

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



**MAPEAMENTO DA MORTALIDADE INFANTIL NO RIO
GRANDE DO SUL: UMA COMPARAÇÃO ENTRE AS
ABORDAGENS EMPIRICA BAYESIANA E TOTALMENTE
BAYESIANA**

Sabrina Letícia Couto da Silva

Orientador: Prof^ª Dr^ª Jandyra Maria Guimarães Fachel

Porto Alegre, 2009.

S586m Silva, Sabrina Letícia Couto da

Mapeamento da mortalidade infantil no Rio Grande Do Sul : uma comparação entre as abordagens bayesiana empírica e totalmente bayesiana / Sabrina Letícia Couto da Silva ; orient. Jandyra Maria Guimarães Fachel. – 2009.

91 f. : il. color.

**Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul.
Faculdade de Medicina. Programa de Pós-Graduação em Epidemiologia.
Porto Alegre, BR-RS, 2009.**

1. Moralidade infantil 2. Rio Grande do Sul 3. Epidemiologia 4. Teorema de Bayes 5. Distribuição espacial da população I. Fachel, Jandyra Maria Guimarães II. Título.

NLM: WA 900

Catálogo Biblioteca FAMED/HCPA

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



DISSERTAÇÃO DE MESTRADO:
MAPEAMENTO DA MORTALIDADE INFANTIL NO RIO GRANDE
DO SUL: UMA COMPARAÇÃO ENTRE AS ABORDAGENS
EMPIRICA BAYESIANA E TOTALMENTE BAYESIANA

Sabrina Letícia Couto da Silva

Orientador: Prof^ª Dr^ª 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.
2009

BANCA EXAMINADORA

Prof^a Dr^a Luciana Neves Nunes, Instituto de Matemática, Departamento de Estatística, UFRGS.

Prof^a Dr^a Suzi Alves Camey, Programa de Pós Graduação em Epidemiologia, UFRGS.

Prof^a Dr^a Sídia Maria Callegari-Jacques, Instituto de Matemática, Departamento de Estatística, UFRGS.

Prof^a Dr^a Jandyra Maria Guimarães Fachel, Programa de Pós Graduação em Epidemiologia, UFRGS (Orientadora).

AGRADECIMENTOS

À minha família pelo amor, ensinamentos, carinho e por estarem sempre ao meu lado fazendo com que os momentos difíceis durante a realização desse projeto fossem atravessados mais facilmente e de modo menos doloroso. Em especial, aos meus pais Clóvis e Gelsa e irmãs Michele e Jaqueline, por fazerem parte da minha vida e por ser a melhor família que uma pessoa pode ter. Ao Fernando, por estar sempre do meu lado nas horas alegres, mas principalmente, nos momentos difíceis.

À professora Dr^a Jandyra Maria Guimarães Fachel pela orientação e pelos ensinamentos ao longo de minha trajetória acadêmica e profissional. Muito obrigada por sua paciência, pela oportunidade de orientação e pelas idéias sempre pertinentes e motivadoras. Às professoras da banca examinadora pelas correções e sugestões relevantes, as quais contribuíram muito para a melhoria do trabalho. À professora Dr^a Patrícia Klarmann Ziegelmann, por ser responsável pela minha paixão pela inferência Bayesiana. Foi quem me apresentou a essa área e me despertou o desejo de saber mais sobre o assunto.

Ao pessoal da CEVS, Carmen, minha querida colega Luciana Sehn, e principalmente, ao Rodrigo pela amizade e pelo auxílio na compreensão do funcionamento do software *TerraView*. Ainda, quero agradecer muito especialmente a Eleonora, da Seção da Saúde da Criança e do Adolescente do Departamento de Ações em Saúde da Secretaria Estadual da Saúde/RS: obrigada pelo apoio técnico, tuas contribuições foram muito relevantes para este projeto.

À minha querida “chefinha”, a Enf^a Laura Berquo: valeu pelas “dicas” e por ter me guiado e incentivado no caminho da Epidemiologia.

À minha estimada colega e amiga Simone Echeveste, Estatística e colega de trabalho na Universidade Luterana do Brasil (ULBRA): obrigada por tua amizade, companheirismo e por ser essa colega e amiga sensacional.

Aos grandes amigos e colegas da Estatística e da Matemática, pelos convites para churrascos e confraternizações em grupo, sempre necessários para aliviar as tensões resultantes da realização de um trabalho dessa magnitude. Aos colegas da Epidemiologia, obrigada pela amizade e pelos momentos alegres durante os intervalos de aulas e nos momentos de estudos em grupo; levarei a todos comigo para sempre! A todos os amigos de toda a vida, os responsáveis pela pessoa que eu sou e por sempre torcerem por mim.

SUMÁRIO

1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1 Mortalidade Infantil	3
2.2 Epidemiologia Espacial	5
2.2.1 Estudos de Correlação Geográfica	7
2.2.2 Mapeamento de Doenças	7
2.2.3 Detecção de Clusters de Doenças	8
2.3 Uso de mapas em Epidemiologia Espacial	8
2.4 Sistemas de Informações Geográficas e Estatística Espacial	10
2.5 Modelagem clássica para o Risco Relativo	11
2.6 Teoria Bayesiana	12
2.6.1 Inferência Bayesiana	13
2.6.2 Simulação estocástica via Cadeias de Markov	15
2.6.3 Amostrador de Gibbs (“ <i>Gibbs Sampler</i> ”)	16
2.6.4 Algoritmo de Metropolis – Hastings	17
2.7 Modelagem Bayesiana em Análise Espacial	18
2.7.1 Método de estimação Bayesiano Empírico	19
2.7.2 Método de estimação Totalmente Bayesiano	21
2.7.3 Comparação entre os métodos de estimação Bayesianos	23
3. REFERÊNCIAS BIBLIOGRÁFICAS	28
4. OBJETIVOS	35
4.1 Objetivo Geral	35
4.2 Objetivos Específicos	35
5. ARTIGO	36
6. CONSIDERAÇÕES FINAIS	62
ANEXOS	65
ANEXO 1 – PROJETO DE PESQUISA	66

1. INTRODUÇÃO	68
1.1 Mortalidade Infantil	68
1.2 Epidemiologia Espacial	69
1.3 Modelagem Clássica para o Risco Relativo	70
1.4 Inferência Bayesiana	72
1.5 Modelagem Bayesiana em Análise Espacial	72
1.6 Método de estimação Totalmente Bayesiano	73
1.7 Método de estimação Bayesiano Empírico	73
2. OBJETIVOS	76
3. METODOLOGIA	77
4. CRONOGRAMA	77
5. REFERÊNCIAS	78
ANEXO II – LINHAS DE COMANDO – WINBUGS	81
ANEXO III – MAPAS DOS RISCOS RELATIVOS	82

RESUMO

A taxa de mortalidade infantil (TMI) tem sido utilizada como um dos principais indicadores da qualidade de vida de uma população e reflete os níveis de saúde e de desenvolvimento sócio-econômico de uma determinada área, sendo considerado um dos mais importantes indicadores epidemiológicos. A análise da dispersão espacial do risco de ocorrência de um evento para dados agregados usualmente é feita via mapas de taxas de incidência, onde as áreas são sombreadas de acordo com os valores calculados para essa taxa. Um grande problema associado ao uso de taxas, porém, é a alta instabilidade que elas possuem para expressar o risco de eventos raros em regiões de população pequena. Alternativamente, existem os métodos de estatística espacial para mapeamento de doenças, denominados estimação Bayesiana Empírica e também estimação Totalmente Bayesiana, que utilizam informações de toda a região ou da vizinhança para estimar o risco de ocorrência do evento em cada área. O presente trabalho aplica e compara os dois métodos de estimação da TMI nos 496 municípios do Rio Grande do Sul através de dados acumulados entre os anos de 2001 a 2004 (dados disponíveis no DATASUS); aponta as vantagens de utilização dos estimadores Bayesianos em relação à taxa bruta e faz a comparação entre as estimativas obtidas através da modelagem bruta e dos métodos Bayesianos. Ao comparar as estimativas obtidas pelas modelagens Bayesianas com as obtidas pelo cálculo bruto, foi possível observar um ganho substancial na interpretação e na detecção de padrões de variação do risco de mortalidade infantil nos municípios do Rio Grande do Sul. Comparando-se os métodos Bayesianos observou-se que as estimativas calculadas através do método Bayesiano Empírico suavizam menos nas áreas de risco alto do que as estimativas Totalmente Bayesianas. O método Bayesiano Empírico pode ser utilizado para reduzir a variação observada na estimação através do método clássico e está implementado em diversos *softwares* de geoprocessamento e Epidemiologia Espacial, que podem ser facilmente utilizados por profissionais da área da Saúde. O método Totalmente Bayesiano, embora tenha também algumas propriedades importantes do ponto de vista estatístico, ainda tem alta complexidade computacional por demandar mais tempo nas análises.

ABSTRACT

The infant mortality rate (IMR) has been used as one of the main indicators of the quality of life of a population. It reflects the levels of health and socioeconomic development in a given area, and is considered one of the most important epidemiological indicators. The analysis of spatial dispersion of the risk of occurrence of an event for aggregate data is usually done by incidence rates maps, where the areas are shaded according to the values calculated for this rate. A major problem associated with the use of rates, however, is their high instability to express the risk of rare events in regions with a small population. Alternately, there are spatial statistical methods to map diseases, called Empirical Bayes estimate, and also Totally Bayesian estimate, which use information from the whole region or surroundings to estimate the risk of occurrence of the event in each area. The present study applies and compares the two methods for IMR estimates in the 496 municipalities of Rio Grande do Sul using data accumulated between 2001 and 2004 (data available in DATASUS); it indicates the advantages of using the Bayesian estimates compared to the gross rate and compares the estimates obtained by gross modeling and the Bayesian methods. When the estimates obtained by Bayesian modeling were compared to those of the gross calculation, a substantial gain could be observed in the interpretation and detection of patterns of variation of infant mortality risk in the municipalities of Rio Grande do Sul. Comparing the Bayesian methods, it is observed that the estimates calculated using the Empirical Bayes method smooth out less in the high risk areas than the Totally Bayesian estimates. The Empirical Bayes method can be used to reduce the variation observed in estimation through the classical method and is implemented in different Geoprocessing and Spatial Epidemiology softwares, which can be easily utilized by health care professionals. Although the Totally Bayesian method has a few important properties from the statistical perspective, it requires highly complex computations.

APRESENTAÇÃO

Este trabalho consiste na dissertação de mestrado intitulada “*Mapeamento da Mortalidade Infantil no Rio Grande do Sul: uma comparação entre as abordagens Bayesiana Empírica e Totalmente Bayesiana*”, apresentada ao Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul. O trabalho é apresentado em três partes, na ordem que segue:

1. Introdução, Revisão da Literatura e Objetivos
2. Artigo
3. Considerações Finais

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

1. INTRODUÇÃO

Os indicadores de saúde são utilizados pela Saúde Pública para avaliar as condições de vida de uma população. A mortalidade infantil é considerado um dos mais sensíveis indicadores e conhecer o seu perfil geográfico é fundamental para a formulação de estratégias que permitam seu controle.

A Epidemiologia Espacial, especialmente as técnicas denominadas de Mapeamento de Doenças tem por objetivo descrever e sintetizar a distribuição e a variação da taxas de mortalidade ou incidência de doenças ou eventos através de mapas. Mapas de taxas de incidências são bastante utilizados em estudos que visam analisar a dispersão espacial do risco de ocorrência de um determinado evento quando os dados estão dispostos em contagens por áreas. Mas o uso de taxas brutas nem sempre configura a melhor opção devido a um grande problema associado ao seu uso que é sua alta instabilidade quando se investiga o risco de eventos raros em pequenas áreas. Para solucionar este problema, novas metodologias baseadas em conceitos de inferência Bayesiana vêm sendo empregadas em estudos desse tipo. Dentre as alternativas existentes, destacam-se os métodos de suavização Bayesiana de taxas, que na sua essência utilizam informações de toda a região ou da vizinhança para estimar a taxa de ocorrência do evento em cada área.

O presente trabalho apresenta no Capítulo 2 a revisão da literatura sobre todos os conceitos a serem utilizados para o desenvolvimento do artigo. Inicia com uma revisão em relação a alguns estudos sobre Mortalidade Infantil, apresenta também uma revisão de conceitos de Epidemiologia Espacial e suas subdivisões e da utilização de mapas em Epidemiologia Espacial. Os principais conceitos de Inferência Bayesiana e de técnicas de suavização Bayesianas são também apresentados no Capítulo 2.

Este trabalho faz parte de uma seqüência de estudos sobre a análise da Mortalidade Infantil no Rio Grande do Sul sob a ótica da metodologia Bayesiana para Mapeamento de Doenças em Epidemiologia Espacial (ver Vieira, 2006 e Kato, 2007, 2009) e introduz resultados sobre o método Bayesiano Empírico, comparando-os com o Método Totalmente Bayesiano. O artigo apresenta a comparação entre os métodos para obtenção e visualização das taxas de mortalidade infantil: compara os resultados da taxa bruta com os resultados das abordagens Bayesiana Empírica (BE) e Totalmente Bayesiana (TB) para a Mortalidade

Infantil (MI) nos 496 municípios do Rio Grande do Sul, através de dados acumulados entre os anos de 2001 a 2004. O trabalho aponta as vantagens do uso dos estimadores Bayesianos na interpretação e visualização espacial dos mapas das taxas de mortalidade. Apesar do estimador Totalmente Bayesiano ter, teoricamente, o melhor desempenho, seu uso é bastante complexo devido à dificuldade de implementação computacional. Para os dados de Mortalidade Infantil no Rio Grande do Sul, os métodos Bayesianos Empíricos apresentaram resultados muito semelhantes àqueles apresentados pelos métodos Totalmente Bayesianos e, ainda, possuem a grande vantagem de serem de fácil utilização por profissionais da área de saúde, destacando igualmente os principais padrões espaciais da Taxa de Mortalidade Infantil no Rio Grande do Sul no período estudado.

2. REVISÃO DE LITERATURA

2.1 Mortalidade Infantil

O Sistema de Informações sobre Nascidos Vivos (SINASC) foi implantado no Brasil em 1990, com o objetivo de dar a conhecer o perfil epidemiológico dos nascimentos, segundo as variáveis de peso ao nascer, duração da gestação, tipo de parto, idade da mãe e número de partos, dentre outras, constantes na Declaração de Nascido Vivo (DNV). O Sistema de Informação de Mortalidade (SIM), por sua vez, já havia sido implantado no Brasil em 1975, tendo seus dados informatizados e disponíveis desde 1976, quando se tornou elemento essencial para o Sistema de Vigilância Epidemiológica Brasileiro. Suas informações vêm de antes da Declaração de Óbito (DO) (Rouquayrol & Almeida Filho, 2003).

A taxa de mortalidade infantil reflete, de modo geral, os níveis de saúde e de desenvolvimento sócio-econômico de uma determinada área, sendo considerado um dos mais importantes indicadores epidemiológicos utilizados internacionalmente (Monteiro & Nazário, 2000). Dessa forma, tem sido utilizada como um dos principais indicadores da qualidade de vida de uma população (Laurenti *et al.*, 1985; Maia & Souza, 2004) além de compor diversos indicadores sócio-econômicos.

Para a Saúde Pública é de fundamental importância que os indicadores calculados a partir do SINASC e do SIM reflitam o perfil epidemiológico da população. Dessa maneira, tem sido enorme o interesse em analisar suas informações, tanto sob o aspecto da cobertura do sistema de saúde quanto a respeito da qualidade e confiabilidade de suas informações (Theme Filha *et al.*, 2004).

A taxa de mortalidade infantil é calculada a partir da razão entre o número de óbitos ocorridos no período e o número de nascidos vivos naquele mesmo período e lugar, multiplicado por mil (Rouquayrol & Almeida Filho, 2003).

Segundo dados do Núcleo de Informações em Saúde (NIS) da Secretaria Estadual da Saúde do Rio Grande do Sul, a taxa de mortalidade infantil no Rio Grande do Sul para o ano de 2008 foi calculada em 12,8 óbitos por mil nascidos vivos. Para o Brasil, segundo dados da Secretaria Estadual de Vigilância em Saúde e IBGE, a taxa de mortalidade infantil

para o ano de 2007 foi estimada em 19,3 óbitos por mil nascidos vivos. Analisando os dados historicamente, percebe-se uma redução importante na Taxa de Mortalidade Infantil (TMI) no Estado e também uma redução no indicador brasileiro. Entretanto, as diferenças regionais acentuadas no território brasileiro, fazem com que esse processo seja mais lento do que o ocorrido no Rio Grande do Sul.

O declínio da mortalidade infantil tem sido atribuído a vários fatores, tais como implementação do Programa de Saúde da Família e Programa Pré-Natal. Outros fatores como avanços da medicina, expansão do acesso a cuidados de saúde, diminuição da taxa de fecundidade, aumento do nível educacional da população, melhoria nutricional e de condições de vida, de modo geral, já eram citados por Oliveira e Mendes, (1995).

As informações utilizadas em estudos ecológicos na área da saúde – como exemplo, a mortalidade infantil – estão disponíveis em bases oficiais como DATASUS, IBGE, entre outras. A qualidade dos registros oficiais no Brasil tenha melhorado nos últimos anos, então as pesquisas epidemiológicas oferecem inúmeras vantagens para um melhor entendimento dos fatores determinantes da mortalidade infantil. Estas incluem a maior fidedignidade dos dados obtidos, a possibilidade de verificar a qualidade da informação sobre a causa de morte e de evidenciar diferenciais conforme diversos fatores de risco (Menezes *et al*, 1996).

Diversos estudos latino-americanos evidenciam estreita relação entre mortalidade infantil e fatores sociais e econômicos (Behm, 1980; Monteiro, 1982; Paim *et al.*, 1987; Yunes, 1983). No Chile, por exemplo, Hollstein *et al.* (1998) identificaram um risco quase cinco vezes maior de mortalidade neonatal para filhos de mães sem instrução, inclusive para causas dificilmente evitáveis, como malformação congênita.

No Brasil também se aponta uma estreita relação entre mortalidade infantil e desigualdade social. Em estudo realizado na cidade de Salvador, Bahia, Costa *et al.* (2001) detectaram que a educação é a variável de maior correlação com a mortalidade infantil.

Shimakura *et al.* (2001) realizaram um estudo para o risco de mortalidade infantil na cidade de Porto Alegre, utilizando a metodologia de estatística espacial divulgada por Kelsall & Diggle (1998). Os autores objetivaram descrever a metodologia por ser inovadora e identificaram a forma como se distribuem os óbitos e seus determinadores de risco no território.

Define-se que a análise estatística é espacial quando os dados são espacialmente localizados e se considera explicitamente a possível importância de seu arranjo espacial na análise ou interpretação dos resultados (Bailey & Gatrell, 1995).

Andrade & Szwarcwald (2001) realizaram um estudo com o objetivo de estabelecer a distribuição espacial da mortalidade neonatal precoce no município do Rio de Janeiro, no período de 1995 a 1996. A metodologia utilizada incluiu as técnicas apresentadas por Cliff & Ord (1981), apropriadas para dados espaciais de área. A análise possibilitou a compreensão da distribuição geográfica da mortalidade neonatal, identificando os problemas relacionados à gravidez em adolescentes e à marginalidade social. Neste mesmo trabalho, comentou-se sobre estudos que têm sido realizados considerando estatística espacial, entre eles o de Sastry (1996) que estudou a mortalidade infantil no Brasil, buscando associação com o contexto ambiental e social em que a criança reside. Estes efeitos foram, em geral, medidos por atributos da mãe, como o nível de escolaridade.

Vieira (2006) e Kato (2007, 2009), em uma seqüência de trabalhos abordando a Mortalidade Infantil no Rio Grande do Sul, utilizaram o método de estimação Totalmente Bayesiano. Vieira (2006) utilizou a metodologia de análise Espaço-Temporal Bayesiana para estimação do risco de mortalidade infantil em cada município gaúcho usando dados de 2001 a 2004. A metodologia utiliza a informação *a priori* tanto dos municípios vizinhos, como de anos próximos. Kato (2007, 2009) realizou um estudo ecológico e utilizou alguns modelos de regressão espacial Bayesianos para estimação do risco relativo com a inclusão de covariáveis sócio-econômicas nos modelos, em dados acumulados de 2001 a 2004. O presente trabalho pertence a esta série de estudos sobre a metodologia Bayesiana para Mapeamento de Doenças em Epidemiologia Espacial e introduz resultados sobre o método Bayesiano Empírico, comparando-os com o Método Totalmente Bayesiano. Os estudos que objetivam associar a mortalidade infantil ou outro evento com a localização geográfica de maneira visual, através de mapas, estão classificados dentro da área da Epidemiologia Espacial.

2.2 Epidemiologia Espacial

O estudo da distribuição geográfica da incidência de eventos relacionados à saúde e da sua relação com fatores de risco potenciais vem ganhando bastante destaque através da

aplicação e desenvolvimento de métodos e modelos estatísticos. Nos últimos anos, foram desenvolvidas ferramentas cada vez mais robustas e de fácil manuseio nesta área de aplicação. Esse tipo de análise vem ganhando cada vez mais adeptos, devido aos avanços na integração de ferramentas do sistema de informações geográficas (SIG) e da análise espacial, em particular, da análise estatística espacial. A análise espacial, com a informação da localização dos dados, é uma técnica que busca descrever os padrões existentes nos dados espaciais e estabelecer, de forma quantitativa ou não, os relacionamentos das várias variáveis georreferenciadas (Nogueira, 2008). A ênfase deste tipo de análise seria a de mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita, incorporando o espaço à análise (Câmara & Monteiro, 2004).

A epidemiologia pode ser definida como o estudo dos fatores que determinam a ocorrência e a distribuição de doenças na população (Jekel *et al.*, 2005). Dessa maneira, apresenta recursos metodológicos que buscam explicitar a desigualdade destas distribuições da saúde nas sociedades, assim como subsidiar o planejamento das ações dos órgãos e serviços de saúde.

O uso de mapas e a preocupação com a distribuição geográfica de dados epidemiológicos são muito antigos. Destaca-se o trabalho pioneiro de John Snow, em 1854, o qual considerou a localização espacial da ocorrência da cólera para tentar explicar o fenômeno da epidemia ocorrida na cidade de Londres. Esse pesquisador utilizou técnicas de mapeamento para localizar os casos dessa doença e os pontos de coleta de água, evidenciando o papel das fontes na ocorrência da doença. A partir desse trabalho, vários outros foram desenvolvidos utilizando a distribuição espacial, descrevendo variabilidades geográficas na distribuição de dados epidemiológicos (Nogueira, 2008).

O crescimento do uso do SIG na área da saúde, bem como das técnicas de análise estatística espacial, tem proporcionado melhores resultados quanto à indicação das áreas de maior risco ou de maior urgência de intervenção da saúde pública.

Em estudos do tipo ecológico, os indicadores são taxas ou riscos. Nesses, a análise geográfica mais simples leva em consideração a taxa bruta denotada por T_i , que representa - para cada unidade espacial i - a razão entre a quantidade observada (doença/incidência) Y_i e a população exposta N_i , tornando possível comparar os resultados entre áreas. Porém,

muitas vezes, além da comparação entre diferentes áreas num mesmo mapa, faz-se necessária a comparação entre doenças/incidências diferentes para a mesma área, ou a comparação da mesma doença/incidência em períodos diferentes, e nestes casos o uso de medidas de risco relativo são mais adequadas.

De acordo com a natureza da distribuição espacial dos dados observados e de acordo com o objetivo da análise, a Epidemiologia Espacial está dividida em três grandes áreas: *Análise Ecológica, Mapeamento de Doenças e Detecção de Clusters de Doenças* (Lawson, 2001a).

2.2.1 Estudos de Correlação Ecológica

Os estudos de correlação ecológica (Elliott *et al.*, 2001) avaliam as relações entre a distribuição espacial da variável em estudo e fatores explicativos. Essas análises são realizadas em unidades espaciais agregadas. Esses modelos são usados, por exemplo, para relacionar eventos de saúde com a distribuição espacial de fatores ambientais, como exposição à radiação solar elevada, cuja exposição pode causar risco, com o intuito de controlar e prevenir maiores riscos à população. Os modelos estatísticos objetivam fazer a predição espacial ou espaço-temporal desses riscos.

2.2.2 Mapeamento de Doenças

Seu principal objetivo é descrever a distribuição geográfica da doença, gerar hipóteses, controlar áreas de risco elevado, estimar o risco da doença controlando suas oscilações e identificar cuidados especiais a serem tomados nas regiões mais afetadas. Para isso, utilizam-se modelos estatísticos que geram mapas sem perturbações externas, buscando-se descobrir a estrutura subjacente dos dados, ou seja, a verdadeira distribuição espacial da população em estudo.

O mapeamento de doenças serve para definir e monitorar epidemias, gerar dados basais para o controle de doenças e revelar mudanças no padrão das mesmas ao longo do tempo (D'Orsi & Carvalho, 1998). Este tópico tem uma proposta extremamente descritiva, fornecendo informações de saúde à população. Novos estudos podem ser contextualizados e o risco pode ser estimado através de um mapa (Fucks *et al.*, 2004).

2.2.3 Detecção de Clusters de Doenças

Avalia se a doença mapeada está formando um agrupamento de casos (*clusters*), em determinados locais ou se aglomerando, de modo geral. Essa formação pode ser monitorada ao longo do tempo (Lawson, 2001b). Há dois tipos de *clusters*: os globais e os locais. A formação de *cluster* global (em todo o mapa) tem como hipótese nula que a distribuição populacional é de forma aleatória, e, como alternativa, uma distribuição probabilística modificada pela existência de um agente (o contágio da doença está em questão). Já a formação de *clusters* locais está relacionada à ocorrência de algum fenômeno ambiental, social ou genético, ocasionando áreas geográficas com risco significativamente alto, e o seu acompanhamento é mais uma questão de vigilância epidemiológica.

2.3 Uso de Mapas em Epidemiologia Espacial

A utilização de mapas em estudos epidemiológicos tem crescido muito nos últimos anos com o advento de programas computacionais que possibilitem o georeferenciamento de áreas. A análise da dispersão espacial do risco de uma doença é feita através de mapas de taxas de incidência ou outra medida epidemiológica de risco.

O objetivo principal de um mapa é demonstrar a distribuição de um fenômeno ou dado no espaço. Através da ilustração dos padrões de distribuição de um evento, como por exemplo, a mortalidade infantil, mapas de incidência podem estimular a formulação de hipóteses etiológicas a serem posteriormente investigadas.

A maior parte dos mapas epidemiológicos é constituída por mapas temáticos, nos quais um conjunto de áreas é sombreado de acordo com seus valores para certa variável de interesse. Na maioria dos casos, as áreas de maiores taxas de incidência recebem uma coloração mais forte, enquanto as de menores taxas recebem uma coloração mais fraca.

Segundo Assunção *et al.* (1998), estes mapas são instrumentos importantes em estudos epidemiológicos e tem três objetivos principais. O primeiro é descritivo e consiste da simples visualização da distribuição espacial da doença na região de interesse. O segundo objetivo, exploratório, é sugerir determinantes locais de doenças e fatores etiológicos desconhecidos que possam ser avaliados em termos de hipóteses a serem investigadas posteriormente. Por fim, um terceiro objetivo é apontar as associações entre

fontes potenciais de contaminação e áreas de risco elevado, tal como o aumento de risco de leucemia infantil nas proximidades de instalações nucleares (Diggle *et al.*, 1990; Elliot *et al.*, 1992). Nesses casos existe uma suspeita prévia de que a vizinhança de um ou mais pontos pré-especificados possuem risco mais elevado e trata-se então de obter evidência da existência desse efeito.

Contudo, existe um grande problema associado ao uso de taxas para a construção de mapas: a alta instabilidade que as taxas brutas possuem para expressar o risco de um determinado evento quando ele é raro e a população da região de ocorrência é pequena. Flutuações aleatórias casuais, como a ocorrência de um ou dois casos de óbito a mais ou a menos numa localidade, causam grandes variações nas taxas brutas se o tamanho de sua população for pequeno, efeito este não verificado em localidades de população grande. Além disso, se não ocorrem óbitos em algumas regiões, a taxa bruta estima o risco de ocorrência do evento como zero, algo irreal tratando-se de dados na área da saúde. Mapas construídos com base nessas estimativas brutas são de difícil interpretação além de, freqüentemente, gerarem conclusões falsas.

O problema de instabilidade nas taxas vem sendo resolvido através de avanços nos métodos estatísticos e computacionais, destacando-se a inferência Bayesiana. Estes avanços possibilitam a confecção de mapas apresentando estimativas mais precisas dos indicadores de saúde, bem como a obtenção de modelos capazes de mensurar o efeito de covariáveis, que podem contribuir para formular hipóteses a respeito da distribuição espacial destes indicadores. O uso das técnicas Bayesianas contribui para a redução da variação, a aleatória e a natural, presente nas taxas brutas. Estas metodologias estimam taxas corrigidas a partir dos valores observados, utilizando conceitos de inferência bayesiana. Dois métodos podem ser citados: o método Bayesiano Empírico (Local e Global) e o método Totalmente Bayesiano.

Esse trabalho apresenta as metodologias Bayesianas citadas acima, calcula as estimativas para as taxas de mortalidade infantil dos 496 municípios do Estado do Rio Grande do Sul, no período acumulado entre os anos 2001 a 2004, e ainda faz a comparação entre as estimativas Bayesianas obtidas, através da visualização de mapas das taxas e via das Diferenças Quadráticas (DQ) entre metodologias.

2.4 Sistemas de Informações Geográficas e Estatística Espacial

Segundo Santos & Noronha (2001), a qualidade da informação de endereço e a eficiência do sistema de informação geográfico (SIG) para localizar os eventos pontualmente são fundamentais para possibilitar as análises dos padrões de distribuição dessas ocorrências.

A disponibilidade de bancos de dados em saúde com a informação do endereço e a estruturação do SIG em diversos municípios brasileiros tem possibilitado o uso crescente da análise de padrões espaciais de mortalidade para diferenciar áreas dentro dos municípios (Lima & Ximenes, 1998; Santos, 1999; Ximenes *et al.*, 1999).

Segundo Mendes (1996) e Teixeira *et al.* (1998), o conhecimento de regiões de maior incidência de casos, ou de maior risco, independentemente da distribuição da população, é fundamental para a adequação de recursos e o direcionamento de ações em saúde. Além disso, o ambiente de SIG dá margem à integração de informações oriundas de diversos setores, as quais poderão proporcionar visão mais abrangente da dinâmica dos processos sociais vigentes nas diferentes áreas delimitadas.

Santos (1999) realizou um estudo utilizando estatística espacial por área com o objetivo de detectar aglomerados espaciais de óbitos violentos na cidade de Porto Alegre. Esse autor usou a técnica de suavização de kernel (Bailey & Gatrell, 1995) para a criação da superfície de risco. Os resultados possibilitaram a identificação de áreas, nas quais a concentração de residências em precárias condições de saneamento foi elevada, relacionando-se com as principais causas externas de morte, sendo de potencial interesse e de grande importância para a implantação de políticas de promoção e de preservação da saúde da população.

Glass *et al.* (1995), produziram um mapa de risco de doença de Lyme a partir de dados epidemiológicos e de um sistema de informação geográfica. Mason (1995) apresenta diversos estudos de campo realizados como consequência das questões ligadas aos determinantes ambientais do câncer levantadas após análise dos diversos atlas editados pelo *National Cancer Institute* americano. Esses estudos abrangem câncer oral (Winn *et al.*, 1981), câncer do intestino (Pickle *et al.*, 1981), câncer de pulmão (Ziegler *et al.*, 1984) e câncer de bexiga (Hoover & Strasser, 1980). Como exemplo de trabalhos que objetivaram apontar associações entre fontes potenciais de contaminação e áreas de risco elevado,

podem ser citados os trabalhos de Diggle *et al.* (1990) e Elliot *et al.* (1992), cujos resultados mostraram que o aumento de risco de leucemia infantil está associado às proximidades de instalações nucleares.

2.5 Modelagem Clássica para o Risco Relativo

O mapeamento de doenças envolve, inicialmente, a escolha de uma medida epidemiológica para ser visualizada através de um mapa (Leyland & Davies, 2005). O mapa permite fazer comparações entre as áreas. Não faz sentido fazer mapas e comparações das contagens brutas já que elas dependem das populações de cada área. Para permitir comparações entre diferentes populações no espaço ou no tempo, as contagens devem ser padronizadas para gerar taxas livres do efeito dos tamanhos das diferentes populações. A medida mais frequentemente utilizada é o risco relativo.

Através do método de máxima verossimilhança, são atribuídas às áreas do mapa estimativas para os riscos relativos de interesse verdadeiros. Eles são estimados usando a SMR (*Standardised Mortality Ratio*) que é a razão entre a taxa de mortalidade da região i e a taxa na região toda. A taxa global de risco em toda a área estudada é dada por:

$$r = \frac{\sum y_i}{\sum N_i}$$

e a SMR da região i é dada por:

$$SMR_i = R_i = \frac{y_i}{e_i}$$

onde, y_i é a contagem de eventos na área i ; N_i é a população em risco na área i ; $e_i = r N_i$ é o número esperado de eventos na área i .

Se $SMR_i = 1$, a área i apresentou tantos casos observados quanto seria esperado caso seu risco fosse idêntico ao de toda área avaliada. A SMR_i varia de zero a infinito.

A variância da $SMR_i (y_i / e_i^2)$ é inversamente proporcional ao número esperado de casos na área i (e_i) e diretamente proporcional ao tamanho da população na área i , isto é, quanto menor o tamanho da população em risco, maior a incerteza na estimativa do risco (Olsen *et al.*, 1996). Essa incerteza causa instabilidade ao modelo, tornando difícil a interpretação dos mapas, e pode se tornar ainda maior em estudos com baixa incidência, por exemplo, mortalidade segundo causa específica. Existem na literatura três métodos alternativos para contornar esse problema: agregar áreas, usar mapas de probabilidade e usar modelagem Bayesiana. Atualmente as duas primeiras não estão sendo mais tão utilizadas em função das vantagens demonstradas pela modelagem Bayesiana (Assunção, 2001).

2.6 Teoria Bayesiana

A realização de inferências é fundamental em Estatística. Alguns dados são observados e deseja-se fazer inferência sobre uma ou mais características desconhecidas do sistema que teria dado origem a esses dados.

A inferência estatística trabalha na presença de observações y , cujos valores são descritos através de uma distribuição de probabilidade com um parâmetro θ . O pesquisador pode ter alguma informação prévia sobre θ e essa informação deve ser incorporada à análise na inferência Bayesiana.

Para Gelman *et al.* (1997), a inferência Bayesiana é o processo de ajustar um modelo de probabilidade para um grupo de dados e resumir o resultado por uma distribuição de probabilidade dos parâmetros do modelo e em quantidades não observáveis como previsões para novas observações.

A inferência Bayesiana é baseada no conceito de probabilidade subjetiva, que mede o grau de desconhecimento que alguém tem da ocorrência de um determinado evento do espaço amostral. Se existe uma incerteza, a descrição dessa incerteza pode ser feita utilizando-se a probabilidade subjetiva. Então, assim como na inferência Clássica, a análise Bayesiana descreve toda a quantidade desconhecida por meio da probabilidade.

A Inferência Estatística Bayesiana é o processo de diminuição de incerteza que se baseia em dados estatísticos e evidências pessoais.

Na Inferência Estatística Bayesiana Paramétrica, se supõe que a forma da distribuição de probabilidade de um vetor de variáveis aleatórias \mathbf{X} é conhecida e dependente de um parâmetro θ ou de um vetor de parâmetros $\boldsymbol{\theta}$. O parâmetro θ ou vetor de parâmetros $\boldsymbol{\theta}$ dessa distribuição são desconhecidos e por isso são considerados como variáveis aleatórias.

Toda a informação que se tenha sobre θ (ou $\boldsymbol{\theta}$) antes de observar os dados pode ser incluída na análise através da distribuição *a priori* para θ . As inferências a respeito dos parâmetros θ são baseadas na distribuição a posteriori de θ . A idéia central da Inferência Bayesiana é: após observar os dados, combiná-los com a informação *a priori* de θ e derivar a distribuição a posteriori para θ .

2.6.1 Inferência Bayesiana

O pesquisador, ao incorporar à análise uma opinião sobre θ por meio de uma densidade de probabilidade $P(\theta)$, está determinando a densidade *a priori*. A densidade *a priori* possui este nome por ser a distribuição de probabilidade de θ antes que se observem os dados amostrais. O parâmetro θ pode ser um escalar ou um vetor de parâmetros.

A densidade conjunta de um grupo de observações y_1, \dots, y_n , examinada como uma função do parâmetro é denominada função de verossimilhança e é representada por $L(y_1, \dots, y_n | \theta)$, sendo n o número de observações.

Se $P(\theta)$ é a densidade *a priori* para θ , então, a densidade a posteriori de θ é dada pelo Teorema de Bayes, ou seja,

$$P(\theta | Y) = \frac{L(Y | \theta)P(\theta)}{\int L(Y | \theta)P(\theta)d\theta}$$

sendo $Y = (y_1, y_2, \dots, y_n)$. Como o denominador não depende de θ , pode-se escrever:

$$P(\theta | Y) \propto L(Y | \theta)P(\theta)$$

onde, α representa “é proporcional a”. Essa expressão acima mostra o “jargão” da Inferência Estatística Bayesiana que diz: “A posteriori é proporcional ao produto da *priori* pela verossimilhança”.

Pode-se pensar no Teorema de Bayes como um mecanismo de atualização da opinião sobre θ . O Teorema de Bayes é um elemento essencial para a análise Bayesiana, pois toda a inferência é feita a partir da distribuição a posteriori.

Por muitos anos, o uso de *priori* subjetiva gerou uma grande controvérsia para a inferência estatística baseada no Teorema de Bayes, pois dependendo da informação *a priori*, a posteriori será muito influenciada por sua derivação depender diretamente da *priori*, principalmente, em um estudo com poucos dados. A escolha da distribuição *a priori* é um problema pertinente na abordagem Bayesiana. Existem dois tipos de *priori*: a *priori* informativa e a *priori* não informativa. Quando o pesquisador possui alguma informação sobre o que está estudando, ele pode usar uma *priori* informativa. De todos os aspectos da inferência Bayesiana, a *priori* informativa é a mais difícil e a que traz maior controvérsia. Por outro lado pode acontecer que o pesquisador tenha pouca ou nenhuma informação a incorporar *a priori*. Quando isso acontece, a distribuição considerada é a não informativa. Existe igualmente controvérsia quanto à *priori* não informativa, pois, muitas vezes, essa *priori* faz com que se tenha uma distribuição a posteriori imprópria. A distribuição *a priori* não informativa mais utilizada é a de Jeffreys (1939), que é invariante quando há reparametrização.

A representação probabilística de todo o conhecimento incerto é a essência da inferência Bayesiana, seja tal conhecimento relacionado ao futuro, a quantidades observáveis ou a parâmetros desconhecidos. Os conceitos de *priori* e *posteriori* são sempre relativos à observação considerada no momento.

Broemeling (1985) afirma que a base da inferência Bayesiana é a distribuição a posteriori de θ , pois, qualquer conclusão é feita a partir dessa distribuição.

A diferença formal entre a inferência Bayesiana e a inferência clássica, frequentista, é a de que na inferência Bayesiana, a incerteza sobre o parâmetro θ é tratada como uma variável aleatória, e para a inferência frequentista, os parâmetros são considerados valores fixos, não sendo possível atribuir a eles uma distribuição de probabilidade.

O'Hagan (1994) comenta o porquê de usar a inferência Bayesiana em vez da inferência clássica. Ele afirma que, em termos gerais, alguns dos argumentos a favor do enfoque Bayesiano são os que a análise Bayesiana é fundamentalmente completa, muito flexível, produz conclusões claras e diretas e faz uso de toda informação disponível. Em contraste, o enfoque clássico ignora a informação/ conhecimento *a priori*.

2.6.2 Simulação estocástica via Cadeias de Markov

O avanço da informática propiciou métodos cada vez melhores para se obter soluções aproximadas dos processos analíticos.

O termo simulação refere-se ao tratamento de problemas reais a partir de reproduções em ambientes ou sistemas controlados pelo pesquisador. Alguns problemas apresentam componentes aleatórios, os quais não podem ser descritos de forma exata e são baseados em informações probabilísticas. O processo de simulação é estocástico, isto é, baseado em distribuições de probabilidades.

As cadeias de Markov são processos estocásticos de tempo discreto, denominados a partir do matemático russo Andrey Andreyevich Markov. Por processos estocásticos entende-se um processo aleatório que ocorre segundo uma função randômica. A variável para esta função é, geralmente, atribuída a intervalos de tempo t . Cada variável $X(t)$, com $t = 0, 1, 2, 3, 4, \dots$, dessa função randômica, assume um valor também randômico, que é determinado por uma probabilidade. O resultado do lançamento de um dado não-viciado, por exemplo, pode assumir os seguintes valores através do tempo: $X(t) = \{1, 2, 3, 4, 5, 6\}$, ou os resultados do lançamento de uma moeda $X(t) = \{\text{Cara}, \text{Coroa}\}$.

Uma cadeia de Markov representa os vários estados possíveis para uma determinada situação e as transições, entre um estado e outro, que ocorrem segundo certa probabilidade.

Dos métodos de simulação que utilizam Cadeias de Markov podem ser citados: o Amostrador de Gibbs e o Metropolis-Hastings (Hastings, 1970). Recentemente, alguns autores (Christensen & Waagepetersen, 2002; Waagepetersen *et al.*, 2008) têm trabalhado com uma variação do algoritmo do Metropolis-Hastings. Este algoritmo é conhecido como Langevin-Hastings e proporciona uma convergência mais rápida na maioria dos casos, segundo Christensen *et al.*, (2001).

2.6.3 Amostrador de Gibbs (“Gibbs Sampler”)

O amostrador de Gibbs é, essencialmente, um esquema iterativo de amostragem de cadeias de Markov, cujo núcleo de transição é formado por distribuições condicionais completas. É uma técnica para gerar variáveis aleatórias de uma distribuição (marginal) sem que se conheça a sua densidade. A idéia do método de Monte Carlo via Cadeias de Markov (MCMC) é similar a um passeio aleatório no espaço do parâmetro θ , o qual converge para uma distribuição estacionária que, em estatística Bayesiana, é a distribuição a posteriori $P(\theta | Y)$.

O amostrador de Gibbs tem sido extremamente útil na resolução de problemas multidimensionais, ou seja, quando há um vetor de parâmetros desconhecidos θ e é definido em termos de subvetores de θ . Seja o vetor de parâmetros θ dividido em p subvetores $(\theta_1, \theta_2, \dots, \theta_p)$, supondo que as distribuições condicionais de cada subvetor (que pode ser um parâmetro ou um vetor desses) θ_i , dados todos os outros, sejam conhecidas. Essas distribuições são denotadas por $f(\theta_1 | \theta_2, \theta_3, \dots, \theta_p, Y)$, $f(\theta_2 | \theta_1, \theta_3, \dots, \theta_p, Y)$, ..., sendo Y o vetor de n observações. Essas distribuições são denominadas distribuições condicionais completas.

Para o amostrador de Gibbs é necessário que essas condicionais completas apresentem uma forma fechada, ou seja, uma forma conhecida de alguma distribuição. Caso contrário, há a necessidade de se utilizar outro amostrador, o Metropolis-Hastings.

Segundo Gamermann & Lopes (2006), o amostrador de Gibbs pode ser descrito da seguinte forma:

Passo 1 – assumir os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ para os parâmetros.

Passo 2 – gerando $m + n$ grupos de números aleatórios, retira-se iterativamente

$\theta_1^{(1)}$ de $f_1(\theta_1 | \theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y)$,

$\theta_2^{(1)}$ de $f_2(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y)$,

↓

$\theta_p^{(1)}$ de $f_p(\theta_p | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{p-1}^{(1)}, Y)$,

obtendo-se na primeira iteração $\theta_p^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{p-1}^{(1)}, Y)^T$.

Passo 3 – descartando-se as primeiras m realizações no passo 2, as n restantes realizações são utilizadas para formar uma amostra aleatória $\{(\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_{p-1}^{(i)})\}_{i=m+1}^{m+n}$ e estima-se a distribuição a posteriori usando essa amostra aleatória.

À medida que o número de iterações aumenta $n \rightarrow \infty$, a sequência se aproxima da sua condição de equilíbrio, ou seja, ela atinge a característica de estacionariedade com um comportamento assintótico. Assim, assume-se que a convergência é atingida em uma iteração cuja distribuição esteja arbitrariamente próxima da distribuição de equilíbrio, ou seja, a densidade marginal desejada, e não no sentido formal e inatingível de número de iterações tendendo para o infinito. Gamermann & Lopes (2006) comentam que, apesar dos resultados teóricos garantirem a convergência do amostrador de Gibbs, sua utilização, na prática, pode ser bastante complicada quando se têm modelos muito complexos. Essa complexidade faz com que a convergência do amostrador de Gibbs seja de difícil caracterização.

2.6.4 Algoritmo de Metropolis-Hastings

Este procedimento é usado em situações em que a distribuição a posteriori condicional completa não apresenta uma forma conhecida, fechada, sendo assim não identificável.

Seja $P(\theta | Y)$ uma distribuição não identificada. O algoritmo de Metropolis-Hastings gera seqüências de pontos aleatórios $(\theta_1, \theta_2, \theta_3, \dots)$ cuja distribuição converge para $P(\theta | Y)$.

O algoritmo é especificado por uma densidade $q(\theta, \beta)$ denominada de núcleo de transição ou densidade auxiliar. Essa densidade é selecionada de uma família de distribuições que requer a especificação de alguma característica do parâmetro, como escalar ou posição. O núcleo de transição representa a probabilidade de movimento de θ para β . Geralmente, o processo se move com mais freqüência de θ para β do que de β para θ , contradizendo a condição de reversibilidade $P(\theta | Y)q(\theta, \beta) = P(\theta | Y)q(\beta, \theta)$. Esse fato pode ser corrigido se diminuirmos o número de movimentos de θ para β . Para isto introduzimos uma probabilidade de movimento p .

Os passos do algoritmo podem ser descritos da seguinte forma:

Passo 1 – inicialize com um valor arbitrário para θ .

Passo 2 – gere um valor de β para $q(\theta, \cdot)$ e um valor u de uma Uniforme $(0, 1)$.

Passo 3 – seja $p = \min \{1, P(\beta | Y)q(\theta^{(j)}, \beta) / P(\theta^{(j)} | Y)q(\beta, \theta^{(j)})\}$.

Passo 4 – se $u \leq p$, faça $\theta^{(j+1)} = \theta^{(j)}$.

Passo 5 – repita os passos 2 e 3 até que a distribuição estacionária seja obtida.

O núcleo de transição q apenas define uma proposta de movimento, a qual deve (ou não) ser confirmada por p . Um problema crítico na eficiência do algoritmo é a escolha do fator escala ou dispersão da densidade q . O fator escala da densidade q afeta o comportamento da cadeia em pelo menos duas dimensões, sendo uma a taxa de aceitação e a outra é a região do espaço amostral que é coberta pela cadeia.

Uma densidade auxiliar que pode ser escolhida é a Normal Multivariada (NM), com a média dada pelo valor correto de θ e com matriz de covariância hI , sendo h uma variância dada pelo usuário ($h > 0$) e I uma matriz identidade, isto é, $q(\theta, \beta)$ tem densidade NM (θ, hI) . Em situações de alta dimensão o algoritmo pode convergir muito lentamente e produzir amostras altamente autocorrelacionadas (Christensen *et al.*, 2001). As características de convergência do processo também são válidas para o algoritmo de Metropolis-Hastings, sendo na maioria dos casos mais lenta.

2.7 Modelagem Bayesiana em Análise Espacial

Os estudos com dados de área foram os que primeiro se beneficiaram do uso de técnicas Bayesianas (Tsutakawa *et al.*, 1985), seguidos de perto pelos que usam dados espacialmente contínuos (Kitanidis, 1986) e processos pontuais (Akman & Raftery, 1986). Clayton & Kaldor (1987) propuseram incluir uma informação *a priori* espacial para levar em consideração a correlação entre as áreas vizinhas. Clayton (1989) propôs uma abordagem Bayesiana hierárquica que não apresentou as desvantagens da SMR.

Os princípios destes métodos foram abordados por Besag *et al.* (1981) que haviam proposto uma abordagem na qual os resultados são o produto de dois estágios. No primeiro estágio, o risco relativo de cada área é a soma de dois componentes representando fatores espaciais e não-espaciais. No segundo estágio, dado esse risco *a priori* e as populações de

risco, a contagem da característica de interesse em cada área segue uma distribuição de Poisson com média especificável. As distribuições *a priori* para os componentes do primeiro estágio podem ser escolhidas de forma que reflitam o conhecimento sobre o evento de interesse na área, incluindo até mesmo variáveis não-observadas, como em Bernadinelli *et al.* (1995a; 1995b). A correlação espacial seria expressa somente por aquelas variáveis com variabilidade de larga escala relativa, enquanto que as com variação de pequena escala, seriam restritas às pequenas áreas sob consideração e não teriam uma estrutura espacial.

Um pouco mais recente é a discussão de métodos Bayesianos aplicados à análise de dados de interação espacial (Banerjee *et al.*, 2000).

Dessa forma, os modelos Bayesianos para análise de dados de área tornaram-se mais populares, principalmente entre os pesquisadores que utilizam mapeamento de taxas. Citando alguns trabalhos, temos Heisterkamp & Doombos (1993), Smith *et al.* (1995), Reis & Assunção (1996), Rytkonen *et al.* (2001), Johnson (2004) e Yang *et al.* (2005).

Uma das razões para esta popularidade está no fato de que os métodos Bayesianos permitem a estimação de taxas para áreas pequenas sem a grande instabilidade presente nas estimativas clássicas. Esses métodos, ao estimarem o risco de uma área pequena, usam a informação das outras áreas (vizinhança) que compõem a região estudada e também de toda a área.

2.7.1 Método de estimação Bayesiano Empírico

A existência de flutuações aleatórias como a ocorrência de um ou dois casos do evento estudado, a mais ou a menos, numa localidade com população pequena, causa variações substanciais nas taxas brutas, efeito que não é verificado em localidades de população grande. Além disso, para situações em que não ocorrem casos do evento em algumas regiões na época estudada, a taxa bruta estima o risco de ocorrência do evento como zero, o que nem sempre é real no contexto de dados epidemiológicos. Mapas de eventos baseados diretamente em estimativas brutas são de difícil interpretação e frequentemente geram falsas conclusões.

Uma alternativa ao uso das taxas brutas em situações como essas - populações relativamente pequenas e eventos raros - foi proposta por Marshall (1991) através dos

estimadores Bayesianos Empíricos. Esses métodos Bayesianos permitem que a informação a respeito do risco contida nos dados das outras áreas possa ser usada para estimar o risco de determinada área. Denotando por $\hat{\theta}$ a estimativa Bayesiana Empírica de certa área, por t a taxa bruta dessa área e por m a taxa média de todas as áreas, podemos escrever,

$$\hat{\theta} = m + c(t - m).$$

O valor de c está entre 0 e 1 e representa o peso dado à diferença entre a taxa bruta da área e a taxa média de todas as áreas. O cálculo de c leva em conta o tamanho da população da área. Quando esta população for relativamente grande, o valor de c é próximo de 1, tornando a estimativa $\hat{\theta}$ muito próxima da taxa bruta, que, nessa área, não sofre o problema da instabilidade. No entanto, quando a população for pequena, o valor de c decresce, dando pouco peso à diferença entre t e m e tornando a estimativa $\hat{\theta}$ mais próxima da taxa média de todas as áreas. Ou seja, quando a população da área é pequena, a informação contida na taxa média tem peso maior na estimativa Bayesiana Empírica. A expressão para $\hat{\theta}$ também pode ser escrita como,

$$\hat{\theta} = ct + (1 - c)m$$

em que se nota que a estimativa $\hat{\theta}$ é a média entre a taxa bruta da área e a taxa média de todas as áreas, ponderada pelo valor c .

O estimador Bayesiano Empírico que utiliza como valor de m a taxa média de todas as áreas em estudo é denominado *estimador Bayesiano Empírico Global*. O estimador Bayesiano Empírico Global calcula uma média ponderada entre a taxa bruta da localidade e a taxa global da região (razão entre o número total de casos e a população total). Existe ainda o *estimador Bayesiano Empírico Local* que inclui o efeito espacial, calculando a estimativa localmente, utilizando somente os vizinhos geográficos da área na qual se deseja estimar a taxa: o valor obtido converge em direção a uma média local ao invés de uma média global. As taxas corrigidas são menos instáveis, pois levam em conta no seu cálculo não só a informação da área, mas também a informação de sua vizinhança.

Segundo Caumo (2006), a modelagem Bayesiana Empírica assume hiperparâmetros desconhecidos (parâmetros indexadores da distribuição *a priori*), porém utiliza estimativas pontuais de máxima verossimilhança para tais hiperparâmetros, como uma forma de aproximação. Esta modelagem, no entanto, pode subestimar a variabilidade dos riscos estimados (além de não lidar bem com a variabilidade das estimativas produzidas), essencialmente quando a priori condicionada aos hiperparâmetros apresentar alta dispersão. Além disso, não permite generalizações para situações mais complexas, como casos de análise espaço-temporal. Ou seja, a estimação Bayesiana Empírica ignora a variabilidade introduzida pelos hiperparâmetros justamente por não considerar a incerteza na estimação dos mesmos (considera-os fixos). A idéia básica consiste na utilização dos próprios dados da amostra para estimar os valores dos hiperparâmetros e então utilizar essas estimativas como valores fixos para os mesmos.

2.7.2 Método de estimação Totalmente Bayesiano

Para superar as dificuldades de instabilidade das estimativas e a autocorrelação espacial na variável resposta, várias modelagens têm sido propostas na literatura, porém as que apresentam melhores propriedades teóricas são os métodos de estimação Totalmente Bayesiano ou Inteiramente Bayesiano (Assunção, 2001). Estes modelos também são conhecidos na literatura como modelos Bayesianos hierárquicos espaciais (Bernadinelli & Montonolli, 1992). Best *et al.* (2005) ilustraram os modelos Bayesianos Hierárquicos mais recentes utilizados no contexto de mapeamento de doenças, com a aplicação de estimadores Totalmente Bayesianos. Em seu trabalho, Knorr-Held (2000) descreve quatro possíveis interações entre espaço e tempo que foram utilizadas para a construção de modelos espaço-temporais para o risco de doenças. Inferências são realizadas utilizando as técnicas MCMC (Markov Chain Monte Carlo) que auxiliam na geração da distribuição a posteriori dos parâmetros do modelo desejado. O amostrador de Gibbs (Gelman, 1997) representa uma solução inovadora para os problemas computacionais até então presentes em todas as áreas da modelagem Bayesiana. Entretanto, as vantagens do amostrador de Gibbs podem não superar a sua implementação computacional, uma vez que os métodos de simulação envolvidos são bastante complexos e soluções analíticas requerem profundo conhecimento computacional (linguagem de programação) e da teoria Bayesiana envolvida.

A interação espacial é usualmente modelada através da média de um modelo CAR (Autoregressivo Condicional) (Waller e Gotway, 2004), de modo que apenas os efeitos das áreas próximas são incluídos. De forma semelhante, o efeito temporal é modelado por um modelo AR (Autoregressivo) que mede a relação dos valores em uma área em diferentes tempos. Por fim, espaço e tempo podem interagir. Porém, ao considerar um ou ambos os efeitos simultaneamente, pode-se ter problemas de confundimento, gerando problemas na identificação dos efeitos, e futuras suposições podem ser necessárias para o uso da técnica.

Supõe-se que o número de eventos observados em cada área possui distribuição Binomial, mas como a maioria dos dados epidemiologicamente mapeados são raros ou com grande variação das taxas/riscos entre diferentes áreas, o modelo Binomial pode ser aproximado pela distribuição *Poisson* (Richardson *et al.*, 2004). Assim, se o número de eventos observados (Y_i) se refere ao número de óbitos no município i ,

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

onde $\mu_i = E_i \theta_i$, sendo θ_i , o risco relativo de óbito infantil na área i e E_i a quantidade esperada de óbitos infantis na i -ésima área sob a hipótese de que o risco seja constante em todas as áreas e igual ao risco geral da região. Sendo assim, o primeiro nível hierárquico do modelo é dado por:

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

No segundo nível do modelo são especificados os componentes utilizados na estimação do logaritmo do risco θ_i .

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

onde β_0 é uma constante, $\beta_p = (\beta_1, \beta_2, \dots, \beta_k)$ é um vetor de constantes que, quando colocados na função exponencial representam os efeitos de cada uma das k covariáveis no $\log(\theta_i)$, X_{ip} é um vetor de covariáveis e u_i e v_i são vetores de efeitos aleatórios.

O modelo é construído de modo a relacionar os componentes com o logaritmo de θ_i e não diretamente com o risco θ_i usando as inúmeras vantagens matemáticas e

computacionais que a transformação logarítmica pode proporcionar, pois a distribuição *Poisson* é uma distribuição de probabilidade enquadrada na família de distribuições Exponenciais.

Depois de obtidas as estimativas de $\log(\theta_i)$ se aplica a função exponencial para a obtenção das estimativas do risco θ_i .

No terceiro nível hierárquico do modelo estão as distribuições *a priori* para cada um dos parâmetros do modelo com β_0 , β_p , u_i e v_i . Em geral ao parâmetro β_0 se atribui uma distribuição *a priori* Uniforme $(-\infty; +\infty)$ e a cada um dos elementos do vetor β_p se atribui independentemente uma distribuição Normal com média igual a zero e parâmetro de dispersão τ_p o menor possível, de modo que estas distribuições *a priori* sejam pouco informativas.

A modelagem Totalmente Bayesiana considera que, por pressupor que os hiperparâmetros são quantidades aleatórias, distribuições hiperprioris devem ser especificadas aos mesmos. Ao estabelecer uma distribuição *a priori* para cada hiperparâmetro, melhores estimativas são produzidas para cada área, mais próximas do verdadeiro valor.

A literatura classifica o método Totalmente Bayesiano como preferível, por considerar toda a variabilidade através da imposição de distribuições de probabilidade para os hiperparâmetros (hiperprioris) e por possuir um processo inferencial mais rico e mais sofisticado, uma vez que podem ser calculados os intervalos de credibilidade para os parâmetros, ou seja, pode-se ter uma medida da precisão das estimativas produzidas (Bernadinelli & Montonolli, 1992). Por esses motivos, vários trabalhos mostram que a modelagem Totalmente Bayesiana é superior, do ponto de vista teórico, à metodologia de estimação Empírica Bayesiana. Os métodos Bayesianos, tanto o totalmente Bayesiano como o Empírico superam as desvantagens do método de agregação de áreas e dos mapas de probabilidades para resolver o problema das pequenas áreas, reduzindo consideravelmente o problema da variabilidade dos riscos estimados.

2.7.3 Comparação entre os métodos de estimação Bayesianos

Alguns estudos apresentam alternativas para a comparação do desempenho dos estimadores clássicos em relação aos estimadores baseados na teoria Bayesiana.

Com base em diversos outros trabalhos, Bernadinelli & Montonolli (1992) fizeram a comparação entre métodos Bayesianos Empíricos e métodos Totalmente Bayesianos na análise da variação geográfica de risco de doenças, utilizando o método MCMC nas análises Totalmente Bayesianas. Em seu trabalho, os autores apontam que o método Totalmente Bayesiano é preferível porque considera a incerteza dos parâmetros do modelo, enquanto que o método Empírico Bayesiano condiciona a estimação em estimativas pontuais dos parâmetros do modelo. Então, os estimadores Bayesianos Empíricos podem ser menos precisos. Por outro lado, os métodos Totalmente Bayesianos são computacionalmente mais intensivos. Se a análise exploratória indica que há uma alta variação de risco entre áreas, a metodologia Totalmente Bayesiana pode ser usada para detectar *clusters* de doenças, por exemplo. Esse seria um trabalho mais detalhado onde o investimento computacional seria conveniente (Bernadinelli & Montonolli, 1992).

O trabalho de Bernadinelli & Montonolli (1992) efetuou a comparação entre duas grandes classes de distribuições *a priori* para o risco relativo. Os mais simples modelos para a priori do risco são os ditos modelos de trocabilidade. Estes modelos determinam um valor médio, em torno do qual as estimativas específicas das áreas avaliadas são mais ou menos deslocadas e esse deslocamento depende da instabilidade intrínseca das estimativas e não da localização da área no mapa. Os modelos de priori mais complexos incorporam a estrutura geográfica do mapa. Esses modelos impõem uma estrutura de independência condicional para o conjunto de riscos relativos, pela qual cada risco relativo é condicionalmente independente de todos os outros, exceto por um conjunto pequeno de vizinhos geográficos. Em outras palavras, a estimativa do risco relativo em uma dada área é fortemente influenciada pela estimativa das áreas geográficas adjacentes, e apenas influenciada indiretamente pelas estimativas de todas as outras áreas do mapa. Como resultado, as estimativas individuais são mais direcionadas para um valor local do que para um valor médio global. Ainda, os autores descrevem a técnica de maximização do logaritmo da verossimilhança penalizada (PLM) empregada no método de estimação Bayesiano Empírico e o emprego do amostrador de Gibbs no método de estimação Totalmente Bayesiano.

A estrutura qualitativa dos relacionamentos entre as variáveis foi mostrada pelos autores através de um gráfico acíclico direto (DAG) mostrado abaixo:

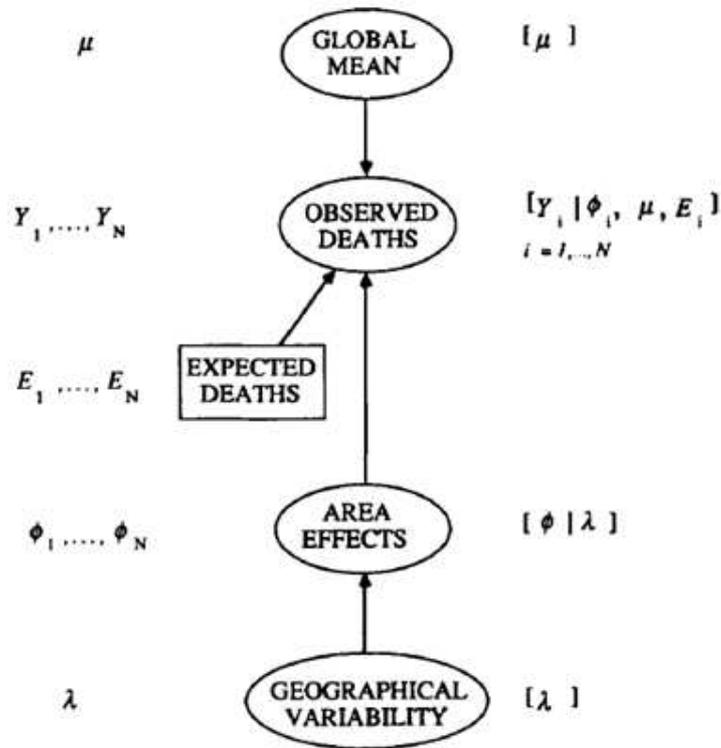


Figura 1. Gráfico acíclico direto (Bernardinelli & Montonolli, 1992)

Os nós desse gráfico representam as variáveis ou grupo de variáveis no modelo. Os nós elípticos representam as quantidades aleatórias, enquanto que os nós retangulares representam as quantidades fixas. Os nós Y_i e E_i são as quantidades conhecidas no modelo, enquanto que os nós μ , λ e ϕ_i são as quantidades desconhecidas a serem estimadas.

O DAG representa a suposição de que cada nó aleatório (variável), condicional a seus parentes, é probabilisticamente independente de todos seus predecessores no gráfico. Então, pela figura acima, o número de mortes observado Y é independente do parâmetro λ , uma vez que conheçamos os efeitos ϕ das áreas específicas. Ainda, a figura representa a suposição abaixo:

$$P(Y, E, \mu, \phi, \lambda) = P(Y | E, \mu, \phi)P(\phi | \lambda)P(\lambda)P(\mu)$$

As distribuições à direita da fórmula acima especificam a distribuição a posteriori conjunta completa $P(\varphi, \lambda, \mu | Y, E)$ para as quantidades desconhecidas do modelo. Mais precisamente, com base na estrutura de independência condicional implícita no gráfico acíclico direto (DAG), essa distribuição pode ser fatorada como o produto:

$$P(\varphi, \lambda, \mu | Y, E) = P(Y | E, \varphi, \mu)P(\varphi | \lambda)P(\lambda)P(\mu)$$

Na análise Totalmente Bayesiana, a derivação computacional da posteriori marginal $P(\varphi, \mu | Y, E)$ envolve integração em relação a λ , o qual é considerado um parâmetro complicado que torna essa integração analiticamente intratável.

A impossibilidade de realizar essa integração de forma analítica tem estimulado o interesse pela abordagem Bayesiana Empírica para o mesmo problema de estimação. A inferência Bayesiana Empírica a respeito do logaritmo do risco relativo desconhecido $\eta_i = \varphi_i + \mu$, não se baseia na verdadeira distribuição a posteriori $P(\varphi, \mu | Y, E)$, mas em sua aproximação $P(\varphi, \mu | Y, E, \hat{\lambda})$, a qual é condicional a uma estimativa adequada $\hat{\lambda}$ do parâmetro λ . De acordo com a estrutura de independência condicional do gráfico acíclico direto (DAG), $P(\varphi, \mu | Y, E, \hat{\lambda})$ produz uma expressão com o produto de termos conhecidos: uma constante de normalização c , a distribuição de verossimilhança dos dados $P(Y | E, \varphi, \mu)$, a priori de φ dado que $\lambda = \hat{\lambda}$, e a priori para μ . Como a média global do logaritmo do risco relativo (μ) segue uma distribuição Uniforme $(-\infty; +\infty)$, as prioris citadas poderão ser omitidas e o produto pode ser definido como:

$$P(\varphi, \mu | Y, E, \hat{\lambda}) = c P(Y | E, \varphi, \mu) P(\varphi | \hat{\lambda})$$

A suposição na análise Bayesiana Empírica é que a localização de $P(\varphi, \mu | Y, E, \hat{\lambda})$ é uma boa aproximação da localização de $P(\varphi, \mu | Y, E)$. Na prática, é conveniente selecionar a expressão de localização acima como uma localização estimada para φ e μ . Portanto, uma análise Bayesiana Empírica envolve: a) selecionar um valor aceitável de $\hat{\lambda}$ para substituir na expressão matemática acima; b) calcular os valores $\hat{\varphi}$ e $\hat{\mu}$ que maximizam a expressão

acima. A partir destes valores de estimativas pontuais, são derivados os logaritmos dos riscos relativos $\hat{\eta}_i = \hat{\phi}_i + \hat{\mu}$, $i = 1, \dots, N$, que serão interpretados através da técnica de maximização da verossimilhança penalizada.

Os estimadores Bayesianos Empíricos têm se mostrado bastante úteis, por exemplo, para fornecer estimativas iniciais do risco da doença no contexto dos Sistemas de Informações Geográfica (SIG) para epidemiologia espacial, onde grandes quantidades de dados estão disponíveis e os algoritmos usados devem ser rápidos do ponto de vista computacional. Os métodos Bayesianos Empíricos são muito mais fáceis de serem utilizados, pois já estão implementados em aplicativos de geoprocessamento existentes, como o TerraView, desenvolvido pelo INPE (Instituto Nacional de Pesquisas Espaciais). Além disso, sua vantagem de uso em relação aos métodos clássicos é bastante grande como abordado em Assunção *et al.*, 1998. Profissionais da área da saúde podem fazer uso do método Bayesiano Empírico em suas pesquisas sem os problemas de programação, o tempo despendido em simulações estocásticas e a complexidade que a metodologia Totalmente Bayesiana apresenta em sua utilização na estimação do risco de um evento na área da saúde.

O estudo de simulação conduzido por Richardson *et al.* (2004) apresentou cenários similares ao estudo de Richardson *et al.* (2003), apontando as similaridades e não similaridades entre os métodos.

Recentemente, Ugarte *et al.* (2009) conduziram um estudo de simulação utilizando a estrutura geográfica dos dados de câncer de lábios na Escócia através das abordagens Bayesiana Empírica e Totalmente Bayesiana. Fizeram então um comparativo entre os métodos e mostram que em casos nos quais se deseja identificar regiões de elevado (alto) risco, o excesso de suavização pode ser um problema porque pode camuflar regiões de risco potencial para o evento analisado. As estimativas obtidas através dos dois métodos foram muito similares. A técnica Bayesiana Empírica por utilizar a técnica de Verossimilhança Quasi-Penalizada (PQL) (Breslow e Clayton, 1993) é mais simples e requer menos recursos computacionais se comparados aos recursos necessários na abordagem Totalmente Bayesiana e por isto é de mais fácil utilização.

3. REFERÊNCIAS BIBLIOGRÁFICAS

Akman, V. E.; Raftery, A. E. (1986) *Bayes Factors for Nonhomogeneous Poisson Processes with Vague Prior Information*. **Journal of the Royal Statistical Society Series B-Methodological**, v. 48, p. 322 – 329.

Andrade, A. L. T.; Szwarcwald, C. L. (2001) Análise especial da mortalidade neonatal. **Caderno de Saúde Pública**, v. 17, p. 1199 – 1210.

Assunção, R. M. (2001) Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia. **Associação Brasileira de Estatística**. Disponível em: <http://www.est.ufmg.br/leste/publicacoes.htm>. Data de acesso: 28/01/2008.

Assunção, R. M.; Barreto, S. M.; Guerra, H. L.; Sakura, E. (1998) Mapas de taxas epidemiológicas: uma abordagem Bayesiana. **Caderno de Saúde Pública**, v. 14, p. 713 – 723.

Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S. (1999) *A National Facility for Small Area Disease Mapping and Rapid Initial Assessment of Apparent Disease Clusters Around a Point Source: The U.K. Small Area Health Statistics Unit*. **Journal of Public Health Medicine**, v. 21, p. 289 – 298.

Bailey, T. C.; Gatrell, A. C. (1995) *Interactive Spatial Data Analysis*. London: Longman.

Banerjee, S. , Gelfand, A. E. , Polasek, W. (2000) *Geostatistical modelling for spatial interaction data with application to postal service performance*. **Journal of Statistical Planning and Inference**, vol. 90, número 1, p. 87 – 105.

Behm, H. (1980) *Determinantes económicas y sociales de la mortalidade en América Latina*. **Revista Cubana de Salud**, v. 6, p. 1 – 30.

Bernadinelli, L.; Montonolli C. (1992) *Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk*. **Statistics in Medicine**; v. 11, p. 983 – 1007.

Bernadinelli, L.; Clayton, D.; Montonolli, C. (1995a) *Bayesian estimates of disease maps: how important are priors?* **Statistics in Medicine**, v. 14, p. 2411 – 2431.

Bernadinelli, L.; Clayton, D.; Pascutto, C.; Montonolli, C.; Ghislandi, M.; Songini, M. (1995b) *Bayesian analysis of space-time variation in disease risk*. **Statistics in Medicine**, v. 14, p. 2433 – 2443.

Besag, J.; York, J.; Mollié, A. (1981) *Bayesian image restoration, with applications in spatial statistics*. **Annals of the Institute of Statistical Mathematics**, v. 43, p. 1 – 59.

Best, N., Richardson, S. and Thomson, A. (2005) *A comparison of Bayesian spatial models for disease mapping*. **Statistical Methods in Medical Research**, v. 14, p. 35 – 59.

- Breslow, N. E., Clayton, D. G. (1993) *Approximate inference in generalized linear mixed models*. **Journal of the American Statistical Association**, v. 88, p. 9 – 25.
- Broemeling, L. D. (1985) *Bayesian analysis of linear models*. New York: M. Dekker, 454 p.
- Câmara, A. G.; Monteiro, A. M. V (2004) Conceitos básicos em ciência da geoinformação. Introdução à ciência da geoinformação. São Jose dos Campos, SP; INPE. Disponível em: <http://www.dpi.inpe.br/gilberto/livro/introd/>. Acesso em 22 de Fevereiro de 2008.
- Caumo, R. B. (2006) Estatística Espacial em Dados de Área: Uma Modelagem Inteiramente Bayesiana para o Mapeamento de Doenças Aplicada à Dados Relacionados com a Natalidade em Mulheres Jovens de Porto Alegre. **Monografia de Conclusão do Bacharelado em Estatística**. Instituto de Matemática, UFRGS.
- Clayton, D.; Kaldor, J. (1987) *Empirical Bayes Estimates of Age-Standardized Relative Risks for use in Disease Mapping*. **Biometrics**, v. 43, p. 671 – 681.
- Clayton, D. (1989) *Hierarchical Bayesian models in descriptive epidemiology*. **Proceedings of the XIVth International Biometrics Conference**, p. 201 – 213.
- Cliff, A. D.; Ord, J. K. (1981) *Spatial process model and applications*. London: Pion.
- Costa, M. C. N.; Azi, P. A.; Paim, J. S.; Silva, L. M. V. (2001) Mortalidade infantil e condições de vida: a reprodução das desigualdades sociais em saúde na década de 90. **Caderno de Saúde Pública**, v. 7, p. 555 – 567.
- Christensen, O. F.; Moller, J. (2001) *Geometric ergocity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models*. **Methodology and Computing in Applied Probability**, v. 3, p. 309 – 327.
- Christensen, O. F.; Waagepetersen, R. P. (2002) *Bayesian prediction of spatial count data using generalized linear mixed models*. **Biometrics**, v. 58, p. 280 – 286.
- Diggle, P. J.; Gatrell, A. C.; Lovett, A. A., (1990) *Modelling the prevalence of cancer of larynx in part of Lancashire: a new methodology for spatial epidemiology*. In: **Spatial Epidemiology** (R. W. Thomas, ed.), p. 153 – 171. London.
- D’Orsi, E.; Carvalho, M.S. (1998) *Birth profile for the city of Rio de Janeiro: a spatial analysis*. **Caderno de Saúde Pública**, vol. 14, p. 367 – 379.
- Elliott, P.; Hills, M.; Beresford, J.; Kleinschmidt, I.; Jolley, D.; Pattenden, S.; Rodrigues, L.; Westlake, A.; Rose, G. (1992) *Incidence of cancer of the larynx and lung cancer near incinerators of waste solvents and oil in Great Britain*. **The Lancet**, vol. 339: p. 854 – 858.

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

Fucks, S.D.; Carvalho, M.S., Câmara, G.; Monteiro, A.M. (2004) *Análise Espacial de Dados Geográficos*. 3ª. Edição, São Jose dos Campos, INPE: EMBRAPA, Cerrados.

Gamerman, D. (1997) *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall, London, 245 p.

Gamermann, D.; Lopes, H. F. (2006) *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. Illinois: Chapman and Hall, 344 p.

Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. (1997) *Bayesian data analysis*. London: Chapman and Hall, 526 p.

Glass, G. E.; Schwartz, B. S.; Morgan, J. M.; Johnson, D. T.; Noy, P. M.; Israel, E. (1995) *Environmental risk factors for Lyme disease identified with geographic information systems*. **American Journal of Public Health**, vol. 85, p. 944 – 948.

Hastings, W. K. (1970) *Monte Carlo sampling methods using Markov Chains and their applications*. **Biometrika**, London, v. 57, p. 97 – 109.

Heisterkamp S. H., Doornbos, G. (1993) *Disease mapping using empirical Bayes and Bayes methods on mortality statistics in the Netherlands*. **Statistics in Medicine**, v. 12, p. 1895 – 913.

Hollstein, R. D.; Veja, J.; Carvajal, Y. (1998) *Desigualdades sociales e salud. Nível socioeconômico y mortalidad infantil Chile, 1985 – 1995*. **Revista de Medicina do Chile**, v. 126, p. 333 – 340.

Hoover, R. N. ; Strasser, P. H. (1980) *Artificial sweeteners and human bladder cancer: preliminary results*. **The Lancet**, vol. 1:, p. 837 – 840.

Jeffreys, H. (1939) *Theory of Probability*. Oxford: Clarendon, 380 p.

Jekel, J. F.; Katz, D. L.; Elmore, J. G. (2005) *Epidemiologia, bioestatística e medicina preventiva*. 2ª edição. Porto Alegre: Artmed.

Johnson, G. (2004) *Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling*. **International Journal of Health Geographics**, v. 3, p. 29.

Kato, S. K. (2007) *Análise de Correlação Ecológica: uma abordagem Inteiramente Bayesiana para a Mortalidade Infantil no Rio Grande do Sul*. **Dissertação de Mestrado em Epidemiologia – Faculdade de Medicina, Universidade Federal do Rio Grande do Sul**.

Kato, S. K.; Fachel, J. M. G.; Vieira, D. M. (2009) Utilização da modelagem inteiramente bayesiana na detecção de padrões de variação de risco relativo de mortalidade infantil no Rio Grande do Sul, Brasil. *Cadernos de Saúde Pública*, 25(7), p. 1501 – 1510.

Kelsall, J. E.; Diggle, P. J. (1998) *Spatial variation in risk of disease: A nonparametric binary regression approach*. *Applied Statistics*, v. 47, p. 559 – 573.

Kitanidis, P. K. (1986) *Parameter Uncertainty in Estimation of Spatial Functions - Bayesian-Analysis*. *Water Resources Research*, v. 22, p. 499 – 507.

Knorr-Held, L. (2000) *Bayesian modelling of inseparable spacetime variation in disease risk*. *Statistics in Medicine*, v. 19, p. 2555 – 2567.

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

Lawson, A. (2001b) Disease map reconstruction. *Statistics in Medicine*, v. 20 (14), p. 2183 – 2204.

Laurenti, R.; Jorge, M.H.P.; Lebrão, M.L. e Gotlieb, S.L.D. (1985) *Estatísticas de Saúde*. São Paulo: EPU/EDUSP.

Leyland, A. H.; Davies, C. A. (2005) *Empirical Bayes methods for disease mapping*. **Statistical Methods in Medical Research** p. 14 – 17.

Lima, M. L. C. de; Ximenes, R. (1998) Violência e morte: diferenciais da mortalidade por causas externas no espaço urbano do Recife. *Caderno de Saúde Pública*, v. 14, p. 829 – 840.

Maia, S.F.; Sousa, T.R.V. (2004) Uma investigação dos determinantes da redução da taxa de mortalidade infantil nos Estados da Região Nordeste do Brasil. In: I Congresso da Associação Latino Americana de População - ALAP, 2004, Caxambu.

Marshall, R. J. (1991) *Mapping disease and mortality rates using empirical bayes estimators*. *Applied Statistics*, v. 40, p. 283 – 294.

Mason, T. J. (1995) *The development of the series of U.S. cancer atlases: implications for the future epidemiologic research*. *Statistics in Medicine*, vol. 14, p. 473 – 479.

Mendes, E. V. (1996) *Uma agenda para a saúde*, São Paulo: Hucitec.

Menezes, A. M. B., Victora, C. G., Barros, F. C. (1996) Mortalidade infantil em duas coortes de base populacional no Sul do Brasil: tendências e diferenciais. *Caderno de Saúde Pública*, v. 12, suppl.1, p. S79 - S86.

Monteiro, C. A. (1982) Contribuição para o estudo do significado do coeficiente de mortalidade infantil no Município de São Paulo, SP (Brasil), nas três últimas décadas (1950 – 1979). *Revista de Saúde Pública*, v. 16, p. 7 – 18.

Monteiro C. A., Nazário C. L. (2000) Evolução da agenda tradicional da saúde pública: Mortalidade infantil e materna. Declínio da mortalidade infantil e equidade social: o caso da cidade de São Paulo entre 1973 e 1993. In: Monteiro C. A. (org). Velhos e novos males da saúde no Brasil: a evolução do país e suas doenças. 2.ed. São Paulo: Hucitec; p. 173-85.

Nogueira, D. A. (2008) Análise espacial da mortalidade infantil no município de Alfenas, MG. *Tese de doutorado – Universidade Federal de Lavras*; 101p.

O'Hagan, A. (1994) *Kendall's and advanced theory of statistics: Bayesian inference* [S.I]: E. Arnold, v. 2b.

Oliveira, L. A. P.; Mendes, M. M. S. (1995) Mortalidade Infantil no Brasil: Uma avaliação de tendências recentes, In: Minayo, M. C. S. (org). Os muitos Brasis: saúde e população na década de 80. São Paulo; Hucitec; p. 291 – 303.

Olsen, S.; Martuzzi, M.; Elliott, P. (1996) *Cluster analysis and disease mapping - Why, when, how? A step by step guide. British Medical Journal*, v. 313, p. 863 – 865.

Paim, J. S.; Costa, M. C. N.; Cabral, V.; Mota, L. A.; Batista-Neves, R. B. (1987) *Spatial distribution of proportional infant mortality and certain socioeconomic variables in Salvador, Bahia, Brazil. Bulletin of the Pan-American Health Organization*, v. 21, p. 225 – 239.

Pickle, L. W.; Greene, M. H.; Ziegler, R. G.; Toledo, A.; Hoover, R.; Lynch, H. T.; Fraumeni Jr., J. F. (1981) *Colorectal cancer in rural Nebraska. Cancer Research*, vol. 44, p. 363 – 369.

Reis, E. A.; Assunção, R. M. (1996) Mapeamento da Malária em Rondônia usando o Estimador Empírico de Bayes. In: X Encontro Nacional de Estudos Populacionais, 1996, Belo Horizonte. *Anais do X Encontro Nacional de Estudos Populacionais*. Belo Horizonte: ABEP, v. 2. p. 1077-1097.

Richardson, S. (2003) Spatial models in epidemiological applications. In: P. Green, N. L. Hjort and S. Richardson (Eds) *Highly Structured Stochastic Systems*. p. 237 – 259. Oxford University Press.

Richardson, S.; Thomson, A.; Best, N.; Elliott, P. (2004) *Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives*; v. 122, p. 1016 – 1025.

Rouquayrol, M. Z.; Almeida Filho, N. (2003) *Epidemiologia e saúde*. Rio de Janeiro; MEDSI.

Rytkonen, M., J. Ranta, *et al.* (2001) *Bayesian analysis of geographical variation in the incidence of Type I diabetes in Finland. Diabetologia*, v. 44, B37-B44.

Santos, S. (1999) Análise da distribuição espacial dos homicídios no ano de 1996 em Porto Alegre, RS. *Dissertação de Mestrado em Epidemiologia – Escola Nacional de Saúde Pública*, Rio de Janeiro.

Santos, S. M.; Noronha, C. P. (2001) Padrões espaciais de mortalidade e diferenciais sócio-econômicos na cidade do Rio de Janeiro. *Caderno de Saúde Pública*, v. 17, p. 1099 – 1110.

Sastry, N. (1996), *Community characteristics individual and household attributes, and child survival in Brazil. Demography*, v. 33, p. 211 – 229.

Shimakura, S. F.; Carvalho, M. S.; Aerts, D. R. G. C.; Flores, R. (2001). Distribuição espacial do risco: modelagem da mortalidade infantil em Porto Alegre, RS, Brasil. *Caderno de Saúde Pública*, v. 17, p. 1251 – 1266.

Smith, T.; Charlwood, J. D.; Takken, W.; Tanner, M.; Spiegelhalter, D. J. (1995) *Mapping the densities of malaria vectors within a single village. Acta Tropica*, v. 59, p. 1 – 18.

Teixeira, C. F.; Paim, J. S.; Vilas Boas, A. L. (1998). SUS, modelos assistenciais e vigilância em saúde. *Informe Epidemiológico do SUS*, v. 7, p. 7 – 28.

Theme Filha, M. M.; Gama, S. G. N.; Cunha, C. B.; Leal, M. do C. (2004). Confiabilidade do Sistema de Informações sobre Nascidos Vivos Hospitalares no município do Rio de Janeiro, 1999 – 2001. *Caderno de Saúde Pública*, v. 20, n. 1, p. 83 – 91.

Tsutakawa, R. K. ; Shoop, G. L.; Marienfeld, C. J. (1985). *Empirical Bayes Estimation of Cancer Mortality Rates. Statistics in Medicine*, v. 4, p. 201 – 212.

Ugarte, M. D., Militino, A. F., Goicoa, T. (2008) Prediction error estimators in empirical Bayes disease mapping. *Environmetrics*, v. 19, p. 287 – 300.

Ugarte, M. D.; Goicoa, T; Militino, A. F. (2009) Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics and Data Analysis*, v. 53, p. 2938 – 2949.

Vieira, D. M. (2006). Análise Espaço - Temporal na Estimção de Taxas de Incidência/ Mortalidade. *Dissertação de Mestrado em Epidemiologia, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul*.

Waller, L. A.; Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, Hoboken, New Jersey.

Ximenes, R. A. A.; Martelli, C. M. T.; Souza, W.V. *et al.* (1999). Vigilância das doenças endêmicas em áreas urbanas: A interface entre mapas de setores censitários e indicadores de morbidade. *Caderno de Saúde Pública*, v. 15, p. 53 – 61.

Waagepetersen, R.; Ibanez, N., *et al.* (2008). A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. *Genetics, Selection, Evolution*, v. 40, p. 161 – 176.

Winn, D. M.; Blot, W. J.; Shy, C. M.; Pickle, L. W.; Toledo, A.; Fraumeni Jr., J. F. (1981). Snuff dipping and oral cancer among women in the southern United States. *New England Journal of Medicine*, vol. 304, p. 745 – 749.

Yang, G. J.; Vounatsou P.; Zhou, X. N. ; Tanner M. ; Utzinger, H. (2005). A Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu province, China. *International Journal for Parasitology*, v. 35, p. 155 – 162.

Yunes, J. C. (1983). Características sócio-econômicas da mortalidade infantil em São Paulo. *Pediatria*, v. 5, p. 162 – 168.

Ziegler, R. G.; Mason, T. J.; Stemhagen, A.; Hoover, R.; Schoenberg, J. B.; Virgo, P.; Waltman, R.; Fraumeni Jr., J. F. (1984). Dietary carotene and vitamin A and risk of lung cancer among white men in New Jersey. *Journal of the National Cancer Institute*, vol. 73, p. 1429 – 1435.

4. OBJETIVOS

4.1 Objetivo Geral

O objetivo deste trabalho é comparar os métodos de Estimção Espacial Bayesiano Empírico e Totalmente Bayesiano entre si e em relação ao estimador clássico (bruto) utilizado para a estimção de taxas na análise estatística espacial de mapeamento de doenças.

4.2 Objetivos Específicos

(1) Mostrar o efeito obtido com a aplicação dos estimadores Bayesianos para a suavização da flutuação aleatória nas taxas, quando se trabalha com pequenas populações e/ou eventos raros.

(2) Ilustrar os métodos estudados através da utilização dos dados de mortalidade infantil em 496 municípios do Rio Grande do Sul, para os dados acumulados de 2001 a 2004, além da comparação entre as estimativas obtidas para as taxas de mortalidade municipais através das modelagens Bayesiana Empírica (Local e Global) e Totalmente Bayesiana.

(3) Apresentar as estimativas das taxas através de mapas temáticos.

5. ARTIGO

Estimação e visualização dos padrões de variação das Taxas de Mortalidade Infantil no Rio Grande do Sul: uma comparação entre as modelagens Bayesiana Empírica e Totalmente Bayesiana

Sabrina Letícia Couto da Silva, mestranda do Programa de Pós-Graduação em Epidemiologia, UFRGS;

Sergio Kakuta Kato, Departamento de Estatística, Faculdade de Matemática, PUCRS;

Sergio Luiz Bassanesi, Programa de Pós-Graduação em Epidemiologia, UFRGS;

Jandyra Maria Guimarães Fachel, Programa de Pós-Graduação em Epidemiologia, UFRGS;
Departamento de Estatística, Instituto de Matemática, UFRGS

A ser enviado para: *Cadernos de Saúde Pública*.

Estimação e visualização dos padrões de variação das Taxas de Mortalidade Infantil no Rio Grande do Sul: uma comparação entre as modelagens Bayesiana Empírica e Totalmente Bayesiana

Sabrina Letícia Couto da Silva ¹

Jandyra Maria Guimarães Fachel ^{1,2}

Sergio Kakuta Kato ³

Sergio Luiz Bassanesi ¹

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
sassalelenash@yahoo.com.br

2 – Departamento de Estatística, Instituto de Matemática, UFRGS

3 – Departamento de Estatística, Faculdade de Matemática, PUCRS

Resumo

Os indicadores de saúde são utilizados pela Saúde Pública para avaliar as condições de vida de uma população. A mortalidade infantil é considerada um dos mais sensíveis indicadores e conhecer o seu perfil geográfico é fundamental para a formulação de estratégias que permitam seu controle. O Mapeamento de Doenças tem por objetivo descrever e sintetizar a distribuição e a variação das taxas de mortalidade ou incidência de doenças através de mapas. Mapas de taxas de incidências são bastante utilizados em estudos que visam analisar a dispersão espacial do risco de ocorrência de um determinado evento quando os dados estão dispostos em contagens por áreas. Mas o uso de taxas brutas nem sempre configura a melhor opção devido a um grande problema associado ao seu uso que é sua alta instabilidade quando se investiga o risco de eventos raros em pequenas áreas. Para solucionar este problema, novas metodologias baseadas em conceitos de inferência Bayesiana vêm sendo empregadas em estudos desse tipo. Dentre as alternativas existentes, destacam-se os métodos de suavização Bayesiana de taxas, que na sua essência utilizam informações de toda a região ou da vizinhança para estimar a taxa de ocorrência do evento em cada área. O presente artigo apresenta a comparação entre os métodos para obtenção e visualização das taxas de mortalidade infantil: taxa bruta, abordagens Bayesiana Empírica (BE) e Totalmente Bayesiana (TB), nos 496 municípios do Rio Grande do Sul, através de dados acumulados entre os anos de 2001 a 2004. O trabalho aponta as vantagens do uso dos estimadores Bayesianos na interpretação e visualização espacial dos mapas das taxas de mortalidade. Apesar do estimador Totalmente Bayesiano ter, teoricamente, o melhor desempenho, seu uso é bastante complexo devido à dificuldade de implementação computacional. Para o problema em estudo, os métodos Bayesianos Empíricos apresentaram resultados muito semelhantes àqueles apresentados pelos métodos Totalmente Bayesianos e ainda, possuem a grande vantagem de serem de fácil utilização por profissionais da área de saúde, destacando igualmente os principais padrões espaciais da Taxa de Mortalidade no Rio Grande do Sul no período estudado.

Palavras chave: Epidemiologia Espacial, Mapeamento de Doenças, Métodos Bayesianos, Mortalidade Infantil.

Abstract

Health indicators are used by Public Health to evaluate the living conditions of a population. Infant mortality is considered one of the most sensitive indicators and it is essential to know its geographic profile in order to formulate strategies to control it. The purpose of Mapping Diseases is to describe and summarize the distribution and variation of mortality rates or incidence of diseases using maps. Incidence rate maps are used a lot in studies with a view to analyzing the spatial dispersion of the risk of occurrence of a given event, when the data are disposed based on counts per areas. But it is not always best to use gross rates, due to a major problem associated with their use, which is their high instability when the risk of rare events in small areas is investigated. In order to solve this problem, new methodologies based on concepts of Bayesian inference have been used in studies of this kind. Outstanding among the existing alternatives are the methods of Bayesian smoothing of rates which, essentially use information from throughout the region or surroundings to estimate the risk of occurrence of the event in each area. The present article shows the method to calculate child mortality rates (crude rate, Empirical Bayes (EB) and Fully Bayesian (FB) approaches) in the 496 municipalities of Rio Grande do Sul, using data accumulated between 2001 and 2004. The estimates obtained are compared by means of the Mean Quadratic Error and by the visual analysis of the maps of estimated rates. The study indicates the advantages of the use of Bayesian estimators in interpreting and viewing the maps of the mortality rates. Although the Fully Bayesian estimator theoretically performs best, its use is rather complex due to the difficulty in computational implementation, therefore a good alternative would be to use the Empirical Bayes estimator, since this has already been implemented in several applications of spatial analysis and its results are much better compared to those of the gross analysis.

Key Words: Spatial Epidemiology: Disease Mapping, Bayesian Methods, Infant Mortality

Introdução

A modelagem estatística espacial tem sido bastante utilizada em diversos ramos da ciência nos quais os dados são coletados em diferentes locais, isto é, cada observação está associada a um ponto ou a uma região no espaço. Os dados podem ser pontuais, quando a exata localização de sua ocorrência é conhecida, ou contagens por área, quando os dados são agregados em áreas geograficamente definidas.

Quando os dados são contagens por áreas, o tamanho da população em risco pode variar consideravelmente entre as áreas sob estudo. Nesses casos, a dispersão espacial do risco de ocorrência de um evento (por exemplo, doença, morte, etc.) é usualmente feita via mapas de incidência de taxas, onde um conjunto de áreas estará sombreado de acordo com os valores da variável de interesse. O objetivo deste mapeamento é inferir a respeito da distribuição geográfica das taxas e então identificar as áreas de maior ou menor incidência. Quando as tonalidades no mapa tendem a mudar suavemente, pode-se chamar isso de padrão espacial. Ou seja, quando a autocorrelação está presente nos dados, regiões próximas tendem a ter um mesmo padrão de ocorrência do evento de interesse, regiões distantes terão padrões distintos. Um exemplo de aplicação do mapeamento de taxas seria o de mapear as taxas de mortalidade infantil do Estado a fim de determinar os padrões de variação das mesmas e identificar as regiões de maior risco de morte infantil.

A Taxa de Mortalidade Infantil reflete os níveis de saúde e de desenvolvimento sócio-econômico de uma determinada área, sendo considerado um dos mais importantes indicadores epidemiológicos utilizados (Monteiro & Nazário, 2000) e também tem sido utilizada como um dos principais indicadores da qualidade de vida de uma população (Laurenti *et al*, 1985; Maia & Souza, 2004). A taxa é dada pela relação entre o número de óbitos de crianças menores de um ano e o número de nascidos vivos, em determinado local e período, calculado na base de 1000 nascidos vivos. A mortalidade infantil é um importante indicador, porque reflete a qualidade com os cuidados pré e pós-natal das crianças, além de indicar uma preocupação na redução das disparidades existentes na pirâmide social. Conhecer o perfil da mortalidade infantil é fundamental para a formulação de estratégias que permitam seu controle, como uma assistência adequada à mulher durante a gravidez e o parto e, principalmente, um acompanhamento cuidadoso das mães em gestação considerada de risco. Segundo dados do Núcleo de Informações em Saúde (NIS)

da Secretaria Estadual da Saúde do Rio Grande do Sul, a taxa de mortalidade infantil no Rio Grande do Sul para o ano de 2008 foi calculada em 12,8 óbitos por mil nascidos vivos. Para o Brasil, segundo dados da Secretaria de Vigilância em Saúde e IBGE, a taxa de mortalidade infantil para o ano de 2007 foi estimada em 19,3 óbitos por mil nascidos vivos. Segundo Oliveira & Mendes (1995), analisando os dados historicamente, percebe-se uma redução significativa na TMI estadual e também uma redução no indicador brasileiro. Ainda, segundo os mesmos autores, o declínio da mortalidade infantil tem sido atribuído a vários fatores, tais como: intervenções ambientais, avanços da medicina, expansão do acesso a cuidados de saúde, diminuição da taxa de fecundidade, aumento do nível educacional da população e melhoria nutricional e das condições de vida de modo geral.

Dentre as alternativas conhecidas, cálculos baseados na taxa bruta são os mais simples para o estudo da ocorrência de um evento, sendo a taxa bruta definida como uma razão entre o número de eventos ocorridos e o número de pessoas expostas à ocorrência desse evento em uma determinada localidade. Entretanto, existe um grande problema relacionado ao uso de taxas brutas que é a sua alta instabilidade para medir o risco de um evento quando o evento é raro ou quando a população da região de ocorrência é pequena. Dessa forma, a alta variabilidade das taxas estimadas pode não estar diretamente relacionada com o fenômeno investigado, mas sim com uma variabilidade associada aos dados. Instabilidades como a ocorrência de um ou dois casos do evento a mais ou a menos numa localidade tem um impacto grande no cálculo das taxas brutas se a população de interesse for pequena, efeito que não se verifica em localidades de população de tamanho grande. Outro problema, em situações nas quais o número de eventos observados for igual a zero, a taxa bruta estima esse risco para o evento como sendo igual a zero, situação esta irreal tratando-se de dados na área da saúde.

Realizando a análise da distribuição espacial de eventos através de mapas construídos com base nas estimativas brutas, frequentemente, a interpretação se torna mais difícil e pode gerar conclusões errôneas. Para corrigir essa instabilidade das taxas brutas existem diversas técnicas para suavizar as taxas calculadas, como os métodos baseados na teoria Bayesiana que estão sendo utilizados com maior frequência em estudos epidemiológicos espaciais. Esses estudos objetivam mapear algum evento de interesse e estudar sua distribuição espacial, bem como analisar fatores que possam estar associados a

sua ocorrência no espaço-tempo. O uso das idéias Bayesianas na área epidemiológica pode ser exemplificado através do Nomograma de Fagan (Lotufo, 2005) que é uma ferramenta bastante utilizada no raciocínio diagnóstico.

Dentre as abordagens Bayesianas existentes para suavização de taxas existem as abordagens Bayesianas Empírica e Totalmente Bayesianas. O crescimento do uso dessas técnicas se deve ao grande avanço nos recursos computacionais disponíveis e também pela necessidade de se obter estimativas mais suavizadas, que reflitam de forma mais adequada a característica subjacente do evento de interesse, sem a influência da alta variabilidade observada nos dados brutos (Assunção, 2001).

A abordagem Totalmente Bayesianas utiliza informações *a priori* e através da informação das áreas vizinhas geograficamente geram distribuições a posteriori que serão usadas para as inferências sobre os parâmetros de interesse. Em geral, a distribuição a posteriori possui uma forma bastante complexa, e muitas vezes não é possível obter sua derivação via cálculos analíticos. Este problema é solucionado através do uso da metodologia de simulação via MCMC (Markov Chain Monte Carlo) (Gelman *et al.*, 1997). Este método é teoricamente preferível pois pode ser generalizado para modelos mais complexos, apesar do enorme esforço computacional necessário para sua aplicação. Alternativamente, há o método Bayesiano Empírico, que se destaca por utilizar os dados observados para estimar os parâmetros da priori.

Bernardinelli & Montonolli (1992) apresentaram um estudo comparativo entre as abordagens Totalmente Bayesianas e Bayesianas Empíricas. Os métodos Bayesianos Empíricos apresentaram resultados muito semelhantes àqueles apresentados pelos métodos Totalmente Bayesianos e ainda, possuem a grande vantagem de serem de fácil integração a ambientes de geoprocessamento. Os estimadores Bayesianos Empíricos têm se mostrado bastante úteis para fornecer padrões do risco da doença no contexto dos Sistemas de Informações Geográficas (SIG) para epidemiologia espacial, onde grandes quantidades de dados estão disponíveis e os algoritmos usados devem ser rápidos. Os métodos Bayesianos Empíricos são muito mais fáceis de serem utilizados do que os métodos totalmente Bayesianos e estão já implementados em aplicativos de geoprocessamento existentes, como o TerraView, desenvolvido pelo INPE (Instituto Nacional de Pesquisas Espaciais). Trata-se

de um programa livre com interface amigável (<http://www.dpi.inpe.br/terraview>). Os métodos Bayesianos apresentam vantagens em relação aos métodos clássicos segundo apontam Assunção *et al.*, 1998. Ugarte *et al.* (2009) fizeram um comparativo entre os métodos Bayesiano Empírico e Totalmente Bayesiano e mostram que em casos que se deseja identificar regiões de elevado risco, o excesso de suavização pode ser um problema por não detectar regiões de risco potencial para o evento analisado. Como as estimativas obtidas através dos dois métodos foram similares, a abordagem Bayesianiana Empírica seria conveniente porque utiliza a técnica de Quase-verossimilhança Penalizada (PQL) (Breslow e Clayton, 1993), que é considerada bastante simples e requer menos recursos computacionais, tanto no que diz respeito à capacidade computacional instalada, como no que diz respeito ao tempo computacional necessário. Segundo Richardson (2003), quando o risco subjacente é elevado e a população é não dispersa, a abordagem Bayesianiana Empírica é mais adequada para suavizar os riscos estimados e na detecção de áreas de alto risco.

O trabalho de Aimsworth e Dean (2006) efetuou a comparação entre as abordagens Bayesianiana Empírica – através do uso da técnica de Quase-verossimilhança Penalizada (PQL) – com a Totalmente Bayesianiana – através do emprego da metodologia MCMC. Os autores mostram que os resultados das duas técnicas foram bastante similares.

Vieira (2006) e Kato (2007, 2009) numa sequência de trabalhos abordando a mortalidade infantil no Rio Grande do Sul utilizaram o método de estimação Totalmente Bayesiano. Vieira (2006) utilizou a metodologia de análise Espaço-Temporal Bayesianiana para estimação do risco de mortalidade infantil em cada município gaúcho para dados de 2001 a 2004. A metodologia utiliza a informação *a priori* tanto dos municípios vizinhos, como de anos próximos. Kato (2007) utilizou alguns modelos de regressão espacial Bayesianos para estimação do risco relativo com a inclusão de covariáveis sócio-econômicas nos modelos para dados do mesmo período. O presente trabalho pertence a essa série de estudos sobre a metodologia Bayesianiana para Mapeamento de Doenças em Epidemiologia Espacial e introduz resultados obtidos através do método Bayesiano Empírico, comparando-os com o Método Totalmente Bayesiano.

O objetivo do artigo é comparar as estimativas para as taxas de mortalidade infantil nos municípios do Rio Grande do Sul, geradas pelos métodos: Clássico (Taxa Bruta), Bayesiano Empírico Local (BEL), Bayesiano Empírico Global (BEG) e Totalmente

Bayesiano (TB), utilizando os dados acumulados dos anos 2001 a 2004. A comparação será feita através da análise de gráficos de sequência para visualizar o comportamento das estimativas geradas por cada método e pelo cálculo das Diferenças Quadráticas (DQ) entre métodos. Finalmente, também foram gerados os mapas das taxas estimadas para identificação do padrão espacial subjacente dos dados.

Metodologia

Foi realizado um estudo epidemiológico ecológico, do tipo análise espacial em saúde. As unidades de análise são os 496 municípios do Estado do Rio Grande do Sul existentes no período de 2001 a 2004. As informações sobre o número de nascimentos e óbitos em recém-nascidos em cada município foram obtidas originalmente através do DATASUS e são as mesmas utilizadas por Vieira (2006) e Kato (2009) para fins de comparação dos métodos. A base de dados com o georreferenciamento dos municípios foi fornecida pelo Centro Estadual de Vigilância em Saúde (CEVS) da Secretaria Estadual de Saúde do Rio Grande do Sul possibilitando a construção dos mapas das taxas estimadas.

Foram calculadas as taxas de mortalidade infantil bruta acumulada para o período de 2001 a 2004. A taxa acumulada foi calculada utilizando o total dos óbitos dividida pelo total de nascidos vivos.

Segundo Marshall (1991), denotando por $\hat{\theta}$ a estimativa Bayesiana Empírica de certa área, por t a taxa bruta dessa área e por m a taxa média de todas as áreas, podemos escrever,

$$\hat{\theta} = m + c(t - m).$$

O valor de c está entre 0 e 1 e representa o peso dado à diferença entre a taxa bruta da área e a taxa média de todas as áreas. O cálculo de c leva em conta o tamanho da população da área. Quando esta população for relativamente grande, o valor de c é próximo de 1, tornando a estimativa $\hat{\theta}$ muito próxima da taxa bruta, que, nessa área, não sofre o problema da instabilidade. No entanto, quando a população for pequena, o valor de c decresce, dando pouco peso à diferença entre t e m e tornando a estimativa $\hat{\theta}$ mais próxima da taxa média de todas as áreas. Ou seja, quando a população da área é pequena, a informação contida na taxa média tem peso maior na estimativa Bayesiana Empírica. A expressão para $\hat{\theta}$ também pode ser escrita como,

$$\hat{\theta} = ct + (1 - c)m$$

em que se nota que a estimativa $\hat{\theta}$ é a média entre a taxa bruta da área e a taxa média de todas as áreas, ponderada pelo valor c .

O estimador Bayesiano Empírico que utiliza como valor de m a taxa média de todas as áreas em estudo é denominado estimador Bayesiano Empírico Global. O estimador Bayesiano Empírico Global calcula uma média ponderada entre a taxa bruta da localidade e a taxa global da região (razão entre o número total de casos e a população total). Existe ainda o estimador Bayesiano Empírico Local que inclui o efeito espacial, calculando a estimativa localmente, utilizando somente os vizinhos geográficos da área na qual se deseja estimar a taxa: o valor obtido converge em direção a uma média local ao invés de uma média global. As taxas corrigidas são menos instáveis, pois levam em conta no seu cálculo não só a informação da área, mas também a informação de sua vizinhança. Mapas baseados nessas estimativas são mais interpretativos e informativos (Santos *et al.*, 2005).

Na abordagem Totalmente Bayesiana, supõe-se que o número de eventos observados em cada área possui distribuição Binomial, mas como a maioria dos dados epidemiologicamente mapeados são raros ou com grande variação das taxas entre diferentes áreas, o modelo Binomial pode ser aproximado pela distribuição Poisson (Richardson *et al.*, 2004). Assim, se o número de eventos observados (Y_i) se refere ao número de óbitos no município i ,

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

onde $\mu_i = E_i \theta_i$, sendo θ_i , o risco relativo de óbito infantil na área i e E_i a quantidade esperada de óbitos infantis na i -ésima área sob a hipótese de que o risco seja constante em todas as áreas e igual ao risco geral da região. Sendo assim, o primeiro nível hierárquico do modelo é dado por:

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

No segundo nível do modelo são especificados os componentes utilizados na estimação do logaritmo do risco θ_i :

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

onde β_0 é uma constante, $\beta_p = (\beta_1, \beta_2, \dots, \beta_k)$ é um vetor de constantes que, quando colocados na função exponencial representam os efeitos de cada uma das k covariáveis no $\log(\theta_i)$, X_{ip} é um vetor de covariáveis e u_i e v_i são vetores de efeitos aleatórios.

O modelo é construído de modo a relacionar os componentes com o logaritmo de θ_i e não diretamente com o risco θ_i usando as inúmeras vantagens matemáticas e computacionais que a transformação logarítmica pode proporcionar, pois a distribuição Poisson é uma distribuição de probabilidade enquadrada na família de distribuições Exponenciais. Depois de obtidas as estimativas de $\log(\theta_i)$, aplica-se a função Exponencial para a obtenção das estimativas do risco θ_i .

No terceiro nível hierárquico do modelo estão as distribuições *a priori* para cada um dos parâmetros do modelo com β_0 , β_p , u_i e v_i . Em geral ao parâmetro β_0 se atribui uma distribuição *a priori* Uniforme $(-\infty; +\infty)$ e a cada um dos elementos do vetor β_p se atribui independentemente uma distribuição Normal com média igual a zero e parâmetro de dispersão τ_p o menor possível, de modo que estas distribuições *a priori* sejam pouco informativas.

A modelagem Totalmente Bayesiana considera que, por pressupor que os hiperparâmetros são quantidades aleatórias, distribuições hiperprioris devem ser especificadas aos mesmos. Ao estabelecer uma distribuição *a priori* para cada hiperparâmetro, melhores estimativas são produzidas para cada área, mais próximas do verdadeiro valor. As estimativas para o método Totalmente Bayesiana para a Mortalidade Infantil foram obtidas por Kato (2007) utilizando o programa WinBugs (gratuito e disponível para download em: www.mrc-bsu.cam.ac.uk/bugs).

No programa computacional TerraView 3.2.0 foram calculadas as estimativas Bayesianas Empíricas Globais e Locais para cada município, usando como base o número de óbitos esperado para cada município a partir das taxas padronizadas e a população do município. A matriz de vizinhança construída e utilizada nas análises feitas pelo TerraView 3.2.0 foi a do critério de contiguidade, ou seja, utiliza o valor 1 (um) para os municípios vizinhos e o valor 0 (zero) quando os municípios não tem fronteira adjacente. O programa também foi utilizado para construção de mapas temáticos para a visualização das taxas de mortalidade infantil no Estado do Rio Grande do Sul (os maiores coeficientes com tonalidades mais escuras e os menores com tonalidades mais claras). Foram construídos

quatro mapas: um mapa para as taxas brutas, outro para as estimativas Bayesianas Empíricas Globais (BEG), um terceiro para as estimativas Bayesianas Empíricas Locais (BEL) e o quarto para as estimativas Totalmente Bayesianas (TB).

Foi feita a análise para verificação da suposição de autocorrelação espacial através do cálculo do Índice de Moran (Moran, 1950) que é uma medida global da autocorrelação espacial e indica o grau de associação espacial presente no conjunto de dados analisados.

Foram calculadas as medidas de tendência central (média e mediana) e de dispersão (amplitude, desvio-padrão) para cada estimador (Taxa Bruta, BEL, BEG e TB) considerando os 496 municípios analisados. A fim de fazer a comparação entre os resultados obtidos em cada metodologia, foram calculadas as Diferenças Quadráticas (**DQ**) dadas por:

$$\mathbf{DQ} = (\mathbf{est}_1 - \mathbf{est}_2)^2,$$

onde **est₁** são as estimativas obtidas pelo primeiro método e **est₂** as estimativas geradas através do segundo método), comparando as abordagens duas a duas. Para efetuar as comparações entre as abordagens também foram gerados gráficos de sequência para visualizar o comportamento seqüencial das estimativas geradas, considerando como referência para todos os gráficos a ordem crescente das taxas brutas dos municípios. Por fim, as estimativas calculadas foram mapeadas para melhor visualização do padrão de mortalidade infantil no Estado, identificando regiões de baixas e altas taxas.

Resultados

O número de nascidos vivos por município no Rio Grande do Sul, no período analisado, variou entre 35 e 79.672 e o número de óbitos em menores de um ano variou entre 0 e 1072. A taxa de mortalidade infantil geral foi de 15,63 óbitos para cada mil nascidos vivos.

O índice de autocorrelação espacial de Moran foi igual a 0,897 ($p < 0,01$). Dessa forma, devemos considerar que existe uma correlação espacial significativa nos dados e a análise BEL torna-se mais adequada para fins de comparação com a TB (Martuzzi *et al.*, 1996).

Na Tabela 1, estão as estatísticas descritivas para cada método de estimação das taxas de mortalidade infantil no Rio Grande do Sul, no período acumulado do ano 2001 ao ano 2004.

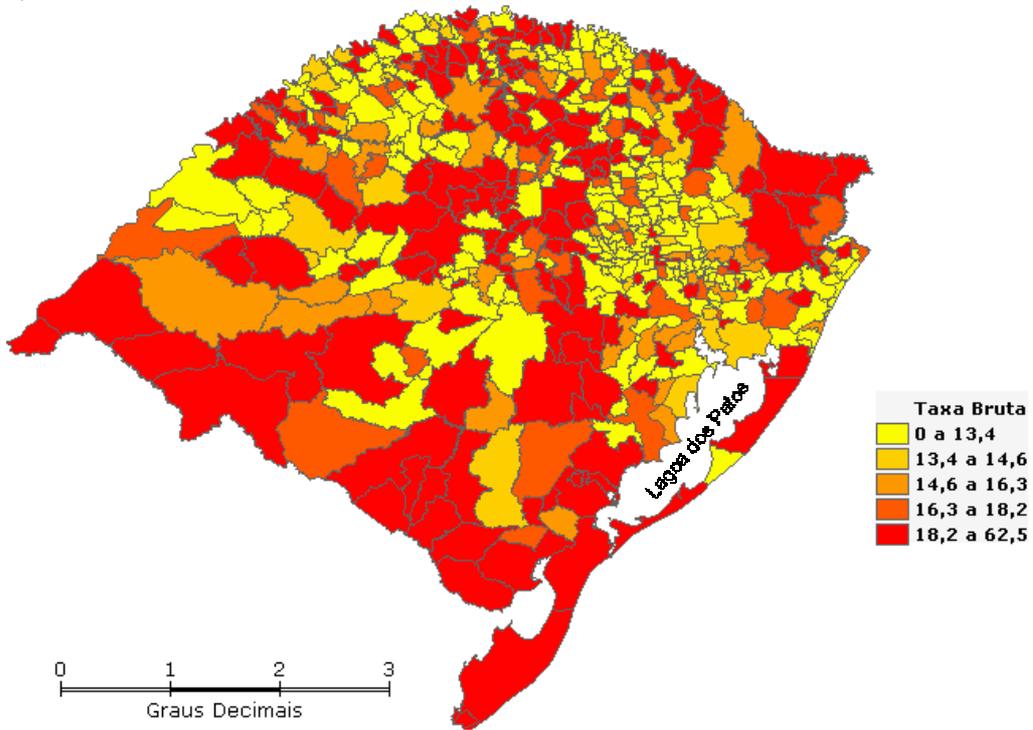
Tabela 1. Estatísticas Descritivas das taxas de mortalidade infantil no Rio Grande do Sul no período acumulado de 2001 a 2004, para os 496 municípios

Estimativa	Mínimo	Máximo	Mediana	Média	Desvio-padrão
Taxa Bruta	0,00	62,50	14,86	15,63	9,52
Taxa BEG	10,10	28,16	15,44	15,70	2,17
Taxa BEL	4,09	46,67	15,11	15,71	4,49
Taxa TB*	9,96	29,89	15,96	16,16	2,98

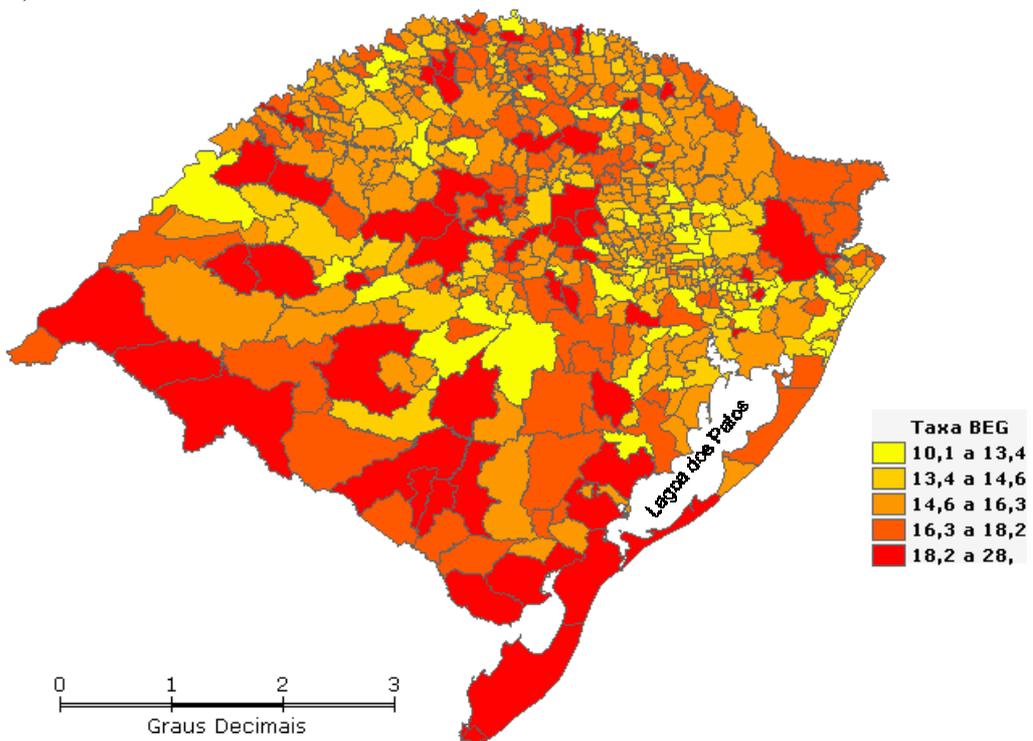
* obtida de Kato (2007).

Através dos resultados da Tabela 1 podemos verificar que todos os métodos possuem médias similares, porém podemos observar a alta variabilidade da taxa bruta, com um desvio-padrão de 9,52 óbitos por mil nascidos vivos, valor esse muito superior aos dos demais métodos. As estimativas geradas para a taxa de mortalidade infantil via método BEL (variação de 4,09 a 46,67) apresentaram a maior variação em torno da média dentre os métodos de suavização utilizados. O valor de 46,67 óbitos a cada mil nascidos vivos corresponde ao município de Cacique Doble que corresponde a uma zona de população indígena. A Figura 1 apresenta os mapas para visualização das taxas de mortalidade infantil para o Estado no período analisado, construídos no aplicativo TerraView 3.2.0, um para cada metodologia estudada (Taxa bruta, BEG, BEL e TB).

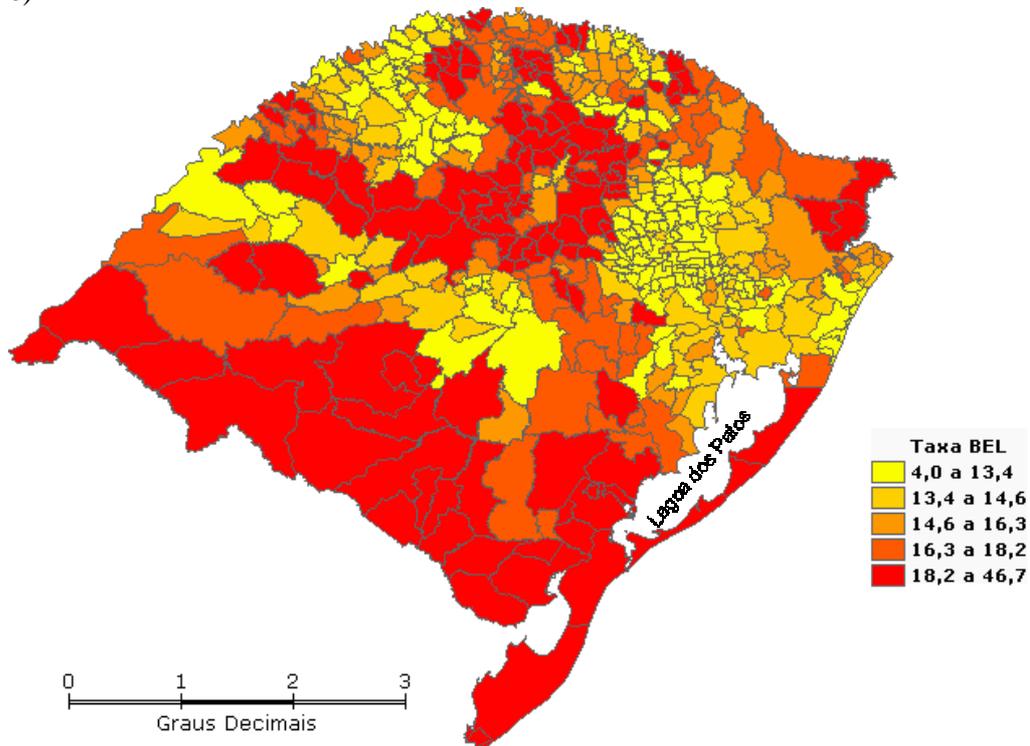
a)



b)



c)



d)

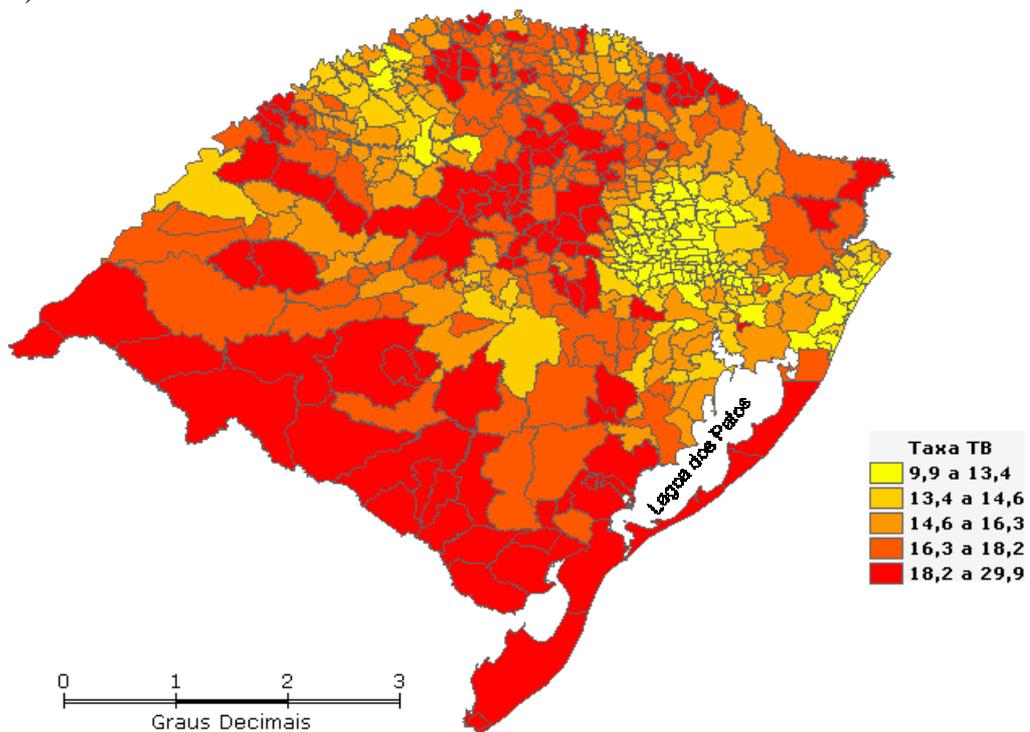


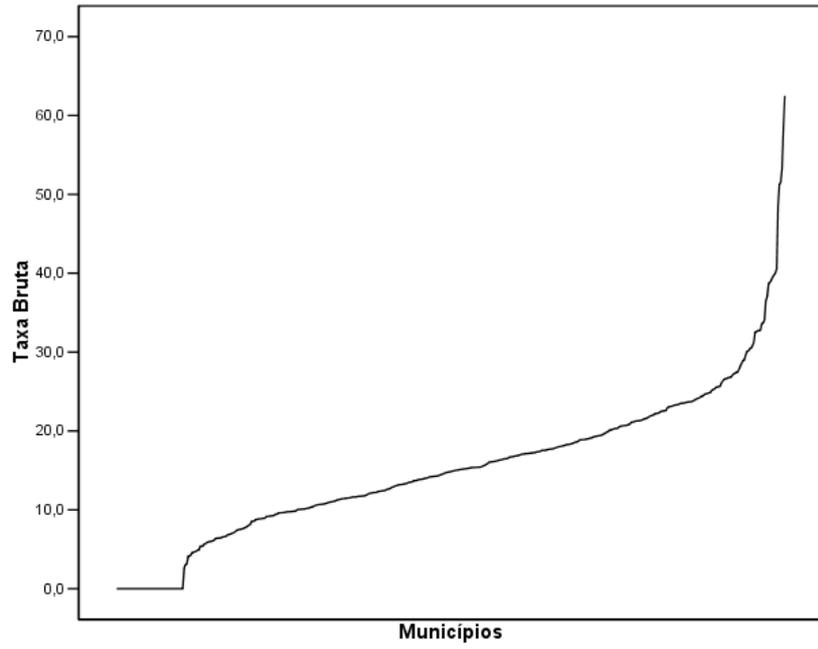
Figura 1. Taxa de Mortalidade Infantil no RS (2001 a 2004) – a) Taxa Bruta, b) Taxa BEG, c) Taxa BEL, d) Taxa TB.

Comprova-se através da visualização dos mapas na Figura 1, que a Taxa Bruta não é uma boa estratégia para expressar a distribuição espacial da taxa de mortalidade infantil, uma vez que o seu mapa apresenta uma aparência de “colcha de retalhos”, não evidenciando de forma clara um padrão espacial para a distribuição das taxas de mortalidade infantil em todo Estado. Assim como a Taxa Bruta, as estimativas Bayesianas Empíricas Globais também não evidenciaram no mapa um padrão espacial para a ocorrência do evento mortalidade infantil. Já as estimativas obtidas pelos métodos Bayesiano Empírico Local e Totalmente Bayesiano produziram mapas muito semelhantes entre si e mostrando claramente um padrão regional que destaca as baixas taxas de mortalidade infantil na Região da Serra do Rio Grande do Sul e as altas taxas na metade-sul do Estado do Rio Grande do Sul. Ainda, devemos destacar que os municípios que tinham estimativas iguais a zero para a taxa de mortalidade infantil bruta, após a suavização dos valores podemos perceber que os mapas não mais apresentam municípios com taxa de mortalidade infantil zero, mas sim, os municípios incorporam nas suas taxas estimadas o padrão subjacente da região na qual estão inseridos.

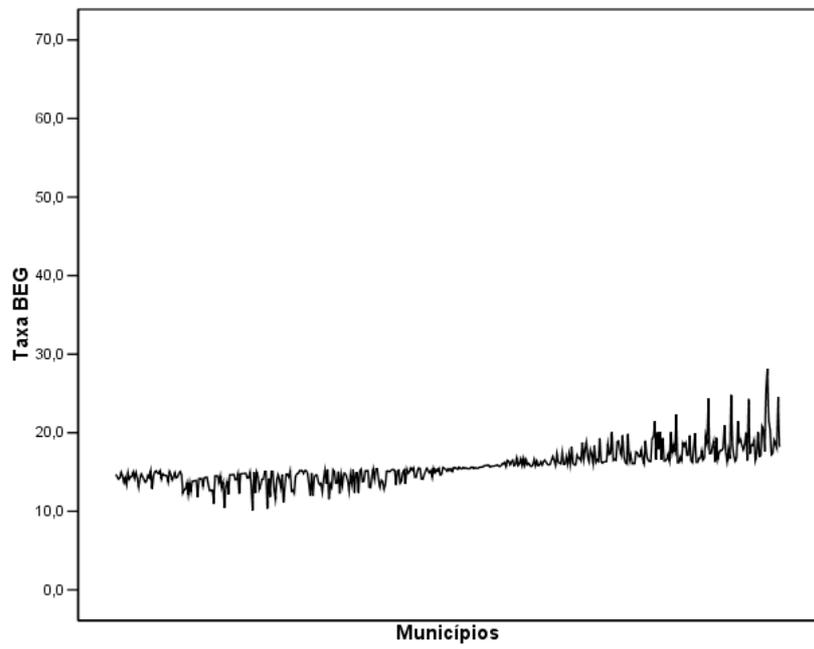
A Figura 2 apresenta os gráficos de sequência para os valores estimados para as taxas de mortalidade infantil – um gráfico para cada metodologia – que servem para comparar os valores estimados e assinalar semelhanças e diferenças entre as abordagens estudadas através da visualização do comportamento seqüencial das estimativas obtidas. Os valores foram ordenados segundo a ordem crescente dos valores das Taxas Brutas dos municípios.

Através da análise das Figuras 2c e Figura 2d, pode-se perceber que a abordagem Bayesiana Empírica Local fornece estimativas bastante similares aos resultados gerados via estimação Totalmente Bayesiana, uma vez que os gráficos seqüenciais estão muito semelhantes um ao outro. Observa-se um “nó” no gráfico das estimativas BEG, que se explica porque essa metodologia de estimação utiliza a taxa média geral (global) como referência.

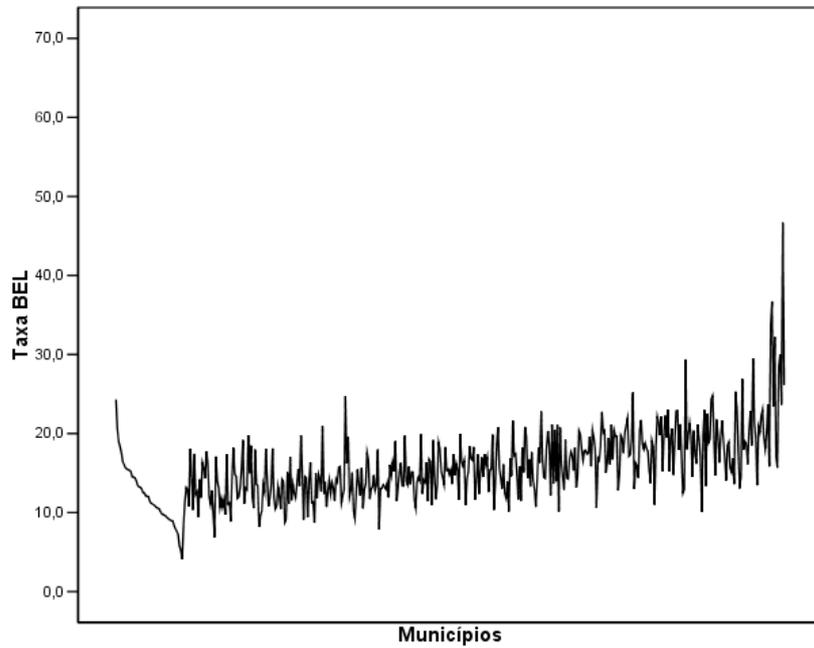
a)



b)



c)



d)

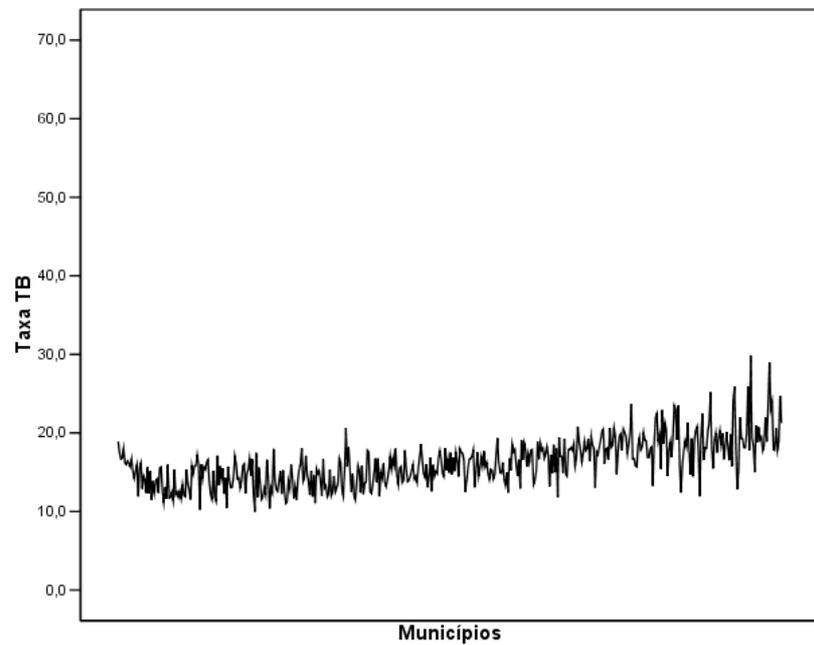


Figura 2. Taxas de Mortalidade Infantil estimadas por município: gráficos dos valores em seqüência dos municípios ordenados pela Taxa Bruta: a) Taxas Brutas, b) Taxas BEG, c) Taxas BEL, d) Taxas TB.

Tabela 2. Estatísticas descritivas para as Diferenças Quadráticas entre as estimativas Bayesianas para a taxa de mortalidade infantil no RS acumulada entre 2001 a 2004.

Diferença Quadrática	N	Mínimo	Máximo	Mediana	Média	Desvio-padrão
BEL versus BEG	496	0,00	487,67	3,11	11,17	28,34
TB* versus BEG	496	0,00	57,13	1,52	3,49	5,80
TB* versus BEL	496	0,00	480,90	1,47	6,25	25,07

* obtidas por Kato (2007).

Conforme os resultados da Tabela 2, o menor valor para a diferença quadrática média foi obtido na comparação entre o estimador BEG e o estimador TB. Na comparação entre os estimadores BEG e BEL, a diferença quadrática média é um valor bastante elevado, evidenciando a diferença entre, a) usar o método que valoriza mais a informação dos municípios vizinhos e dar menor peso à taxa média global (estimação Local) e b) usar o método que dá maior peso para a média global e menor importância para as taxas das regiões vizinhas (estimação Global). Através desses achados, podemos concluir que os métodos BEG e TB são mais similares, porém, na presença de autocorrelação espacial (situação presente nos dados) essa similaridade pode não ser de fato correta, sendo mais adequado utilizar os valores gerados através da abordagem BEL para comparação com os valores obtidos através do método TB. O valor máximo de 487,67 encontrado na diferença quadrática entre os métodos BEL e BEG corresponde ao município de Cacique Doble; o valor máximo de 57,13 encontrado na diferença quadrática entre os métodos TB e BEG corresponde ao município do Chuí e o valor máximo de 480,90 encontrado na diferença quadrática entre os métodos TB e BEL corresponde ao município de Cacique Doble.

Tabela 3. Dez municípios com maiores Taxas de Mortalidade Infantil pela estimativa de cada um dos métodos: BEG, BEL e TB

Município	Taxa BEG	Município	Taxa BEL	Município	Taxa TB
Redentora	28,16	Cacique Doble	46,67	São José do Norte	29,89
Fontoura Xavier	25,45	Redentora	36,70	Redentora	28,97
São Gabriel	24,81	Fontoura Xavier	33,71	Hulha Negra	25,92
Cacique Doble	24,58	Braga	32,24	Chuí	25,90
Uruguaiana	24,36	Vanini	29,99	Uruguaiana	25,18
São José do Norte	24,29	São José do Norte	29,53	Cacique Doble	24,73
Santana do Livramento	22,32	São José do Herval	29,37	São Gabriel	24,12
Manoel Viana	21,48	Engenho Velho	28,50	Braga	23,96
Bagé	21,47	Barros Cassal	26,89	Barra do Quaraí	23,70
Barros Cassal	21,47	Alto Alegre	26,14	Aceguá	23,49

Analisando a Tabela 3 podemos verificar em quais municípios se encontram as dez maiores taxas de mortalidade infantil segundo cada abordagem estudada. Podemos ver por que o município de Cacique Doble é o responsável pela maior diferença quadrática comparando os métodos BEL e BEG; as estimativas geradas pelos métodos são muito diferentes, sendo de 24,58 óbitos a cada mil nascidos vivos segundo a abordagem BEG e de 46,67 de acordo com a metodologia BEL.

Discussão e Conclusões

Os resultados corroboram a necessidade do uso de métodos de suavização para melhor representar a verdadeira taxa em cada município. Essa é a principal justificativa para o uso da metodologia Bayesiana de suavização das taxas para reduzir a alta variação que ocorre devido a flutuações aleatórias do cálculo bruto das taxas. A alta variabilidade nos valores calculados através da taxa bruta está relacionada ao tamanho da população sob risco e às instabilidades presentes na essência do indicador.

O município de Caciue Doble evidenciado como uma das áreas de alta mortalidade infantil pelos métodos estudados corresponde a uma área de população indígena, cujos recursos são mais escassos. A taxa de mortalidade infantil elevada pode ser explicada pela dificuldade de acesso do município a serviços na área da saúde bem como pelas baixas condições de desenvolvimento sócioeconômico próprias de regiões com essas características.

Os mapas da Figura 1c e da Figura 1d evidenciam regiões de altas taxas de mortalidade infantil ao sudoeste e centro-oeste do Estado, e ainda, ao norte também foram detectadas regiões de elevados valores para o indicador. A região centro-oeste do Estado é onde fica a sede da décima Coordenadoria Regional de Saúde (10^a CRS), que estava desprovida de UTI neonatal até Abril de 2009. Segundo informações da Secretaria Estadual de Saúde do Rio Grande do Sul, as elevadas taxas localizadas nessa região poderiam ser reflexo da ausência de leitos especializados para os casos de gestações e nascimentos de risco. Sendo assim, foram implantados no ano de 2009 dez leitos de UTI neonatal nessa região. Esse investimento foi demandado devido, basicamente, a óbitos por falta de acesso a leito de cuidado intensivo - óbitos neonatais, precoces, vinculados a prematuridade. A região possui sete dos trinta e cinco municípios prioritários para a redução do óbito infantil no RS. Ainda, algumas regiões do Estado que apresentam taxas de mortalidade infantil altas correspondem a municípios de área indígena ou de assentamentos onde há grande vulnerabilidade social, com a ocorrência de óbitos no período infantil tardio, vinculados a fatores de risco conhecidos e identificados nesta população devido à desorganização da rede de atenção básica de saúde, da falta ou pelo precário acompanhamento pré-natal e ausência de programas adequados de atenção hospitalar.

Existem diversos aplicativos utilizados para a análise estatística de dados espaciais que já possuem implementado o método de estimação Bayesiano Empírico, o qual pode ser empregado através de interfaces bem mais acessíveis e amigáveis se comparados com os procedimentos computacionais utilizados para o método de estimação Totalmente Bayesiano. No caso particular do programa TerraView, uma desvantagem é que esse não fornece uma medida da precisão das estimativas geradas, o software não calcula os Intervalos de Confiança (Ugarte *et al.*, 2008) para os riscos estimados. Já para as taxas Totalmente Bayesianas, estimadas com o auxílio do aplicativo WinBugs, pode-se obter o cálculo de intervalos de credibilidade para poder medir a precisão das estimativas geradas. Esse aplicativo requer um conhecimento mais apurado de sua linguagem, uma vez que a definição das distribuições *a priori* e outros passos da análise devem ser realizados mediante linhas de programação.

No estudo de Ugarte *et al.* (2009) também foi feito um comparativo entre os dois métodos Bayesianos e as estimativas obtidas foram muito similares. Os autores perceberam que em casos nos quais se desejam identificar regiões de elevado risco, o excesso de suavização produzida pelo método Totalmente Bayesiano pode ser um problema no sentido de não detectar regiões de potencial risco para o evento analisado.

A semelhança dos resultados dos métodos Empírico Bayesiano Local e Totalmente Bayesiano para as estimativas da Mortalidade Infantil no Rio Grande do Sul reproduz aqui outros achados semelhantes sobre a similaridade dos métodos. Nesse trabalho, assim como no trabalho de Bernadinelli e Montonolli (1992), o método Bayesiano Empírico Local apresentou resultados similares aos apresentados pela abordagem Totalmente Bayesiana.

Assim, na falta de recursos computacionais mais sofisticados, e também pela dificuldade de implementação e o tempo de computação despendido para geração de estimativas Totalmente Bayesianas, as estimativas calculadas através do programa TerraView são resultados que devem ser considerados úteis para os profissionais da área da saúde – em especial da Epidemiologia – para obtenção do padrão subjacente dos eventos estudados, salientando regiões com baixas ou altas taxas do evento de interesse.

A identificação das regiões de elevado risco de mortalidade infantil é um passo importante para uma melhor alocação dos recursos financeiros e humanos necessários para redução de tal indicador pelos gestores em saúde.

De acordo com os estudos citados e com nossos achados nesse trabalho, podemos sugerir o uso da metodologia Bayesiana Empírica Local já que as estimativas para a taxa de mortalidade infantil, obtidas pelo método Bayesiano Empírico Local, não se apresentaram muito diferentes das estimativas obtidas por Kato (2007, 2009) utilizando o método Totalmente Bayesiano, e ainda, com a grande vantagem de minimizar o tempo despendido nas análises, tornando-as mais acessíveis aos pesquisadores da área da saúde (secretarias estaduais, municipais, etc.), sendo esses os maiores interessados na melhor gestão da saúde, particularmente quando necessitam administrar os recursos de forma mais eficaz para se reduzir a taxa de mortalidade infantil.

Referências Bibliográficas

Aimsworth, L. M.; Dean, C. B. (2006) Approximate inference for disease mapping. *Computational Statistics & Data Analysis*, v. 50, p. 2552 – 2570.

Assunção, R. M. (2001). *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*. Associação Brasileira de Estatística. Disponível em: <http://www.est.ufmg.br/leste/publicacoes.htm>. Data de acesso: 28/01/2008.

Assunção, R. M.; Barreto, S. M.; Guerra, H. L.; Sakura, E. (1998) Mapas de taxas epidemiológicas: uma abordagem Bayesiana. *Caderno de Saúde Pública*, v. 14, p. 713 – 723.

Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., et al. (1999) A National Facility for Small Area Disease Mapping and Rapid Initial Assessment of Apparent Disease Clusters Around a Point Source: The U.K. Small Area Health Statistics Unit. *Journal of Public Health Medicine*, v. 21, p. 289 – 298.

Bernadinelli, L.; Montonolli, C. (1992) Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine*; v. 11, p. 983 – 1007.

Breslow, N. E., Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, v. 88, p. 9 – 25.

Gelman, A.; Carlin, J. B.; Stern, H. S.; Rubin, D. B. (1997) *Bayesian data analysis*. London: Chapman and Hall, 526 p.

Kato, S. K. (2007). *Análise de Correlação Ecológica: uma abordagem Inteiramente Bayesiana para a Mortalidade Infantil no Rio Grande do Sul*. Dissertação de Mestrado em Epidemiologia – Faculdade de Medicina, Universidade Federal do Rio Grande do Sul.

Kato, S. K.; Fachel, J. M. G.; Vieira, D. M. (2009) Utilização da modelagem inteiramente bayesiana na detecção de padrões de variação de risco relativo de mortalidade infantil no Rio Grande do Sul, Brasil. *Cadernos de Saúde Pública*, Rio de Janeiro, 25(7), p. 1501 – 1510.

Laurenti, R.; Jorge, M. H. P.; Lebrão, M. L.; Gotlieb, S. L. D. (1985) *Estatísticas de Saúde*. São Paulo: EPU/EDUSP.

Lotufo, P. A. (2005) *Epidemiologia: Abordagem Prática*. Editora Sarvier, p. 231.

Maia, S. F.; Sousa, T. R. V. (2004) Uma investigação dos determinantes da redução da taxa de mortalidade infantil nos Estados da Região Nordeste do Brasil. In: *I Congresso da Associação Latino Americana de População - ALAP, 2004, Caxambu*.

Marshall, R. J. (1991) Mapping disease and mortality rates using empirical Bayes estimators', *Applied Statistics*, vol. 40, p. 28 – 294.

Martuzzi, M.; Elliott, P. (1996) Empirical Bayes estimation of small area prevalence of non-rare conditions. *Statistics in Medicine*, vol. 15, p. 1867 – 1873.

Monteiro C. A.; Nazário C. L. (2000) Evolução da agenda tradicional da saúde pública: Mortalidade infantil e materna. Declínio da mortalidade infantil e equidade social: o caso da cidade de São Paulo entre 1973 e 1993. In: Monteiro C. A. (org). *Velhos e novos males da saúde no Brasil: a evolução do país e suas doenças*. 2.ed. São Paulo: Hucitec; p. 173-85.

Moran, P.A.P. (1950) Notes on continuous stochastic phenomena. *Biometrika*, vol. 37, p. 17 – 23.

Oliveira, L. A. P.; Mendes, M. M. S. (1995) Mortalidade Infantil no Brasil: Uma avaliação de tendências recentes, In: Minayo, M. C. S. (org). *Os muitos Brasis: saúde e população na década de 80*. São Paulo; Hucitec; p. 291 – 303.

Richardson, S. (2003) Spatial models in epidemiological applications. In: P. Green, N. L. Hjort and S. Richardson (Eds) *Highly Structured Stochastic Systems*. p. 237 – 259. Oxford University Press.

Richardson, S., Thomson, A., Best, N., Elliot, P. (2004) Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, v. 112, p. 1016 – 1025.

Santos, A. E. dos; Rodrigues, A. L.; Lopes, D. L. (2005) Aplicações de estimadores Bayesianos Empíricos para Análise Espacial de Taxas de Mortalidade. VII Simpósio Brasileiro de Geoinformática, Campos do Jordão, Brasil, 20 a 23 de Novembro 2005, INPE, p. 300 – 309.

Ugarte, M. D.; Goicoa, T; Militino, A. F. (2009) Empirical Bayes and Fully Bayes procedures to detect high-risk areas in disease mapping. *Computational Statistics and Data Analysis*, v. 53, p. 2938 – 2949.

Ugarte, M. D.; Goicoa, T; Militino, A. F. (2008) Prediction error estimators in empirical Bayes disease mapping. *Environmetrics*, v. 19, p. 287 – 300.

Vieira, D. M. (2006). *Análise Espaço - Temporal na Estimção de Taxas de Incidência/ Mortalidade*. Dissertação de Mestrado em Epidemiologia, Faculdade de Medicina, Universidade Federal do Rio Grande do Sul.

6. CONSIDERAÇÕES FINAIS

Os resultados do trabalho corroboram a necessidade do uso de métodos de suavização para melhor representar a verdadeira taxa de mortalidade infantil em cada município. Essa é a principal justificativa para o uso da metodologia Bayesiana de suavização das taxas para reduzir a alta variação que ocorre devido a elevada instabilidade presente no cálculo bruto dessas taxas. Essa alta variabilidade está relacionada ao tamanho da população sob risco e às instabilidades presentes na essência do indicador.

Verificou-se que o município de Cacique Doble ficou evidenciado como uma área de alta mortalidade infantil ao olharmos para as estimativas obtidas pelos métodos estudados. Essa região corresponde a uma área de população indígena, cujos recursos são mais escassos. A taxa de mortalidade infantil elevada pode ser explicada pela dificuldade de acesso do município a serviços na área da saúde bem como pelas baixas condições de desenvolvimento sócioeconômico próprias de regiões com essas características. A análise dos mapas permitiu destacar as que as regiões de altas taxas de mortalidade infantil encontram-se ao sudoeste e centro-oeste do Estado, e também ao norte, nas proximidades do município de Cacique Doble, foram detectadas regiões de alto risco. A região centro-oeste do Estado é onde fica a sede da décima Coordenadoria Regional de Saúde (10ª CRS), que estava desprovida de UTI neonatal até Abril de 2009. Para diminuir as elevadas taxas causadas pela ausência de leitos especializados, foram implantados no ano de 2009, 10 leitos de UTI neonatal nessa região. Esse investimento foi demandado devido, basicamente, a óbitos por falta de acesso a leito de cuidado intensivo devido à inexistência deste serviço na região - óbitos neonatais, precoces, vinculados a prematuridade. A região abarcava cerca de sete dos cinquenta e quatro municípios prioritários para a redução do óbito infantil no RS. As gestantes de alto risco e os prematuros de muito baixo peso precisavam acessar os serviços de saúde em Santa Maria (no Centro) ou Bagé (no extremo Sul).

Ainda, na região mais ao norte do Estado existem alguns municípios de área indígena ou de assentamentos onde há grande vulnerabilidade social, com a ocorrência de óbitos no período infantil tardio, vinculados a fatores de risco conhecidos e identificados nesta população devido à desorganização da rede de atenção básica de saúde, da falta ou

pelo precário acompanhamento pré-natal e ausência de programas adequados de atenção hospitalar.

Uma vantagem a ser destacada é a existência de vários aplicativos utilizados para a análise estatística de dados espaciais que já possuem implementado o método de estimação Bayesiano Empírico, o qual pode ser empregado através de interfaces bem mais acessíveis e amigáveis se comparados com os procedimentos computacionais utilizados para o método de estimação Totalmente Bayesiano. No caso particular do programa TerraView, uma desvantagem é que esse não fornece uma medida da precisão das estimativas geradas, o software não calcula os Intervalos de Confiança as taxas estimadas. Já para as taxas Totalmente Bayesianas, estimadas com o auxílio do aplicativo WinBugs, pode-se obter o cálculo de intervalos de credibilidade para poder medir a precisão das estimativas geradas. Esse aplicativo requer um conhecimento mais apurado de sua linguagem, uma vez que a definição das distribuições *a priori* e outros passos da análise devem ser realizados mediante linhas de comando e programação bem específicas.

Notou-se que se o objetivo é detectar/ identificar regiões de elevado risco, o excesso de suavização produzida pelo método Totalmente Bayesiano pode ser um problema no sentido de não detectar regiões de potencial risco para o evento analisado. No trabalho foi possível perceber que as estimativas geradas através do método Totalmente Bayesiano se apresentaram suavizadas demasiadamente em alguns casos, como exemplo, podemos citar o município de Cacique Doble, que apresentou uma estimativa Totalmente Bayesiana igual a 24,73 óbitos por mil nascidos vivos, enquanto sua estimativa através do método Bayesiano Empírico Local foi igual a 46,67 óbitos a cada mil nascidos vivos.

A identificação das regiões de elevado risco de mortalidade infantil é um passo muito importante para uma melhor alocação dos recursos financeiros e humanos, por parte dos gestores em saúde, necessários para redução de tal indicador.

A semelhança dos resultados dos métodos Empírico Bayesiano Local e Totalmente Bayesiano para as estimativas da Mortalidade Infantil no Rio Grande do Sul reproduz aqui outros achados sobre a similaridade dos métodos. Nesse trabalho, assim como nos trabalhos de Bernadinelli e Montonolli (1992), Ugarte *et al.* (2009), o método Bayesiano Empírico Local apresentou resultados similares aos apresentados pela abordagem Totalmente Bayesiana.

Assim, na falta de recursos computacionais mais sofisticados, e também pela dificuldade de implementação e o tempo de computação despendido para geração de estimativas Totalmente Bayesianas, as estimativas calculadas através do programa TerraView são resultados que devem ser considerados úteis para os profissionais da área da saúde – em especial da Epidemiologia – para obtenção do padrão subjacente dos eventos estudados, salientando as regiões com baixas ou altas taxas do evento de interesse.

Uma limitação a ser citada em nosso estudo foi que não efetuamos o cálculo dos intervalos de confiança para medir a precisão das estimativas geradas.

Finalizamos enfatizando que deveria haver um aumento da utilização das técnicas de Estatística Espacial, mais precisamente um aumento da utilização das técnicas de Análise Espacial Bayesianas em Epidemiologia Espacial, através da utilização mais simplificada (redução do tempo despendido nas análises) e igualmente eficiente do método Bayesiano Empírico, de modo que sejam mais acessíveis aos pesquisadores da área da saúde, secretarias estaduais, municipais, etc.

ANEXOS

ANEXO I – PROJETO DE PESQUISA

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

PROJETO DE PESQUISA

**MAPEAMENTO DA MORTALIDADE INFANTIL NO RIO GRANDE
DO SUL: UMA COMPARAÇÃO ENTRE AS ABORDAGENS
EMPIRICA BAYESIANA E TOTALMENTE BAYESIANA**

AUTOR: Sabrina Letícia Couto da Silva

ORIENTADORA: Prof^a Dr^a Jandyra Maria Guimarães Fachel

Porto Alegre, 2008.

SUMÁRIO

1. INTRODUÇÃO	64
1.1 Mortalidade Infantil	64
1.2 Epidemiologia Espacial	2
1.2.1 Análise Ecológica	2
1.2.2 Mapeamento de Doenças	3
1.2.3 Detecção de Clusters de Doenças	3
1.3 Modelagem Clássica para o Risco Relativo	3
1.4 Inferência Bayesiana	4
1.5 Modelagem Bayesiana em Análise Espacial	5
1.6 Método de estimação Totalmente Bayesiano	6
1.7 Método de estimação Empírico Bayesiano	6
2. OBJETIVOS	9
2.1 Objetivo Geral	9
2.1 Objetivo Específico	9
3. METODOLOGIA	10
4. CRONOGRAMA	10
5. REFERÊNCIAS	11

1. INTRODUÇÃO

1.1 Mortalidade Infantil

A taxa de mortalidade infantil reflete, de modo geral, os níveis de saúde e de desenvolvimento sócio-econômico de uma determinada área, sendo considerado um dos mais importantes indicadores epidemiológicos utilizados internacionalmente (Monteiro e Nazário, 2000), dessa forma, tem sido utilizada como um dos principais indicadores da qualidade de vida de uma população (Laurenti *et al*, 1985; Maia e Souza, 2004) além de compor diversos indicadores sócio-econômicos.

As taxas de mortalidade infantil vêm diminuindo progressivamente em muitos países em desenvolvimento. Conforme estimativas diretas obtidas do Registro Civil de nascimentos e óbitos infantis, a tendência da mortalidade infantil no Brasil na década de 1980 a 1990 foi de declínio nos primeiros dois anos, de estagnação nos dois anos seguintes e novamente de redução de 1984 para 1990 (Simões e Monteiro, 1995). Segundo dados do IBGE, a mortalidade infantil no Rio Grande do Sul apresentou um sensível decréscimo de 48,4 óbitos por mil nascidos vivos em 1970, para 13,1 em 2006. No Brasil, também vem ocorrendo uma redução significativa da mortalidade infantil. A taxa passou de 48,0 óbitos por mil nascidos vivos em 1990, para 34,6 em 1999, e 27,8, em 2002. Entretanto, as diferenças regionais acentuadas fazem com que esse processo seja mais lento que o ocorrido no Rio Grande do Sul. Segundo o DATASUS, em 2005, o coeficiente de mortalidade infantil estimado para o Brasil era de 17,0 e para o Rio Grande do Sul era de 13,7. Na literatura existem diversos estudos que apresentam ações ou intervenções para a redução da mortalidade infantil. Dentre esses estudos, estão os chamados estudos ecológicos nos quais o objeto de comparação são grupos ao invés de pessoas. Os estudos ecológicos são em geral rápidos, de menor custo, ideais para amostras integrais (como exemplos, poluição, altitude, clima, etc.), além de úteis para levantar hipóteses. As informações utilizadas em estudos desse tipo, focados à área da saúde – como exemplo, a mortalidade infantil – estão disponíveis em bases oficiais como DATASUS, IBGE, etc. Ainda, informações sobre mortalidade infantil são provenientes de censos, pesquisas demográficas e estudos epidemiológicos. Embora a qualidade dos registros oficiais no Brasil venha melhorando nos últimos anos, as pesquisas epidemiológicas oferecem

inúmeras vantagens para um melhor entendimento dos fatores determinantes da mortalidade infantil. Estas incluem a maior fidedignidade dos dados obtidos, a possibilidade de verificar a qualidade da informação sobre a causa de morte e de evidenciar diferenciais conforme diversos fatores de risco (Menezes *et al*, 1996).

Para entender melhor a distribuição espacial da mortalidade infantil, alguns estudos têm sido realizados (Assunção *et al.*, 1998; Santos e Noronha, 2001; Kato, 2007), inclusive no Rio Grande do Sul, através de estimação espaço-temporal (Vieira e Fachel, 2006).

Os estudos que visam associar a mortalidade com a localização geográfica de maneira visual, através de mapas, estão classificados dentro da área da Epidemiologia Espacial.

1.2 Epidemiologia Espacial

O estudo da distribuição geográfica da incidência de doenças e da sua relação com fatores de risco potenciais vem constituindo um terreno fértil para a aplicação e desenvolvimento de métodos e modelos estatísticos. Nos últimos anos, foram desenvolvidas ferramentas cada vez mais poderosas e versáteis nesta área de aplicação.

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

Em estudos do tipo ecológico, os indicadores são taxas ou riscos. Nesses, a análise geográfica mais simples leva em consideração a taxa bruta denotada por T_i , que representa - para cada unidade espacial i - a razão entre a quantidade observada (doença/incidência) Y_i e a população exposta N_i , tornando possível comparar os resultados entre áreas. Porém, muitas vezes, além da comparação entre diferentes áreas num mesmo mapa, faz-se necessária a comparação entre doenças/incidências diferentes para a mesma área, ou a comparação da mesma doença/incidência em períodos diferentes, e nestes casos o uso de medidas de risco relativo são mais apropriadas.

De acordo com a natureza da distribuição espacial dos dados observados e de acordo com o objetivo da análise, a Epidemiologia Espacial está dividida em três grandes áreas (Lawson, 2001): *Análise Ecológica, Mapeamento de Doenças e Detecção de Clusters de Doenças*.

1.2.1 Análise Ecológica

Podendo também ser chamada de estudo de correlação ecológica (Elliott *et al.*, 2001), a análise ecológica tem características muito similares ao contexto do Mapeamento de Doenças, ilustrando a distribuição da doença através de um mapa, mas seu diferencial está no objetivo do estudo. Os principais objetivos dos estudos ecológicos são a identificação de regiões de sobre-risco em relação à média global do processo estudado e a busca de fatores potencialmente explicativos dos diferenciais de incidência encontrados, seja em uma análise exploratória, ou mapeando doenças, ou na busca de modelos, identificando diferenças de risco e apontando medidas preventivas.

1.2.2 Mapeamento de Doenças

O Mapeamento de Doenças tem como objetivo descrever e sintetizar a distribuição e a variação da taxas de mortalidade ou incidência de doenças no mapa. Os modelos fornecem estimativas para o parâmetro de interesse associado à doença em estudo e permitem geração de mapas informativos, gerando informações sintéticas e objetivas, para identificarem diferenciais de risco, realizar previsões de epidemias, suspeitar da etiologia da doença, auxiliar em tomadas de decisões e levantar hipóteses.

1.2.3 Detecção de Clusters de Doenças

Consiste em identificar um agrupamento geográfico com elevada ocorrência da doença, sem que haja hipóteses etiológicas pré-definidas. Seu objetivo principal é detectar um foco de doença no mapa. Também podem ser chamados de “Avaliação de Fatores de Risco” (Elliott *et al.*, 2001) e são utilizados quando, por exemplo, suspeita-se que uma fonte, como por exemplo, uma indústria, pode causar um acréscimo no risco de ocorrência da doença.

1.3 Modelagem Clássica para o Risco Relativo

O mapeamento de doenças envolve, inicialmente, a escolha de uma medida epidemiológica para ser visualizada através de um mapa (Leyland e Davies, 2005). O mapa permite fazer comparações entre as áreas. Não faz sentido fazer mapas e comparações das contagens brutas já que elas dependem das populações de cada área. Para permitir

comparações entre diferentes populações no espaço ou no tempo, as contagens devem ser padronizadas para gerar taxas que tiram o efeito das diferentes populações. A medida mais frequentemente utilizada é o risco relativo.

Através do método de máxima verossimilhança, estimativas para os verdadeiros riscos relativos de interesse são atribuídas às áreas do mapa, a partir da SMR (*Standardised Mortality Ratio*) que é a razão entre a taxa da região i e a taxa da região toda. A SMR da região i , é dada por:

$$SMR_i = R_i = \frac{y_i}{e_i}$$

onde, y_i é a contagem de eventos na área i ; $e_i = r N_i$ é o número esperado de eventos na área i e N_i é a população em risco na área i .

A taxa global de risco em toda a área estudada é dada por:

$$r = \frac{\sum y_i}{\sum N_i}$$

A SMR varia de zero a infinito, sendo que, se $R_i = 1$ a área i teve tantos casos observados quanto seria esperado caso seu risco fosse idêntico ao de toda área avaliada.

A variância da SMR_i (y_i / e_i^2) é inversamente proporcional ao número esperado de casos na área i (e_i) e diretamente proporcional ao tamanho da população na área i (N_i), isto é, quanto menor o tamanho da população em risco, maior a incerteza na estimativa do risco (Olsen *et al.*, 1996). Essa incerteza causa instabilidade ao modelo, tornando difícil a interpretação dos mapas, e pode se tornar ainda maior em estudos com baixa incidência, por exemplo, mortalidade segundo causa específica. Existem na literatura três métodos alternativos para contornar esse problema: agregar áreas, mapas de probabilidade e modelagem Bayesiana. Atualmente as duas primeiras não estão sendo mais utilizadas em função das vantagens demonstradas pela modelagem Bayesiana (Assunção, 2001).

1.4 Inferência Bayesiana

A estatística Bayesiana pode ser utilizada como método de inferência na análise estatística de dados espaciais. Em geral, o objetivo da inferência estatística é a estimação dos parâmetros de uma certa população e/ou testar hipóteses sobre eles, através de uma amostra extraída dessa população. Em inferência clássica, toda informação sobre o parâmetro a ser estimado está contida na amostra e expressa através da chamada função de verossimilhança.

Na inferência Bayesiana, nem toda a informação sobre o parâmetro de interesse é proveniente da amostra. Parte dela está contida na amostra (função de verossimilhança) e outra parte reside num conhecimento *a priori* sobre o parâmetro (antes de se conhecer os dados) e pode ser expressa através de uma distribuição de probabilidade (chamada de distribuição *a priori* do parâmetro). O peso de cada informação, da amostra ou da *priori*, é medido através de quão *informativa* é a distribuição *a priori* atribuída ao parâmetro. As duas partes da informação são combinadas com o auxílio do teorema de Bayes, resultando na distribuição *a posteriori* do parâmetro, utilizada nas inferências sobre ele. De posse dessa distribuição, é possível calcular probabilidades, percentis, média, entre outras propriedades (Carlin e Louis, 1996). Em problemas mais complexos, o cálculo analítico das distribuições *a posteriori* é inviável. As inferências são feitas através de amostras destas distribuições, obtidas por simulação via *Monte Carlo Markov Chain* (MCMC) (Gelman, 1997).

1.5 Modelagem Bayesiana em Análise Espacial

Os estudos com dados de área foram os que primeiro se beneficiaram do uso de técnicas Bayesianas (Tsutakawa *et al.*, 1985), seguidos de perto pelos que usam dados espacialmente contínuos (Kitanidis, 1986) e processos pontuais (Akman e Raftery, 1986). Um pouco mais recente é a discussão de métodos Bayesianos aplicados à análise de dados de interação espacial (Banerjee *et al.*, 2000). Em 1989, Clayton abordou o uso modelos Bayesianos hierárquicos (abordagem Totalmente Bayesiana) em epidemiologia descritiva. Com base nestes e em outros trabalhos, Bernardinelli e Montomoli (1992) fizeram uma comparação entre métodos Bayesianos Empíricos e Totalmente Bayesianos na análise da variação geográfica do risco de doenças, utilizando a metodologia MCMC nas análises

Totalmente Bayesianas. Desde então, os modelos Bayesianos para análise de dados de área tornaram-se muito populares, principalmente entre os pesquisadores que lidam com mapeamento de taxas. Citando alguns trabalhos, temos Clayton e Kaldor (1987), Heisterkamp *et al* (1993), Smith *et al* (1995), Reis e Assunção (1996), Rytönen *et al.* (2001), Johnson (2004) e Yang *et al.* (2005).

Uma das razões para esta popularidade está no fato de que os métodos Bayesianos permitem a estimação de taxas para pequenas áreas sem a grande instabilidade presente nas estimativas clássicas. Esses métodos, ao estimarem o risco de uma pequena área, têm como idéia o uso de informação das outras áreas (vizinhança) que compõem a região estudada. Dessa maneira, diminui-se o erro quadrático médio total da estimação dos riscos.

1.6 Método de estimação Totalmente Bayesiano

Best *et al.* (2005) ilustraram os mais recentes modelos Bayesianos Hierárquicos que têm sido utilizados no contexto de mapeamento de doenças, com a aplicação de estimadores Totalmente Bayesianos. Knorr-Held (2000) descreve quatro possíveis interações entre espaço e tempo que foram utilizadas para a construção de modelos espaço-temporais para o risco de doenças. Inferências são realizadas utilizando as técnicas MCMC (Markov Chain Monte Carlo) que auxiliam na geração da distribuição a posteriori dos parâmetros do modelo desejado.

A interação espacial é usualmente modelada através da média de um modelo CAR (Autoregressivo Condicional) (Waller e Gotway, 2004), de modo que apenas os efeitos das áreas próximas são incluídos. De forma semelhante, o efeito temporal é modelado por um modelo AR (Autoregressivo) que mede a relação dos valores em uma área em diferentes tempos. Por fim, espaço e tempo podem interagir. Porém, ao considerar um ou ambos efeitos simultaneamente, pode-se ter problemas de confundimento gerando problemas na identificação dos efeitos e futuras suposições podem ser necessárias para o uso da técnica.

1.7 Método de estimação Empírico Bayesiano

A existência de flutuações aleatórias como a ocorrência de um ou dois casos do evento estudado, a mais ou a menos, numa localidade, causam variações substanciais nas taxas brutas se a população do local for pequena, efeito que não é verificado em localidades

de população grande. Além disso, para situações em que não ocorrem casos do evento em algumas regiões, a taxa bruta estima o risco de ocorrência do evento como zero, um dado pouco real no contexto de dados epidemiológicos. Mapas de eventos baseados diretamente em estimativas brutas são de difícil interpretação e frequentemente geram falsas conclusões.

Uma alternativa ao uso das taxas brutas em situações como essas - populações relativamente pequenas e eventos raros - foi proposta por Marshall (1991) através dos estimadores Bayesianos Empíricos. Esses métodos Bayesianos permitem que a informação a respeito do risco contida nos dados das outras áreas possa ser usada para estimar o risco de determinada área. Denotando por B a estimativa Bayesiana Empírica de certa área, por t a taxa bruta dessa área e por m a taxa média de todas as áreas, podemos escrever,

$$B = m + c(t - m).$$

O valor de c está entre 0 e 1 e representa o peso dado à diferença entre a taxa bruta da área e a taxa média de todas as áreas. O cálculo de c leva em conta o tamanho da população da área. Quando a população for relativamente grande, o valor de c é próximo de um, tornando a estimativa B muito próxima da taxa bruta, que, nessa área, não sofre o problema da instabilidade. No entanto, quando a população for pequena, o valor de c decresce, dando pouco peso à diferença entre t e m e tornando a estimativa B mais próxima da taxa média de todas as áreas. Ou seja, quando a população da área é pequena, a informação contida na taxa média tem peso maior na estimativa Bayesiana Empírica. A expressão para B também pode ser escrita como,

$$B = ct + (1 - c)m,$$

em que se nota que a estimativa B é a média entre a taxa bruta da área e a taxa média de todas as áreas, ponderada pelo valor c . O estimador Bayesiano Empírico, que utiliza como valor de m a taxa média de todas as áreas em estudo é denominado *estimador Bayesiano Empírico Global*. O estimador Bayesiano Empírico Global calcula uma média ponderada entre a taxa bruta da localidade e a taxa global da região (razão entre o número total de casos e a população total). Existe ainda o estimador Bayesiano Empírico Local que inclui o

efeito espacial, calculando a estimativa localmente, utilizando somente os vizinhos geográficos da área na qual se deseja estimar a taxa, convergindo em direção a uma média local ao invés de uma média global. As taxas corrigidas são menos instáveis, pois levam em conta no seu cálculo não só a informação da área, mas também a informação de sua vizinhança. Bernardinelli e Montomoli (1992) fizeram uma comparação entre os procedimentos de estimação Empírico Bayesiano e o Totalmente Bayesiano com modelos que tratam com a variação geográfica da doença. Os autores apontam que o método Totalmente Bayesiano é preferível porque considera a incerteza dos parâmetros do modelo, enquanto que o método Empírico Bayesiano condiciona a estimação em estimativas pontuais dos parâmetros do modelo. Então, os estimadores Empíricos Bayesianos podem ser menos precisos. Por outro lado, os métodos Totalmente Bayesianos, usualmente, são computacionalmente mais intensivos, no entanto, os estimadores Bayesianos Empíricos têm se mostrado úteis, por exemplo, para fornecer estimativas iniciais do risco da doença no contexto dos Sistemas de Informações Geográfica (SIG) para epidemiologia espacial, onde grandes quantidades de dados estão disponíveis e os algoritmos usados devem ser rápidos e não muito intensivos computacionalmente (Aylin *et al.*, 1999). Se a análise exploratória indica que há uma alta variação de risco entre áreas, a metodologia Totalmente Bayesiana pode ser usada para detectar clusters de doenças, por exemplo. Esse seria um trabalho mais detalhado onde o investimento computacional seria conveniente.

2. OBJETIVOS

2.1 Objetivo Geral:

O objetivo geral deste trabalho é comparar os Métodos de Estimação Espacial Bayesianos (Empírico e Totalmente Bayesiano) em Estudos Ecológicos, para estimação do risco relativo, através do cálculo do Erro Quadrático Médio (EQM) além de utilizar mapas para visualização da distribuição do risco estimado pelos dois métodos.

2.2 Objetivo Específico:

Ilustrar os dois métodos propostos através da utilização dos dados de mortalidade infantil em 496 municípios do Rio Grande do Sul, para dados acumulados entre os anos de 2001 a 2004, além da comparação entre as estimativas obtidas para as taxas de mortalidade através das modelagens Empírica Bayesiana e Totalmente Bayesiana.

3. METODOLOGIA

Serão utilizados os dados referentes aos óbitos em menores de 1 ano e de nascidos vivos em 496 municípios do Rio Grande do Sul no período de 2001 a 2004, publicados no site do DATASUS.

Para o cálculo das estimativas do risco de mortalidade infantil, serão utilizados os softwares WinBugs e R (modelagem Totalmente Bayesiana) e o software Terra View (modelagem Empírica Bayesiana). Para ilustração dos achados serão feitos mapas no TerraView.

4. CRONOGRAMA

<i>Atividade</i>	<i>2007/2</i>	<i>2008/1</i>					<i>2008/2</i>				
		<i>mar</i>	<i>abr</i>	<i>mai</i>	<i>jun</i>	<i>jul</i>	<i>ago</i>	<i>set</i>	<i>out</i>	<i>nov</i>	<i>dez</i>
<i>Revisão da literatura</i>											
<i>Familiarização softwares</i>											
<i>Análise dos dados</i>											
<i>Redação</i>											
<i>Defesa Preliminar</i>											
<i>Correções</i>											
<i>Defesa Final</i>											

5. REFERÊNCIAS BIBLIOGRÁFICAS

Akman, V. E. and A. E. Raftery (1986). Bayes Factors for Nonhomogeneous Poisson Processes with Vague Prior Information. *Journal of the Royal Statistical Society Series B-Methodological*, vol. 48, n. 3, pp. 322-329.

Assunção, R.M. (2001). *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*. São Carlos. ABE. Disponível em:
<http://www.est.ufmg.br/leste/publicacoes.htm>. Data de acesso: 28/01/2008.

Assunção, R.M.; Barreto, S.M; Guerra, H.L.; Sakura, E. (1998). Mapas de taxas epidemiológicas: uma abordagem Bayesiana. *Cadernos de Saúde Pública*, Rio de Janeiro, 14(4):713-723.

Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., Jarup, L., Arnold, R., Wheeler, G. and Elliot, P. (1999). A National Facility for Small Area Disease Mapping and Rapid Initial Assessment of Apparent Disease Clusters Around a Point Source: The U.K. Small Area Health Statistics Unit. *Journal of Public Health Medicine*, 21(3):289 – 298.

Bailey, T.C.; Gatrell, A.C. (1995). *Interactive Spatial Data Analysis*. London: Longman.

Banerjee, S. , Gelfand, A. E. , Polasek, W. (2000). Geostatistical modelling for spatial interaction data with application to postal service performance. *Journal of Statistical Planning and Inference*, vol. 90, número 1, pp. 87-105.

Bernardinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* 1992; 11: 983 – 1007.

Best, N., Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* 14(1), 35 – 59.

Carlin, B. e Louis, T. A. (1996) *Bayes and Empirical Bayes Methods for Data Analysis*. Ed. Chapman & Hall, London, 399 p.

Clayton, D.; Kaldor, J. (1987). Empirical Bayes Estimates of Age-Standardized Relative Risks for use in Disease Mapping. *Biometrics*, vol. 43, pp. 671 – 681.

Clayton, D. (1989). Hierarchical Bayesian models in descriptive epidemiology. *Proceedings of the XIVth International Biometrics Conference*, pp. 201-213.

Devine, O.J. and Louis, T.A. A constrained empirical bayes estimator for incidence rates in areas with small populations. *Statistics in Medicine*, 13:1119 – 1133.

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

Gamerman, D. (1997) Markov Chain Monte Carlo: stochastic simulation for Bayesian inference. Chapman and Hall, London, 245 p.

Heisterkamp SH, Doornbos G, Gankema M. Disease mapping using empirical Bayes and Bayes methods on mortality statistics in the Netherlands. *Statistics in Medicine* 1993; 12: 1895 – 913.

Johnson, G. (2004). Small area mapping of prostate cancer incidence in New York State (USA) using fully Bayesian hierarchical modelling. *International Journal of Health Geographics* 3(1): 29.

Kato, S. K. (2007). Análise de Correlação Ecológica: uma abordagem Inteiramente Bayesiana para a Mortalidade Infantil no Rio Grande do Sul. Dissertação Mestrado em Epidemiologia – Universidade Federal do Rio Grande do Sul.

Kitanidis, P. K. (1986). Parameter Uncertainty in Estimation of Spatial Functions - Bayesian-Analysis. *Water Resources Research*, vol 22. n. 4, pp 499-507.

Knorr-Held, L. (2000). Bayesian modelling of inseparable spacetime variation in disease risk. *Statistics in Medicine* 19, 2555 – 2567.

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

Laurenti, R.; Jorge, M.H.P.; Lebrão, M.L. e Gotlieb, S.L.D., 1985. *Estatísticas de Saúde*. São Paulo: EPU/EDUSP.

Leyland, A H.; Davies, C. A. Empirical Bayes methods for disease mapping. *Stat Methods Med Res* 2005; 14; 17

Maia, S.F.; Sousa, T.R.V. (2004). Uma investigação dos determinantes da redução da taxa de mortalidade infantil nos Estados da Região Nordeste do Brasil. In: I Congresso da Associação Latino Americana de População - ALAP, 2004, Caxambu. I Congresso da Associação Latino Americana de População - ALAP. Caxambu : ALAP.

Marshall, R.J. (1991). Mapping disease and mortality rates using empirical bayes estimators. *Applied Statistics*, 40(2):283 – 294.

Menezes, A.M. B., Victora, C.G., Barros, F.C. *et al.* Mortalidade infantil em duas coortes de base populacional no Sul do Brasil: tendências e diferenciais. *Cad. Saúde Pública*, 1996, vol.12 suppl.1, p.S79-S86. ISSN 0102-311X.

Monteiro CA, Nazário CL. Evolução da agenda tradicional da saúde pública: Mortalidade infantil e materna. Declínio da mortalidade infantil e equidade social: o caso da cidade de São Paulo entre 1973 e 1993. In: Monteiro C.A, organizador. Velhos e novos males da

saúde no Brasil: a evolução do país e suas doenças. 2.ed. São Paulo: Hucitec; 2000. p. 173-85.

Olsen, S.; Martuzzi, M.; Elliott, P. (1996). Cluster analysis and disease mapping - Why, when, how? A step by step guide, *British Medical Journal*, 313, 863-865.

Reis, E. A. ; Assunção, R. M. Mapeamento da Malária em Rondônia usando o Estimador Empírico de Bayes. In: X Encontro Nacional de Estudos Populacionais, 1996, Belo Horizonte. Anais do X Encontro Nacional de Estudos Populacionais. Belo Horizonte: ABEP, 1996. v. 2. p. 1077-1097.

Rytkonen, M., J. Ranta, et al. (2001). Bayesian analysis of geographical variation in the incidence of Type I diabetes in Finland. *Diabetologia* 44: B37-B44.

Santos, S.M.; Noronha, C.P. (2001). Padrões espaciais de mortalidade e diferenciais sócio-econômicos na cidade do Rio de Janeiro. *Cadernos de Saúde Pública*, Rio de Janeiro, 17(5):1099-1110.

Simões, C.C.S. ; Monteiro, C.A., 1995. Tendência secular e diferenciais regionais da mortalidade infantil no Brasil. In: *Velhos e Novos Males da Saúde no Brasil. A Evolução do País e de Suas Doenças* (C. A. Monteiro, org.), pp. 151 –156, São Paulo: Hucitec.

Smith, T.; Charlwood, J.D.; Takken, W.; Tanner, M.; Spiegelhalter, D.J.(1995) Mapping the densities of malaria vectors within a single village. *Acta Tropica*. Vol. 59, número 1, pp. 1-18.

Tsutakawa, R. K. , Shoop, G. L. e Marienfeld, C. J (1985). Empirical Bayes Estimation of Cancer Mortality-Rates. *Statistics in Medicine*, vol. 4, n. 2, pp 201-212.

Vieira, D.M. (2006). Análise Espaço -Temporal na Estimção de Taxas de Incidência/ Mortalidade. Dissertação Mestrado em Epidemiologia – Universidade Federal do Rio Grande do Sul.

Waller, L.A. and Gotway, C.A. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, Hoboken, New Jersey.

Yang, G.J.; Vounatsou P.; Zhou, X.N. ; Tanner M. ; Utzinger, H. (2005). A Bayesian-based approach for spatio-temporal modeling of county level prevalence of *Schistosoma japonicum* infection in Jiangsu province, China. *International Journal for Parasitology* 35(2): 155-162.

ANEXO II

LINHA DE COMANDO ESTIMAÇÃO TOTALMENTE BAYESIANA (Kato, 2007)

WinBugs - Modelo – Com efeito espacial e sem covariáveis

```
model
{
for (i in 1:m)
{
# Verossimilhança
y[i]~dpois(mu[i])
log(mu[i])<-log(e[i])+alpha+v[i]+u[i]
# Risco Relativo
theta[i]<-exp(v[i]+u[i])
# Taxa de incidência
tx[i]<- 15.632*theta[i]
}

# Priori Car Normal
u[1:m]~car.normal(adj[],weights[],num[],tau.u)

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

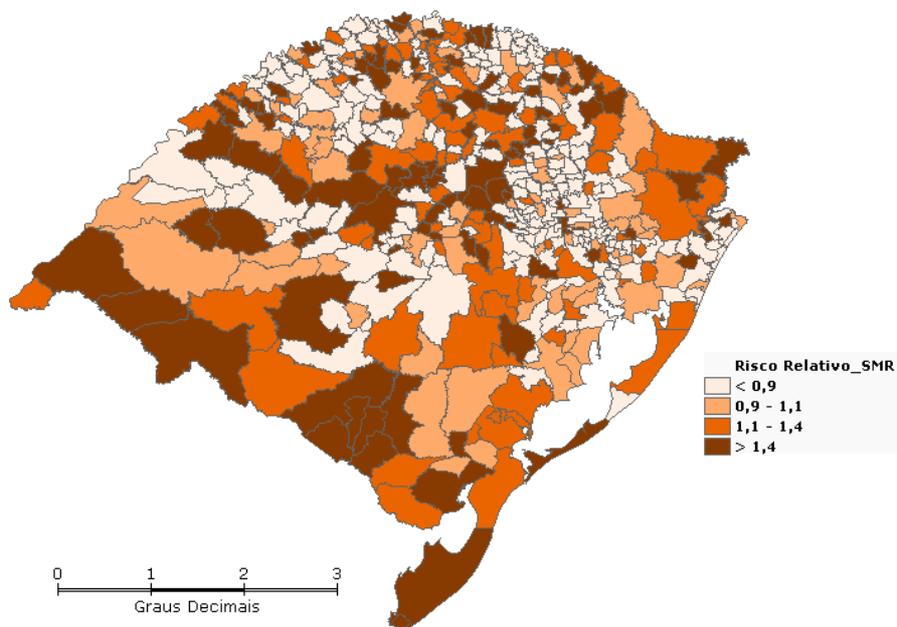
#Weights
for(j in 1:sumNumNeig)
{
weights[j]<-1
}

alpha~dflat()

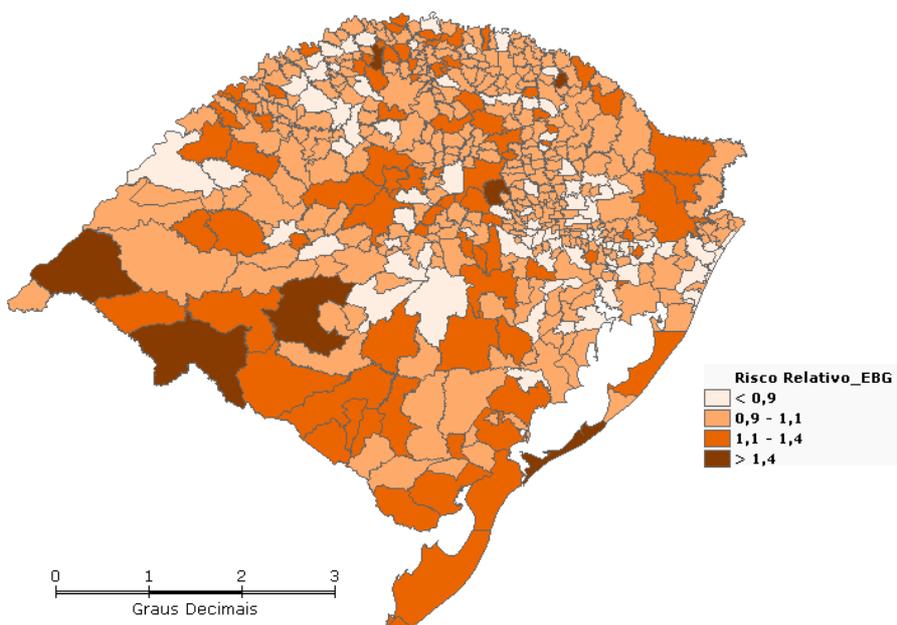
# Hiperprioris
tau.v~dgamma(0.5, 0.0005)
tau.u~dgamma(0.5, 0.0005)
}
```

ANEXO III

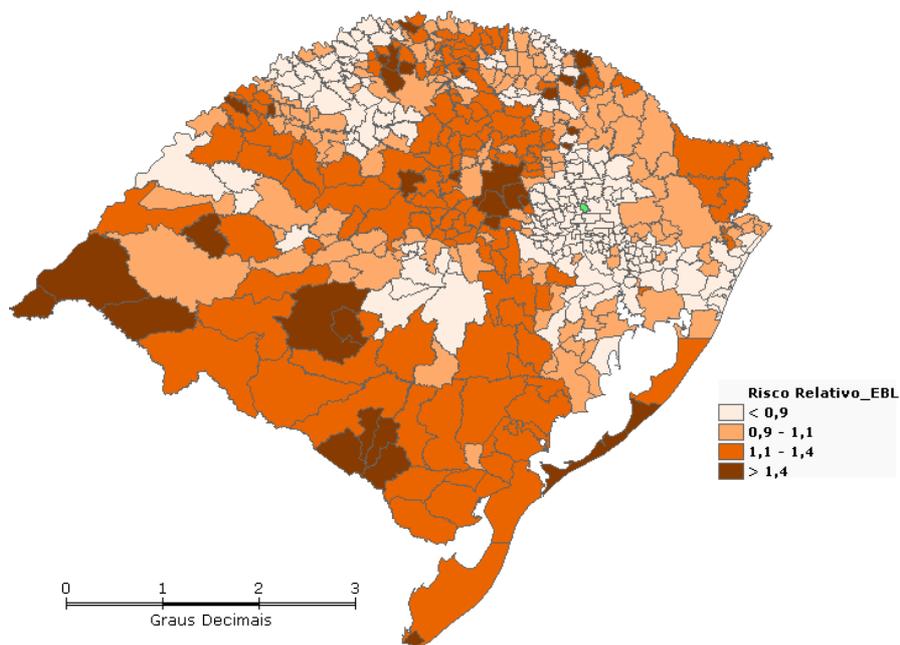
RISCOS RELATIVOS DE MORTALIDADE INFANTIL NO RS – 2001 a 2004



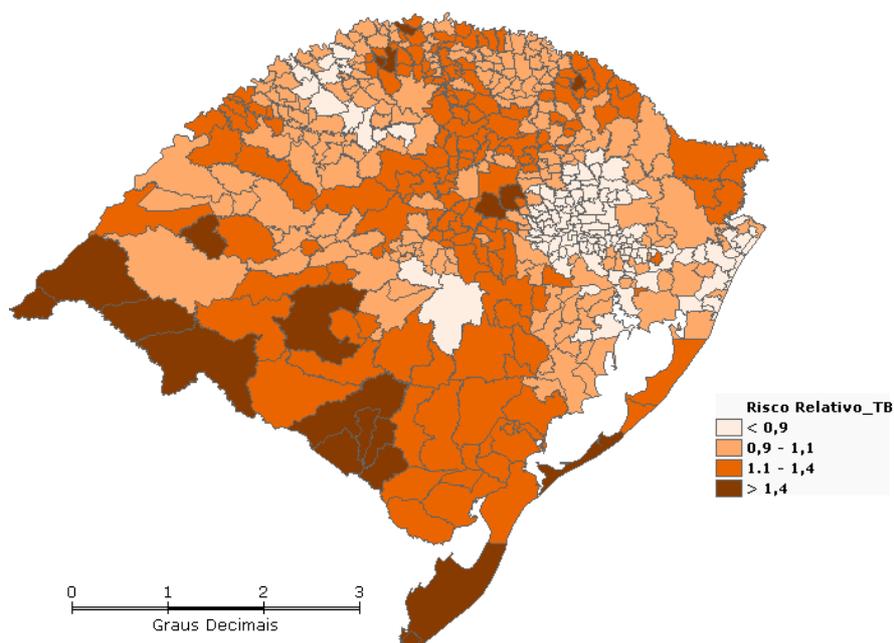
Mapa 1. Risco relativo de Mortalidade Infantil, segundo a metodologia SMR (*Standardized Mortality Ratio*)



Mapa 2. Risco relativo de Mortalidade Infantil, segundo a metodologia BEG (Bayesiana Empírica Global)



Mapa 3. Risco relativo de Mortalidade Infantil, segundo a metodologia BEL (Bayesiana Empírica Local)



Mapa 4. Risco relativo de Mortalidade Infantil, segundo a metodologia TB (Totalmente Bayesiana)