

Testes de detecção de Conglomerados Espaciais

Renato M.
Assunção

Departamento de Estatística - Universidade Federal de Minas Gerais

Resumo

Este capítulo trata do problema de testar a presença de relativamente pequenas sub-regiões com parâmetros substancialmente diferentes que o restante do mapa quando os dados disponíveis são da forma de dados de área. As sub-regiões com suspeita de valores diferentes podem ser identificadas a priori (testes focados) ou podem não ser identificadas (testes genéricos).

Palavras-Chave: testes de detecção de conglomerados, testes focados.

1 Introdução

Profissionais de saúde pública frequentemente são chamados a intervir em regiões que parecem possuir uma taxa de ocorrência de agravos à saúde acima do que é esperado num dado período de tempo. Geralmente, no cálculo do número esperado de ocorrências, leva-se em consideração os fatores de risco que sabidamente afetam a incidência dos eventos, tais como: a distribuição por idade e sexo da população residente, o tempo de exposição a algum agente etiológico conhecido, etc. Procurando orientar ações de saúde pública nestas situações, os estudos de detecção de conglomerados de doenças ganharam bastante atenção dos pesquisadores nos últimos anos. Por exemplo, o problema da agregação espacial de leucemias e linfomas têm sido estudado com freqüência. De interesse especial tem sido a tentativa de verificar a hipótese de que estas doenças possuem uma causa viral através de testes de que a distribuição espacial dos casos é aleatória ou possui conglomerados de alta incidência. Está implícita a suposição de que a hipótese de origem virótica leva imediatamente a conglomerados, suposição esta que tem sido contestada mais recentemente, tendo em vista resultados de pesquisa sobre os mecanismos de transmissão viral e a existência de longos períodos de latência que amortecem a formação de conglomerados ao longo do tempo. Entretanto, os estudos de conglomerados são importantes pois é de interesse identificar áreas de risco mais elevado em meio a regiões maiores de risco aproximadamente constante para realizar estudos mais pormenorizados nas áreas identificadas. Um melhor entendimento da estrutura de conglomerados espaciais de uma doença leva a uma análise mais objetiva de qualquer suspeita de excesso de risco.

Em particular, a observação de um número bem maior do que o esperado de casos de leucemia infantil ao redor de uma instalação nuclear em Sellafield (Inglaterra) levou a uma grande controvérsia e serviu como poderoso argumento para a criação de um sistema inglês para coletar dados de incidência e mortalidade em pequenas áreas geográficas. Este caso gerou grande insatisfação entre epidemiólogos e estatísticos pois o conglomerado não foi detectado pelo sistema de vigilância epidemiológica existente e chegou ao público através de um programa de televisão. A principal motivação era a ocorrência de 4 casos de leucemia infantil registrados no pequeno *ward* de Seascale entre 1968 e 1982, quando eram esperados apenas 0.2 casos neste período. Este episódio foi amplamente discutido num número do *Journal of the Royal Statistical Society* (Gardner, 1989) e muitos desenvolvimentos estatísticos apareceram no contexto deste caso específico.

As áreas de risco significativamente elevado mas não esperado (ou não explicado) recebem o nome de conglomerados. Eles podem ser puramente temporais (se o espaço é ignorado) ou podem ser analisados levando-se em conta tanto o espaço quanto o tempo. Knox (1988) define um conglomerado (*cluster*, em inglês) como *um grupo de ocorrências geograficamente limitado de tamanho e concentração tais que seja improvável de ocorrer por mero acaso*. Infelizmente, esta definição não possui um caráter operacionalizável

Notas de aula do curso de Estatística Espacial, 2003. Agradeço a Mônica S. Monteiro-de-Castro a leitura cuidadosa de uma primeira versão desse texto e as várias sugestões de melhoria. E-mail: assuncao@est.ufmg.br

embora ela sugira métodos baseados em testes de hipóteses para identificar os conglomerados. De fato, a maioria dos métodos existentes são baseados em testes de significância ou testes de hipóteses. A literatura no assunto é vasta e vamos apresentar apenas alguns dos principais métodos propostos até o momento. Os métodos de detecção de conglomerados dividem-se em duas categorias gerais. O primeiro grupo de métodos é aquele em que os dados são coletados para testar a hipótese de um possível excesso ao redor de uma fonte suspeita e esta fonte é identificada *antes de observar os dados*. Eles são chamados de *testes focados de conglomerados*. O segundo grupo de métodos procura identificar as áreas geográficas com um risco significativamente elevado sem especificar de antemão quais e quantas áreas seriam estas. Eles são chamados de *testes genéricos de conglomerados*. As diferenças entre os dois grupos de métodos pode ser melhor entendidas considerando as duas perguntas que os seus testes de hipóteses associados procuram responder:

- Os testes focados procuram responder, com um *sim* ou um *não*, a seguinte pergunta: existe evidência nos dados de que ao redor da fonte suspeita, localizada na posição x_0 , há um aumento significativo do risco de certa doença? O fundamental é que a posição x_0 seja conhecida *antes* de se observar os dados.
- Os testes genéricos procuram responder, com um *sim* ou um *não*, a esta outra pergunta: existe evidência nos dados de que uma ou mais regiões do mapa, não identificadas previamente, possuem um risco significativamente maior que o valor médio observado no mapa como um todo?

Para ambos os tipos de testes o modelo nulo é o mesmo, o de ausência de conglomerados com o padrão espacial da doença sendo totalmente casual. Qual o sentido preciso desta afirmação? É o de que a taxa de incidência por pessoa é constante em todos os locais, implicando que o número esperado de casos em um local é proporcional ao número de pessoas em risco morando neste local. Suponha que a região de estudo é dividida em n áreas e denote o número observado de casos em cada uma delas por y_i e o número esperado de casos por E_i , $i = 1, \dots, n$. Seja N_i o número total de pessoas-ano em risco na área i e seja λ a taxa anual individual (ou per capita) de ocorrência da doença. O modelo nulo de aleatoriedade completa ou ausência de conglomerados é expresso como sendo a hipótese nula

$$H_0 : y_i \sim \text{Poisson}(E_i = \lambda N_i), \text{ independentes, } i = 1, \dots, n$$

Os testes são especificados tendo em mente as alternativas de interesse. No caso dos testes focados, espera-se que a taxa individual λ não seja constante mas que varie com i , aumentando quanto mais próximo for a área i da fonte suspeita. Os testes genéricos assumem apenas que existem uma ou mais regiões onde o valor de λ é mais elevado que no restante do mapa. Observe que, neste segundo caso, não está excluída a hipótese alternativa de heterogeneidade espacial, a simples variabilidade das taxas entre as áreas. Entretanto, estaremos mais interessados, e isto estará refletido nas estatísticas de teste, nas alternativas mais realistas de autocorrelação espacial em que as taxas de áreas vizinhas são similares. Exemplo Foi registrado por volta de 1984 um elevado número de casos de leucemia de 0 a 19 anos ao redor de 12 depósitos de lixo tóxico contaminados com um composto orgânico volátil chamado tricloroetileno. O Departamento de Saúde do Estado de Nova York iniciou uma série de estudos para averiguar o caso e criar uma política pró-ativa de saúde, em contraste com uma política reativa. Estudou-se o período de 1978 - 1982 numa região com pouco mais de 1 milhão de pessoas entre 0 e 19 anos. A região é dividida em 790 áreas e registrou 592 casos de leucemia no período. Apenas a área de residência do caso era conhecida.

Vamos utilizar este exemplo ao longo deste capítulo para ilustrar alguns dos métodos apresentados. A questão de interesse é saber se existem conglomerados de risco mais elevado na área (teste genérico) e, em particular, saber se existe evidência de risco mais elevado em volta dos 12 depósitos de lixo tóxico (teste focado).

Parte dos dados para a análise estatística consiste de uma tabela com 790 linhas, uma por área, e cinco colunas: identificador de área, coordenadas x e y , tamanho da população residente na área e número de casos de leucemia observados na área. A Tabela 1 mostra as cinco primeiras linhas da tabela de dados. A outra parte dos dados corresponde as posições geográficas dos 12 depósitos de lixo.

Exemplo Monteiro-de-Castro et al. (2001) analisaram os registros de residências de doadores de sangue da Fundação Hemominas de Belo Horizonte que estavam infectados com o vírus HTLV I/II. Cada doador

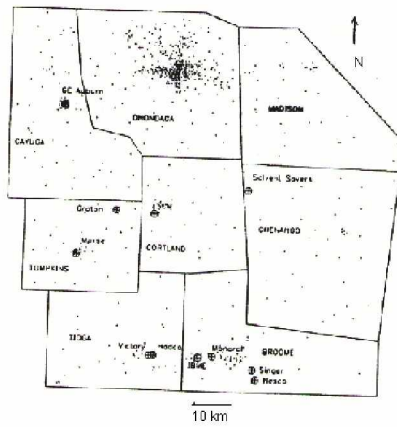


Figura 1: Mapa de região ao redor de Nova Iorque com posições dos centróides das áreas indicadas por um ponto e os locais de processamento de lixo contendo triclorotileno indicado por x (Waller et al., 1992)

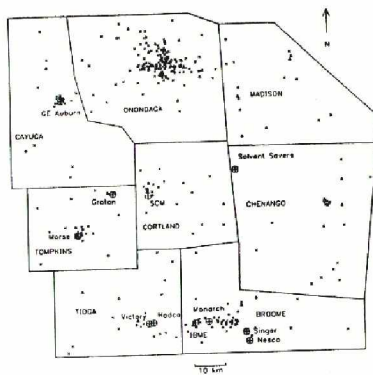


Figura 2: Mapa de região ao redor de Nova Iorque com posições dos centróides das áreas em que havia pelo menos um caso indicadas por um ponto. Áreas onde casos não foram registrados não são mostradas. (Waller et al. 1992)

Área	x	y	população	casos
1	4.07	-67.35	3540	3
2	4.64	-66.86	3560	4
3	5.71	-66.98	3739	1
4	7.61	-66.00	2784	1
5	7.32	-67.32	2571	3

Tabela 1: Tabela com as cinco primeiras linhas de dados referentes ao exemplo de leucemia no norte do estado de Nova York (Waller et al., 1992).

infectado, denominado caso, teve sua residência marcada como um evento no mapa de Belo Horizonte. Ao mesmo tempo, uma amostra de doadores que não eram infectados pelo vírus, denominados indivíduos de *controle*, também tiveram suas residências marcadas como eventos. Uma pergunta de interesse neste problema é saber se o padrão espacial de residências dos infectados é diferente do padrão espacial de residências dos indivíduos controles. Uma diferença no padrão espacial entre os dois grupos, casos e controles, pode existir se, por exemplo, existir algum fator associado à infecção que é distribuído na população de forma espacialmente diferenciada. Se fosse observado que o grau de concentração de casos na periferia é maior que aquele observado na amostra de controles, uma possível explicação para isto poderia estar associada à dinâmica migratória da cidade. Suponha que a probabilidade de um indivíduo ser infectado tem relação com seu passado migratório, migrantes do Norte de Minas Gerais tendo uma chance mais alta de serem infectados. Os migrantes não estão distribuídos de forma totalmente casual na cidade pois existem relativamente mais migrantes na periferia da cidade de Belo Horizonte. Assim, o padrão espacial da infecção poderia ser explicado por um fator, a migração, que está distribuída de forma espacialmente diferenciada. Encontrar conglomerados de infectados era portanto de interesse epidemiológico para traçar o perfil deste vírus no Brasil.

Um exemplo de teste focado também foi considerado neste trabalho. Um indivíduo testado é considerado positivo para o vírus HTLV se ele é positivo num primeiro teste (Elisa) e é positivo também num segundo teste (Western Blot) que ele faz num segundo momento. Nem todos os indivíduos positivos no primeiro teste (Elisa) retornaram ao Hemocentro para realizar o segundo teste. Como o hemocentro era localizado no centro da cidade de Belo Horizonte, surgiu a hipótese de que as pessoas que moravam na periferia, pobres em sua maioria, teriam um maior custo de retornar ao hemocentro para uma segunda visita (custo de transporte, perda de um dia de trabalho, etc.). Isto poderia ser testado através de um teste que supõe que a proximidade ao hemocentro seria um fator aumentava a chance de um indivíduo vir a retornar para um segundo exame. Isto é, na linguagem apresentada neste capítulo, o hemocentro seria um foco.

A análise em Monteiro-de-Castro et al. (2001) faz uso de testes e métodos diferentes daqueles apresentados aqui pois os dados são do tipo caso e controle e sob a forma de processos pontuais. Não vamos tratar em detalhes esta situação neste capítulo e o artigo em questão deve ser consultado para maiores informações pelo leitor interessado.

Assim, os testes de detecção de conglomerados procuram responder se existe evidência de aglomeração de casos acima da aglomeração que aparece naturalmente devido a distribuição geográfica não-homogênea das populações. Usualmente, o cálculo dos valores esperados em cada área é feito padronizando-se pela distribuição de idade e sexo da sua população, como já explicamos antes. A maior parte dos estudos de conglomerados que tem sido publicada envolve uma forma ou outra de câncer que atingem uma população de forma diferenciada por idade e sexo. Assim, se a distribuição etária e por sexo está disponível e taxas regionais ou nacionais estão disponíveis, a padronização é um cuidado básico que deve ser tomado.

2 Testes genéricos

Existem n áreas, y_i é o número observado de casos e E_i é o número esperado de casos, $i = 1, \dots, n$. Seja N_i o número total de pessoas-ano em risco na área i . Se não existe padronização por idade e sexo, $E_i = N_i$. Seja $\lambda = y_i / N_i$ a estimativa da taxa anual individual (ou per capita) de ocorrência da doença. Em todos os métodos, toma-se $\lambda = \bar{y}$ ignorando-se o erro de estimação envolvido já que a variância de \bar{y} é bastante pequena relativamente a variabilidade de y_i . A localização da área i é identificada com o seu centróide denotado por x_i .

2.1 Método GAM de Openshaw

Openshaw et al. (1988) propôs um método intensivo computacionalmente e com grande apelo visual chamado *Geographical Analysis Machine*, abreviado por GAM. Para aplicar o método, vamos considerar $E_i = N_i$, isto é, sem padronização por idade e sexo (a extensão para esta situação é imediata). Imagine uma grade regular bem fina sobreposta ao mapa, com nós regularmente espaçados, e indexada por $j, j = 1, \dots, J$. Associe os valores de cada área aos seus centróides. O procedimento segue o algoritmo descrito abaixo:

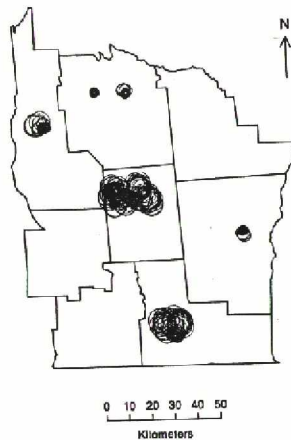


Figura 3: Resultados do método GAM de Openshaw et al. (1988) com círculos de raios 1, 2, e 4 km que foram significativos ao nível nominal de 0.002 (extraído de Turnbull et al., 1990)

- Selecione um raio r (por exemplo, $r = 1, 2,$ ou 4 km).
- Em cada nó j da grade regular, fixe um círculo C_{jr} de raio r .
- Calcule

$$O_{jr} = \sum_{i=1}^X y_i I_{[x_i, C_{jr}]} \text{ e } N_{jr} = \sum_{i=1}^X N_i I_{[x_i, C_{jr}]}$$

o número de casos e pessoas-ano das áreas cujos centróides caem dentro do círculo C_{jr} , respectivamente.

- Calcule o p -valor p_{jr} associado com O_{jr} sob a hipótese nula:

$$p_{jr} = P(X \geq O_{jr}) \text{ onde } X \sim \text{Poisson}(\lambda N_{jr})$$

- Desenhe o círculo C_{jr} se $O_{jr} \geq 2$ e se $p_{jr} \leq 0.002$.
- Refaça o procedimento acima aumentando escolhendo outro raio r para o círculo.

O resultado final é a identificação de áreas de risco por emaranhado de círculos sobrepostos, todos altamente significantes e com pelo menos dois eventos.

Exemplo Usando três valores de r (1,2, e 4 km), Turnbull et al. (1990) examinaram 83587 círculos com os dados de Nova York. A Figura 3 mostra os 503 círculos (0.6% do total) que foram significativos ao nível nominal de 0.002.

Embora cada círculo seja julgado individualmente significativo ao nível 0.002, um nível de significância para todos os círculos simultaneamente não é conhecido. A razão é o conhecido fenômeno de testes simultâneos não independentes que, neste caso, é particularmente difícil devido ao imenso número de testes efetuados. São realizados R testes em cada nó de uma grade fina de tamanho J , supostamente bastante grande, onde R é o número de raios experimentados (veja exemplo). Observe que, se um círculo é desenhado, vários outros círculos vizinhos serão desenhados, muitos contendo exatamente os mesmos casos. A escolha pouco comum do ponto de corte 0.002 para o p -valor é arbitrária e procura compensar, de alguma forma, este problema de testes simultâneos mas esta escolha carece de qualquer base teórica. Infelizmente, a escolha deste ponto de corte para o p -valor é crucial. Um ponto de corte muito grande (tal como o usual 0.05) levaria a um grande número de falsos positivos (isto é, a muitos falsos conglomerados) enquanto que um p -valor muito pequeno levaria a muitos falsos negativos (isto é, a perder muitos

conglomerados reais). Não existe solução satisfatória para este problema e o que pode ser recomendado é que o usuário do método experimente com diferentes pontos de corte para verificar a estabilidade dos resultados obtidos. Existe software gratuito disponível na internet para implementar este procedimento em ambiente UNIX e Windows. (INDICAR ???)

As vantagens e desvantagens deste método podem ser resumidas da seguinte maneira: é um método simples de entender, com grande apelo intuitivo, e visualmente impressionante; é um método exploratório e não inferencial devido ao problema de muitos testes simultâneos e dependentes implicando em níveis de significância nominais; é intensivo computacionalmente; os círculos não são inteiramente comparáveis entre si pois as variáveis aleatórias envolvidas possuem diferentes distribuições.

A última desvantagem da lista acima pode ser minorada se fixarmos o numerador ou denominador das taxas de cada círculo. Esta é a proposta do método de Besag e Newel, descrito a seguir.

2.2 Método de Besag e Newel

O método GAM de Openshaw fixa o raio r do círculo e calcula o p -valor da contagem que encontra dentro do círculo C_{jr} . Em termos gerais, o método de Besag e Newell (1991) fixa o número k de eventos que devem ser buscados e calcula o raio necessário para englobá-los. No círculo resultante, calcula p -valor e, procedendo como GAM, desenha apenas os círculos significativos (p -valor ≤ 0.002). A seguir, varia k para verificar a estabilidade dos resultados.

Mais especificamente, suponha que, em toda a região de estudo, existem $C = \sum_i y_i$ casos e $M = \sum_i N_i$ pessoas-ano em risco. Suponha que a área i possua pelo menos um caso. Seja L a variável aleatória que conta o número de outras áreas necessárias para acumular os k primeiros casos mais próximos de i . Seja l o valor observado de L e seja m o número de pessoas em risco nessas l áreas. Sob o modelo nulo, o número esperado de casos nessas l áreas é dado por mCM . O número N de casos nas primeiras l áreas possui distribuição de Poisson com valor esperado mCM . Assim, $P(L \leq l) = 1 - P(L > l + 1)$. Mas isto é simplesmente 1 menos a probabilidade de que as l primeiras áreas possuam menos que k casos. Desse modo, o p -valor é dado por

$$P(N = j) = 1 - \sum_{j=1}^{k-1} \frac{m^j C^j M^j}{j!} e^{-mCM}$$

Turnbull et al. (1994) usou este método nos dados de Nova York tomando $k = 8$ e considerando um conglomerado como significativo se ele tivesse p -valor menor que 0.05. Eles encontraram 31 conglomerados significativos. Comparando com os resultados do método GAM, eles observam que existe uma concordância geral entre os mapas com muitos dos conglomerados anteriormente sendo também detectados por este procedimento atual.

Por causa do mesmo problema de testes simultâneos que afeta o método GAM, o método de Besag e Newel vai sempre apresentar alguns conglomerados como significativos mesmo que a hipótese nula seja verdadeira. Para aliviar um pouco este problema, eles propuseram um teste estatístico global associado com o número total T_{BN} de conglomerados detectados. No exemplo acima, foram detectados 31 conglomerados nos dados de Nova York. Eles calculam um p -valor para esta estatística sob a hipótese nula de ausência de conglomerados. O valor exato da esperança de T_{BN} é fácil de ser calculada. Seja I_j uma função indicadora de que o círculo j cobrindo os k casos mais próximos é significativo.

Então $E(T_{BN}) = \sum_{j=1}^n E(I_j) = \sum_{j=1}^n \alpha_j$. Para calcular α_j , aplique a fórmula do p -valor acima para obter $P(L \leq h)$ para $h = 0, 1, \dots$ até que o nível de significância exceda α para $h = h_j$. O valor de $P(L \leq h_j)$ é o valor de α_j .

No caso dos dados de Nova York, encontra-se $E(T_{BN}) = 24.22$ que é bastante inferior ao número observado 31 sugerindo que um maior número de conglomerados foi observado do que o valor esperado sob a hipótese nula. Infelizmente, a variância de T_{BN} não pode ser calculada já que as variáveis I_j não são independentes. Entretanto, um teste de hipótese pode ser efetuado baseando-se na distribuição da estatística T_{BN} por simulação de Monte Carlo usando o seguinte algoritmo:

1. Gere S conjuntos de dados independentemente, onde cada réplica aleatória possui o mesmo número de casos C que o conjunto de dados original. Estes C casos são distribuídos ao acaso entre as n áreas de acordo com a hipótese nula. Isto é, são obtidos como realizações de uma distribuição multinomial $M(C, \pi_1, \dots, \pi_n)$ onde $\pi_j = N_j / M$.

2. Em cada um dos S conjuntos de dados gerados, rode todo o procedimento de Besag e Newel e calcule a estatística T_{BN} obtendo T_1, \dots, T_S .
3. Ordene o valor de T_{BN} do conjunto de dados original e dos S conjuntos simulados. Denote o posto da estatística T_{BN} associada ao conjunto de dados original por R . Se R estiver entre os $100(1 - \alpha)\%$ maiores postos, rejeite a hipótese nula ao nível de significância α . O p -valor associado com este teste é $1 - R/(S + 1)$.

Usando 999 réplicas, Turnbull et al (1994) encontraram um p -valor de 0.18 para $k = 8$ e $T_{BN} = 31$ nos dados de Nova York. Isto indica que não há evidência de conglomerados nestes dados de Nova York utilizando este teste genérico. Observe que nenhuma informação sobre as posições dos depósitos de lixo foi utilizada neste teste.

Embora este teste baseado na estatística T_{BN} seja extremamente útil ele não ajuda a identificar quais são os conglomerados significativos que deveriam ser analisados. Em princípio, poderíamos pensar em usar aqueles dentre os contados em T_{BN} que tiveram os menores α_j . No entanto, não sabemos quantos desses considerar e também sabemos que o p -valor depende do tamanho da população sob consideração (ver Capítulo 2).

As vantagens e desvantagens do método de Besag e Newel são as seguintes: ele estabiliza mais as estatísticas de teste locais; continua a existir o problema de testes simultâneos e portanto o método é exploratório também; é visualmente agradável como o GAM; como GAM, ele identifica os conglomerados.

2.3 Adaptação do método de Cuzick e Edwards

Cuzick e Edwards (1990) fizeram uma proposta que trabalha com uma amostra de casos e outra de controles sob a forma de um processo pontual. Esta proposta pode ser facilmente adaptada para trabalhar com dados de área como mostramos a seguir, representando uma pequena variação em relação aos testes propostos por Besag e Newel (1991) e por Turnbull et al. (1990). Como Besag e Newel (1991), eles começam fixando o número de casos r . A seguir, em torno do centróide de cada área i que possui pelo menos um caso, trace um círculo que vai se expandindo seu raio até que o círculo contenha uma população para a qual *esperamos observar* r eventos. A seguir, verifique quantos eventos O_i foram de fato observados e calcule então

$$U_r = \sum_{i=1}^K (O_i - r) I_{[y_i > 0]}$$

Cuzick e Edwards derivaram as fórmulas dos momentos desta estatística sob a hipótese nula e mostraram que ela possui uma distribuição assintoticamente normal possibilitando assim realizar um teste. Entretanto, recomendamos que seja feito um teste baseado na distribuição de Monte Carlo obtida simulando S padrões obtidos através da distribuição multinomial.

Este teste apresenta as mesmas vantagens e desvantagens daquele teste proposto por Besag e Newel (1991) mas não existe estudo comparativo envolvendo este método.

2.4 Método de Varredura de Kulldorf

Kulldorf e Nagarwalla (1995) e Kulldorf (1997) apresentaram um método que permite resolver o problema de testes simultâneos de forma bastante elegante envolvendo um número mínimo de parâmetros não cruciais para serem escolhidos pelo usuário. Ele possui várias extensões e está **implementado** no software SaTScan, disponível gratuitamente na internet no endereço

<http://dcp.nci.nih.gov/BB/SaTScan.html>

Vamos adotar a mesma notação de Kulldorf e Nagarwalla (1995) embora ela fique um pouco confusa, às vezes. Seja N o número total de indivíduos na população em risco e C o número total de casos observado na região de estudo. A análise é feita condicionalmente no número total de casos e assim C é tratado como uma constante conhecida.

Denote por Z o conjunto das áreas z candidatas a formarem um conglomerado. Estes candidatos z são os círculos de raio r arbitrário centrados em cada um dos n centróides x_i . Em tese, existe um número infinito de tais círculos mas, na prática, estes círculos podem ser restritos apenas aqueles centrados em x_i e com raios iguais as distâncias entre x_i e os demais centróides. A razão é que círculos com raios ligeiramente diferentes vão conter, em geral, as mesmas áreas e portanto definem o mesmo possível conglomerado. O número de círculos $z \in Z$ pode ser ainda mais reduzido se colocarmos um limite no raio de modo que nenhum candidato a conglomerado z contenha mais do que uma certa porcentagem, tal como 20%, da população total da área. Afinal, não faz sentido propor um conglomerado que contenha 50% da população total.

O teste de varredura de Kulldorf é baseado no método de máxima verossimilhança. O parâmetro neste caso é (z, p, r) onde z denota o círculo em Z , possivelmente parametrizado pelo seu raio e coordenadas do centro, p é a probabilidade de que um indivíduo qualquer dentro de z seja um caso enquanto que r é a mesma probabilidade fora de z . A hipótese de interesse que os autores consideram é que existe uma única zona z tal que $p > r$. Isto é, a hipótese alternativa é: $H_1 : z \in Z, p > r$. A hipótese nula é que $H_0 : p = r$, a aleatoriedade completa com cada indivíduo sendo igualmente provável de se tornar um caso.

Seja n_z o número de indivíduos na zona z , seja C_z a variável aleatória denotando o número de casos na zona z e seja c_z o número observado de C_z nos dados. A função de log-verossimilhança para o parâmetro (z, p, r) é dada por

$$L(z, p, r) = p^{c_z}(1-p)^{(n_z-c_z)} r^{(C-c_z)}(1-r)^{(N-n_z-C+c_z)}$$

O valor do parâmetro que maximiza esta função não é necessariamente aquele correspondendo ao círculo com a maior taxa $p = c_z/n_z$ nem aquele com o maior número de casos c_z . A razão de verossimilhança é

$$\lambda = \frac{\sup_{z \in Z, p > r} L(z, p, r)}{\sup_{p=r} L(z, p, r)}, \quad p, r \in [0, 1]$$

O denominador desta razão é simplesmente

$$\sup_{p \in (0,1)} p^C(1-p)^{N-C} = \frac{C^C(N-C)^{N-C}}{N^N} = L_0$$

o qual não depende da zona z nem da distribuição espacial dos casos mas apenas do número total de casos.

O numerador da razão de máxima verossimilhança λ é obtido em duas etapas de maximização sucessivas:

$$\sup_{z \in Z, p > r} L(z, p, r) = \sup_z \sup_{p > q} L(z, p, q)$$

Isto é, fixamos $z \in Z$ e encontramos a solução $p(z)$ e $r(z)$ que maximiza a verossimilhança e, a seguir, trabalhamos com $L(z, p(z), r(z))$ para obter a solução que maximiza em z . Assim, para um possível candidato a conglomerado z , temos

$$L(z, p(z), r(z)) = \sup_{p > r} p^{c_z}(1-p)^{(n_z-c_z)} r^{(C-c_z)}(1-r)^{(N-n_z-C+c_z)}$$

cuja solução é dada por

$$\frac{c_z}{n_z} > \frac{C-c_z}{N-n_z}, \quad \text{se } \frac{c_z}{n_z} > \frac{C-c_z}{N-n_z}$$

$$\frac{c_z}{n_z} \leq \frac{C-c_z}{N-n_z}, \quad \text{se } \frac{c_z}{n_z} \leq \frac{C-c_z}{N-n_z}$$

A seguir, o conglomerado mais verossímil é aquela zona b para a qual $L(z, p(z), r(z))$ é maximizada: $L(b, p(b), r(b)) \geq L(z, p(z), r(z))$ para todo $z \in Z$.

A distribuição de λ depende da distribuição da população e é virtualmente impossível de ser obtida analiticamente. Além disto, a aproximação assintótica usual por uma distribuição qui-quadrado da transformação $-2 \log \lambda$ não é válida pois as condições de regularidade não são satisfeitas. Assim, a saída proposta por Kulldorf e Nagarwalla (1995) é utilizar um procedimento de simulação Monte Carlo para obter a distribuição exata de λ condicionada no número total de casos observado C . O algoritmo é o seguinte:

1. Gere S conjuntos de dados independentemente, onde cada réplica aleatória possui o mesmo número de casos C que o conjunto de dados original. Estes C casos são distribuídos ao acaso entre as n áreas de acordo com a hipótese nula. Isto é, são obtidos como realizações de uma distribuição multinomial $M(C, \pi_1, \dots, \pi_n)$ onde $\pi_j = N_j / \sum_j N_j$, proporcional a população de cada área. Alternativamente, π_j é proporcional ao número esperado de eventos em cada área, caso variáveis de confundimento (tais como idade e sexo) estejam sendo usadas.
2. Em cada um dos S conjuntos de dados gerados, calcule a estatística do teste da razão de verossimilhança obtendo $\lambda_1, \dots, \lambda_S$.
3. Ordene o valor de λ do conjunto de dados original e dos S conjuntos simulados. Denote o posto da estatística λ associada ao conjunto de dados original por R . Se R estiver entre os $100(1 - \alpha)\%$ maiores postos, rejeite a hipótese nula ao nível de significância α . O p -valor associado com este teste é $1 - R/(S + 1)$.
4. Caso a hipótese nula tenha sido rejeitada, salve a zona b associada com o máxima verossimilhança do modelo não nulo. Este é o conglomerado mais verossímil.

Para o modelo em que os dados possuem distribuição de Poisson, seja $E(\Psi)$ o número de casos esperado em Ψ sob o modelo nulo (geralmente, $E(\Psi) = \lambda N_\Psi$ onde N_Ψ é o número de pessoas-ano dentro do candidato a conglomerado Ψ). Então

$$\frac{L(\Psi)}{L_0} = \frac{O_\Psi}{E(\Psi)}^{\alpha} \frac{C - O_\Psi}{M - E(\Psi)}^{1-\alpha}$$

se $O_\Psi > E(\Psi)$, e 1 caso contrário.

O teste baseado na estatística de varredura espacial possui as seguintes vantagens:

- leva em conta a densidade populacional não constante na região e também qualquer número de variáveis de risco conhecidas tais como idade e sexo.
- Procura conglomerados sem especificar de antemão sua localização e tamanho.
- Se a hipótese nula é rejeitada, o teste fornece a localização do conglomerado mais verossímil que levou a rejeição.
- Evita o problema de teste múltiplos forçando um p -valor real. Esta é sua principal vantagem.

As desvantagens do teste de varredura espacial são as seguintes:

- O método fixa a forma geométrica dos candidatos a conglomerados como círculos (ou, alternativamente, alguma outra forma simples tal como elipses ou retângulos). Isto tende a criar conglomerados compactos englobando muitas vezes áreas que, de fato, não fazem parte do conglomerado. Aplicações deste método em Belo Horizonte apresentaram este problema de forma bastante clara. (REF ?????)
- O método deve ter um poder baixo contra alternativas com um grande número de pequenos clusters localizados em posições bastante diferentes.

3 Testes focados

Os testes focados são aqueles em que, antes de observar os dados, identificam-se k posições no mapa que são fontes suspeitas de aumentar o risco em sua vizinhança. A hipótese nula é a mesma de antes mas agora a alternativa é mais precisamente especificada e, para este problema de testes focados de conglomerados, é possível definir testes mais poderosos do que aqueles apresentados anteriormente para testes de conglomerados genéricos. De fato, a experiência de saúde pública tem mostrado que em situações de vigilância (*surveillance*), um baixo poder é a tônica. Uma forma típica de aumentar o poder em testes

estatísticos é reduzir o conjunto de alternativas e procurar estatísticas de teste que sejam poderosas contra este conjunto reduzido de alternativas. Reduzindo o poder contra alternativas improváveis ou sem interesse prático, um maior poder pode ser obtido contra as alternativas que permanecem de interesse. Geralmente, os testes focados procuram verificar se o risco aumenta com a proximidade da fonte suspeita. A dificuldade técnica neste problema é que não é conhecida a forma funcional da relação entre risco e distância à fonte. Stone (1988) propôs um teste que não depende do conhecimento desta forma funcional. Como antes, a hipótese nula é que todas as pessoas possuem o mesmo risco de ser um caso, independentemente dos outros casos e da posição de sua residência. Mais especificamente, supõe-se que $y_i \sim \text{Poisson}(E_i)$ onde, no caso mais simples, $E_i = \lambda N_i$, com N_i igual ao número de pessoas em risco na área i e λ a taxa global na região em estudo. É comum que λ seja estimado a partir dos próprios dados como $\hat{\lambda} = \sum y_i / \sum N_i$ com a variabilidade resultante sendo ignorada a seguir. Nas situações com mais informação, o valor de E_i é calculado ajustando-se por fatores de risco conhecidos e diferentes entre as áreas, tais como idade e sexo. A hipótese alternativa agora é que o risco aumenta com a proximidade de locais pré-especificados mas sem estabelecer qual a forma funcional deste aumento.

O problema da escala é um aspecto prático importante nestes estudos. Um conglomerado é semelhante a um *outlier* num estudo estatístico. Uma observação é considerada um *outlier* no contexto de um modelo estatístico. Do mesmo modo, a noção do que é um conglomerado depende da região sob consideração. Uma região muito grande fatalmente será heterogênea em aspectos que estão relacionados com o risco mas que não estão disponíveis. É comum até que grandes diferenças regionais em taxas sejam relacionadas a diferenças na cobertura do registro de casos. Assim, os estudos de conglomerados trabalham com regiões relativamente homogêneas e variando entre 10 e 25 km de raio em torno do suposto foco. No Reino Unido, os estudos da Small Area Health Statistics Unit usam um raio de 10 km em torno das fontes suspeitas como sendo a região de estudo.

Exemplo No exemplo de Turnbull et al. (1994), 12 locais são possíveis fontes de contaminação. Estes locais são os depósitos de lixo tóxico contaminados pelo composto orgânico volátil tricloroetileno (TCE). Este composto contamina o solo e tem sido associado com incidência de leucemia, embora a evidência seja fraca. Lençóis de água contaminada frequentemente contêm TCE bem como outros compostos cujo efeito cancerígeno é mais conhecido. Assim, a presença de TCE foi considerado uma proxy para a exposição a compostos cancerígenos em geral no solo.

Exemplo Elliott et al. (1996) estudaram a incidência de câncer do estômago ao redor de incineradores municipais de lixo sólido. Os dados consistem do número de casos de câncer de estômago em cada Distrito de Enumeração (*Enumeration District*, em inglês), que são similares aos Setores Censitários brasileiros. Além destas observações aleatórias, eles também possuíam o número esperado de casos padronizado por idade e sexo e um índice de qualidade de vida chamado Carstairs.

3.1 Adaptação dos métodos anteriores

Uma forma simples de atacar este problema é adaptar alguns dos métodos anteriores de teste genéricos a este problema de testes focados. Assim, Besag e Newell (1991) propuseram determinar a região necessária para obter pelo menos r casos em torno apenas das fontes suspeitas. Se as fontes não forem muitas, o efeito de testes simultâneos podem ser controlados através de procedimentos simples tais como correções de Bonferroni. Embora atrativo, este procedimento sofre da desvantagem de ignorar a distância até a fonte ao considerar do mesmo modo em r os casos muito próximos da fonte e aqueles um pouco mais afastados. Além disto, ele necessita da determinação de r antes de observar os dados o que nem sempre é fácil de ser feito. Em geral, r é escolhido de forma intuitiva levando-se em conta a densidade populacional e a escala na qual os conglomerados são antecipadamente esperados. No entanto, uma escolha mal feita do valor de r , um r muito maior do que o necessário, por exemplo, pode levar a não detecção de um conglomerado pequeno que aparece submerso numa região maior que foi necessário acumular em torno da fonte.

Exemplo Usando os dados de Nova York, pode-se calcular o p -valor associado com cada um dos 12 locais onde há depósito com TCE utilizando $r = 6, 8, 10$ (ver Tabela 2). Apenas dois locais possuem p -valor menor que 0.05, um foco quando $k = 6$ e outro foco quando $k = 10$. Os dois focos foram diferentes. Assim, este método não encontra evidência suficiente nos dados para afirmarmos que existe efeito de aumento de risco ao redor das fontes. O teste global é baseado na estatística T_{BN} e seu p -valor foi

Foco	$r = 6$	$r = 8$	$r = 10$
Monarch	0.069	0.072	0.013
IBM	0.227	0.056	0.072
Singer	0.665	0.725	0.470
Nesco	0.665	0.725	0.470
GE Auburn	0.108	0.133	0.143
Solvent Savers	0.551	0.254	0.168
Smith Corona	0.337	0.258	0.120
Victory Plaza	0.013	0.492	0.237
Hadco	0.769	0.490	0.284
Morse Chain	0.843	0.726	0.879
Groton	0.847	0.875	0.722
Teste Global	0.315	0.602	0.334

Tabela 2: Tabela com os resultados de aplicação do teste de Besag e Newel com $k = 6, 8, 10$ nos dados de leucemia no norte do estado de Nova York (Waller *et al.*, 1992).

calculado usando-se a distribuição Monte Carlo desta estatística.

3.2 Teste de escore

Sem perda de generalidade, rotule as n áreas pela distância à fonte de modo que $i = 1$ é a mais próxima da fonte e $i = n$ é a mais distante. Suponha que as observações são as contagens $y_i \sim \text{Poisson}(\lambda_i E_i)$ e que $\sum_{i=1}^n y_i = C$. Suponha que os riscos relativos de cada área são conhecidos sob as duas hipóteses, nula e alternativa, e são escritos como $\lambda_{10}, \lambda_{20}, \dots, \lambda_{n0}$ sob a nula e como $\lambda_{11}, \lambda_{21}, \dots, \lambda_{n1}$ sob a alternativa. O logaritmo da razão de verossimilhanças sob as duas hipóteses é dado por

$$\log L = \sum_{i=1}^n [y_i \log (\lambda_{i1} / \lambda_{i0}) - E_i (\lambda_{i1} - \lambda_{i0})]$$

Para valores fixos e conhecidos dos riscos, o segundo termo acima é constante e portanto, pelo lema de Neyman e Pearson, o teste mais poderosos de H_0 versus H_1 é da forma

$$\sum_{i=1}^n y_i \log (\lambda_{i1} / \lambda_{i0}) > c$$

para alguma constante c . Assumindo que, para todo i , $\lambda_{i0} = \lambda_0$, desconhecido, o melhor teste é dado por

$$\sum_{i=1}^n y_i \log \lambda_{i1} - \sum_{i=1}^n y_i \log \lambda_0 > c$$

Se os valores esperados foram calculados de tal forma que podemos assumir $\lambda_0 = 1$, então o teste acima é facilmente calculado quando conhecemos os valores λ_{i1} . No entanto, muitas vezes, λ_0 é desconhecido pois os valores esperados podem ter sido calculados usando-se valores de uma região muito maior, tal como um estado ou todo o país, que a estudada. A razão para isto será o pequeno número de casos da região estudada para efetuar uma padronização indireta. Entretanto, como C é estatística ancilar para λ_{i1} , a análise apropriada é condicionar no estimador de máxima verossimilhança $\lambda_0 = C / \sum_{i=1}^n E_i$ e usar a razão de log-verossimilhança baseada na distribuição multinomial

$$\sum_{i=1}^n y_i \log \lambda_{i1} - C \log \frac{\prod_{i=1}^n \lambda_{i1}^{E_i}}{E_i^{E_i}}$$

onde o segundo termo é uma constante. Isto é equivalente portanto a trabalhar com um teste baseado na soma ponderada de todos os casos usando como peso um escore que é a taxa λ_{i1} da área de residência dos casos.

Entretanto, os valores λ_{1i} usualmente não serão conhecidos. Existe a alternativa de impor uma forma paramétrica a estes riscos relativos mas isto não é uma estratégia muito adotada. A razão é que não existem elementos suficientes para assumir esta forma paramétrica e, em geral, C e cada y_i será muito pequeno para permitir uma boa estimação de parâmetros que modelem o efeito de fonte que não deve ser muito grande.

Entretanto, esta análise levou clássica motivou Waller et al. (1992) e Lawson (1993) que, independentemente propuseram o mesmo teste localmente mais poderoso baseado na estatística escore considerando a seguinte particular hipótese alternativa:

$$H_1 : \lambda_{1i} = 1 + \alpha^2 g_i$$

onde α^2 é uma constante positiva mas desconhecida e g_i são constantes conhecidas e representando proxies para a exposição na área i . É comum, por exemplo, tomar $g_i = 1/d_i$ ou $g_i = \exp(-d_i/K)$ onde d_i é a distância do centróide da área i até a fonte e K é fixo e conhecido. Note que este é um modelo com único parâmetro, α^2 , e que a hipótese nula corresponde a situação em que $\alpha^2 = 0$. O ponto importante é que a estatística do teste de escore localmente mais poderoso não depende de α^2 .

$$T = \frac{\partial \log L}{\partial \alpha^2} \Big|_{\alpha^2=0} = \sum_{i=1}^n g_i (y_i - E_i)$$

Assim, estes autores propuseram o teste baseado na estatística T definida acima que, sob a hipótese nula, possui esperança igual a zero e variância igual a informação de Fisher, que é aproximadamente igual a

$$\text{Var}(T) \approx \sum_{i=1}^n g_i^2 E_i$$

A estatística padronizada $T = U / \sqrt{\sum_{i=1}^n g_i^2 E_i}$ possui distribuição assintótica $N(0, 1)$. Esta aproximação é geralmente muito boa.

Se várias fontes estão presentes e são relativamente próximas, é comum definir g_i usando a menor distância da área i a cada uma das fontes presentes na região. Vamos apresentar um exemplo com os dados de Nova York no final da próxima seção.

3.3 Teste de Stone (1988)

Stone (1988) propôs um teste para uma única fonte pontual suspeita que tornou-se bastante popular. Um grande número de estudos de casos têm sido baseado neste teste (Bithell, 1990; Bithell e Stone, 1989; Elliott et al., 1992; Turnbull et al., 1990; Waller et al., 1992, entre outros). Este método tem sido considerado o melhor teste focado para dados de área até o momento, com algumas adaptações aparecendo ocasionalmente (Morton-Jones et al., 1999). Ele é o método preferido pela importante agência britânica Small Area Health Statistics Unit (SAHSU).

Sem perda de generalidade, rotule as n áreas pela distância à fonte de modo que $i = 1$ é a mais próxima da fonte e $i = n$ é a mais distante. Suponha que as observações são as contagens y_i Poisson($\lambda_i E_i$) e que $\lambda_i y_i = C$. Nesse teste, as hipóteses nula e alternativa são as seguintes:

$$H_0: \lambda_1 = \lambda_2 = \dots = \lambda_n$$

$$H_A: \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$$

A estatística de teste é o máximo do risco relativo estimado para uma região em torno da fonte suspeita:

$$T_{\text{Stone}} = \max_{1 \leq j \leq n} \frac{P_j}{\sum_{i=1}^{j-1} y_i / E_i}$$

O p -valor do teste associado com o valor observado $T_{\text{Stone}} = t$ é $P(T_{\text{Stone}} > t) = 1 - P(T_{\text{Stone}} \leq t)$, como explicaremos abaixo. Embora o teste seja intuitivamente simples, ele não é um teste uniformemente mais poderoso (UMP) para um acréscimo de risco monótono com a distância a fonte, como vimos na

seção anterior. Entretanto, raramente é o caso em que existem dados suficientes em cada área (isto é, com pequena variância relativamente à média) para que uma forma paramétrica específica (tal como um decaimento linear ou exponencial a partir da fonte) seja adotado. Assim, o teste proposto por Stone (1988) continua sendo uma boa solução caso não haja informação adicional disponível. Seja E o maior inteiro menor ou igual a $t \sum_{i=1}^n E_i$. Então,

$$P(T_{\text{Stone}} \leq t) = P \left(\prod_{k=1}^E y_k \leq t \prod_{k=1}^E E_k : i = 1, \dots, n \right) = \sum_{j=0}^E q(n, j)$$

e

$$q(v, j) = P \left(\prod_{k=1}^j y_k = j \cap \prod_{k=1}^v y_k \leq t \prod_{k=1}^v E_k : i = 1, \dots, v \right)$$

As probabilidades $q(v, j)$ são interpretadas como a probabilidade de que as v áreas mais próximas da fonte contenham j casos e que todas as razões acumuladas até v de casos sobre número esperado é menor que t . Estas probabilidades são obtidas de forma recursiva a partir das equações *backward* associadas com um passeio aleatório no plano e, sob a hipótese nula, são dadas por

$$q(v, j) = \begin{cases} P_j \prod_{k=1}^j \frac{1}{E_k} & j = 0, \dots, E \\ 0 & j > E \end{cases} \quad \text{com } q(v-1, j-k) \exp$$

com $q(0, 0) = 1$ e E o maior inteiro menor ou igual a $t \sum_{i=1}^n E_i$.

Este procedimento pode ser visto como um meio de ajustar o nível de significância nominal por causa do efeito de selecionar uma região ao redor da fonte que possui o maior risco relativo observado.

Um programa de computador para realizar estes cálculos pode ser obtido diretamente do autor. No entanto, na prática, é sempre mais simples obter a distribuição da estatística de teste por meio de simulação Monte Carlo sob a hipótese nula condicionada no número total C de casos observado.

1. Gere S conjuntos de dados independentemente, onde cada réplica aleatória possui o mesmo número de casos C que o conjunto de dados original. Estes C casos são distribuídos como realizações de uma distribuição multinomial $M(C, \pi_1, \dots, \pi_n)$ onde $\pi_j = E_j$.
2. Em cada um dos S conjuntos de dados gerados, calcule a estatística de teste T_{Stone} obtendo T_1, \dots, T_S , além do valor t calculado com os dados originais.
3. Ordene os valores $\{t, T_1, \dots, T_S\}$ e denote o posto de t por R . Se R estiver entre os $100(1 - \alpha)\%$ maiores postos, rejeite a hipótese nula ao nível de significância α . O p -valor associado com este teste é $1 - R/(S + 1)$.

Uma simples modificação do teste é feita quando existe mais de um foco suspeito presente na região e esses focos não estão muito distantes entre si. Ordenam-se as áreas pelas suas distâncias ao foco mais próximo e assim realizar um teste simultâneo de que o risco aumenta quanto mais próxima a área está de qualquer um das fontes suspeitas. Observe que a alternativa considerada aqui é que o risco aumenta em torno de todas as fontes, embora não precise ser da mesma forma em cada uma delas. No entanto, o teste terá menos poder na situação em que algumas fontes aumentam os riscos enquanto outras não o fazem.

Quando F focos estão em posições tão separadas que a análise é feita separadamente para cada foco, a alternativa para obter um p -valor agregado é calcular a tradicional estatística proposta por Fisher. Neste caso, combinamos os p -valores p_k de cada fonte individual obtendo a estatística $-2 \sum_{k=1}^F \log p_k \sim \chi^2_{2F}$.

A principal consideração que deveria nos guiar na escolha de um teste apropriado é o seu poder relativamente a outros testes candidatos com relação a hipóteses alternativas de interesse. Infelizmente, não existem resultados analíticos sobre o poder da estatística T_{Stone} . A distribuição nula pode ser aproximada por uma mistura de distribuições qui-quadrado mas com os pesos da mistura desconhecidos. Parece inevitável que conclusões sejam baseadas em estudos de simulação mas não existem resultados conclusivos ainda.

Foco	T_{Stone}	T
Monarch	0.011	≤ 0.001
IBM	0.036	≤ 0.001
Singer	0.245	0.007
Nesco	0.245	0.007
GE Auburn	0.156	0.020
Solvent Savers	0.502	0.386
Smith Corona	0.162	0.005
Victory Plaza	0.102	0.024
Hadco	0.572	0.031
Morse Chain	0.784	0.782
Groton	0.385	0.209
Teste Global	0.207	≤ 0.001

Tabela 3: Tabela com os resultados de aplicação do teste de Stone e de escore nos dados de leucemia no norte do estado de Nova York (Waller *et al.*, 1992).

Exemplo No exemplo do estado de Nova York, aplicando-se os métodos de Stone e de escore, encontramos os resultados da Tabela 3. Fica claro que o teste global de escore detectou forte evidência de que o risco é mais elevado ao redor dos depósitos de lixo enquanto que o teste de Stone não foi capaz de detectar este efeito. Ao mesmo tempo, o teste de escore parece ser mais sensível que o de Stone quando aplicado em cada um dos locais separadamente.

4 Conclusões

Testes de comparação entre estes métodos, genéricos ou focados, existem mas em geral não são totalmente conclusivos. Uma das dificuldades é a dependência da distribuição da estatística de teste na estrutura da população em risco, da forma de afastamento da hipótese nula, do nível da taxa de ocorrência de casos e da divisão política adotada. A dependência de tantos fatores mais o número muito grande de testes que continuam sendo propostos, dos quais cobrimos apenas os principais, faz com que sugestões quanto ao teste a ser adotado muito firmes sejam difíceis de serem dadas.

Uma das comparações mais importantes feitas com testes de conglomerados genéricos foi aquela realizada por Alexander e Boyle (1996). Estes pesquisadores convenceram os principais autores dos métodos propostos até aquele momento a submeterem-se a um experimento. Os autores receberam um grande número de mapas onde deveriam aplicar seus métodos para detectar possíveis conglomerados. Alguns mapas não possuíam nenhum conglomerado enquanto que outros mapas possuíam um número variável de conglomerados que também variavam em tamanho. Os autores não sabiam nada sobre os mapas a não ser o procedimento genérico adotado para gerá-los. Eles eram livres para escolher da forma que preferissem os parâmetros associados com seus métodos (tal como o p-valor no caso do GAM de Openshaw ou de Besag e Newel). Após aplicar os seus métodos, os autores declaravam se havia conglomerado no mapa, quantos eram e onde estavam.

Várias análises dos resultados obtidos foram feitas e, embora Alexander e Boyle (1996) sejam cautelosos ao sugerir conclusões, fica claro a partir do experimento que os métodos GAM de Openshaw e o de Besag e Newel são os melhores. Como estes métodos exigem uma série de escolhas de valores de parâmetros para os seus algoritmos, esta superioridade pode não se manter nas mãos de pessoas sem a devida experiência com os métodos e com o problema concreto sob análise.

Duas importantes qualificações são necessárias. A primeira é notar que houve uma importante ausência na comparação de Alexander e Boyle (1996), a do método de Kulldorf e Nagarwalla (1995) que estava sendo publicado junto com o livro acima. Isto é lamentável pois este método é o único que controla o erro tipo I da forma correta. A outra qualificação diz respeito a uma limitação deste estudo. Não era difícil imaginar que, dos muitos mapas a serem analisados, uma proporção substancial, talvez mais de 50% deles, deveriam conter pelo menos um conglomerado. Assim, os autores que anteciparam corretamente este fato e que então implicitamente usaram erros tipo I mais generosos tendiam a ser favorecidos. Isto é, uma escolha de um p-valor não muito pequeno para cada círculo testado individualmente seria benéfico

aos métodos GAM de Openshaw e o de Besag e Newel.

Quanto aos testes focados, alguns estudos de poder tem mostrado que ambos, o teste de Stone e o teste de score, são os melhores testes disponíveis (Waller e Lawson, 1995; Waller e Poquette, 1999).