um fator relevante na explicação da ocorrência de determinado fenômeno, como exemplo pode-se citar doenças contagiosas onde unidades mais próximas no espaço possuem maiores chances de contágio e à medida que esta distância aumenta, esta probabilidade de contágio diminui.

A análise de dados espaciais, aplicada ao contexto epidemiológico, também conhecida como Epidemiologia Espacial, vem apresentando crescente utilização por parte dos pesquisadores da área da saúde como uma das ferramentas de desenvolvimento de informação quanto ao comportamento de doenças e construção de indicadores de saúde (2).

Este recente crescimento se deve em grande parte ao desenvolvimento dos sistemas geográficos de informação (GIS), aos inúmeros avanços metodológicos na área da estatística, bem como a combinação destas duas ferramentas.

A Epidemiologia Espacial pode ser dividida em três grandes áreas (3):

- Mapeamento de doenças: que tem como objetivo avaliar a variação espacial de taxas de mortalidade ou incidência de doenças, visando identificar diferenciais de risco, realizar predições de epidemias, orientar ações em saúde e levantar hipóteses.
- Estudos ecológicos: tem por objetivo estudar a relação entre potenciais fatores etiológicos e a incidência de doenças ou mortalidade. Essencialmente

constroem modelos onde se busca a explicação das variações das taxas de incidência através de indicadores sócio-econômicos e comportamentais além da sua localização espacial.

- Detecção de cluster: o objetivo principal desta análise é estabelecer a significância de um sobre-risco em um determinado espaço, que em geral pode ser pensado com um foco da doença.

Este trabalho está situado dentro do mapeamento de doenças, onde a análise da dispersão espacial do risco de uma determinada doença ou de mortalidade é realizada através da construção de mapas de incidência ou mortalidade.

Porém, estes mapas, construídos a partir das taxas observadas podem não representar o verdadeiro processo subjacente que gerou estas taxas, devido ao fato de que taxas provenientes de regiões com população reduzida, geralmente apresentam alta variabilidade, apresentando muito facilmente, valores extremamente maiores, ou menores do que a taxa média, sem que estas regiões realmente se caracterizem como regiões de alto ou baixo risco (4).

Existem diversos métodos para diminuir o efeito da aleatoriedade na estimação destas taxas, tais como controlar as diferenças através de indicadores demográficos, mapas de probabilidade e a utilização de métodos bayesianos. Os dois primeiros métodos estão cada vez mais caindo em desuso dado o ganho circunstancial que os métodos bayesianos incorporam à análise, pois este diferentemente dos demais utiliza diretamente a localização na estimação das taxas (5).

## **1.2 Modelos Espaço-Temporais**

Padrões de mortalidade ou morbidade, propagação de epidemias, transmissão sexual de doenças e a transferência de comportamentos ou valores podem ser explicados através de uma abordagem que além de considerar os grupos, estude o espaço e o tempo.

Uma análise puramente espacial em epidemiologia apresenta a distribuição espacial

de uma determinada taxa de incidência em um momento específico. Porém, dependendo do evento em estudo o maior ganho é obtido quando é analisada a evolução desta taxa no espaço ao longo do tempo, agregando à análise a possibilidade de previsão de epidemias, determinação da direção e velocidade de propagação de doenças.

A análise que utiliza as informações da localização no espaço e no tempo conjuntamente é uma extensão da análise espacial e é denominada análise espaço-temporal (5). Este tipo de análise pressupõe que unidades próximas no espaço tendem a ser mais semelhantes e unidades próximas no tempo também tendem a ser semelhantes. Desta forma o objetivo deste tipo de análise é estudar a interação entre espaço e tempo, ou seja, definir o comportamento temporal da taxa em cada uma das regiões do espaço.

Neste trabalho serão utilizados dados de contagem agrupados em área em consecutivos espaços de tempo discretos, onde se assume que:

$y_{it} \sim \text{Poisson}(e_{it}r_{it})$

onde  $y_{it}$  representa o número de casos novos no tempo  $t$  na região  $i$ ,  $r_{it}$  representa o risco do evento e  $e_{it}$  representa o número esperado de casos na região  $i$  no tempo  $t$ .

Será construído um modelo onde a variável dependente será o  $\log(r_{it})$  e as variáveis independentes serão a localização e o tempo. Este modelo será construído através de métodos bayesianos. A forma deste modelo depende da estrutura dos dados e do tipo de problema que se deseja estudar.

Este modelo será utilizado na construção dos mapas, de modo que estes apresentem informações de maneira mais clara do que os mapas realizados com as taxas brutas.

## **2. OBJETIVOS**

### **2.1 Objetivo Geral:**

O objetivo geral deste trabalho é apresentar o método para obtenção dos estimadores inteiramente bayesianos na estimação de taxas de incidência ou mortalidade para dados agregados em área e em unidades de tempo, analisando de maneira adequada sua distribuição espaço-temporal, através da construção de modelos estatísticos, além da utilização de mapas para a representação desta distribuição.

### **2.1 Objetivos Específicos:**

- 1) Apresentar modelos de suavização das estimativas de incidência ou mortalidade em regiões de pequena população através da utilização de dados de localização espacial e temporal dos casos.
- 2) Apresentar uma aplicação dos modelos espaço temporais aos dados de mortalidade infantil das cidades do Rio Grande do Sul.



## 5. REFERÊNCIAS

(1) BAILEY, T. C. & GATRELL, A. C. (1995). *Interactive Spatial Data Analysis*. Essex: Longman.

(2) CARVALHO, M. S. & SOUZA-SANTOS, R. (2005) Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cad. Saúde Pública*, vol.21, no.2, p.361-378.

(3) LAWSON A. *Statistical methods in spatial epidemiology*. Sussex: John Wiley & Sons; 2001.

(4) ELLIOTT P, WAKEFIELD J, BEST N, BRIGGS D. *Spatial epidemiology: methods and applications*. London: Oxford University Press; 2001.

(5) ASSUNÇÃO, R.M. (2001). *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*, São Carlos-SP, Associação Brasileira de Estatística. Disponível em <http://www.est.ufmg.br/leste/publicacoes.htm>. Data de acesso: 04/06/2005



**ANEXO 2 – LINHAS DE COMANDO****WinBugs - Modelo 1**

```
model
{
  for (i in 1:m)
  {
    for (k in 1:T)
    {
      y[i,k]~dpois(mu[i,k])
      log(mu[i,k])<-log(e[i,k])+alpha+v[i]+(delta+beta[i])*t[k]
      theta[i,k]<-exp(alpha+v[i]+(delta+beta[i])*t[k])
      tx[i,k]<- 15.63190*theta[i,k]
    }
  }

  for(i in 1:m)
  {
    v[i]~dnorm(0,tau.v)
    beta[i]~dnorm(0,tau.beta)
  }

  alpha~dflat()
  delta~dflat()

  tau.v~dgamma(0.001, 0.001)
  tau.beta~dgamma(0.001, 0.001)
}
```

**WinBugs - Modelo 2**

```

model
{
  for (i in 1:m)
  {
    for (k in 1:T)
    {
      # Verossimilhança
      y[i,k]~dpois(mu[i,k])
      log(mu[i,k])<-log(e[i,k])+alpha+u[i]+(beta+delta[i])*t[k]
      # Risco Relativo em cada area e período de tempo
      theta[i,k]<-exp(alpha+u[i]+(beta+delta[i])*t[k])
      # Taxa em cada area e período de tempo
      tx[i,k]<- 15.63190*theta[i,k]
    }
  }

  # Prioris CAR Normal
  u[1:m]~car.normal(adj[],weights[],num[],tau.u)
  delta[1:m]~car.normal(adj[],weights[],num[],tau.delta)

  #Matriz de Pesos
  for(j in 1:sumNumNeig)
  {
    weights[j]<-1
  }

  # Prioris Uniformes

  alpha~dflat()
  beta~dflat()

  # Hiperprioris
  tau.delta~dgamma(0.001, 0.001)
  tau.u~dgamma(0.001,0.001)
}

```

**WinBugs - Modelo 3**

```

model
{
for (i in 1:m)
{
    for (k in 1:T)
    {
        y[i,k]~dpois(mu[i,k])
        log(mu[i,k])<-log(e[i,k])+u[i]+delta[i]*tt[k]
        theta[i,k]<-exp(u[i]+delta[i]*tt[k])
        tx[i,k]<- 15.63190*theta[i,k]
    }
}
for(k in 1:T){ time[k]<-exp(tt[k])}

u[1:m]~car.normal(adj[],weights[],num[],tau.u)
delta[1:m]~car.normal(adj[],weights[],num[],tau.delta)

tt[1]~dnorm(0,tau.t)
for(k in 2:T) {
tt[k] ~dnorm(tt[k-1],tau.t)
}
for(k in 1:sumNumNeig)
{
    weights[k]<-1
}

tau.delta~dgamma(0.001,0.001)
tau.t~dgamma(0.001,0.001)
tau.u~dgamma(0.001,0.001)
}

```

## Obtenção da matriz de vizinhanças – Software R

*Necessária a instalação dos pacotes SPDEP e Maptools*

```
x<-read.shape(system.file("shapes/municipio_05.shp", package="maptools")[1])
```

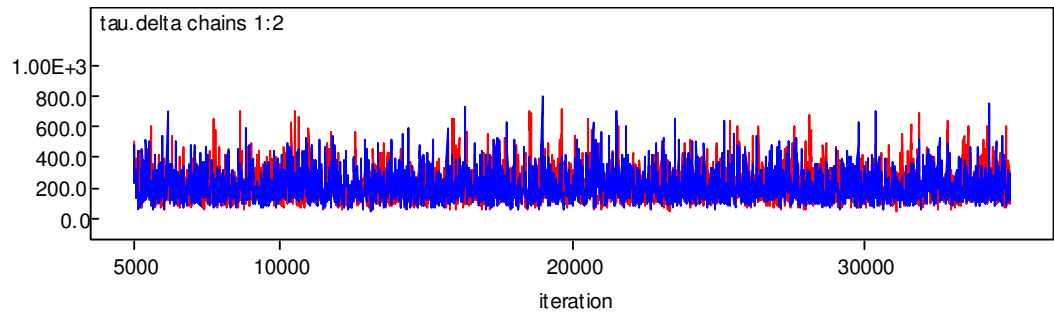
```
y<-Map2poly(x, as.character(x$att.data$NEIGNO), quiet=FALSE)
```

```
y1<-poly2nb(y, row.names = NULL, snap=sqrt(.Machine$double.eps),  
queen=TRUE)
```

```
write.nb.gal(y1,"c:/vizinhanca2005.txt")
```

## Gráficos de convergência e Distribuições a Posteriori

Monitoramento das duas cadeias utilizadas na estimação de  $\tau_\beta$



Distribuições a Posteriori das taxas de mortalidade para o município de Alvorada entre 2001 e 2004

