



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Faculdade de Medicina
Programa de Pós-graduação em Epidemiologia



Métodos Bayesianos em Epidemiologia Espacial: Avaliando a Taxa de Fecundidade em Mulheres de 10 a 19 anos em Porto Alegre

Mariana Teixeira Carballo

Orientadora: Jandyra M. G. Fachel

Dissertação de Mestrado

Porto Alegre, Dezembro de 2005

Agradecimentos

Nesse momento tão importante na minha vida, preciso agradecer a algumas pessoas que contribuíram e me ajudaram para essa realização.

Em primeiro lugar, gostaria de agradecer à minha orientadora, Jandyra, pelo companheirismo, conselhos e desafios colocados ao longo da minha caminhada, tanto no mestrado como na minha graduação, e agora na minha vida profissional. Espero que continues a me guiar por essa longa jornada.

À Denise Aerts, agradeço a proposta de trabalho, comentários e sugestões e principalmente a credibilidade depositada em mim.

À Equipe de Informação em Saúde / CWS / SMS - POA, em especial ao Rui Flores, agradeço a ajuda na construção do banco de dados juntando os logradouros da cidade com as informações do SINASC e a disposição das declarações de nascimentos e a base de dados da Secretaria.

Ao Gustavo Ferreira, meu colega de graduação, por me ajudar a compreender o funcionamento do WinBUGS.

Aos professores Bruce Duncan e Maria Inês Schmidt, agradeço as oportunidades e a credibilidade depositada.

Aos colegas de mestrado, em especial a Fabiana e a Luciane, agradeço as discussões epidemiológicas e estatísticas que tivemos. Com certeza nessas conversas não só vocês quem lucraram, deixei um pouco de lado a significância estatística e passei a ver a relevância clínica.

Às minha amigas e colegas Locimara Ramos e Iracema Santos, agradeço o apoio e incentivo nas horas em que mais precisei de socorro.

Aos meus familiares, em especial, aos meus pais e irmãos pelo apoio e pela confiança.

Resumo

Introdução - A forma tradicional para se estudar a variação espacial de taxas é feita considerando a contagem do número de pessoas em cada área de um mapa em relação à população exposta, denominada taxa bruta. Entretanto, essa representação nem sempre é a melhor, principalmente na presença de população pequena. **Objetivos** - Utilizar métodos de Epidemiologia Espacial para representar a taxa de fecundidade específica em mulheres de 10 a 19 anos, utilizando como exemplo a cidade de Porto Alegre no ano de 2003 e investigar sua distribuição geográfica por meio de técnicas de Mapeamento de Eventos de Saúde. **Métodos** - Utilizando-se a base de dados do SINASC e o Censo do IBGE para o ano de 2000, considerando-se as divisões de bairro da cidade de Porto Alegre, construiu-se a taxa de fecundidade específica na adolescência. Métodos de suavização para controlar as flutuações aleatórias do risco foram usados, como o Estimador Bayesiano Empírico Local de Marshall e o Estimador Totalmente Bayesiano. **Resultados** - No ano de 2003 ocorreram 28 nascidos vivos por mil mulheres de 10 a 19 anos na cidade. Em alguns bairros, a representação bruta revelou valores quase três vezes maiores. Com os métodos de suavização, apresenta-se uma distribuição espacial subjacente mais real e taxas com menor heterogeneidade espacial. Alguns bairros com maiores taxas de maternidade na adolescência encontram-se em regiões com piores condições sócio-econômicas. **Conclusões** - O método Totalmente Bayesiano mostrou ser o melhor para estimar a variabilidade da taxa de fecundidade específica em mulheres de 10 a 19 anos nos bairros de Porto Alegre, principalmente nos locais onde a população exposta era muito pequena.

Abstract

Introduction: The traditional way to study the spatial variability in rates is done by the gross rate: dividing the number of people in the area by the population at risk. However, the variability between and within neighborhoods and a small population at risk might cause a random noise in this representation. **Objective:** to present the specific fecundity rate for teenagers mothers using Spatial Epidemiology methods in the year of 2003 using Porto Alegre as example, and to investigate their geographic distribution through techniques of Health Event Mapping. **Methods:** the adolescent fecundity rate was built using the SINASC database and the year 2000 census. Besides the gross rate, a standard rate was built considering differences in population size. To evaluate spatial variation in risk, the Bayesian approach was used. The estimation methods used were the Empirical Local Bayesian and the Full Bayesian Estimators. The Full Bayesian estimator was build using hierarchical models. **Results:** In 2003, there were 28 births per thousand women with age 10 to 19. In some neighborhoods, the gross rate showed values three times larger than the average rate for the entire city. Through the methods for disease mapping studies, we summarize spatial variation, in order to assess the amount of true spatial heterogeneity of adolescent mothers in Porto Alegre; the areas with elevated risk also have an economic level lower than the average. **Conclusion:** the Full Bayesian estimator obtained best results in reducing the fecundity rate variability.

Índice

I Introdução	6
II Revisão Bibliográfica	10
2.1 Epidemiologia Espacial	10
2.1.1 Mapeamento de Doença	13
Representação estatística - métodos exploratórios	14
Métodos Informais de Suavização	16
Modelagem no mapeamento de doença	19
2.1.2 Detecção de <i>Clusters</i>	24
Testando a hipótese inicial – análise do padrão espacial	27
Métodos Tradicionais – Testes de Homogeneidade	28
Métodos que se baseiam na distância – Investigação de <i>Cluster</i> Global	28
Métodos circulares - Investigação de <i>Clusters</i> Locais	32
Modelagem na detecção de <i>clusters</i>	33
2.1.3 Estudos de Correlação geográfica	34
Modelagem na correlação geográfica	34
2.2 Sistema de Informação Geográfica	35
2.3 Taxa de Fecundidade na adolescência	36
2.3.1 Indicadores Demográficos	38
2.3.2 Indicadores Socioeconômicos	39
III Referências	41
IV Artigo	45
Resumo	46
Abstract	47
Introdução	48
Material e Métodos	49
Resultados	55
Discussão	59
V Considerações Finais	64
Anexos	66
Anexo A - Projeto	67
Anexo B - Etapas de aplicação do método de Epidemiologia Espacial	75
Anexo C - Comandos para o modelo totalmente Bayesiano no WinBUGS	77

I Introdução

O aumento de mães adolescentes (mães com idade entre 10 a 19 anos) em países em desenvolvimento tem despertado interesse de pesquisadores e profissionais de saúde, uma vez que esse evento está ligado a pobreza, baixa escolaridade e baixa condição de vida (1-8).

O conhecimento desta população criará alguns indicadores demográficos, a saber: a taxa de fecundidade específica por idade da mãe e o número de nascidos vivos por faixa etária da mãe.

A saúde materno-infantil vem ganhando espaço em diversos trabalhos que consideram dados populacionais na investigação da distribuição geográfica, através de métodos de análise espacial (9). Os métodos estatísticos identificam áreas onde os indicadores demográficos são significativamente superiores ao esperado para aquela população, ou seja, estabelecem quais são as áreas prioritárias para atuação da equipe de saúde, buscando uma melhoria na saúde da mãe e da criança.

A disponibilidade de índices demográficos e de dados populacionais, os avanços computacionais, o desenvolvimento de Sistemas de Informação Geográfica (SIG) e as metodologias espaciais permitem uma avaliação real das variações nas taxas (índices demográficos, incidências de eventos de saúde ou risco de doenças), principalmente em pequenas áreas, através de sua localização espacial. Esse tipo de estudo está ganhando cada vez mais destaque nos estudos epidemiológicos e de saúde pública, sendo denominado de *Epidemiologia Espacial* (10;11).

As informações referentes aos dados espaciais podem estar associadas a localizações pontuais, obtidas pela localização espacial exata através do endereço ou das coordenadas fornecidas através de um SIG, podem ser mapeadas por áreas geográficas (uma região espacialmente definida representando unidades geográficas como município, bairro, setor censitário, etc) ou podem apresentar-se como uma medida observada numa superfície espacial contínua.

Dados em pontos podem ser investigados por uma modelagem onde a relação entre as características dos indivíduos (sexo, idade, fatores genéticos, estilo de vida e exposição a poluentes ambientais) e a ocorrência do evento será avaliada, verificando-se se a distribuição espacial do evento é ou não aleatória (11).

Os dados em áreas são identificados pelas coordenadas geográficas do centro da região obtidas por SIG. Eles representam contagens de casos populacionais em divisões administrativas através das quais obtêm-se taxas supondo-se homogeneidade interna (dentro de um mesmo município, por exemplo). Nessa situação, a modelagem usada controla a variabilidade das taxas em áreas onde a população é pequena. Em situações onde há grande contraste social, os grupos podem estar agregados em uma mesma região, causando heterogeneidade interna. Faz-se necessário ajustes e correções para este problema (12). Entretanto, considerando os dados em nível de área, perde-se precisão por não se ter à disposição a informação individual, e sim a informação agregada em unidades geográficas.

Já na avaliação contínua a unidade amostral fornece a localização para alguns pontos do espaço, fazendo-se a inferência para toda a região em relação ao fenômeno em estudo. Este tipo de estudo é comum, por exemplo, quando coletamos algumas amostras dentro de um rio e, a partir dessas amostras, procura-se estimar o teor de determinadas substâncias em toda a área contínua, como, por exemplo, um trecho de um rio. A esse tipo de avaliação denomina-se de Geoestatística, e representa medições de parâmetros ambientais avaliados em pontos amostrais obtidos nessa superfície com o objetivo de reconstruir a superfície na qual se mediram as amostras. Os estudos que envolvem Geoestatística não serão descritos nesse texto e, para maiores detalhes, recomenda-se autores específicos (13;14).

A investigação de dados espaciais é habitualmente denominada de Estatística Espacial, diferenciando-se da Epidemiologia Espacial. A Estatística Espacial avalia a completa aleatoriedade espacial ou não do processo, através de pressupostos como a homogeneidade dos eventos e estacionariedade de primeira ordem. Um processo estocástico é chamado de estacionário quando apresenta média constante em toda a região, não havendo tendências. Já na Epidemiologia Espacial os modelos tiveram que ser modificados, pois as populações são heterogêneas, não há estacionariedade de primeira ordem e há outras covariáveis interferindo no processo espacial. Assim, na Epidemiologia Espacial, admite-se que a probabilidade de um indivíduo se tornar caso em um determinado lugar é independente da de outro lugar.

Considerando a Epidemiologia Espacial, o foco de estudo será identificar a distribuição espacial de eventos de saúde, contribuindo, assim, para a formulação de hipóteses e para a especificação dos métodos de análise a serem usados.

As hipóteses formuladas poderão ser classificadas em três categorias: mapeamento de doença, análise ecológica e *cluster* de doença. No mapeamento de doença, a distribuição da doença será descrita num mapa, tendo como objetivo remover as variações extras. A análise ecológica avalia a relação entre a distribuição espacial da doença com fatores explanatórios medidos em nível agregado. O *cluster* de doença investiga agregações não usuais da doença, ou seja, áreas de maior incidência, podendo ocorrer na presença de doenças infecciosas, contágio, ou estando associado a determinados locais (15).

A avaliação da distribuição geográfica do evento de saúde é feita em etapas. Após a determinação da população em risco, faz-se o mapeamento da taxa de fecundidade específica. Na presença de populações muito pequenas, flutuações aleatórias podem aparecer; então, procura-se estimar a verdadeira taxa através de modelos estatísticos, considerando informações sobre a distribuição probabilística dos dados, a localização geográfica e os fatores que podem estar relacionados com o evento em estudo. Várias são as propostas na literatura para tratar estas modelagens, gerando melhores estimativas: uma alternativa para o seu controle é utilizar, além de métodos de Estatística Espacial, os Bayesianos e os de *kernel*.

Já está estabelecido que os estimadores Bayesianos eliminam vieses através da redução da variância dos estimadores, principalmente em locais onde a amostra é pequena. São feitas as modificações na variabilidade do processo considerando informações da vizinhança presente na estrutura hierárquica que considera a relação entre áreas próximas (correlação entre vizinhos) e dentro das áreas (efeito espacial) (16).

Nos estudos de desigualdades, a variabilidade social é bem-vinda; entretanto, a variabilidade gerada por flutuações aleatórias não é bem-vinda, pois generalizações errôneas a respeito do fenômeno serão feitas. A separação da dependência espacial da variabilidade total das taxas gerará estimativas mais precisas, reduzindo as causas de variação não-explicadas.

O presente estudo pretende comparar diferentes métodos de representação da taxa de fecundidade na adolescência, usando como exemplo o município de Porto Alegre, no ano de 2003, e investigar a distribuição geográfica da maternidade na adolescência nesse período através do seu mapeamento. Também se pretende relacionar alguns fatores socioeconômicos (percentual de chefes de família analfabetos em cada bairro e percentual de famílias com renda menor do que um salário mínimo) com a incidência de maternidade na adolescência.

Como essa abordagem espacial ainda é relativamente recente, uma breve revisão será apresentada, visando descrever os últimos desenvolvimentos metodológicos e como essa metodologia pode ser utilizada na formulação de hipóteses relacionadas a eventos de saúde com o intuito de descobrir sua etiologia. Todos os procedimentos serão apresentados; no entanto, somente o mapeamento de doenças será exemplificado, considerando a taxa de fecundidade no ano de 2003 na cidade de Porto Alegre.

Nessa dissertação será apresentada uma breve revisão de literatura para apresentar a Epidemiologia Espacial ao leitor. A Epidemiologia Espacial é dividida em três áreas: Mapeamento de Doença, Detecção de Conglomerados Espaciais e Correlação Ecológica. Todas elas serão brevemente comentadas na revisão, embora a aplicação será feita somente com a técnica de Mapeamento de Doença. Sua aplicação será discutida em forma de artigo. Uma breve revisão sobre as taxas de fecundidade, evidenciando o panorama brasileiro, será apresentado.

No artigo, a técnica de mapeamento de doença será apresentada utilizando a taxa de fecundidade como exemplo para se demonstrar as questões metodológicas comentadas. Os principais achados do trabalho estão presentes no artigo.

Em anexo encontra-se o projeto de pesquisa submetido e aprovado pelo Comitê de Ética da Universidade Federal do Rio Grande do Sul e pelo Comitê Interno da Secretaria de Saúde da Prefeitura Municipal de Porto Alegre, bem como a programação computacional, para aqueles que desejarem utilizar essa metodologia com outros desfechos.

II Revisão Bibliográfica

2.1 Epidemiologia Espacial

A distribuição geográfica de eventos relacionados à saúde humana tem uma longa tradição em epidemiologia, e o mapeamento de incidências (prevalências) de doenças ou eventos de saúde avaliados espacialmente, juntamente com aspectos ambientais, comportamentais ou socioeconômicos, tem recebido considerável desenvolvimento metodológico nas duas últimas décadas.

Entretanto, ainda não há um padrão para a nomenclatura utilizada na classificação desses estudos. Há autores que os classificam como estudos de Epidemiologia Espacial (15;17), outros usaram inicialmente o termo “Epidemiologia Geográfica” (18) ou ainda “Epidemiologia Ambiental” (19). Há também autores que utilizam o termo mais geral “Estatística Espacial”, e o exemplificam com aplicações em Epidemiologia (9;20). Atualmente, o termo “Epidemiologia Espacial” está predominando na bibliografia revisada, como, por exemplo, Bailey (21). Nesta dissertação, adotaremos o termo “Epidemiologia Espacial”, por acharmos este termo mais apropriado ao conjunto de técnicas que vamos descrever.

A Epidemiologia Espacial tem por objetivo incorporar variações espaciais no risco ou na incidência / prevalência de um determinado evento de saúde, descrevendo seu padrão geográfico ou sua distribuição espacial.

Nesse tipo de estudo, a localização espacial do evento é a principal informação a ser investigada. Na maioria dos estudos espera-se que pontos próximos se assemelhem mais do que os afastados. Segundo alguns autores (22;23), quando a localização de um determinado fenômeno modifica sua distribuição de probabilidade, propiciando que pontos vizinhos apresentem probabilidades similares de ocorrência, estamos num caso especial de um processo estocástico denominado de Processo Estocástico Espacial.

Na Epidemiologia Espacial, através da avaliação da distribuição geográfica e de informações relativas ao padrão de proximidade, pretende-se descrever, quantificar e principalmente explicar as variações nas taxas. A questão central desse tipo de estudo será investigar a dependência espacial que poderá ainda estar associada a causas sociais ou econômicas. Essa investigação contribuirá para uma melhor compreensão da etiologia da doença.

Em alguns tipos de eventos de saúde, considera-se que indivíduos residentes próximos uns dos outros sejam mais suscetíveis à contaminação de uma doença, e então a busca por conglomerados ou *clusters* de casos ou áreas pode ser interessante.

Conforme Elliott e Wartenberg (11), a partir de estudos espaciais urbanos, observa-se que as pessoas tendem a agrupar-se de forma sistemática. Por exemplo, pessoas com alto padrão sócio-econômico tendem a morar próximas umas das outras, em bairros com alta renda familiar, com boas escolas e também com melhores condições de saneamento básico. E, similarmente, pessoas de baixo padrão socioeconômico tendem a morar em locais com piores condições. Esta organização socioespacial auxilia na predição do risco de doenças e, portanto, será necessário considerá-la para avaliar e auxiliar a compreensão de determinada doença.

Há, assim, grande interesse na epidemiologia em estudar a distribuição espacial de eventos inerentes à saúde e relacioná-los às características da população em estudo, buscando uma possível explicação para o problema em questão. A distribuição geográfica de um determinado evento será estudada através de mapas considerando informações populacionais. Será fundamental definir quem é a população em risco ou exposta para se estimar os casos esperados da doença naquela localização e depois obter a taxa global. No processo de obtenção dos dados espaciais, uma revisão da digitalização será necessária, retirando-se os casos repetidos e mal relatados. Também é importante definir fronteiras temporais e espaciais. Estimativas para áreas desconhecidas também poderão ser realizadas. Segundo Assunção (20), os mapas são mais úteis quando podemos descrevê-los, compará-los e interpretá-los, e, para isso, a inferência estatística será essencial.

A Inferência Estatística também é necessária na investigação do verdadeiro modelo biológico. Esse modelo considera a relação entre o desfecho (distribuição geográfica da doença), com as características da população representadas pela amostras, e os fatores socioeconômicos presentes no local para estimar o verdadeiro risco. Os estudos tradicionais da epidemiologia também avaliam essa relação, porém não consideram a dependência espacial. Dessa forma, a epidemiologia espacial pode servir como hipótese inicial para os outros estudos.

Um exemplo pioneiro na utilização de técnicas de análise espacial, e também precursor da Epidemiologia Espacial, é o estudo realizado no século XIX por John Snow. A epidemia de cólera foi melhor compreendida depois que ele relacionou em um mapa o local da residência dos óbitos ocasionados pela doença com as bombas de abastecimento de água na cidade (12;24).

A epidemiologia espacial pode ser dividida em três grandes áreas segundo a classificação presente no livro denominado *Statistical methods in Spatial Epidemiology* (15):

- **Mapeamento de doenças (*Disease Mapping*)** - 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 áreas mais afetadas. Para isso, utiliza-se modelos estatísticos que geram mapas limpos de perturbações externas, buscando-se descobrir a estrutura subjacente dos dados, isto é, a verdadeira distribuição espacial da população em estudo. Serve para definir e monitorar epidemias, gerar dados basais para o controle de doenças e revelar mudanças no padrão das doenças ao longo do tempo (25). 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 (17).
- **Detecção de *clusters* ou identificação de *clusters* de doença (*Cluster Detection, Disease Clustering*)** - avalia se a doença mapeada está formando *clusters*, agrupamentos de casos, em determinados locais ou se aglomerando, de forma geral. Essa formação pode ser monitorada ao longo do tempo (26). Há dois tipos de *clusters*: os globais e os locais. A formação de *cluster* global (em todo o mapa) tem como hipótese nula a distribuição populacional aleatória, e, como alternativa, uma distribuição probabilística modificada pela presença 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 (susceptibilidade genética na população) ocasionando áreas geográficas com risco significativamente elevado, e seu monitoramento é mais uma questão de vigilância epidemiológica.
- **Estudos de Correlação Geográfica (*Geographical Correlation Studies*)** - avaliam relações ou correlações geográficas entre a distribuição espacial da

variável em estudo e várias medidas ou fatores explicativos. Estas análises são feitas normalmente em nível de unidades espaciais agregadas. Esses modelos são usados, por exemplo, para relacionar eventos de saúde com a distribuição espacial de fatores ambientais relevantes cuja exposição pode causar risco, com o intuito de controlar e prevenir maiores riscos à população. Os modelos estatísticos têm por objetivo fazer a predição espacial ou espaço-temporal desses riscos. Estes modelos são também denominados de Estudos de Correlação Ecológica, ou simplesmente Estudos Ecológicos (18).

No livro de Elliott et al. (17), há uma quarta área na classificação para Epidemiologia Espacial, a qual é a **avaliação do risco relacionado a um ponto ou a uma fonte específica**. Esta área está associada à investigação do aumento do risco em determinadas regiões em relação a uma fonte, que pode ser um ponto (como uma indústria química), uma linha (uma estrada) ou uma região. Pode-se investigar a existência de *clusters* locais, como será descrito mais detalhadamente na seção específica de *clusters*, mas, a nosso ver, esta quarta área está já classificada nas três grandes áreas definidas acima.

2.1.1 Mapeamento de Doença

Segundo diversos autores (17;18;20;21;26), o mapeamento de doenças modela os dados na forma de taxas, construídas através da contagem de casos pela população em risco em cada área, juntamente com seus descritores sócio-econômicos e com medidas de exposição desses grupos. Assim, a distribuição geográfica subjacente e o padrão da doença serão levados em consideração.

O processo de monitoramento de epidemias para identificar alterações preliminares ou não usuais a partir de valores extremos nos indicadores conduz naturalmente ao estudo do mapeamento de doenças. No entanto, em populações muito pequenas, valores extremos podem ocorrer por várias causas, como variações populacionais em relação a sexo, idade, tamanho da área, entre outras, denominadas de flutuações aleatórias. Essas flutuações levam a uma interpretação equivocada do problema.

Para evitar esses vieses de interpretação, é necessário mensurar o denominador do indicador adequadamente, ou seja, definir adequadamente a população em risco, pois

quando a definição dos limites da área em estudo não está adequada (divisões muito pequenas), a população em risco torna-se muito pequena. Como Bailey (21) definiu, controlar as flutuações aleatórias por métodos estatísticos permite o mapeamento do verdadeiro risco da doença sem vieses aleatórios.

Como exemplos do uso do mapeamento de doenças, temos o estudo de Shimakura et al. (27), que descreveu a mortalidade infantil em Porto Alegre, o trabalho de d'Orsi et al. (28), que avaliou a taxa de natalidade no Rio de Janeiro, e o trabalho de Mollié (29), que relata a doença de Hoodkin na França.

Após a identificação do tipo de dado a ser investigado (pontos ou áreas), inicia-se o mapeamento, com a avaliação do risco através de um mapa. Normalmente adota-se cores mais escuras para os locais onde há maior risco ou incidência. Entretanto, se com esse método exploratório não conseguirmos identificar áreas com maior risco ou com maior incidência, será preciso utilizar modelos inferenciais para identificar as áreas com riscos estatisticamente significativos.

Representação estatística - métodos exploratórios

As análises exploratórias podem ser usadas para resumir a estrutura dos dados e gerar hipóteses relevantes. Existem diversas maneiras de se explorar a distribuição espacial do número de casos em uma determinada região. Segundo Lawson (30), a forma mais simples de se descrever a distribuição espacial de um evento de saúde é apresentar os casos num mapa, sendo essa uma representação bruta. Entretanto, esse mapa não revela corretamente a variação espacial do risco, pois as variações na densidade populacional não são controladas, ocasionando riscos espúrios em determinados locais. A melhor representação será obtida quando o efeito da população em risco for considerado através da construção da taxa específica de fecundidade.

A medida-resumo normalmente utilizada nesses processos é a razão de riscos padronizada, sendo interpretada na forma de um risco relativo. Ela será descrita utilizando a definição apresentada por Assunção (20). Considere inicialmente as diferenças no tamanho entre as populações das áreas. Divide-se o mapa em n área indexadas por i , $i=1, \dots, n$. Seja y_i o número de eventos na área i e N_i sua população em risco. A taxa per capita na área i , será definida por $r_i = \frac{y_i}{N_i}$.

A razão entre valores observados e esperados para dados de mortalidade é chamada de razão padronizada de mortalidade ou morbidade (*Standardised Mortality*

Ratio, SMR) e equivale a um risco relativo médio estimado para cada região. Os valores observados representam a contagem de eventos em cada área do mapa (y_i) e os valores esperados (E_i) são estimados através do risco médio total da região, $r = \frac{\sum_i y_i}{\sum_i N_i}$, multiplicado pela população total da área, (N_i), $E_i = r N_i$. A SMR para cada área i é calculada por $\hat{\lambda}_i = \frac{Y_i}{E_i}$. Quando a razão padronizada for igual a um, o número de casos observados naquela região é igual ao esperado. Quando for maior que um, indica um número maior de casos do que o esperado naquele local. Se a $SMR_i = 2$, por exemplo, então a área i teria um número de casos duas vezes maior que o esperado se seu risco fosse idêntico ao risco médio da região. O erro padrão da estimativa de SMR_i é inversamente proporcional ao tamanho da população estudada na área em questão.

Segundo Assunção (20), quando há variação no risco através da idade e sexo, é necessário considerá-la gerando uma padronização *indireta*. Seja i o índice da área, j o índice da classe de idade-sexo, y_{ij} o número de eventos que ocorreram entre pessoas da classe j na área i e N_{ij} o número de pessoas da classe j na área i . A taxa global em todo o mapa que refere-se apenas à classe de idade-sexo j é dada por $r = \frac{\sum_i y_{ij}}{\sum_i N_{ij}}$. Então, $E_{ij} = r_j N_{ij}$ é o número esperado de eventos na classe j na área i se o risco na classe j fosse constante no espaço. O número total de eventos esperados na área i se o risco de cada classe de idade-sexo é constante no espaço é dado por $E_i = \sum_j E_{ij}$, a soma dos números esperados na área i nas diferentes classes de idade-sexo. A SMR é então calculada como a razão entre o número observado de eventos e o número esperado, caso o risco fosse constante no espaço: $\hat{\lambda}_i = \frac{Y_i}{E_i}$.

As SMRs são bastante usadas na apresentação de mapas de doenças, com ênfase na taxa de mortalidade. Contudo, apresentam restrições, mencionadas e enfatizadas por alguns autores (20;26). A principal restrição é a possibilidade de ocorrer flutuações aleatórias em locais onde a população é pequena. Quando trabalhamos com unidades amostrais muito pequenas, como municípios pouco populosos ou unidades de área exíguas e com baixa densidade populacional, a população em risco é pouco representativa, fazendo com que o valor esperado de cada área seja pequeno. Valores esperados próximos de zero fazem com que a razão seja grande, independente do valor

observado. Nessa situação, as taxas de incidência serão muito instáveis. Conforme Assunção (20), o acréscimo ou decréscimo de um caso nestas áreas poderá causar mudanças drásticas nas taxas.

As oscilações do risco relativo não estarão associadas com as variações no risco subjacente, mas serão apenas flutuações aleatórias. Conforme Olsen et al. (31), quanto menor a população em risco, mais as estimativas serão influenciadas por flutuação aleatória.

Métodos Informais de Suavização

A primeira tentativa para corrigir as flutuações aleatórias da SMR ou das taxas brutas seria agregar mais as áreas, tornando-as mais populosas. Porém, com isso, perde-se a informação local do evento. Outras possibilidades seriam utilizar mapas de probabilidade, o Estimador Bayesiano Empírico Local de Marshall ou métodos não-paramétricos, como o *kernel*.

Os mapas de probabilidade são apresentados por Assunção (20) e são construídos com base na comparação do número de eventos observados em relação à média global de toda a região considerada constante. O número de casos y_i segue uma distribuição de probabilidade de Poisson com média igual a E_i , representando um risco constante no espaço em cada classe de idade-sexo. Seja X uma variável com a mesma distribuição que y_i , isto é, $X \sim \text{Poisson}(E_i)$. Define-se então:

$$\rho_i = \begin{cases} P(X \geq y_i) \text{ se } y_i \geq E_i \\ P(X \leq y_i) \text{ se } y_i < E_i \end{cases}$$

Assim, um valor de ρ_i muito próximo de zero indica que a taxa é muito alta ou muito baixa em relação ao valor esperado. A vantagem desse método apontada pelo autor é que ele considera a natureza estocástica dos valores observados, considerando sua variabilidade. Deste modo, taxas muito extremas baseadas em excessos aleatórios não serão consideradas. Para essas áreas, admite-se que o risco seja igual ao valor esperado. Como o autor coloca, há restrições na construção de mapas de probabilidades próximas de zeros em áreas muito grandes (ou muito populosas), dificultando sua interpretação, pois eles consideram a área homogênea. Entretanto, em muitas situações isso não acontece.

Os mapas ignoram o valor da taxa, de modo que duas áreas que apresentem a mesma SMR podem ter ρ_i muito diferentes. Diante dessas limitações, são pouco usados.

O método não-paramétrico de *kernel* também pode ser usado para controlar flutuações aleatórias, em especial nos processos pontuais. Ele considera o risco como uma medida de variação contínua na região de interesse, ou seja, o processo estocástico em questão é contínuo. Basicamente, calcula médias ponderadas dos valores vizinhos com base na distância entre eles (30). A distância entre pontos ou áreas vizinhos será avaliada pela largura da banda (h), sendo determinada conforme a distribuição espacial e os objetivos da análise. Quando os dados são agregados, h representa o diâmetro médio das áreas de interesse. Habitualmente, h pode ser determinado com informações *a priori* sobre os dados ou por métodos de validação cruzada, como apresentado por Shimakura et al. (27).

Segundo Câmara e Carvalho (32), a função de *kernel* é uma função bidimensional que compõe uma superfície sobre as unidades investigadas cujo valor será proporcional à quantidade de amostras por unidade de área. Essa função realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse. Outros autores (26;33) recomendam também o uso da função *kernel* para se estimar o risco em locais onde não há observação de casos.

O Estimador Bayesiano Empírico Local de Marshall consiste também em utilizar informações das áreas vizinhas para gerar estimativas médias do risco naquela região, de tal forma que o risco subjacente seja convergente para a média local da vizinhança. Este método é semelhante aos métodos frequentistas, e cada área terá seu valor reestimado através de uma média ponderada entre o valor medido e a taxa média local, sendo este peso inversamente proporcional à população da região (9). Esse estimador já está implementado no pacote estatístico R dentro da função “*Spded*”, recebendo o nome de “*EBlocal*”.

Ambos os processos, Bayesiano Empírico de Marshall e *kernel*, consideram a proximidade entre as áreas para suavizar a estimação do risco, admitindo que locais próximos apresentem as mesmas características, ou seja, os riscos subjacentes são similares na vizinhança. Assim, a estrutura de vizinhança deverá ser considerada, sendo representada na forma matricial.

Além do método de *kernel* ser um método não-paramétrico, ele depende só da proximidade. Já o método Bayesiano Empírico Local de Marshall, além da proximidade (denominada de autocorrelação espacial), também depende da variabilidade da taxa na região, e é um método paramétrico. Os métodos paramétricos são sempre preferidos em relação aos não-paramétricos.

Quando os dados estão apresentados na forma de pontos, estipula-se uma distância máxima para considerar pontos próximos e calcula-se a distância entre todos os pontos. Aqueles que apresentarem distâncias menores que a estipulada serão considerados vizinhos. Para dados em áreas, a proximidade será avaliada através da distância entre os centróides da região, considerando, assim, centros geográficos próximos como vizinhos.

Em termos epidemiológicos, essa proximidade será importante para o estudo, pois caracteriza o processo de contaminação da doença. Espera-se encontrar um maior número de pessoas expostas a uma doença de alto contágio em áreas geograficamente próximas umas das outras.

O método Bayesiano Empírico de Marshall traz vantagens em relação ao método não-paramétrico de *kernel*, principalmente na avaliação de dados em áreas. Entretanto, não pode considerar covariáveis no controle da variabilidade do processo. Outra proposta seria aplicar um modelo hierárquico considerando estimação totalmente Bayesiana, que leva em conta as características basais do processo contidas nas distribuições *a priori*.

A modelagem tem como suposição básica a homogeneidade da taxa naquela região específica. Estatisticamente, investiga-se qual a probabilidade da taxa extrema ser afetada por flutuação aleatória. Uma vantagem em relação às outras técnicas de suavização (Bayesiana Empírica de Marshall ou *kernel*) é que ela ainda pode considerar variáveis explanatórias no controle da variabilidade. A modelagem Bayesiana será descrita na próxima sessão.

Modelagem no mapeamento de doença

Os métodos descritivos são úteis nos estágios iniciais da avaliação, mas quando hipóteses mais substantivas estão disponíveis, a modelagem passa a ter mais vantagens. Segundo alguns autores (15;26;33), os modelos básicos consideram o princípio de verossimilhança. Eles normalmente seguem uma das três pressuposições: (i) a população em estudo é independente da propensão da doença, ou seja, não há autocorrelação; nessa situação, o risco da doença é considerado homogêneo; (ii) a população em risco apresenta uma distribuição espacial contínua; o processo não apresenta estacionariedade de primeira ordem pela ação de fatores ambientais ou outras covariáveis; (iii) os eventos (casos) são únicos e não estão fixos em determinados locais. Quando se considera a pressuposição (i), os eventos são modelados através da verossimilhança, condicionados na realização do número de eventos e é normalmente válida para doenças não infecciosas. Na presença de autocorrelação, esta deve ser considerada na análise. A pressuposição (ii) permite que a verossimilhança seja construída com base na população em risco e será válida conforme a escala utilizada. A pressuposição (iii) está relacionada com doenças raras e será violada quando o domicílio tiver mais de um caso. Na pressuposição (i) considera-se que a variação espacial foi gerada por um processo discreto e a pressuposição (ii) admite um processo estocástico contínuo.

Na avaliação de dados pontuais considera-se um modelo de variação espacial contínua — pressuposição (ii) — admitindo-se um processo estocástico de pontos amostrais medidos em todos os locais. Os modelos serão usados com a fim de dissolver os limites urbanos impostos erroneamente aos dados, passando-se a trabalhar com superfícies contínuas (11). Um exemplo seria desconsiderar a divisão por setor censitário e considerar regiões urbanas densamente povoadas. O modelo considerado é denominado Modelo Aditivo Generalizado (GAM), como proposto por Shimakura et al (27). Esse modelo também é recomendado para um desenho de caso-controle.

As estimativas espaciais do risco serão geradas através do logaritmo do risco em uma determinada localidade s , $\rho(s) = \log\left\{\frac{\lambda_1(s)}{\lambda_2(s)}\right\}$, a partir de uma constante aditiva proposta em um modelo simples atribuído a variáveis binárias. Matematicamente, esse modelo será representado por:

$$\log\left\{\frac{\rho(\mathbf{s})}{1-\rho(\mathbf{s})}\right\} = \frac{q_1\lambda_1(\mathbf{s})}{q_1\lambda_1(\mathbf{s}) + q_2\lambda_2(\mathbf{s})} + c,$$

onde q_1 e representa a proporção de casos registrados, q_2 a de controles, λ_1 e λ_2 representa a intensidade de casos e dos controles respectivamente, $c = \log\left(\frac{q_1}{q_2}\right)$ constante.

Essa constante aditiva c representa uma função suave das coordenadas espaciais s , considerando a informação da vizinhança. O modelo tem como hipótese nula que o risco é constante em toda a região. Quando essa função aditiva for igual a zero, o modelo se reduz a um modelo de regressão logística usual. A estimativa dessa função pode ser obtida através do método de *kernel*, como demonstrada por Shimakura et al. (27), ou admitir que ela seja um processo estocástico, o qual necessita ser estimado por modelos geoestatísticos conforme Diggle (33) sugere.

O modelo básico para dados agregados considera a região dividida em n subáreas (bairros, setores censitários, municípios, entre outros). A variável resposta será o número de casos em cada bairro, setor censitário ou município. Admite-se que os números de casos de cada sub-área (Y_i) são mutuamente independentes e apresentam uma distribuição discreta de probabilidade. Os modelos são denominados de CAR (*Conditional Auto Regressive*) e se equivalem a uma regressão linear com o acréscimo de um efeito aleatório condicionado na vizinhança (9). No entanto, esses modelos exigem que a variável resposta tenha distribuição normal. Por se tratar de contagens, outras distribuições devem ser usadas, e o modelo passa a ser visto na sua forma genérica de um Modelo Linear Generalizado (MLG). Nesse modelo pode-se considerar simultaneamente a estimativa dos parâmetros relacionados com os fatores socioeconômicos e a estabilização dos riscos.

Segundo Cordeiro e Lima Neto (35), os MLG podem ser usados para avaliar a relação entre o risco da doença em determinadas áreas e os fatores socioeconômicos. Um MLG é definido pela distribuição de probabilidade da variável resposta pertencente à família exponencial, pelo conjunto de variáveis independentes que descrevem a estrutura linear do modelo e pela função de ligação que relaciona a variável resposta com as variáveis independentes. O MLG será formado por duas componentes: uma aleatória, representando a variável resposta por meio de um vetor de médias e sua distribuição subjacente, e outra sistemática, composta pelas variáveis explicativas, mais

uma função monotônica diferenciável, conhecida como função de ligação, que relaciona as duas componentes.

A parte aleatória de um MLG no caso de mapeamento de doenças, em geral, supõe que a variável resposta Y tenha como distribuição subjacente a distribuição de Poisson (doenças raras) ou a distribuição binomial. A função de ligação a ser usada quando Y tem distribuição de Poisson será a logarítmica. Já no caso da distribuição binomial, será usada a função logística (35).

No modelo de Poisson admite-se que a média do processo seja condicionada ao número de casos nos bairros e que a população em risco já é previamente definida. Nesse modelo admite-se que os casos sejam mutuamente independentes, mas quando o modelo binomial é utilizado, admite-se que a variável resposta é binária.

Em estudos de pequenas áreas, a estimação do verdadeiro risco é dominada pela variabilidade amostral que é formada por duas componentes: o efeito aleatório não-espacial, u_i , e o efeito espacial, b_i , na área i . A primeira parte, u_i , representa efeitos de pequena escala, diferenças dentro de um mesmo bairro e variações não mensuradas. Estes efeitos não são compartilhados pela vizinhança, sendo modelados como efeitos aleatórios independentes. A segunda parte, b_i , possui uma estrutura espacial e representa efeitos de larga escala, variação entre bairros ou devido a variáveis não observadas. Espera-se que o risco em bairros vizinhos seja semelhante. Na ausência da componente espacial, a variabilidade passa a ser não-estruturada, e, na sua presença, estruturada. A utilização das duas componentes gera uma suavização global e outra local para o risco.

Uma alternativa para controlar a parte da variação não mensurada da taxa em cada bairro é considerar a diferença populacional dos bairros no modelo como um *offset* (variável explanatória com coeficiente de regressão conhecido, constante). A substituição apresentada por Diggle (33) de se utilizar o número de casos esperados no bairro (E_i) ao invés do tamanho da população em risco (N_i) passa a investigar o risco de se apresentar o evento de saúde dado a sua localização espacial. O estimador de máxima verossimilhança $\hat{\lambda}_i$ do risco relativo λ_i é a SMR que possui variância inversamente proporcional ao número de eventos esperados.

Devido à composição da variabilidade do risco, um modelo hierárquico é adotado, gerando um estimador mais robusto para os parâmetros, suavizando, assim, as estimativas do risco. O modelo hierárquico proposto por Wakefield et al. (36) vem sendo o mais usado no mapeamento de doenças.

Freqüentemente, a estimação dos parâmetros em modelos é feita considerando o método de máxima verossimilhança, abordagem clássica (21;30). Porém, no contexto de mapeamento de doença, a modelagem é mais complexa com um grande número de parâmetros a ser estimado, dificultando a utilização dos métodos tradicionais. Uma alternativa bastante adotada é usar os princípios de Inferência Bayesiana para estimação dos parâmetros do modelo.

Descreveremos o modelo hierárquico Bayesiano proposto por Wakefield et al. (36) para o risco, utilizando o detalhamento e as definições matemáticas apresentadas por Assunção (20). Esse modelo é resolvido em três estágios: primeiro, modela-se os casos observados como uma função do risco relativo de cada bairro. O segundo estágio modela a variação espacial do risco através da distribuição conjunta de todos os riscos com variáveis explanatórias. O segundo estágio depende, portanto, de parâmetros desconhecidos. No terceiro estágio ocorre a determinação de funções de probabilidade, conhecidas como *hiperprioris*, para os parâmetros desconhecidos, denominados de *hiperparâmetros*.

No primeiro estágio, no caso da taxa de fecundidade, o número de nascidos vivos observados em cada bairro i (y_i) é considerado como uma variável aleatória que apresenta uma distribuição de probabilidade de Poisson para sua ocorrência com valor esperado (μ_i), específico em cada bairro. Se a taxa de fecundidade na adolescência fosse constante em toda a cidade, a média de cada bairro seria representada pelo produto entre a taxa média global e a população de cada bairro, $\mu_i = r N_i$. Entretanto, ela não é a mesma, pois a estrutura populacional é diferente. Considerando essas diferenças, y_i terá distribuição de Poisson com média proporcional ao risco em cada área, $\mu_i = \lambda_i N_i$. O parâmetro λ_i , $i = 1, \dots, n$ quantifica o risco de cada área i . Supõe-se, então, os valores observados em cada bairro, y_1, \dots, y_n , condicionados no seu risco, $\lambda_1, \dots, \lambda_n$, apresentam distribuição de Poisson com média $e_i \lambda_i$.

$$Y_i | \lambda_i, e_i \sim \text{Poisson}(e_i \lambda_i) \quad i = 1, \dots, n$$

Y_i representa os casos; λ_i representa o risco relativo e e_i representa o valor esperado em cada bairro.

No segundo estágio, a modelagem será realizada considerando a dependência espacial entre os riscos e a distribuição conjunta de todos os riscos com variáveis explanatórias, de modo que as regiões próximas tenham riscos semelhantes.

Para eliminar problemas nas soluções matemáticas, o logaritmo da média do risco será avaliado. O resíduo de cada área representará um fator de risco não medido, comum a todos os indivíduos nesse bairro i que não exibe um padrão espacial. Matematicamente, essa relação será escrita como:

$$\log(\lambda_i) = \beta_0 + \beta X + u_i + b_i,$$

onde λ_i representa o risco relativo no bairro i , β_0 é o intercepto do modelo, representando o logaritmo do risco relativo global, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ representam os coeficientes das covariáveis, u_i é o efeito aleatório e b_i é o efeito espacial.

Em se tratando de um processo Bayesiano, todos os parâmetros do modelo acima serão considerados como variáveis aleatórias e apresentam uma distribuição de probabilidade. Essa distribuição de probabilidade será especificada pelo pesquisador, considerando seus conhecimentos iniciais do processo, sendo denominada de distribuição de probabilidade *a priori*.

A função de verossimilhança e a distribuição *a priori* representam duas fontes de informação para o processo e combinadas caracterizam a distribuição *a posteriori*. A distribuição de probabilidade *a posteriori* tem como objetivo revelar o comportamento do parâmetro de interesse depois que os dados foram observados, gerando pressuposições atualizadas a respeito dos parâmetros. A função de verossimilhança revelará informações do parâmetro com base nos dados, enquanto a distribuição de probabilidade *a priori* informa crenças iniciais a respeito dos parâmetros. Na presença de amostras pequenas, as informações *a priori* dominam os resultados *a posteriori*. Já na presença de grandes amostras, a função de verossimilhança dos dados dominará os resultados *a posteriori*.

Será necessário determinar a distribuição *a priori* dos parâmetros u_i e b_i . Supõe-se que os u_i são independentes e apresentam distribuição normal com média 0 e variância $\frac{1}{\tau_\phi}$. O parâmetro τ_ϕ é geralmente desconhecido e controla o grau de dispersão dos efeitos aleatórios não estruturados espacialmente. Se τ_ϕ for pequeno, haverá grande variabilidade dos efeitos u_i em torno de uma média comum igual a zero, ou seja, uma grande variabilidade dos riscos relativos dentro de um mesmo bairro. Alternativamente, quando τ_ϕ for grande, a variabilidade dos riscos relativos é menor.

Já o segundo componente, b_i , tem uma estrutura espacial e é definido de modo a estimar áreas vizinhas sendo mais semelhantes em termos de risco relativos. Essa estimação é feita considerando uma distribuição condicional que avaliará a dependência espacial, atribuindo pesos positivos para o conjunto de bairros vizinhos e zeros para os demais bairros. Sua distribuição é um campo aleatório de Markov conhecido por CAR. A distribuição CAR do efeito de vizinhança também dependerá de um parâmetro adicional τ_0 que mede o grau de similaridade espacial entre os bairros vizinhos.

O terceiro estágio inicia-se com a determinação das *hiperprioris* para os *hiperparâmetros*: o intercepto β_0 , os coeficientes de regressão β , e a variabilidade do efeito aleatório τ_ϕ e do efeito espacial τ_0 . Esses parâmetros adicionais serão estimados com os dados e também serão considerados como variáveis aleatórias, apresentando distribuições de probabilidade. Habitualmente, escolhe-se uma distribuição Gama com parâmetros fixos e conhecidos, $\Gamma(\varepsilon, \varepsilon)$, com ε sendo escolhido bem pequeno, tal como $\varepsilon=10^{-3}$, para τ_ϕ e τ_0 , além de supor que eles são independentes. Para os β supõe-se uma distribuição normal com média zero e variância grande.

Um exemplo dessa modelagem Bayesiana foi proposto por Bailey et al. (37) para identificar áreas de sub-registro de hanseníase em Olinda, Pernambuco, utilizando uma covariável que representa um indicador de carência social.

Segundo Assunção (20), a modelagem Bayesiana para o mapeamento de doenças é a mais usada, pois ela modela situações extremamente complexas incorporando restrições ou dependência estocástica de forma simples e possui um método computacional de estimação: os algoritmos MCMC (Simulação de Monte Carlo via Cadeias de Markov), recentemente desenvolvidos para otimizar o esforço computacional necessário para a estimação dos parâmetros.

2.1.2 Detecção de *Clusters*

A avaliação de *clusters* ou seja, aglomerados de unidades espaciais, pode revelar áreas de risco elevado em meio a regiões de risco constante, indicando quais locais necessitam maior atenção em nível de saúde pública, estando frequentemente associados com uma fonte causadora, como uma indústria, ou agente causador doenças ou epidemias. Utilizando-se técnicas estatísticas de análise de padrão de pontos, pode-se evidenciar espacialmente as causas de uma determinada doença.

Segundo Olsen et al. (31), esse tipo de análise é feito *a posteriori* e não deve ser usada para o desenvolvimento de hipóteses iniciais. Seu principal objetivo é avaliar se a distribuição espacial de pontos exibe um comportamento sistemático ou estão aleatoriamente distribuídos no espaço.

Segundo a literatura (31), “*Cluster* é um grupo geográfico limitado, de ocorrências suficientes em tamanho e concentração, que não pode ter ocorrido por acaso”. Um *cluster* espacial representa um conjunto de eventos ou de taxas no espaço que são semelhantes em locais próximos entre si (18). Como, por hipótese, admite-se que o risco é constante em toda a região, suspeita-se que há formação de *clusters* quando o número de casos está acima do valor esperado para aquela região. Entretanto, a maneira inferencial de se avaliar a formação ou não de *clusters* é através de testes estatísticos. Os testes avaliam se a distribuição observada da doença pode ser representada pela completa aleatoriedade espacial (hipótese nula) ou não. Quando a hipótese nula é aceita, ou seja, não há formação de *clusters* em determinadas áreas, a observação de muitos casos em determinado ponto é considerada normal e o estudo passa a ser do tipo de mapeamento de doença como comentado anteriormente.

Na década passada, a investigação de *clusters* de eventos era um grande desafio para a saúde pública. Centros de Controle de Doenças eram procurados para tratar de agrupamentos de doenças locais, normalmente quando a epidemia já se encontrava em um estágio mais avançado. Muitas dessas investigações começaram a desenvolver um novo conceito em saúde pública, envolvendo uma ampla gama de doenças, como leucemia e até suicídio. A relação da doença com o meio ambiente e a exposição ocupacional passou a ser considerada, tornando assim a avaliação de *clusters* mais complexa, se comparada com a avaliação de uma doença infecciosa (38).

Um exemplo pioneiro na investigação de *clusters* espaciais que, segundo Assunção (20), causou insatisfação entre os estatísticos e epidemiologistas, foi a investigação de leucemia infantil ao redor de uma instalação nuclear de Sellafield (Inglaterra). A insatisfação ocorreu devido ao fato de que sua aglomeração não ser diagnosticada pelo serviço de vigilância epidemiológica, mas sim por um programa de televisão. A principal motivação foi a ocorrência de 4 casos de leucemia infantil registrados no local entre 1968 e 1982, quando eram esperados apenas 0,2 casos neste período. O aumento de casos na região levou a criação de um sistema de coleta de dados de incidência e mortalidade em pequenas áreas geográficas.

Conforme diversos autores relatam (33;39;40), há dois principais motivos para a ocorrência de um *cluster*: presença de um número suficientemente grande e concentrado de casos em um determinado local, não ocorrendo por acaso, devido à presença de uma fonte física (agente poluidor, p.ex.), ou investiga-se uma possível relação entre os pontos de aglomeração e um mecanismo social ou biológico ou até genético local.

Ambas as situações apresentarão uma dependência positiva. Assim, a probabilidade de existir um caso num local x onde já tenha um caso próximo dele será maior do que num outro local onde não apresente caso algum. Um evento formará *cluster* quando há grande variação espacial no risco.

Os métodos para identificação de *clusters* são divididos em dois grupos. No primeiro, os dados são coletados para investigar se o excesso de casos está sendo produzido por uma fonte poluente. Nessas condições, os testes estatísticos avaliam a presença de *clusters* locais, considerando, como hipótese nula, a completa aleatoriedade entre os *clusters*. Já no segundo grupo, o objetivo é identificar áreas de maior risco, sem especificação prévia de quais ou quantas áreas estão presentes. São chamados de testes globais de *clusters*.

Em ambos os métodos, os dados utilizados podem ser pontos ou áreas. É indispensável controlar fatores como a distribuição populacional e outras covariáveis que podem gerar *clusters*. Na prática, como Elliot e Wartenberg (11) salientam, quando não se tem muita informação a respeito dos dados em áreas, modela-se pontos. Segundo Bailey (18), os *clusters* são usados em estudos de seguimento.

Dessa forma, a avaliação dos *clusters* será feita em etapas. Inicia-se com uma idéia de que determinado agente está aumentando o número de casos da doença naquela região. Dessa forma, as hipóteses serão formuladas e testadas. As estimativas de risco são representadas num mapa revelando os locais onde há aumento do risco. Utiliza-se a forma usual de se estimar o risco, obtendo-se valores esperados para aquela região. Compara-se os valores observados com os esperados, calculando-se as razões padronizadas (SMR) conforme comentado anteriormente. A escolha de qual método utilizar depende do objetivo da investigação. Se o interesse está na investigação de um possível aumento de casos em um local, utilizam-se os métodos para detecção de *clusters* locais. Já quando não se tem informação sobre um agente, mas pretende-se investigar um aumento de casos em determinadas áreas, considerando o risco constante em toda a região, utilizam-se os métodos globais de avaliação.

Testando a hipótese inicial – análise do padrão espacial

Em termos estatísticos, avalia-se o padrão espacial dos pontos associados à ocorrência de um evento como sendo um processo estocástico. A localização dos eventos será baseada em suas coordenadas geográficas (latitude e longitude ou representação cartesiana). No processo estocástico espacial, a dependência espacial é a principal relação a ser investigada, estando presente em todas as direções da área estudada, ficando mais fraca à medida que aumenta a distância na localização dos dados (23). Assim, o padrão espacial ou pode ser descrito por *clusters*, ou ser aleatório ou estar distribuído de forma regular. Um padrão de *clusters* máximo será aquele no qual todos os pontos apresentam a menor distância. Alternativamente, um padrão regular será aquele onde os pontos estão igualmente espaçados através da área em estudo (41). E no padrão aleatório não há evidências nem de máxima aglomeração, nem de regularidade espacial.

Como toda análise estatística, a análise de padrão de pontos exige um modelo teórico sobre o qual são feitas todas as pressuposições estatísticas. O modelo teórico mais simples é admitir que os pontos exibam um comportamento espacial aleatório. Assim, considera-se que as ocorrências em qualquer sub-área não estão correlacionadas e são homogêneas, estando associadas a uma distribuição de probabilidade de Poisson. Nesse caso, os eventos ocorrem de forma independente uns dos outros, tendo igual probabilidade de ocorrência em toda a região (23).

Diversos métodos avaliam o padrão espacial dos pontos. Esses métodos podem ser divididos em quatro grupos: métodos tradicionais, métodos que se baseiam na distância, métodos circulares e métodos que estimam o risco residual espacial. Os métodos tradicionais avaliam a superdispersão na série de casos através da dependência entre casos, utilizando principalmente o cálculo do valor esperado controlado pelos fatores de risco (sexo e idade quando modificam o risco).

Os métodos que se baseiam na distância avaliam a dependência entre casos através de um conjunto de dados, produzindo medidas estatísticas que a descrevem. Os métodos circulares avaliam se o excesso de casos pode ter ocorrido por acaso.

Métodos Tradicionais – Testes de Homogeneidade

Os métodos tradicionais são assim denominados por serem utilizados desde os primórdios da Estatística. Eles avaliam a hipótese nula de completa aleatoriedade espacial, onde a taxa é constante em toda a região, contra a hipótese da não aleatoriedade, onde a taxa é diferente em pelo menos uma sub-área. A avaliação dessa hipótese é feita considerando uma estatística de teste que compara os valores observados e esperados avaliando a variabilidade total dos eventos na região, a estatística de Qui-quadrado, também conhecida como teste de homogeneidade. Entretanto, este teste não revela nada sobre a dependência espacial do evento; só revela que o risco de se ter o evento é maior em pelo menos uma sub-área. Ele também não revela a localização exata do *cluster*. Grandes valores da estatística de teste indicarão a heterogeneidade espacial da região (31;40). O método proposto por Potthoff e Whittinghill é considerado o teste uniformemente mais poderoso para avaliar a heterogeneidade, pois considera o risco como sendo um efeito aleatório, através de uma distribuição conjunta de probabilidade gama (40).

Os métodos tradicionais só indicam que naquela região há formação de *cluster*, mas, para se identificar onde eles ocorrem (avaliar a dependência espacial), é necessária a avaliação através de outros testes mais apropriados.

Métodos que se baseiam na distância – Investigação de *Cluster* Global

Os métodos que se baseiam na distância avaliam o padrão espacial dos pontos e podem ser escritos como um produto cruzado de duas matrizes, onde uma reflete o vizinho mais próximo, a distância ou a relação adjacente, e a outra revela as informações de saúde (42). As funções de segunda ordem, chamadas de testes em conjunto, também serão usadas. Segundo Ripley (43), esses métodos são os preferidos dos estatísticos pela sua simplicidade.

Atualmente existem diversos métodos para se avaliar a dependência entre casos e a dependência espacial. No entanto, o método pioneiro iniciou-se com Ripley (43), através das funções K , que utilizam a estrutura dos momentos para avaliar o processo espacial. Com o passar dos anos, outros métodos, baseados na proximidade (distância) entre áreas, surgiram. Eles estão relacionados com as funções de segundo momento sugeridas por Ripley (43).

A avaliação da distância é feita através da matriz de vizinhança, isto é, diz-se que dois pontos são vizinhos quando a distância entre eles for pequena. Os métodos mais simples se baseiam na vizinhança binária e os mais complexos consideram outras funções para construir a matriz de vizinhança.

As funções K utilizam a estrutura dos momentos de primeira e segunda ordem do processo para avaliar sua distribuição espacial num mapa, considerando o processo como um processo estocástico. Assim, o número de pontos esperados em cada área descreve a estrutura de primeiro momento de um processo estacionário (43).

Em termos epidemiológicos, a utilização de estimadores de primeiro e segundo momento significa avaliar a distribuição espacial do evento através da investigação de possíveis modificações no número de eventos esperados (média do processo) e possíveis modificações locais, considerando-se para isso a estrutura de correlação espacial, ou seja, alterações na dependência espacial. A dependência espacial será calculada semelhantemente à covariância entre pares de eventos (23;32). Porém, pequenas modificações na média do processo são aceitáveis, quando se percorre a vizinhança de um determinado local.

Diz-se que um processo é estacionário quando apresenta média e variância constante em toda a região em estudo, não apresentando diferenças entre municípios, áreas ou pontos, ou seja, não há tendência na série. Em situações onde a covariância entre dois pontos quaisquer depende somente da sua localização, ou seja, a dependência espacial é a mesma em toda a região, diz-se que o processo está em isotropia.

Para um processo estacionário, os efeitos de segunda ordem dependem somente da direção e da distância entre os pontos, e não da sua localização. Assim, pode-se dizer que há uma ligação entre as propriedades de segunda ordem e a distribuição de distâncias entre pares de eventos. Dessa forma, encontrar a distribuição de distâncias entre vários eventos será uma boa ferramenta para se descrever a dependência espacial dos pontos (23).

A forma matemática de se quantificar a dependência espacial é feita através da autocorrelação espacial, ou seja, avalia-se uma possível correlação da mesma variável aleatória, tomada em diferentes locais no espaço. Como cada sub-área pode estar correlacionada com muitas outras, é necessário examinar-se a autocorrelação entre elas (31). A autocorrelação espacial dependerá da definição de vizinhança do local, podendo percorrer todas as direções (23).

Esse processo se assemelha aos modelos de séries temporais auto-regressivos da Teoria de Box & Jenkins, quando se comparam valores de uma amostra com os valores de seus vizinhos. No caso de séries temporais, as medições em um determinado período são comparadas com o momento anterior, identificando possíveis mudanças.

Existem diversos métodos para se medir a autocorrelação espacial. Um local ou ponto qualquer será definido como vizinho de outro se a sua distribuição conjunta de probabilidade, considerando todos os valores, depender somente daquele ponto, ou seja, a distância entre os dois pontos será a menor possível (23).

A distância média entre todos os pontos será a estatística de teste. Isso significa que o valor da distância do vizinho mais próximo serve como medida para se identificar o padrão. Se os pontos estiverem num padrão de *cluster*, a distância média entre vizinhos será pequena. Já num padrão aleatório entre pontos, a distância média entre vizinhos não se modifica, ficando próxima da média, e num comportamento regular, a distância entre os vizinhos é maior. Outra forma bastante usada considera o inverso da distância entre os centróides das áreas. A escolha do método dependerá do tipo de dependência espacial que se está buscando.

Na vizinhança binária, considera-se como vizinho somente aqueles bairros/áreas que fazem fronteira entre si. Esses bairros recebem o valor 1 na matriz de vizinhança e 0 os outros. Outros métodos mais sofisticados consideram o tamanho da fronteira ou as barreiras naturais (20;32;40;41).

Diversos métodos estatísticos foram desenvolvidos para se identificar *clusters* de taxas altas ou baixas em um mapa, melhorando sua significância estatística. Muitas áreas vizinhas dividem as mesmas características socioeconômicas, às quais a taxa da doença está normalmente associada. Todos eles consideram, de alguma forma, a autocorrelação e são descritos por diversos autores (23;34;40;44), conforme quadro 1.

O índice de Moran (45) é um dos mais usados para identificação de *clusters* globais na região. Apesar de ter sido desenhado para avaliar dados contínuos, esse índice também pode ser aplicado a dados agregados.

Quadro 1: Testes estatísticos para avaliar presença de *cluster* global

Nome	Características	Estatística de teste
Moran's I	-Semelhante ao coeficiente de correlação; - Medida de similaridade entre locais próximos; -Avalia modificações na taxa de cada área em relação à média global.	H ₀ : indep. espacial Dist. subjacente normal
Gaery	-Avalia dependência espacial através da proximidade entre áreas.	H ₀ : indep. espacial Dist. subjacente normal
Teste Não-paramétrico	- Construção de um ranking da taxa observada para avaliar dependência espacial; - Não é recomendado usar para a SMR.	H ₀ : indep. espacial Significância avaliada por Monte Carlo
Ipop	- Identifica desvios de uma distribuição aleatória, incorporando a variabilidade populacional; - Composto pela variação dentro e entre áreas.	
David-Moore (Variância relativa)	-Mede a superdispersão; -Calculado com os dados nos quadrantes e comparado com a igualdade entre média e desvio padrão (dist. Poisson).	H ₀ : $\sigma^2/\mu = 0$, ou seja, processo regular $D = \sigma^2/\mu$
Whittemore	- Mede superdispersão no caso de existir casos correlacionados; - Avalia a presença de risco constante utilizando a distância média entre pares de casos.	
Tango	- Mede proximidade entre áreas; -Utilizado na Geoestatística para descrever a estrutura de correlação do processo gaussiano.	

σ^2 – variância, μ – média

Métodos circulares - Investigação de *Clusters* Locais

Uma outra forma de se avaliar a autocorrelação é a estatística de rastreamento do tipo “*Scan*”. Essas técnicas investigam a possível existência de *clusters* em pequenas áreas de uma determinada região. Um dos primeiros métodos de rastreamento foi o GAM (*Geographical Analysis Machine*), criado em 1987. Ele se baseia na construção de círculos em torno da região em estudo e avalia, através de testes locais, a presença ou não de *clusters* espaciais (44;46).

Os métodos de *Scan* consideram o número máximo de eventos encontrados nos círculos de mesmo tamanho, como estatística de teste, comparando com uma intensidade subjacente uniforme, através do teste de Monte Carlo, sob distribuição aleatória, dentro do círculo. A estatística é calculada várias vezes, modificando-se o tamanho da população dentro do círculo, gerando um conjunto de testes independentes (40).

Um problema em utilizar esta técnica de quadrantes ou círculos, apresentada por Kuldorff e Nagarwalla (46), é que este método não apresenta um teste único para avaliar sua significância. Os agrupamentos encontrados estão correlacionados e o procedimento de Bonferroni, aplicado a múltiplos testes, é muito conservador, além do que, esse tipo de teste pode ser comparando aos testes de identificação de agrupamentos gerais na região, sendo, assim, tendencioso.

Então Kuldorff & Nagarwalla (46) propuseram um dos métodos mais usados atualmente, que considera o problema da inferência na identificação de *cluster*, não tem restrições quanto ao seu tamanho, utiliza métodos de máxima verossimilhança, apresenta uma hipótese alternativa bem definida, não avalia o local separadamente e pode ser aplicado a dados agregados ou não. O método identifica a localização de possíveis agrupamentos de casos em uma população com densidade espacial não homogênea e testa sua significância estatística, considerando uma população fixa. Esse processo consiste basicamente em calcular todos os conjuntos de círculos possíveis em torno de um ponto e, em cada círculo, a probabilidade do caso estar dentro (p) ou fora (q) do círculo é calculada. Se p é maior que q , então o círculo pode ser visto como um *cluster*. Com um teste estatístico, pode-se avaliar quão provável esse agrupamento é. Habitualmente, o valor p será calculado através de métodos de simulação, implementados nos pacotes estatísticos. O processo de simulação é descrito por Gómez-Rubio et al. (44).

Com a grande diversidade de testes estatísticos para avaliar a autocorrelação espacial, escolhe-se aquele que é mais poderoso para a identificação de aglomerados espaciais. Song & Kulldorff (47) elaboraram um estudo onde testaram o poder de diversos testes estatísticos para identificação de *clusters* espaciais e, segundo eles, deve-se utilizar testes tipo *Scan* para identificação de *clusters* locais. O método de *Scan* foi recentemente usado para se fazer vigilância epidemiológica de uma determinada doença. Para esse tipo de investigação, é necessário o monitoramento ao longo de vários anos, identificando-se, dessa forma, um possível aumento de casos em alguma sub-área (48).

Modelagem na detecção de *clusters*

Na maioria dos estudos, o objetivo é avaliar a formação de *clusters* na região, e, para isso, utilizam-se métodos de avaliação global. Além de avaliar a formação, um segundo passo é identificar em quais locais há *clusters* de casos, o que acontece normalmente próximos aos agentes. Isso será feito através dos métodos de avaliação local dos *clusters*. Entretanto, a investigação de *clusters* não é apropriada para descrever a causa na variabilidade da taxa, uma vez que, para isso, os estudos de mapeamento de doenças são mais apropriados.

Na busca pela causa das variações no risco, quando há outras variáveis atuando no modelo, a modelagem se faz necessária. Como esse não é o objetivo nessa investigação, ela nem sempre é realizada. Para aqueles estudos que pretendem investigar a causa das variações no risco na presença de um agente causador dos *clusters*, essa variação deve ser incorporada ao modelo. O agente causador será tratado como uma variável *dummy*. As áreas onde ele se localiza recebem o valor 1 e 0 as demais (49).

Os modelos não são muito usados, pois a presença de *clusters* faz com que os pontos observados sejam dependentes entre si, e considera-se que a taxa é constante em toda a região. Isso dificulta os procedimentos utilizados, pois a maioria dos modelos supõe aleatoriedade dos casos. Uma alternativa seria construir um modelo considerando efeitos mistos (aleatórios e fixos). Assim, a variabilidade será decomposta em uma variação dentro e entre. Os efeitos aleatórios representaram a variação extra, a má localização do caso, a superdispersão ou a correlação espacial entre as unidades espaciais (autocorrelação). A superdispersão será gerada pela presença dos *clusters*,

fazendo com que o desvio padrão global seja maior que a média global. O efeito aleatório pode ser estimado por métodos probabilísticos ou por interpolação (21;26).

A suavização da taxa poderá ser realizada pelo método de *kernel* ou pelo método não-paramétrico independente da presença dos *clusters*. O método de *kernel* avalia modificações no logaritmo do *odds* (calculado através da razão de risco avaliada nos casos e os não-casos) sobre a região. Esse método pode ser substituído pelo método da regressão binária (40), que é aplicado a dados pontuais.

2.1.3 Estudos de Correlação geográfica

Os estudos de correlação geográfica, também denominado de Estudos Ecológicos, têm por objetivo examinar relações entre variáveis ou entre fatores e desfechos, analisando-se as variações geográficas em grupos populacionais ou em unidades geográficas expostas. Estes fatores podem ser variáveis ambientais, socioeconômicas, demográficas e de estilo de vida, relacionados a desfechos de saúde. Tanto o desfecho como seus preditores, revelando a exposição ambiental, devem ser medidos a nível da unidade geográfica.

Às vezes, o erro de medida no efeito da exposição é menor quando abordado em grupo. No entanto, a correlação geográfica é também afetada pelos mesmos problemas encontrados no mapeamento de doenças, juntamente com complicações nas possíveis correlações entre as variáveis explanatórias (11).

Freqüentemente são usados modelos de regressão para explicar a variação na incidência da doença através de outras variáveis (11;32). Conclusões equivocadas podem acontecer por inúmeras razões, como: viés de especificação na suposição de um modelo não-linear para o risco de exposição individual em nível de grupo; presença de variáveis confundidoras dentro e entre grupo; erros nas variáveis; modificação de efeito e padronizações (50).

Modelagem na correlação geográfica

Os modelos para avaliar a correlação geográfica serão semelhantes aos modelos usados no mapeamento de doença. Porém, nesse tipo de estudo temos o desfecho agregado por área e informações limitadas sobre a exposição. A principal dificuldade é avaliar a adequabilidade do modelo (51).

Como exemplo de exposição considerada no modelo, podemos citar a exposição ambiental (medida no ar, na água ou no solo) e o estilo de vida medido em nível agregado. A exposição será considerada constante em cada área. A escolha de qual modelo utilizar, se um modelo aditivo ou um modelo multiplicativo, dependerá da especificação da doença e de seus fatores de exposição (50).

No modelo inicial, consideram-se as unidades de uma população divididas em sub-áreas. A cada unidade atribui-se um valor para a variável de exposição e um para a variável de desfecho, e o interesse geral é prever o desfecho em relação à variável de exposição em cada sub-área. Para estimar os parâmetros dessa regressão, não dispomos de informações individuais, mas sim médias gerais que representam a informação média de cada sub-área. A regressão ecológica é o procedimento usado para estimar os parâmetros desse modelo de médias, com apenas uma observação para cada sub-área. Assim, substituem-se os coeficientes locais de regressão correspondentes a cada indivíduo por parâmetros agregados (51).

Os modelos ecológicos serão usados em duas situações: compreender a relação a as variações geográficas entre o desfecho e o preditor, e estimar uma interação entre eles. A primeira situação será representada pela da estimação do coeficiente linear (β) de cada sub-área, e a segunda pela estimação do risco de cada sub-área, controlando pelas covariáveis (51).

2.2 Sistema de Informação Geográfica

Um Sistema de Informação Geográfica (SIG), segundo alguns autores (12;19;20), representa o conjunto de equipamentos e programas de computação que integram mapas e gráficos com uma base de dados sobre um espaço geográfico. Esses equipamentos são capazes de coletar, armazenar, manejar, analisar e visualizar informações georeferenciadas localizadas na superfície terrestre e representadas numa projeção cartográfica.

O funcionamento de um SIG ocorre em etapas. Inicia-se com o armazenamento em meio digital do dado na forma de um mapa específico. Em seguida, ele é referenciado por meio de uma coordenada geográfica, definindo sua posição. Utilizando-se de funções analíticas, os mapas e os atributos são manipulados, gerando duas saídas: uma em forma de tabela que representa um resumo dos dados e a outra cartográfica que representa os mapas encontrados (19).

Para representar bem o comportamento espacial do fenômeno, é necessária uma boa representação espacial dos dados, uma escala correta e acurácia. A representação dos dados depende das unidades amostrais que serão usadas pelo SIG, podendo estar na forma de *rasters*, pixels de imagens de um satélite usados no sensoriamento remoto, ou por vetores representando pontos, linhas ou polígonos, baseados na geometria contínua do espaço. A escolha da escala a ser utilizada requer cuidados especiais, uma vez que afeta o padrão espacial dos dados, podendo interferir na capacidade do estudo de identificar a relação entre a exposição e o desfecho. As escalas podem ser classificadas quanto ao tipo, sendo cartográficas (estão relacionadas com o tamanho da região em relação ao mapa real), geográficas (referente ao tamanho da área em estudo, sendo regional ou global), resolução espacial (representam as menores unidades perceptíveis) e escala operacional (escala a qual o processo de interesse ocorre). Já a acurácia identifica quão bem o SIG representa a realidade em termos de localização (real representação dos dados), atributos (quão corretas estão as informações contidas nos dados) e acurácia temporal (19). As informações populacionais podem ser obtidas junto ao Instituto Brasileiro de Geografia e Estatística (IBGE), Ministério da Saúde e Secretarias Municipais.

2.3 Taxa de Fecundidade na adolescência

A maternidade na adolescência não se constitui em fenômeno novo no cenário brasileiro. Acompanhando uma tendência internacional, ela assume o status de problema social de bastante interesse para a saúde pública. O aumento da gravidez e da maternidade na adolescência, em países em desenvolvimento, vem despertando interesse de pesquisadores.

Entre países industrializados, os Estados Unidos apresentam os maiores índices de maternidade na adolescência. Já a Itália e o Japão apresentam os menores. No ano de 1999, nos Estados Unidos, ocorreram 55 nascimentos para cada 1000 mulheres de 15 a 17 anos, enquanto que no Japão e na Itália esses números ficam em torno de 5 nascimentos para 1000 mulheres nessa faixa etária (52;53).

O Brasil encontra-se num período de transição demográfica, onde as taxas de mortalidade e de fecundidade vêm diminuindo ao longo dos anos. A taxa de fecundidade total reduziu 57% nos últimos 30 anos.

Em 1970, as mulheres brasileiras em período fértil tinham em média 5,8 filhos. No ano de 2002, esta taxa baixou para 2,15 filhos em média. Essa queda na fecundidade é justificada pela participação das mulheres no mercado de trabalho e pela utilização de métodos contraceptivos, gerando implicações na estrutura etária do Brasil. Nesse mesmo período, as taxas específicas de fecundidade por grupos de idade da mãe no período reprodutivo diminuíram acentuadamente em quase todas as faixas, com exceção da faixa de 15 a 19 anos. A taxa específica de fecundidade no grupo de mulheres de 15 a 19 variou de 75 nascimentos para cada 1000 mulheres em 1970 a 90 nascimentos em 1000 mulheres em 1999. No ano de 2002, a taxa de fecundidade registrada para o país foi 76 nascimentos por mil mulheres em idade fértil (54).

No ano de 2001, o Sistema de Informações de Nascimento (SINASC) registrou um total de 3.115.474 de nascidos vivos em todo o território nacional, sendo 724.886 (23,3%) filhos de mães adolescentes entre 10 e 19 anos de idade. Do total de nascimentos, 0,9% correspondem a nascimentos de mães com 10 a 14 anos e 22,4% dos nascimentos correspondem a filhos de mães de 15 a 19 anos. Já no ano de 2002, os filhos de mães adolescentes representam 22,7% do total de nascimentos naquele ano, sendo 0,9% dos nascimentos em mães com 10 a 14 anos e 21,8% dos nascimentos em mães entre 15 a 19 anos (54).

As taxas específicas de fecundidade variam conforme a região. As maiores taxas são encontradas nas regiões Norte, Nordeste e Centro-Oeste. Quando considerada a faixa etária de 15 a 19 anos, a região Norte apresentou 123,6 nascimentos em mil mulheres, a Nordeste, 94,8 nascimentos em mil mulheres e a Centro-Oeste, 92,7 nascimentos em mil mulheres. A região Sul apresenta resultados bastante semelhantes com a região Sudeste: 72,2 nascimentos em mil mulheres e 73,6 nascimentos em mil mulheres de 15 a 19 anos, respectivamente. Essa relação manteve-se em 2002 (54).

Analisando por Estados, para o ano de 2001, observou-se que Roraima apresentou a maior taxa de fecundidade específica em mil mulheres de 15 a 19 anos: 157,2 nascimentos. Os Estados de São Paulo (68,5 nascimentos em mil mulheres), Rio Grande do Sul (67,5 nascimentos em mil mulheres) e Santa Catarina (68,6 nascimentos em mil mulheres) apresentam as menores taxa de fecundidade específica em mil mulheres de 15 a 19 anos. Em 2002, a maior taxa de fecundidade específica em mil mulheres de 15 a 19 anos foi registrada em Tocantins (131,80 nascimentos) (54).

A taxa específica de fecundidade de mães adolescentes vem se mantendo ao longo da última década na cidade de Porto Alegre. Para as mulheres de 10 a 14 anos, oscila em torno de 2,4 a 3,59 nascimentos por 1000 mulheres nessa faixa etária. Seu ápice foi no ano de 2000, quando chegou a ser 4,1 nascimentos por 1000 mulheres. Para as mulheres de 15 a 19 anos, ocorreu uma diminuição ao longo dos anos, ficando em torno de 52,8 a 79,3 nascimentos por mil mulheres nessa faixa. Essa diminuição é contrária ao restante do país, embora seja ainda preocupante saber que mais de 50 meninas tornam-se mães a cada ano na cidade. Apesar de ter havido uma diminuição em Porto Alegre, a taxa ainda é considerada alta (55). Esses números subestimam a real prevalência de gestações com menos de 19 anos, pois são considerados apenas nascidos vivos. Os casos em que a gestação foi interrompida de modo espontâneo ou provocada e os natimortos foram desconsiderados.

2.3.1 Indicadores Demográficos

Segundo a Organização Pan-Americana de Saúde - OPAS (56), a taxa de fecundidade específica representa o número médio de filhos nascidos vivos por mulher de uma determinada faixa etária do período reprodutivo, na população residente em um determinado espaço geográfico. Pode dizer então que a taxa de fecundidade será um mediador da intensidade dos fatores reprodutivos a que as mulheres estão expostas. Usaremos essa taxa para detectar variações em um dos grupos de maior risco reprodutivo, as adolescentes, e ainda subsidiar processos de planejamento, gestão e avaliação da atenção materno-infantil, calculando-se através da divisão do número de nascidos vivos em relação a uma determinada faixa etária da mãe, e a população feminina total dessa faixa etária.

$$\text{Taxa de Fecundidade} = \frac{\text{n}^\circ \text{ de nascidos vivos de mães em determinada faixa etária}}{\text{população feminina total dessa faixa etária}} * 1000$$

Para a construção dessa taxa, é necessário ter o conhecimento da população residente através do Censo do IBGE e a informação dos nascidos vivos divulgados pelo Ministério da Saúde através da base do Sistema de Informação sobre Nascidos Vivos (SINASC).

O SINASC é um sistema de informações de âmbito nacional, sob responsabilidade das Secretarias Estaduais e Municipais de Saúde, concebido para diminuir a sub-notificação e melhorar a qualidade da informação sobre nascidos vivos. Este sistema é alimentado pelas declarações de nascimentos (DN), documento oficial emitido pelo hospital onde ocorre o nascimento (55).

Meninas de 10 a 14 anos não fazem parte das avaliações reprodutivas, mas torna-se cada vez mais evidente a necessidade de sua inclusão em programas de saúde sexual. No ano de 2000, do total de adolescente na faixa etária de 10 a 14 anos, 0,43% tinha filho e 17% delas tinha mais de um filho. Ainda que essa faixa etária não apresente um impacto significativo sobre a taxa de fecundidade, seu peso relativo passou de 0,24% em 1991, para 0,57% em 2000 (5).

Para enfatizar essa questão social, incluiremos as adolescentes de 10 a 14 anos nos cálculos da taxa de fecundidade específica por idade para a cidade de Porto Alegre.

Variações geográficas e temporais globais da natalidade são investigadas pela taxa bruta de natalidade, que representa a frequência geral de nascimentos na população (56). A taxa de natalidade bruta mede o número de nascidos vivos por mil habitantes na população residente em determinado espaço geográfico.

$$\text{Taxa de Natalidade} = \frac{\text{n}^\circ \text{ de nascidos vivos total no ano}}{\text{população residente total}} * 1000$$

Para seu cálculo, também é necessário informações a respeito da população residente obtida nos Censos do IBGE e do número de nascidos vivos gerados pelo SINASC.

2.3.2 Indicadores Socioeconômicos

Estudos em diversos países mostraram associações entre fatores sócio-econômicos e condições neonatais, relacionando baixo peso ao nascer com classe social, renda, idade materna, número de filhos vivos, estado marital da mãe, tipo de cuidados de saúde e cuidados pré-natais durante a gravidez (57).

Existem diversos indicadores socioeconômicos para se avaliar o desenvolvimento humano de um determinado local. Os indicadores referentes a educação e renda avaliam as condições dos indivíduos residentes naquele local. Já os indicadores a respeito das condições de moradia descrevem como está o ambiente em que o indivíduo está morando, em relação às condições básicas de saneamento.

A Organização Pan-Americana de Saúde (OPAS) recomenda que a educação seja avaliada em relação à taxa de analfabetismo e pelos níveis de escolaridade. A taxa de analfabetismo revelaria as dificuldades de acesso à educação pela população. Os níveis de escolaridade descrevem a distribuição percentual da população de 15 anos ou mais por grupos de anos de estudo. O nível de instrução inferior a quatro anos de estudo tem sido utilizado como a *proxy* do analfabetismo funcional (56). Juntos, esses indicadores expressam o grau de influência da escolaridade como fator condicionante da situação econômica da população. O perfil da escolaridade nacional atual acentua as desigualdades regionais e afeta negativamente o índice de desenvolvimento humano do país (54).

No Brasil, há uma grande desigualdade educacional. Em média, a população com mais de 25 anos tem 6 anos de estudo e é constituída por 16% de analfabetos. As regiões Norte e Nordeste apresentam as menores médias de escolaridade para as pessoas de 25 anos ou mais. No Rio Grande do Sul, a população de 25 anos ou mais apresenta 6,42 anos de estudos em média e é constituída de 8% de analfabetos (58).

O nível de riqueza padronizado é medido pelo Produto Interno Bruto (PIB) e a distribuição de rendas é medida pela razão de renda e pela proporção de pobre (56). O PIB representa o valor médio agregado por indivíduo em moeda corrente dos bens e serviços finais produzidos em determinado espaço, medindo, dessa forma, a produção por habitante. A razão de renda representa o número de vezes que a renda dos 20% mais ricos é superior à renda dos 20% mais pobres. Quanto mais elevado o valor, maior desigualdade de renda entre os grupos populacionais. A proporção de pobres é medida através da proporção da população com renda inferior a um salário mínimo, revelando a proporção em estado de pobreza.

A observação apenas da qualidade de vida média da população tende a obscurecer os diferenciais existentes nas condições socioeconômicas das distintas camadas sociais, sabendo-se, através das evidências apontadas por diversos estudos, que essa situação tem uma incidência espacial. Determinados territórios da cidade são mais acentuadamente marcados pela pobreza e pela exclusão social (59).

III Referências

- (1) Amazarray MR, Machado PS, Oliveira VZ, Gomes WB. A experiência de assumir a gestação na adolescência: um estudo fenomenológico. *Psicol Reflex Crit* 1998; 11 no 3:431-440.
- (2) Aquino EM, Heilborn ML, Knauth D, Bozon M, Almeida MC, Araujo J. Adolescence and reproduction in Brazil: the heterogeneity of social profile. *Cad Saude Publica* 2003; 19 Suppl 2:S377-S388.
- (3) Gama SGN, Szwarcwald CL, Leal MC, Theme Filha MM. Gravidez na adolescência como fator de risco para baixo peso ao nascer no Município do Rio de Janeiro, 1996 a 1998. *Rev Saude Publica* 2001; 35(1):74-80.
- (4) OPAS. La salud de los adolescentes y los jóvenes en las Américas: escribiendo el futuro. Washington: Comunicación para la Salud, 6, 1995.
- (5) Rede Feminista de Saúde. Adolescentes: Saúde Sexual e Reprodutiva, dossiê. Belo Horizonte: 2004.
- (6) Ribeiro ER, Barbieri MA, Bettiol H, da Silva AA. Comparison between two cohorts of adolescent mothers in municipality of the Southeastern, Brazil. *Rev Saude Publica* 2000; 34(2):136-142.
- (7) Simões VFM, Silva AAM, Bettiol H, Lamy-Filho F, Tonial SR, Mochel EG. Características de Gravidez na Adolescência em São Luiz, Maranhão. *Rev Saúde Pública* 2003; 37(5):559-65.
- (8) Tomasi E, Barros CF, Victora C. As mães e suas gestações: comparação de duas coortes de base populacional no Sul do Brasil. *Cad Saude Publica* 1996; 12 (Supl.1):21-25.
- (9) Carvalho MS, Souza-Santos R. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. *Cad Saude Publica* 2005; 21(2):361-378.
- (10) Elliott P, Wakefield J, Best N, Briggs D. *Spatial Epidemiology: Methods and Application*. London: Oxford University Press, 2000.
- (11) Elliott P, Wartenberg D. *Spatial epidemiology: current approaches and future challenges*. *Environ Health Perspect* 2004; 112(9):998-1006.
- (12) Câmara G, Monteiro AM, Fucks SD, Carvalho MS. Análise Espacial e Geoprocessamento. *Análise Espacial de Dados Geográficos*. Brasília: EMBRAPA, 2004.
- (13) Diggle P, Ribeiro PJ. *Model-based Geostatistics*. Mini-curso apresentado no 14º SINAPE (Simpósio Nacional de Probabilidade e Estatística) ed. ABE - Associação Brasileira de Estatística, 2000.
- (14) Fucks SD, Carvalho MS, Câmara G, Monteiro AM. *Análise Espacial de Dados Geográficos*. 3 ed. São José dos Campos, INPE: EMBRAPA Cerrados, 2004.

- (15) Lawson AB. Statistical methods in spatial epidemiology. New York: John Wiley, 2001.
- (16) Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* 2005; 14(1):35-59.
- (17) Elliott P, Wakefield J, Best N, Briggs D. *Spatial Epidemiology: Methods and Application*. London: Oxford University Press, 2001.
- (18) Bailey TC. Spatial statistical methods in health. *Cad Saude Publica* 2001; 17(5):1083-1098.
- (19) Nuckols JR, Ward MH, Jarup L. Using geographic information systems for exposure assessment in environmental epidemiology studies. *Environ Health Perspect* 2004; 112(9):1007-1015.
- (20) Assunção RM. *Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia*. Departamento de Estatística, Universidade Federal de Minas Gerais . 2001. ABE. 10-4-2005.
- (21) Bailey TC. An introduction to spatial and spatial-temporal modeling of small area disease rates. VI Congresso Brasileiro de Epidemiologia. Recife: ABRASCO, 2004.
- (22) Bailey TC, Gatrell AC. *Interactive Spatial Data Analysis*. 1995.
- (23) Cressie NAC. *Statistics for Spatial Data*. New York: John Wiley, 1993.
- (24) Pereira MG. *Conceitos Básicos de Epidemiologia. Epidemiologia Teoria e Prática*. Rio de Janeiro: Editora Guanabara Koogan S.A., 1995: 1-15.
- (25) Jarup L. Health and environment information systems for exposure and disease mapping, and risk assessment. *Environ Health Perspect* 2004; 112(9):995-997.
- (26) Lawson AB. Disease map reconstruction. *Stat Med* 2001; 20(14):2183-2204.
- (27) Shimakura SE, Carvalho MS, Aerts DR, Flores R. [Spatial risk distribution: modeling infant mortality in Porto Alegre, Rio Grande do Sul State, Brazil]. *Cad Saude Publica* 2001; 17(5):1251-1261.
- (28) d'Orsi E, Carvalho MS. Birth profile for the city of rio de janeiro: a spatial analysis. *Cad Saude Publica* 1998; 14(2):367-379.
- (29) Mollié A. Bayesian mapping of Hodgkin's disease in France. In: Elliott P, Wakefield JC, Best N, Briggs D, editors. *Spatial Epidemiology: Methods and Applications*. Oxford University Press, 2001: 267-285.
- (30) Lawson AB. On the analysis of mortality events associated with a pre-specified fixed point. *J R Stat Soc Ser A Stat Soc* 1993; 156(3):363-377.

- (31) Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping-why, when, and how? A step by step guide. *BMJ* 1996; 313(7061):863-866.
- (32) Câmara G, Carvalho MS. *Análise Espacial de Eventos. Análise Espacial de Dados Geográficos*. Brasília: EMBRAPA, 2004.
- (33) Diggle PJ. Overview of statistical methods for disease mapping and its relationship to cluster detection. In: Elliott P, Wakefield J, Best N, editors. *Spatial Epidemiology Methods and Application*. London: OXFORD University Press, 2001: 87-103.
- (34) Câmara G, Carvalho MS, Cruz OG, Correa V. *Análise Espacial de Áreas. Análise Espacial de Dados Geográficos*. Brasília: EMBRAPA, 2004.
- (35) Cordeiro GM, Lima Neto EA. Modelos Paramétricos. 16º SINAPE - Simpósio Nacional de Probabilidade e Estatística ed. ABE - Associação Brasileira de Estatística, 2004.
- (36) Wakefield J, Best N, Waller S. Bayesian approaches to disease mapping. In: Elliott P, Wakefield JC, Best N, Briggs D, editors. *Spatial Epidemiology Methods and Application*. London: Oxford University Press, 2001: 105-127.
- (37) Bailey TC, Carvalho MS, Lapa TM, Souza WV, Brewer MJ. Modeling of Under-detection of Cases in Disease Surveillance. *Ann Epidemiol* 2005; 15(5):335-343.
- (38) Rothenberg RB, Steinberg KK, Thacker SB. The public health importance of clusters: a note from the Centers for Disease Control. *Am J Epidemiol* 1990; 132(1 Suppl):S3-S5.
- (39) Elliott P, Wakefield J. Disease clusters: should they be investigated, and, if so, when and how? *J R Statist Soc A* 2001; 164(Part 1):3-12.
- (40) Wakefield JC, Kelsall JE, Morris SE. Clustering, cluster detection, and spatial variation in risk. In: Elliott P, Wakefield JC, Best N, editors. *Spatial Epidemiology Methods and Application*. London: OXFORD University Press, 2001: 128-152.
- (41) Ross A, Davis S. Point pattern analysis of the spatial proximity of residences prior to diagnosis of persons with Hodgkin's disease. *Am J Epidemiol* 1990; 132(1 Suppl):S53-S62.
- (42) Jacquez GM. Disease cluster statistics for imprecise space-time locations. *Stat Med* 1996; 15(7-9):873-885.
- (43) Ripley BD. Modelling Spatial Patterns. *Journal of the Royal Statistical Society* 1977; B 39:172-192.
- (44) Gómez-Rubio V, Ferrándiz J, López A. *Detecting Clusters of Diseases with R. Clusters of Disease, software R*. 2003.

- (45) Moran PA. Notes on continuous stochastic phenomena. *Biometrika* 1950; 37(1-2):17-23.
- (46) Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med* 1995; 14(8):799-810.
- (47) Song C, Kulldorff M. Power evaluation of disease clustering tests. *Int J Health Geogr* 2003; 2(1):9.
- (48) Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistics. *J R Statist Soc A* 2001; 164:61-72.
- (49) Morris SE, Wakefield J. Assessment of disease risk in relation to a pre-specified source. *Spatial Epidemiology: Methods and Applications*. London: Oxford University Press, 2001: 153-183.
- (50) Wakefield J, Salway R. A statistical framework for ecological and aggregate studies. *J R Statist Soc A* 2001; 164:119-137.
- (51) Gelman A, Park DK, Ansolabehere S, Price PN, Minnite LC. Models, assumptions and model checking in ecological regressions. *J R Statist Soc A* 2001; 164:101-118.
- (52) ALAN Guttmacher Institute. Teenage Sexual And Reproductive Behavior In Developed Countries : Can More Progress Be Made? *Pregnancy Birth* . 2001. 6-4-2005.
- (53) Ventura SJ, Abma JC, Mosher WD, Henshaw S. Revised pregnancy rates, 1990-97, and new rates for 1998-99: United States. *Natl Vital Stat Rep* 2003; 52(7):1-14.
- (54) Ministério da Saúde. *Saúde Brasil 2004: Uma análise da situação de saúde*. 2004.
- (55) Ministério da Saúde. DATASUS. Ministério da Saúde, Governo Federal . 2005. 21-11-2005.
- (56) OPAS. *Indicadores Básicos para a saúde no Brasil: conceitos e aplicações*. Brasília: Rede Interagencial de Informações para a Saúde, 2002.
- (57) d'Orsi E, Carvalho MS, Cruz OG. Similarity between neonatal profile and socioeconomic index: a spatial approach. *Cad Saude Publica* 2005; 21(3):786-794.
- (58) Ipea, PNDU, Fundação João Pinheiro. *Atlas do Desenvolvimento Humano no Brasil*. 2000.
- (59) Furtado A, Costa BM, Macedo CEG, Germano LRRGBN, Macerata MA, Silva MEM et al. *Mapa da Inclusão e Exclusão Social de Porto Alegre*. Porto Alegre: Prefeitura de Porto Alegre, 2004.

IV Artigo

Epidemiologia Espacial exemplificada pela taxa de fecundidade específica em mulheres de 10 a 19 anos na cidade de Porto Alegre

Resumo

Introdução - A presença de população pequena em unidades geográficas estudadas prejudica a avaliação da distribuição espacial do risco, pois pode gerar alta variabilidade. **Objetivos** - Utilizar métodos de Epidemiologia Espacial para representar a taxa de fecundidade em mulheres de 10 a 19 anos na cidade de Porto Alegre no ano de 2003 e investigar sua distribuição geográfica. **Metodologia** - Utilizando-se a base de dados do SINASC e o Censo do IBGE do ano de 2000, considerando divisões por bairro, a taxa é construída. Métodos de suavização, como Estimador Bayesiano Empírico Local de Marshall e um modelo hierárquico totalmente Bayesiano foram usados para controlar flutuações aleatórias. **Resultados** - Em 2003 ocorreram 28 nascidos vivos por mil mulheres de 10 a 19 anos na cidade. Em alguns bairros, a representação bruta revelou valores quase três vezes maiores do que o esperado. Os métodos Bayesianos reduziram a variabilidade do risco. **Conclusão** - Métodos de Suavização Bayesiana geram resultados mais fidedignos, sendo possível identificar áreas com elevadas taxas que realmente necessitam cuidados especiais.

Palavras-chave: Epidemiologia Espacial, estimativa espacial Bayesiana, maternidade na adolescência.

Abstract

Background: The traditional way to study the spatial variability in rates is done by the gross rate: dividing the number of people in the area by the population at risk. However, the variability between and within neighborhoods and a small population at risk might cause a random noise in this representation. **Objective:** to present the specific fecundity rate of woman with age 10 to 19 in Porto Alegre, in 2003 as an example of the Spatial Epidemiology methods. **Methods:** Using the database SINASC and the Census for the year 2000, considering the city's districts, the fecundity rate was built. Besides the gross rate, a standard rate was built considering differences in size. To summarize spatial variation in the risk, the Bayesian approach was also used. The estimation methods used here were empirical and full Bayesian estimators. **Results:** in 2003, there were 28 births by a thousand women with age 10 to 19. In some areas, the gross rate was higher than the average rate. Those areas showed also poor social conditions. **Conclusion:** the Full Bayesian estimator obtained best results for reducing the variability in the fecundity rate.

Keywords: Spatial Epidemiology, Bayes estimator, adolescent maternity

Introdução

A utilização de dados populacionais para avaliar a distribuição geográfica de alguns eventos de saúde por meio de métodos de Estatística Espacial vem ganhando destaque nos últimos anos.

A disponibilidade de dados populacionais, os avanços computacionais, o desenvolvimento dos Sistemas de Informação Geográfica - SIG (conjunto de equipamentos e programas de computação que integram mapas e gráficos com uma base de dados sobre um espaço geográfico) e o aprimoramento das técnicas de análise espaciais estimulam o desenvolvimento da *Epidemiologia Espacial* (1;2).

O foco da Epidemiologia Espacial é o estudo da distribuição espacial de eventos, contribuindo para a formulação das hipóteses, sendo dividida em três categorias: mapeamento de doença, correlação ecológica e detecção de conglomerados espaciais. Na categoria Mapeamento de Doenças, a distribuição espacial de eventos de saúde é descrita em mapas, através de estimativas para a verdadeira taxa de risco em cada região geográfica. As variações aleatórias em locais com população pequena são controladas, bem como as variações temporais. É utilizada para definir e monitorar epidemias, gerar dados basais para o padrão de doenças e revelar mudanças ao longo do tempo. A correlação ecológica ou correlação espacial avalia a relação entre a distribuição espacial da doença com fatores explanatórios medidos em nível espacial, geralmente agregados. Por fim, o método de detecção de conglomerados de doença investiga concentrações não usuais da doença, ou seja, conglomerados ou *clusters* locais, o que ocorre geralmente na presença de doenças infecto-contagiosas ou contaminações determinadas por fontes locais específicas (1-5).

Além do mapeamento das doenças infecto-contagiosas, há interesse em outros temas, como eventos relacionados à saúde materno-infantil e aos serviços de saúde (6).

Por meio da distribuição espacial de eventos de saúde, pode-se gerar hipóteses, do ponto de vista espacial, sobre a ocorrência do fenômeno. No Brasil há interesse sobre as questões reprodutivas, possibilitando melhores diretrizes para a elaboração de políticas populacionais e de desenvolvimento socioeconômico. Considerando a informação em nível de área, alguns indicadores de grande interesse epidemiológico podem ser construídos.

Dentre todos os indicadores relacionados com a reprodução, a taxa de fecundidade tem grande importância, pois revela como a estrutura etária da população será constituída (7), além de apontar regiões de menor desenvolvimento social.

A taxa de fecundidade no Brasil vem apresentando nos últimos anos um declínio devido às mudanças no comportamento reprodutivo das mulheres, à crescente participação das mulheres no mercado de trabalho e a uma maior utilização de métodos contraceptivos. Ocorre, assim, uma diminuição do número de filhos por mulher em idade fértil. Por outro lado, a população de mulheres adolescentes tem mostrado uma taxa de fecundidade com comportamento diferenciado em relação a dos outros grupos etários femininos (7).

O principal objetivo desse estudo é apresentar métodos de Epidemiologia Espacial, utilizando-se como exemplo a representação espacial da taxa de fecundidade específica na adolescência para o município de Porto Alegre no ano de 2003. Também se pretende correlacionar alguns fatores socioeconômicos (percentual de chefes de família analfabetos em cada bairro e percentual de famílias com renda menor do que um salário mínimo) com a incidência de maternidade na adolescência.

Material e Métodos

Realizou-se um estudo transversal que investigou a maternidade na adolescência durante o ano de 2003 com a ajuda do Sistema de Informações Geográficas de Porto Alegre, que conta com bases cartográficas de logradouros, setores censitários, malha de bairros e limites do município fornecidos pela Equipe de Informação em Saúde da Coordenadoria Geral de Vigilância em Saúde da Secretaria Municipal de Saúde de Porto Alegre (Equipe de Informação em Saúde – EI/SMS). Estas bases de dados georeferenciados integram o SIG de Porto Alegre e foram analisadas como o auxílio do MapInfo.

Os registros de nascidos vivos de mães com idade inferior a 20 anos foram selecionados no Sistema de Informação de Nascimento (SINASC) do Ministério da Saúde, fornecidos pela EI/SMS. O endereço das mães adolescentes, registrado na declaração de nascimento, foi comparado com o endereço que constava na base de dados dos maiores hospitais da cidade. Foram verificados 55% do total de nascimentos naquele ano, com a finalidade de se constatar a fidedignidade dos dados. Observou-se um erro de digitação em menos de 1% dos casos. Comprovando-se a qualidade da digitação dos dados, procedeu-se ao georeferenciamento dos mesmos. O

georeferenciamento das residências das mães adolescentes foi realizado por meio da comparação entre o endereço que constava na declaração de nascimento com o cadastro de logradouros.

As coordenadas geográficas de cada ponto foram obtidas pelo SIG, utilizando-se a interpolação de números em trechos dos logradouros, permitindo a visualização da distribuição dos eventos na base cartográfica, onde cada ponto corresponde a um recém-nascido de mãe adolescente (Figura 1). Em 11% dos casos, devido a não identificação correta do número do domicílio no logradouro, o ponto foi colocado no centro geográfico do logradouro. Os dados então foram agregados por bairros, a partir do número de nascidos vivos de mães adolescentes existente em cada bairro oficial, segundo o Censo do Instituto Brasileiro de Geografia e Estatística (IBGE) do ano de 2000 (8). Porto Alegre contava, em 2000, com 82 bairros oficiais. A base cartográfica da cidade digitalizada é representada por 109 polígonos, pois a região do estuário do Guaíba, denominada de Arquipélago, é formada por 25 ilhas. Existiam duas áreas da cidade ainda não cadastradas junto ao IBGE em 2000 que não apresentavam população residente.

A população considerada como exposta foram as mulheres com idades entre 10 e 19 anos, residentes em Porto Alegre, e os dados referentes ao tamanho da população foram obtidos no Censo de 2000.

No ano de 2003, ocorreram 3537 nascimentos de mães adolescentes, sendo que 38 foram gemelares. Nos casos gemelares contabilizou-se o endereço apenas uma vez, excluindo-se um dos casos. Em 289 declarações (8,17%), o endereço não foi identificado corretamente. Ao todo, 308 declarações de nascimento foram excluídas, restando 3229 para a análise.

No mapeamento de doenças, a avaliação é feita em etapas (9). Para dados agregados em bairros, inicia-se com a representação visual da taxa bruta num mapa. A população exposta deve ser determinada antes do cálculo da taxa bruta. A taxa bruta foi calculada dividindo-se o número de nascidos vivos de mães com 10 a 19 anos pela população feminina de 10 a 19 anos em cada bairro.

Outra forma de se avaliar a taxa de fecundidade seria avaliá-la em termos de risco relativo, através de uma padronização. A padronização pode controlar eventuais diferenças em relação a sexo e idade.

Os riscos padronizados (SMR) foram calculados pela divisão do número de casos observados (y_i) pelo número de casos esperados em cada bairro (e_i). O número de casos esperado foi calculado considerando que o risco é constante na região e admitindo que o número de casos em cada bairro tem distribuição de Poisson. A estimativa gerada para o valor esperado em cada bairro foi calculada considerando a taxa de fecundidade de 10 a 19 anos global ($\sum y_i / \sum N_i$) ponderada pela população de mulheres adolescentes em cada bairro (N_i), através da seguinte expressão:

$$e_i = \frac{\sum y_i}{\sum N_i} \times N_i$$

Na presença de populações pequenas, flutuações aleatórias podem ocorrer. Uma forma de se controlar essas perturbações é utilizar técnicas de suavização. Algumas técnicas são aplicadas com o intuito de suavizar as taxas brutas ou as padronizadas, podendo-se considerar informações sobre a distribuição de probabilidade dos dados e/ou sua localização geográfica. Tais técnicas são: *kernel*, Alisamento Bayesiano Empírico Local de Marshall ou uma modelagem (1;5;9-11).

O Estimador Bayesiano Empírico Local de Marshall utiliza informações dos bairros vizinhos para gerar estimativas médias do risco. Em cada bairro, o risco de ocorrer um nascimento de uma mãe adolescente em Porto Alegre foi re-estimado através de uma média ponderada entre o valor observado e o risco médio local daquela região, sendo este peso inversamente proporcional à população de cada bairro (10).

A matriz de vizinhança foi construída considerando a fronteira natural dos bairros de forma automática dentro do pacote estatístico R (12). Dessa forma, os bairros próximos receberam o valor 1 e os mais afastados recebem valor zero. Essa construção de vizinhança recebe o nome de vizinhança binária.

O Estimador Bayesiano Empírico possui fácil aplicação, mas tem como desvantagem considerar a variabilidade introduzida na estimação dos parâmetros. Desta forma, a suavização nem sempre será alcançada com sucesso. Uma alternativa para controlar esse problema é a modelagem (10;13). O modelo hierárquico proposto por Wakefield et al. (14) vem sendo o mais usado no mapeamento de doenças.

Em estudos de pequenas áreas, a estimação do verdadeiro risco é dominada pela variabilidade amostral que é formada por duas componentes: o efeito aleatório não-espacial (representando efeitos de pequena escala, como as diferenças dentro de um mesmo bairro, ou variações não mensuradas) e o efeito espacial (representando efeitos em larga escala, como a variação entre bairro, ou a devida às covariáveis não mensuradas). Para a modelagem dos dados de Porto Alegre, modelos hierárquicos foram construídos tendo como desfecho de interesse o risco relativo, o qual identifica áreas com alta incidência de nascimentos em mães adolescentes.

Pelo número de parâmetros envolvidos e pela complexidade da modelagem, o método Bayesiano de estimação foi escolhido. Assim, o risco foi interpretado como uma variável aleatória, sendo caracterizado por uma distribuição de probabilidade. Essa distribuição é especificada pelo pesquisador no início do processo e é denominada de *distribuição a priori* (11). A distribuição *a priori* informa, ao modelo, crenças iniciais a respeito dos parâmetros, ao contrário da Inferência Clássica, por máxima verossimilhança, onde o risco em cada local é considerado fixo.

Os modelos hierárquicos Bayesianos são resolvidos em estágios: no primeiro, avaliou-se a relação entre os casos observados e o risco dentro de um mesmo bairro por meio de uma função de verossimilhança, sendo representado pela seguinte expressão:

$$Y_i | \lambda_i, e_i \sim \text{Poisson}(e_i \lambda_i) \quad i = 1, \dots, 109,$$

onde Y_i representa os nascimentos de mães adolescentes, λ_i representa o risco de ocorrer um nascimento de uma mãe adolescente na cidade e e_i representa o número de nascimentos esperados para o bairro i . Os valores esperados foram calculados da mesma forma que na padronização.

A função de verossimilhança (como a apresentada acima) e a distribuição *a priori* representam duas fontes de informação do processo e combinadas caracterizam a distribuição *a posteriori*. A distribuição *a posteriori* tem como objetivo revelar o comportamento do parâmetro de interesse depois que os dados foram coletados. Já a função de verossimilhança revela informação dos parâmetros com base nos dados. Em amostras pequenas, as informações *a priori* dominam a análise, e, em grandes amostras, a função de verossimilhança dos dados domina (11).

No segundo estágio, modelou-se a variação espacial do risco considerando a dependência espacial entre os riscos e a distribuição conjunta de todos os riscos com variáveis explanatórias, de modo que bairros próximos tenham riscos semelhantes. A

distribuição conjunta dependerá de parâmetros desconhecidos, denominados de *hiperparâmetros*.

No modelo utilizado avaliou-se a relação do risco (λ_i) utilizando-se uma função de ligação logarítmica mais um efeito aleatório (u_i) e um efeito espacial (b_i). Como não há covariáveis no modelo, toda a variação devida a elas foi atribuída ao efeito espacial. Já a variabilidade não explicada permanece no erro aleatório. Essa relação é descrita da seguinte forma:

$$\log(\lambda_i) = u_i + b_i$$

Em se tratando de um processo Bayesiano, todos os parâmetros do modelo acima foram considerados como variáveis aleatórias, sendo necessário atribuir distribuições de probabilidade *a priori* a eles. O efeito aleatório é independente dos demais e apresenta uma distribuição de probabilidade normal com média zero e variância constante desconhecida. A relação entre bairros foi considerada no efeito espacial, admitindo uma distribuição *a priori* um campo aleatório de Markov, denominado de distribuição condicional auto-regressiva normal (CAR), levando em conta as informações da matriz de vizinhança, como média e a variância constante desconhecida

No terceiro estágio, os *hiperparâmetros*, a variabilidade do efeito espacial e a variabilidade do efeito aleatório, foram definidos como variáveis aleatórias, atribuindo-se a elas uma distribuição de probabilidade com base nos dados. Uma alternativa comumente usada na literatura é usar distribuições *a priori* para os *hiperparâmetros* Gama não informativas. Entretanto, a estimação do risco ficará totalmente dependente da distribuição *a priori*. Kelsall e Wakefield propuseram a escolha de uma distribuição Gama $\Gamma(0,5; 0,0005)$ como sendo sensível no controle dessa dependência (15). Assim, atribui-se uma *hiperpriori* Gama $\Gamma(0,5; 0,0005)$ para a variância do efeito aleatório e para a variância do efeito espacial.

Para a obtenção da distribuição *a posteriori*, utilizou-se o método de simulação de Monte Carlo (MCMC) (16), com 10000 amostras do processo realizadas no software WinBUGS versão 1.4 (17). As 1000 primeiras amostras foram retiradas das distribuições *a posteriori*.

Uma tentativa para se controlar a variabilidade não mensurada é considerar as diferenças populacionais entre os bairros através do valor esperado de cada bairro como um *offset*. Em ambos os modelos, os valores esperados foram considerados.

A correlação não espacial entre as covariáveis: percentual de chefes de famílias analfabetos e o percentual de famílias com renda inferior a um salário mínimo com a taxa de fecundidade foi avaliada pela correlação não-paramétrica de Spearman.

Esse projeto foi submetido ao Comitê de Ética da Universidade Federal do Rio Grande do Sul e ao Comitê Interno da Secretaria de Saúde da Prefeitura de Porto Alegre, sendo aprovado pelos dois órgãos. Após a aprovação junto ao Comitê de Ética, a utilização dos dados foi liberada pela EI/SMS. Os autores do mesmo assinaram um termo de compromisso, comprometendo-se com a utilização dos dados apenas para essa pesquisa.

Resultados

Dos 3229 nascimentos analisados, 144 (4,5%) ocorrem entre mães com idade de 10 a 14 anos. Os nascimentos analisados distribuídos nos bairros de Porto Alegre estão apresentados na Figura 1.

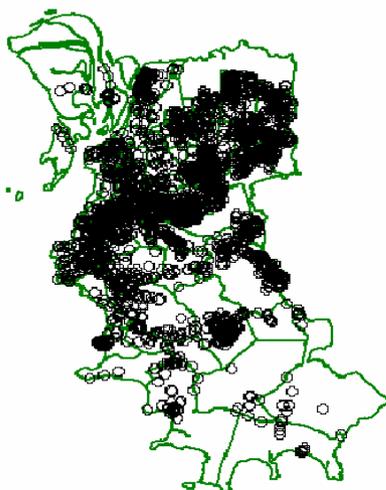


Figura 1: Distribuição espacial dos nascidos vivos de mães adolescentes, Porto Alegre, 2003

A distribuição espacial do percentual de chefes de família analfabetos e o percentual de famílias com renda inferior a um salário mínimo, em cada bairro, estão apresentados na Figura 2.

A taxa bruta estimada para o ano de 2003 foi de 3 nascidos vivos por mil mulheres de 10 a 14 anos e 50 nascidos vivos por mil mulheres de 15 a 19 anos. Considerando o grupo todo, a taxa bruta foi de 28 nascidos vivos por mil mulheres de 10 a 19 anos, e sua distribuição espacial por bairro está apresentada na Figura 3.

A taxa de fecundidade apresenta uma grande variabilidade na presença de populações pequenas, evidenciando a presença de flutuações aleatórias. Bairros com populações pequenas aparecem com altas taxas de fecundidade. Será necessário utilizar algum método de suavização.

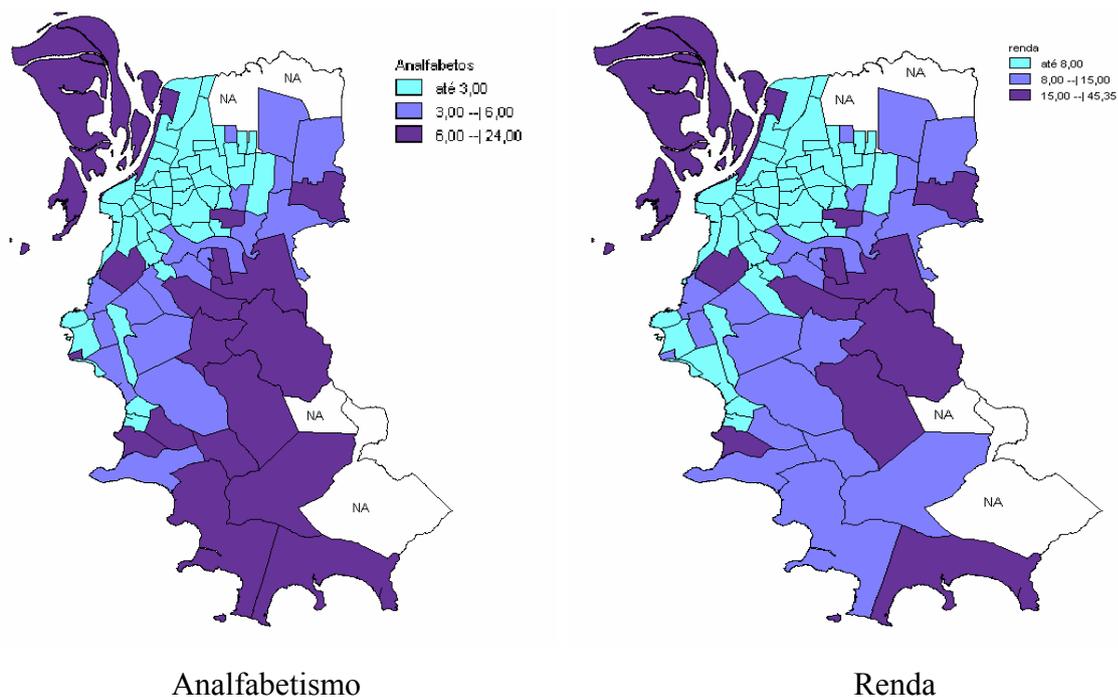


Figura 2: Distribuição espacial da proporção de chefes de família analfabetos e da proporção de famílias com renda inferior a 1 salário-mínimo por bairros, Porto Alegre, 2000.

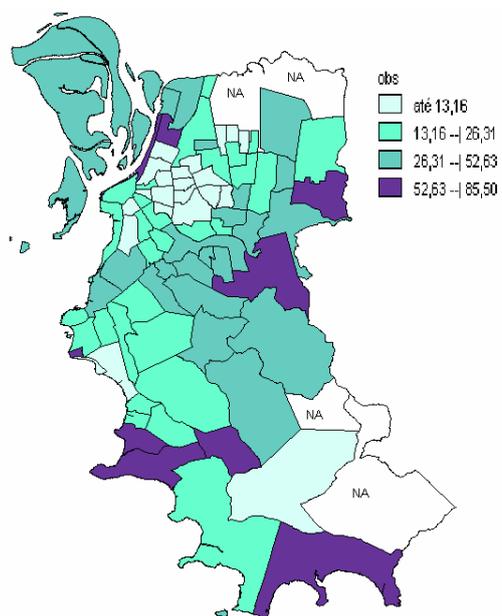


Figura 3: Distribuição espacial da taxa de fecundidade específica por idade de 10 a 19 anos, Porto Alegre, 2003.

As estimativas suavizadas do risco (Bayesiano Empírico e Totalmente Bayesiano sem covariáveis) estão apresentadas na Figura 4. Verifica-se uma redução na variabilidade com a redução da amplitude dos intervalos finais da legenda. Nos bairros com risco inferior a 1, o modelo (Figura 4C) e o alisamento Bayesiano Empírico de Marshall (Figura 4B) superestimaram o risco relativo ou não modificam o risco de forma relevante. A melhor suavização ocorre nos bairros com risco superior a 1. Padronizando-se as escalas de todos os mapas com a suavização, os bairros se mantiveram nas mesmas faixas de riscos quando comparamos o método Bayesiano Empírico de Marshall (Figura 4B) e o modelo hierárquico sem covariáveis (figura 4C). Só alguns bairros, como Navegantes e Serraria, modificaram suas faixas de risco.

O modelo (Figura 4C) apresentou estimativas *a posteriori* suavizadas para o risco de nascimentos de filhos de mães adolescentes. Em uma análise Bayesiana, o Intervalo de Credibilidade é usado para avaliar a significância estatística dos parâmetros em cada modelo. Quando o valor zero estiver no intervalo, o parâmetro é não significativo. Neste modelo totalmente Bayesiano sem covariáveis, somente o efeito espacial é significativo, considerando o intervalo de 95% de credibilidade.

Observando a distribuição espacial do percentual de chefes de família analfabetos e o percentual de famílias com renda inferior a um salário-mínimo (Figura 2), verifica-se que há uma correlação positiva com a incidência de maternidade na adolescência ($r_s=0,824$; $r_s=0,801$, respectivamente). Bairros com maior percentual de chefes de família analfabetos e com maior percentual de famílias com renda inferior a um salário mínimo apresentam maior taxa de fecundidade específica de 10 a 19 anos.

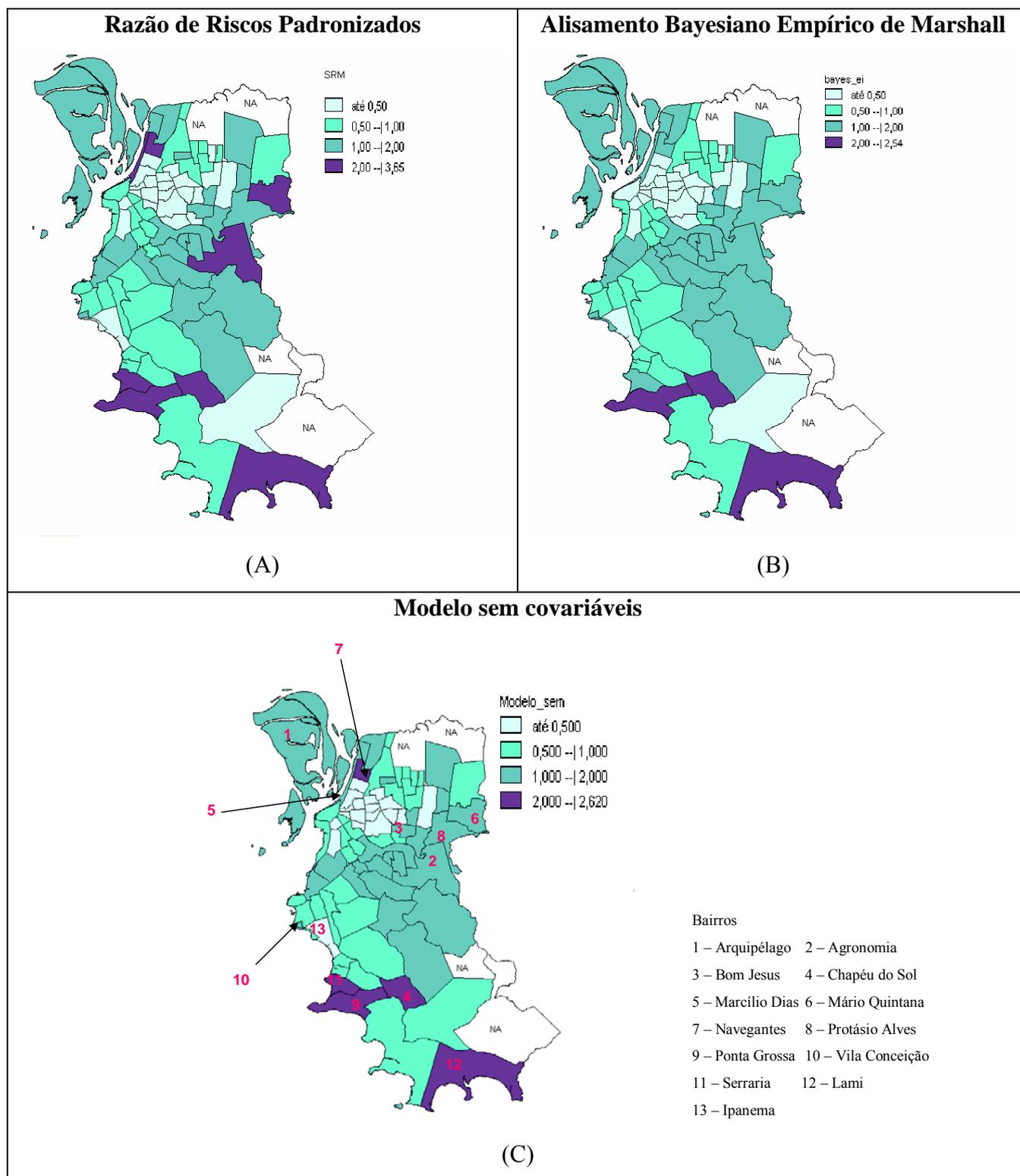


Figura 4: (A) Risco bruto de maternidade na adolescência estimado - razão de riscos padronizada; (B) Taxa de fecundidade específica suavizada pelo método Bayesiano empírico; (C) Risco estimado pela modelagem Bayesiana hierárquica.

Discussão

A representação bruta dos nascimentos para a cidade de Porto Alegre não revelou claramente o padrão espacial da maternidade na adolescência. Para uma avaliação correta, foi necessário calcular a taxa de fecundidade específica, a qual denominou-se taxa bruta, para diferenciar este resultado da razão de riscos padronizada e das taxas suavizadas que foram apresentadas.

A taxa de fecundidade específica de 10 a 14 anos em Porto Alegre no ano 2002 foi de 3 por mil mulheres, idêntica à taxa nacional. A taxa de fecundidade específica de 15 a 19 anos para a cidade de Porto Alegre em 2002 foi de 56 para mil mulheres e está abaixo da média nacional de 76 nascimentos para cada mil mulheres de 15 a 19 anos (19).

Quando desagregada por bairros, a taxa de fecundidade de 10 a 19 anos em Porto Alegre apresenta valores com alta variabilidade, mostrando bairros com valores superiores à média da cidade, equiparando-se a cidades do Norte e Nordeste do país, onde se encontram as maiores taxas regionais.

A suavização da taxa de fecundidade específica por idade foi utilizada para corrigir flutuações aleatórias decorrentes da alta variabilidade dentro e entre bairros e pela presença de populações pequenas em alguns desses, gerando, assim, um mapa livre de flutuações aleatórias. Várias propostas foram apresentadas para a correção das estimativas, como a Suavização Bayesiana Empírica Local de Marshall e a Suavização Totalmente Bayesiana sem covariáveis, considerando um modelo hierárquico Bayesianos. A presença de áreas não habitadas influenciaram na suavização do risco em alguns pontos da cidade. Uma possível solução para esse problema seria considerar outro tipo de vizinhança que não fosse a vizinhança binária por fronteira, como um método que considere alguma variável socioeconômica na sua construção.

A representação padronizada é a mais utilizada no contexto de mapeamento de doenças, mas é baseada em um estimador de razão que sofre grandes modificações quando pequenas alterações nos valores esperados do denominador são feitas, por exemplo, quando se têm valores esperados próximos de zero, podendo ocorrer estimativas artificialmente altas (10;11). Ela apenas controla um possível confundimento entre as variáveis, não corrigindo as flutuações aleatórias.

Outro problema das representações padronizadas é gerar uma dependência espacial nos riscos relativos. Essa correlação poderá ser explicada por covariáveis não

mensuradas. A solução apresentada para controlar essa situação foi considerar o risco relativo como uma variável aleatória, especificando-se uma distribuição *a priori* para ele (15) e gerar estimativas através dos pressupostos da Inferência Bayesiana.

Já é conhecido que os procedimentos Bayesianos eliminam vieses e reduzem a variância do risco relativo principalmente em áreas com população pequena. Essa redução na variância é realizada com a utilização da informação das áreas vizinhas (13).

A estimação Bayesiana foi apresentada pelos estimadores Totalmente Bayesianos e pelo estimador Empírico Local de Marshall. Os estimadores Totalmente Bayesianos são melhores que o estimador Bayesiano Empírico Local de Marshall. As principais vantagens do método Totalmente Bayesiano são: considerar toda a variabilidade do processo a partir dos efeitos aleatórios introduzidos ao modelo, avaliar modificações no risco em nível de espaço e tempo quando ocorrem, além de gerarem a verdadeira distribuição *a posteriori* para o risco. No caso de Porto Alegre, os dois métodos (Totalmente Bayesiano e o Empírico de Marshall) apresentam as mesmas estimativas para o risco. Somente o bairro Navegantes não foi suavizado pelo método Totalmente Bayesiano, e o bairro Serraria ficou subestimado pelo método Empírico Local de Marshall.

Todavia, a abordagem Bayesiana, de forma geral, não está sendo ainda muito utilizada na saúde pública, por envolver conhecimentos estatísticos avançados na definição das distribuições de probabilidade *a priori* e da distribuição conjunta do evento com o risco. Ainda que complexos, os métodos Bayesianos vêm se desenvolvendo, com a disponibilidade de alguns programas computacionais que facilitam seu uso, principalmente na Epidemiologia Espacial. Outros autores estão fazendo o uso dessas ferramentas para suavizar as taxas investigadas, avaliando outros desfechos e obtendo também bons resultados (20-23).

Em relação ao padrão espacial da taxa de fecundidade, destacam-se áreas com maior risco apontadas com cores mais fortes (risco acima de 1). Entretanto, para se compreender o processo espacial presente, é preciso conhecer o processo de segregação urbano da cidade. Esse fenômeno expressa-se na recente migração de grupos populacionais para a periferia da cidade, a qual é carente de infra-estrutura. Além disso, a estrutura de bairros da cidade apresenta um padrão heterogêneo com bairros de baixa inserção socioeconômica ao lado de bairros com melhores condições. A presença de bairros ricos próximos de bairros pobres prejudicou as estimativas Bayesianas que consideram essa informação de vizinhança e proximidade entre bairros para gerar as

estimativas *a posteriori*. Isso explicaria a mudança de nível de risco apresentados nos mapas em alguns bairros como Ipanema e Vila Conceição.

Quando os valores estimados pelo modelo (método Totalmente Bayesiano) são muito diferente dos observados, há evidências de que o modelo não está bem ajustado ou alguma suposição estatística não está sendo satisfeita (24).

Observa-se que as áreas de maior risco de nascimentos na adolescência, após a aplicação dos métodos, também apresentam maior proporção de chefes de famílias analfabetos e de chefes de família recebendo menos de um salário-mínimo. Esse resultado sugere que a maternidade na adolescência está associada às baixas condições de vida da população, sendo evidenciado pela modelagem hierárquica Bayesiana. Uma forma de mensurar esta dependência seria adicionar covariáveis ao modelo, como renda e analfabetismo, buscando controlar o efeito espacial. Confirmando-se a nulidade do efeito espacial no modelo, pode-se dizer que não há uma dependência espacial da maternidade na adolescência entre os bairros de Porto Alegre. Uma baixa dependência espacial da maternidade na adolescência também foi encontrada na cidade do Rio de Janeiro em 1998 (25).

Os alisamentos Bayesiano Empírico Local de Marshall e o Totalmente Bayesiano conseguiram suavizar numericamente as taxas nos bairros Ponta Grossa, Chapéu do Sol, Serraria e Lami, pois esses bairros estão situados próximos a áreas sem população e apresentam valores esperados pequenos, embora no mapa não mudam de classe. Nas áreas não habitadas e sem casos, considerou-se a população exposta composta por apenas um indivíduo, para evitar os problemas matemáticos de estimação; assim o método estimou a taxa como zero. Uma tentativa para se melhorar a suavização seria considerar outra forma de agregação que não os bairros, como setores censitários.

Com a suavização empregada, alguns bairros destacaram-se por apresentarem ainda grande incidência de nascimentos de filhos de mães adolescentes, como Marcílio Dias, Navegantes, o Arquipélago e os bairros Agronomia, Mário Quintana e Bom Jesus. Nesses bairros, os indicadores socioeconômicos apontam para piores condições de vida. Para essas regiões, ações de qualidade de vida oriundas de órgãos governamentais podem ser originadas após constatação de altas taxas de fecundidade específica na adolescência.

Embora mais complexos, os estimadores Bayesianos devem ser usados nesse contexto, a fim de minimizar as flutuações aleatórias geradas por populações pequenas, eliminando os vieses decorrentes do processo de estimação. Para a validade completa

dessa análise, o estatístico se faz necessário, para juntamente com o pesquisador, obter as especificações das distribuições de probabilidades *a priori* do processo.

Referências

- (1) Elliott P, Wakefield J, Best N, Briggs D. Spatial Epidemiology: Methods and Application. London: Oxford University Press, 2001.
- (2) Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect 2004; 112(9):998-1006.
- (3) Bailey TC. Spatial statistical methods in health. Cad Saude Publica 2001; 17(5):1083-1098.
- (4) Jarup L. Health and environment information systems for exposure and disease mapping, and risk assessment. Environ Health Perspect 2004; 112(9):995-997.
- (5) Lawson AB. Statistical methods in spatial epidemiology. New York: John Wiley, 2001.
- (6) Carvalho MS, Souza-Santos R. Análise de dados espaciais em saúde pública: métodos, problemas, perspectivas. Cad Saude Publica 2005; 21(2):361-378.
- (7) Ministério da Saúde. Saúde Brasil 2004: Uma análise da situação de saúde. 2004.
- (8) IBGE. Censo Demográfico 2000: Características da População e dos Domicílios: Resultados do universo. 2000.
- (9) Lawson AB. Disease map reconstruction. Stat Med 2001; 20(14):2183-2204.
- (10) Assunção RM. Estatística Espacial com Aplicações em Epidemiologia, Economia e Sociologia. Departamento de Estatística, Universidade Federal de Minas Gerais . 2001. ABE. Disponível em site: www.est.ufmg.br/%7Eassuncao/cursos/espacial/aplicado/index.html. Acessado em 10-4-2005.
- (11) Lawson AB, Browne WJ, Rodeiro CLV. Disease Mapping with WinBUGS and MLwiN. Wiley, 2003.
- (12) R: A Language for Environment for Statistical Computing. 2004.
- (13) Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. Stat Methods Med Res 2005; 14(1):35-59.
- (14) Wakefield J, Best N, Waller S. Bayesian approaches to disease mapping. In: Elliott P, Wakefield JC, Best N, Briggs D, editors. Spatial Epidemiology Methods and Application. London: Oxford University Press, 2001: 105-127.

- (15) Silva SA, Melo LLM, Ehlers R. Spatial Analysis of Incidence Rates: a Bayesian Approach. Departamento de Estatística - UFPR . 2004. Disponível em site: <http://www.est.ufpr.br/rt/ehl04.pdf>. Acessado em 15-11-2005.
- (16) Gilks WR, Richardson S, Spiegelhalter DJ. Introducing Markov Chain Monte Carlo. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. Markov Chain Monte Carlo in Practice. London: Chapman & Hall, 1995.
- (17) WinBUGS. MRC Biostatistics Unit, 1999.
- (18) Spiegelhalter DJ, Best N, Carlin BP, Van den Linde A. Bayesian Measures of Model Complexity and Fit(with discussion). Journal of the Royal Statistical Society 2002; Series B(64):1-34.
- (19) Ministério da Saúde. DATASUS. Ministério da Saúde, Governo Federal . 2005. Site: <http://w3.datasus.gov.br/datasus/datasus.php>. Acessado em 21-11-2005.
- (20) Bailey TC, Carvalho MS, Lapa TM, Souza WV, Brewer MJ. Modeling of Under-detection of Cases in Disease Surveillance. Ann Epidemiol 2005; 15(5):335-343.
- (21) Devine OJ, Louis TA. A constrained empirical Bayes estimator for incidence rates in areas with small populations. Stat Med 1994; 13(11):1119-1133.
- (22) Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. J R Stat Soc Ser C Appl Stat 1991; 40 (2):283-294.
- (23) Souza WV, Barcellos CC, Brito AM, Carvalho MS, Cruz OG, Albuquerque MF et al. [Empirical bayesian model applied to the spatial analysis of leprosy occurrence]. Rev Saude Publica 2001; 35(5):474-480.
- (24) Stern HS, Cressie N. Posterior predictive model checks for disease mapping models. Stat Med 2000; 19(17-18):2377-2397.
- (25) d'Orsi E, Carvalho MS. Birth profile for the city of Rio de Janeiro: a spatial analysis. Cad Saude Publica 1998; 14(2):367-379.

V Considerações Finais

De uma maneira geral, a investigação da distribuição espacial ou geográfica de uma determinada doença ou evento de saúde revelará locais onde as equipes de saúde necessitam ter mais atenção. Para gerar resultados representativos, um controle de qualidade nos dados deve ser feito para evitar a duplicidade de registros.

A escolha dos limites geográficos deve ser respeitada para se ter uma boa definição da população em risco. Se não houver essa escolha, padrões equivocados da distribuição do evento/doença poderão ser revelados, pois os valores esperados são calculados para cada local com base na população em risco, gerando as estimativas da razão de riscos padronizada.

Técnicas exploratórias são devem ser utilizadas para visualizarmos espacialmente as taxas brutas e as taxas padronizadas. Em Epidemiologia Espacial, das técnicas de mapeamento de doenças são utilizadas para estimar a verdadeira distribuição espacial do evento, cujas taxas brutas podem ser influenciadas por flutuações aleatórias decorrentes de estudos em pequenas populações. Técnicas de epidemiologia Espacial controlam as diferenças ou semelhanças na estrutura geográfica do evento, revelando possíveis fatores de exposição e contaminação, e gerando hipóteses sobre sua etiologia, revelando áreas de maior risco. Habitualmente, estas técnicas são aplicadas a dados resultantes de contagens de casos ou de indicadores, representando taxas de incidência ou risco relativo. No numerador, aparece o número de casos incidentes observados naquela área, e, no denominador, informações referentes à população sob risco.

Um cuidado essencial deve ser adotado na interpretação de mapas, pois um padrão aparente poderá ser visualizado ou a dependência espacial poderá não ser identificada, dependendo da forma pelos quais os limites serão analisados, da escala geográfica utilizada e das cores escolhidas. Mapas dos mesmos dados em diferentes escalas podem apresentar padrões visuais que não existem na realidade. Assim, é importante escolher o tamanho da unidade e o método de agregação a ser utilizado. A homogeneidade na escolha de grupos agregados será de muito valor na interpretação.

Em estudos de pequenas áreas busca-se a melhor suavização na taxa para que o verdadeiro risco da doença seja representado num mapa. Essa suavização pode ser feita por modelagem, pelo Alisamento Bayesiano Empírico Local de Marshall ou *kernel*.

Na presença de um agente causador do *cluster*, considera-se sua existência na hora da modelagem. No entanto, em muitas investigações de existência de *clusters*, pretende-se apenas mostrar sua formação e suavizar as taxas, sem explicar relações causais. Nessas ocasiões, a modelagem se faz desnecessária, pois não existem outras covariáveis interagindo no processo.

Já no mapeamento de doenças e nos estudos de correlação geográfica, a investigação causal faz parte do processo e é realizada juntamente com a modelagem. Os métodos estatísticos utilizados nos estudos do tipo mapeamento de doenças e os de correlação geográfica são muito semelhantes, embora o segundo está mais interessado na etiologia da doença.

Uma vez definida a questão de pesquisa, inicia-se o método exploratório e a modelagem. Para auxiliar na sua avaliação espacial da doença em estudo, elaborou-se um roteiro conforme quadro no Anexo B. Dependendo da hipótese, o tipo de estudo será definido e sua avaliação será realizada, buscando-se identificar e explicar o padrão espacial do evento.

Concluindo, a Epidemiologia Espacial mostrou ser uma ferramenta bastante eficaz no mapeamento e investigação da taxa de fecundidade específica de mulheres de 10 a 19 anos em Porto Alegre, no ano de 2003, principalmente em pequenas áreas.

Anexos

Anexo A – Projeto de pesquisa

Anexo B – Etapas de aplicação do método de Epidemiologia Espacial

Anexo C – Comandos para a realização do método de estimativa totalmente Bayesiano no WinBUGS

Anexo A - Projeto

Título: Epidemiologia Espacial: Avaliando a Taxa de Fecundidade em Mulheres de 10 a 19 anos em Porto Alegre

1 Local de origem do projeto

Este projeto corresponde a uma dissertação de mestrado do Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul, sendo parte de uma parceria estabelecida entre a Prefeitura Municipal de Porto Alegre com o mesmo programa.

2 Introdução

Maternidade na adolescência não constitui um fenômeno novo no cenário brasileiro. Acompanhando uma tendência internacional, ela assume o status de problema social de bastante interesse para a saúde pública (2).

Nos Estados Unidos, o número de nascimentos para cada 1000 adolescentes grávidas vem diminuindo nas últimas décadas, embora ainda seja considerado alto. Só no ano de 1999, houve 55 nascimentos para cada 1000 mulheres com idade entre 15 e 17 anos(9).

Apesar do Brasil se encontrar em transição demográfica, com queda nas taxas de fecundidades decorrente da diminuição do número de nascimentos, ainda é grande o número de nascidos vivos em mulheres com menos de 19 anos. A taxa de fecundidade total reduziu muito nos últimos 30 anos, passando de 5,8 filhos em média no ano de 1970 para 1,82 filho em média no ano de 2002. As taxas de fecundidade específicas por grupos etários da mãe diminuíram em quase todos os grupos etários, com exceção da faixa dos 15 aos 19 anos. A taxa de fecundidade específica nesse grupo etário vai desde 75 nascimentos para cada 1000 mulheres em 1970 a 90 nascimentos por 1000 mulheres em 1999. No ano de 2002, a taxa de fecundidade específica ficou em 76 nascimentos por 1000 mulheres.

As maiores taxas estão registradas nos estados das regiões Norte, Nordeste e Centro-Oeste. No Rio Grande do Sul, durante o ano de 2001, registrou-se 67,5 nascimentos em mil mulheres de 15 a 19 anos. Já na cidade de Porto Alegre, a taxa teve seu ápice em 1995, chegando a 79,3 nascimentos em mil mulheres. Em 2002 foram registrados 55,66 nascimentos em mil mulheres. A taxa de fecundidade específica vem diminuindo na cidade de Porto Alegre, embora ainda seja considerada alta.

Entre as situações que determinam a ocorrência desse fenômeno, encontram-se as questões socioeconômicas, as quais necessitam de uma atenção específica por parte dos profissionais da área de saúde, dos poderes públicos e da sociedade em geral.

A gravidez na adolescência pode gerar complicações no parto e problemas psicológicos nas mães, pois elas são imaturas e não estão fisiologicamente preparadas para terem uma criança (6).

A adolescência compreende o período de idade que vai dos 10 aos 19 anos (3). É nesse período que ocorre as principais transformações nas esferas física, psíquica e de relacionamentos. Nessa fase ocorre a busca pela própria identidade, a tendência grupal, a necessidade de intelectualizar e fantasiar, a evolução sexual, as contradições sucessivas em todas as manifestações de conduta, a separação progressiva dos pais e as constantes flutuações do humor e do estado de ânimo (2).

Durante todo esse período, o corpo está sendo preparado. No período que vai dos 10 aos 14 anos, há a maturação física sexual e desperta o interesse pelo corpo. Já o período que vai dos 15 aos 17 anos é marcado pela menarca no sexo feminino. Nesse período ocorre a busca por um local na sociedade e satisfação sexual. Após os 18 anos, ocorre a maturidade física completa.

Diz-se que uma menina é fisicamente preparada para ser mãe após a menarca. Porém, com as melhoras com os cuidados com a saúde infantil e a nutrição, a adolescência tem iniciado mais precocemente. Além desses fatores, as mudanças de costumes, a maior participação da juventude nas decisões e os meios de comunicação têm favorecido uma erotização precoce por parte das meninas (2).

Historicamente, a maternidade na adolescência não era um fato problemático. Tal fato era visto como supervalorização da reprodução. Com a industrialização, a adolescência emergiu como um estágio distinto da vida, no qual a juventude foi educada para ter uma identidade. Assim, a maternidade começou a ser vista como prejudicial, pois colocaria em risco a trajetória à vida adulta. Na década de 60, os problemas obstétricos e pediátricos, decorrentes de uma gestação com menos de 19 anos, passaram a ser considerados. Nos dias de hoje, a maternidade nesse período é apontada como sendo um desperdício de oportunidades, pois as mães adolescentes desmerecem as chances oferecidas aos jovens em geral e não lutam pela emancipação feminina. Sem contar que a maternidade na adolescência começa a ir contra a transição demográfica do país (2;6).

A crença de que a maternidade precoce resulta em aspectos negativos para a vida do adolescente e seu filho tem fundamentado as políticas de saúde e programas de atenção aos adolescentes.

A Saúde Coletiva, incorporando conceitos da Medicina Preventiva e da Medicina Integral, mudou o enfoque sobre maternidade na adolescência e passou a considerar dimensões psicológicas, sociológicas, econômicas, educacionais e demográficas, ao invés de se restringir somente às dimensões biológicas (2).

Segundo a pesquisa qualitativa desenvolvida por Heilborn, intitulada Pesquisa GRAVAD (Estudo Multicêntrico sobre Jovens, Sexualidade e Reprodução, desenvolvida em três capitais brasileiras: Porto Alegre, Rio de Janeiro e Salvador), os adolescentes desenvolvem trajetórias extremamente heterogêneas quanto à classe social e gênero para a escolha de carreiras escolar-profissional, familiar-conjugal e contraceptivas. Em todas essas trajetórias, a gravidez é um fator que atrapalha, pois há enormes disparidades econômicas, sociais e culturais entre as classes no Brasil.

Conforme Heilborn mostrou no seu estudo, as condições sócio-econômicas são fundamentais para determinar os cuidados com a adolescente. A localização espacial das mães adolescentes está associada com essas condições sócio-econômicas, supondo-se variações dessa situação dentro da cidade (polis). A Estatística Espacial fornecerá ferramentas para investigar essa dependência espacial entre as gestantes e seus locais de moradia (6).

A disponibilidade de índices demográficos e dados populacionais, os avanços computacionais, o desenvolvimento de sistemas de informação geográfica e as metodologias espaciais permitem uma avaliação real das variações nas taxas (índices demográficos, incidências de eventos de saúde ou risco de doenças), principalmente em pequenas áreas. Esse tipo de estudo está ganhando cada vez mais destaque nos estudos epidemiológicos e de saúde pública, sendo denominado de Epidemiologia Espacial, a qual se utiliza de métodos estatísticos que consideram a localização espacial como um elemento fundamental no processo de avaliação (4;5).

A Geografia da Saúde poderá auxiliar no planejamento familiar, feito nos postos de saúde administrados pela prefeitura, por meio da detecção de grupos de gestantes adolescentes através da identificação da distribuição espacial desses nascimentos. Uma possível melhora no planejamento familiar é de grande interesse para a Secretaria Municipal de Saúde.

A utilização de mapas e a preocupação com a distribuição geográfica de diversas doenças são antigas. Em 1768, James Lind publicou o livro *An Essay on Diseases Incidental to Europeans in Hot Climates*, no qual procurava explicações para a distribuição de doenças, chegando a inclusive determinar áreas geográficas específicas

(8). Desde então, diversos trabalhos foram escritos na Geografia Médica e da Saúde, descrevendo variações geográficas na distribuição de doenças. Um exemplo clássico na área espacial é o estudo de John Snow em Londres, referente a epidemia do cólera, no qual utiliza técnicas de mapeamento para localizar os casos dessa doença e pontos de coleta de água, mostrando o papel da contaminação da água na ocorrência do cólera (8).

A distribuição geográfica da doença tem um importante papel nos diferentes estudos epidemiológicos e de saúde pública. Para lidar com esse tipo de informação, há três tipos de propostas. O primeiro seria mapear a doença de forma a produzir um mapa da verdadeira distribuição geográfica subjacente do risco da doença, sugerindo hipóteses para controlar o problema. A segunda seria através de estudos ecológicos, onde se avalia a relação entre o risco da doença e demais covariáveis medidas em nível de grupo, e não de indivíduo. Por fim, a terceira seria investigar a formação de grupos geográficos (*clusters*) representando áreas significativas para o risco da doença (1).

A disponibilidade de bancos de dados em saúde com informação de endereço e a estrutura de Sistema de Informações Geográficas (SIG) em diversos municípios brasileiros têm possibilitado o uso da Estatística Espacial por padrões de pontos e áreas para diferenciar regiões dentro do mesmo município.

Os SIGs são sistemas computadorizados que permitem a coleta, armazenamento, manipulação, análise, representação gráfica e produção de relatórios de dados georreferenciados (7).

3 Objetivos

O principal interesse desse trabalho é investigar a distribuição geográfica da maternidade na adolescência durante o ano 2003 no município de Porto Alegre, utilizando-se métodos e modelos de Epidemiologia Espacial. Também se pretende investigar possíveis relações com alguns indicadores de exclusão social, explicar variações nas taxas padronizadas e buscar melhores estimadores para a taxa de fecundidade.

4 Método

4.1 Delineamento

Trata-se de um estudo ecológico transversal que investigará como desfecho principal a maternidade na adolescência durante o ano de 2003. Um Sistema de Informação Geográfica será usado para relacionar o endereço do caso com o logradouro da cidade.

Serão consideradas informações pontuais, através do endereço do caso, e por área, através do número de casos em cada bairro da cidade. Alguns fatores sócioeconômicos, como a proporção de analfabetos em cada bairro e a proporção de famílias com renda inferior a um salário mínimo no bairro, também serão utilizados, com o objetivo de diminuir a variabilidade no risco.

4.2 Amostra

Participarão do estudo todas as declarações de nascimento de mães com idade entre 10 a 19 anos que estão inseridas no banco de dados SINASC (Sistema de Informações de Nascimento) no período de 2003.

Serão excluídas das análises aquelas declarações que tiverem endereços mal-preenchidos, e será excluído de forma aleatória um dos gêmeos em partos gemelares.

4.3 Logística

Após a aprovação do projeto pelo Comitê de Ética, a base de dados da Secretaria Municipal de Saúde (SMS) estará liberada para o uso.

Serão consideradas as declarações de nascimentos de gestantes adolescentes, emitidas pelos hospitais da capital, já cadastradas e armazenadas na base de dados da SMS.

O controle de qualidade dos dados é feito inicialmente na SMS, onde as declarações com informações incoerentes são revisadas. O processo de limpeza é feito eletronicamente. Após essa fase, a base de dados estará pronta para análises.

Pretende-se utilizar o Sistema de Informações Geográficas (SIG), que conta com bases cartográficas de logradouros, dos setores censitários e da malha de bairros e dos limites do município de Porto Alegre cedido pela Secretaria Municipal de Saúde.

Para fazer o georeferenciamento das residências de mães adolescentes, pretende-se comparar o endereço que consta na declaração de nascimento com o cadastro de

logradouros. As coordenadas geográficas serão interpoladas por SIG, que permitirá a visualização da distribuição dos eventos na base cartográfica, onde cada ponto corresponderá a uma adolescente grávida.

5 Análise dos dados

Pretende-se utilizar técnicas da Epidemiologia Espacial para estudar a distribuição geográfica da maternidade na adolescência a partir do número de casos presentes em cada bairro para, assim, se estimar o risco, eliminando as flutuações aleatórias presentes na região.

Como método exploratório, faz-se uma análise gráfica através de mapas. Assim, inicia-se com o mapeamento da taxa de fecundidade específica em mulheres de 10 a 19 anos (1;7). O mapeamento dos dados brutos pode revelar flutuações aleatórias, valores observados muito altos, decorrentes das pequenas populações. Para controlar essas flutuações, o alisamento Bayesiano empírico será empregado, juntamente com a modelagem hierárquica Bayesiana.

6 Cronograma

atividade	2004										2005									
	mar	abr	mai	jun	jul	ago	set	out	nov	dez	Jan	fev	mar	jun	jul	ago	set	out	nov	
Elaboração do projeto																				
Aperfeiçoamento do projeto																				
Aprovação no Comitê de Ética																				
Revisão bibliográfica																				
Apresentação do projeto																				
Preparação da base de dados																				
Análise dos dados																				
Apresentação do projeto II																				
Elaboração da dissertação																				
Aprovação do projeto																				
Defesa Preliminar																				
Sessão pública																				
Homologação																				

7 Questões éticas

Como a base de dados investigada é secundária, não se faz uso de termo de consentimento informado aos participantes. No entanto, os autores do projeto se comprometem com o sigilo das informações contidas na base de dados, bem como se responsabilizam para uso exclusivamente dessa pesquisa (Anexo 1). O projeto já foi aprovado pelo Comitê de Ética.

8 Recursos financeiros

Esse projeto não recebe apoio financeiro de nenhuma agência financiadora. Os gastos com material de escritório e descolamento dos autores decorrerão por conta dos mesmos. Somente a pesquisadora principal tem bolsa Capes para realizar seu trabalho.

9 Referências

- (1) Bailey TC. Spatial statistical methods in health. *Cad Saude Publica* 2001; 17(5):1083-1098.
- (2) Chamis NMA. Estudo da Maternidade na Adolescência em Campo Bom. Universidade Luterana do Brasil - ULBRA, 2002.
- (3) Comissão Nacional de População e Desenvolvimento (CNPD). Jovens no Brasil: diagnóstico nacional. 1997.
- (4) Elliott P, Wakefield J, Best N, Briggs D. *Spatial Epidemiology: Methods and Application*. London: Oxford University Press, 2001.
- (5) Elliott P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect* 2004; 112(9):998-1006.
- (6) Heilborn ML, Salem T, Rohden F, Brandão E, Knauth D, Victora C et al. Aproximações Sociantropológicas sobre a Gravidez na Adolescência. *Horizontes Antropológicos*, ano 8. Porto Alegre: 2002: 13-45.
- (7) Nobre FF. Introdução a Análise de Dados Espaciais. 32 Reunião Regional da ABE; Associação Brasileira de Estatística, 2001.
- (8) Santos SM, Barcellos C, Carvalho MS, Flores R. [Spatial clusters detection of violent deaths in Porto Alegre, Rio Grande do Sul, Brazil, 1996]. *Cad Saude Publica* 2001; 17(5):1141-1151.
- (9) Ventura SJ, Abma JC, Mosher WD, Henshaw S. Revised pregnancy rates, 1990-97, and new rates for 1998-99: United States. *Natl Vital Stat Rep* 2003; 52(7):1-14.

Anexo B - Etapas de aplicação do método de Epidemiologia Espacial

Quadro 2: Etapas da avaliação de Epidemiologia Espacial

Etapas	Mapeamento de doença	Avaliação de <i>Clusters</i>	Correlação Geográfica
1. Objetivos	-Descrever a distribuição espacial; -Gerar hipóteses; - Estimar a variabilidade do risco.	- Avaliar a distribuição dos pontos; - Verificar se aumento é ou não aleatório.	- Examinar relações com variáveis ambientais e sócio-econômicas , considerando a localização geográfica.
2. Formulação de hipóteses	Como é a distribuição geográfica da doença?	Há aumento de casos em alguns locais na presença de um agente?	Há associação entre o risco da doença e os fatores sócio-econômicos medidos?
3. Seleção da Região Geográfica	Definir unidade geográfica (município, bairro, setor censitário), população em risco, confundidores (fatores socioeconômicos), exposição ambiental.		
4. Qualidade dos dados	Geocodificação, correta notificação dos casos, verificação de duplicidade.		
5. Estimação do Risco	Através da relação entre observado e esperado (SRM)	Através da relação entre observado e esperado (SRM)	
6. Análise Exploratória	Mapeamento do risco e avaliação da hipótese, observando possíveis flutuações aleatórias	Mapeamento do risco e avaliação da hipótese observando possíveis flutuações aleatórias e locais com risco elevado	
7. Teste de hipóteses		Avaliar o padrão de pontos ou áreas e comprovar a presença ou não de <i>cluster</i>	
8. Modelagem	Modelagem clássica usando MLG com distribuição subjacente Poisson ou Binomial; Modelagem Bayesiana hierárquica	Modelagem considerando a presença de <i>cluster</i>	Modelagem clássica usando MLG com distribuição subjacente Poisson ou Binomial; Modelagem Bayesiana hierárquica

SMR – taxa de mortalidade padronizada

MLG – modelo ailinear generalizado

Anexo C - Comandos para o modelo totalmente Bayesiano no WinBUGS

Modelo sem covariáveis

```
model {  
  b [1:109] ~ car.normal (adj[], weights[], num[], taub)  
  b.mean <- mean(b[])  
  for (i in 1 : 109) {  
    casot[i] ~ dpois(mu[i])  
    log(mu[i]) <- log(e[i]) + u[i] + b[i]  
    RR[i] <- exp(u[i] + b[i])  
    u[i] ~ dnorm(0,tauu)  
  }  
  taub ~ dgamma(0.5,0.0005)  
  sibmab <- 1 / sqrt(taub)  
  tauu ~ dgamma(0.5,0.0005)  
  sigmau <- 1 / sqrt(tauu)  
}
```