

Estudos de agregados de doença no espaço-tempo: conceitos, técnicas e desafios

Studies on space-time disease clusters:
concepts, techniques, and challenges

Guilherme L. Werneck ¹
Claudio J. Struchiner ²

¹ Departamento de Medicina Preventiva, Faculdade de Medicina, Universidade Federal do Rio de Janeiro. Av. Brigadeiro Trompovsky s/nº, HCCFF, 5º andar, Ilha do Fundão, Rio de Janeiro, RJ 21949-900, Brasil.
² Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz. Rua Leopoldo Bulhões 1480, Rio de Janeiro, RJ 21041-210, Brasil.

Abstract *Methods for the analysis of space-time clusters have been used frequently in epidemiology and public health in recent decades. However, they have generally been applied without an appropriate critique of their limits and usefulness. This article reviews some of the most important issues involved in improving the interpretability and utility of such methods in this context. We begin with a brief discussion about the epidemiological meaning of the concept of space-time clustering. We then describe five commonly used methods, including the procedures for their use. We also present a conceptual and statistical generalization of the techniques. Finally, we summarize some important issues for use of these methods in epidemiology and public health.*

Key words *Space-Time Clustering; Epidemiologic Methods; Disease Clusters; Epidemiology*

Resumo *Os estudos de agregados espaço-temporais têm sido freqüentemente utilizados em epidemiologia e saúde pública nas últimas décadas, muitas vezes sem a adequada compreensão de seus limites e potencialidades. O objetivo geral deste artigo é rever alguns tópicos essenciais para aprimorar a interpretação e utilidade de estudos neste contexto. Inicia-se com uma discussão sobre o significado epidemiológico do conceito de agregação espaço-temporal. A seguir, descrevem-se cinco métodos analíticos, assim como os procedimentos necessários para sua utilização. Apresenta-se também uma proposta de generalização conceitual e estatística das técnicas analisadas. Por fim, discute-se uma gama de tópicos relevantes para a otimização do uso destes estudos em epidemiologia e saúde pública.*

Palavras-chave *Agregação Espaço-Temporal; Métodos Epidemiológicos; Agregados de Doença*

Introdução

A distribuição de doenças no espaço e no tempo é tema dos mais explorados em epidemiologia. Nos manuais clássicos da disciplina, por exemplo, a tríade “pessoa-tempo-lugar” é concebida como ferramenta básica para a descrição dos fenômenos epidemiológicos (MacMahon & Pugh, 1970; Lilienfeld & Lilienfeld, 1980). Na década de 80, surge um renovado interesse no estudo dos padrões espaciais e temporais de doenças, conforme salienta a extensa literatura publicada em periódicos de diferentes áreas, incluindo importantes revisões (Williams, 1984; Gesler, 1986; Smith, 1982; Hills & Alexander, 1989; Marshall, 1991; Knox, 1991; Richardson, 1992). Dentre os desenhos epidemiológicos utilizados neste contexto, destacam-se os denominados estudos de agregados (*clusters*, na língua inglesa) (Kleinbaum et al., 1982).

De modo geral, o termo agregado refere-se a uma inesperada aglomeração, no espaço e/ou tempo, de eventos relacionados à saúde (CDC, 1990). Agregação espacial de doença pode ser atribuída a fatores demográficos, genéticos, ambientais ou sócio-culturais superpostos geograficamente ao padrão de ocorrência observado. Agregação na dimensão temporal, como certos padrões sazonais, endêmicos ou epidêmicos, pode contribuir para o esclarecimento dos mecanismos responsáveis pela geração dos casos de doença. No entanto, ainda que a análise puramente espacial ou temporal seja justificável no estudo de enfermidades com comportamento estável em uma destas dimensões, o estabelecimento de nexos espaço-temporais entre eventos é um componente essencial para a investigação de processos dinâmicos, como doenças infecciosas ou decorrentes de fatores ambientais transitórios (Marshall, 1991). Maior ênfase em análise espaço-temporal tem sido sugerida também para o estudo de doenças crônicas não infecciosas (Marshall, 1991).

Tendo em vista o crescente interesse pelo estudo de agregados espaço-temporais, pelo menos duas lacunas na literatura justificam a apresentação deste artigo: 1) a ausência de uma sistematização crítica que amplie o horizonte de aplicações destas técnicas para além da investigação etiológica de doenças raras e oriente o investigador em saúde pública na escolha da técnica mais adequada aos seus objetivos; 2) a potencial utilidade de uma unidade teórica para a exploração de novas metodologias e campos de aplicação. Por conseguinte, propomo-nos a explorar o significado epide-

miológico do conceito de agregação espaço-temporal, rever cinco técnicas utilizadas na investigação de agregados espaço-temporais, discutir seus principais limites e usos potenciais e apresentar os procedimentos para testes de hipótese. Pretendemos também enfatizar similitudes e diferenças entre elas, assim como descrevê-las por meio de um modelo estatístico geral. Por fim, discutimos alguns tópicos pouco explorados, mas de relevância para a otimização do uso destes estudos em epidemiologia.

Agregação espaço-temporal: conceitos e usos

Agregação espaço-temporal pode ser entendida como uma forma de não-aleatoriedade na distribuição da doença em que, entre eventos próximos no tempo, existe um excesso não esperado de eventos que estão também próximos no espaço (McAulliffe & Afifi, 1984). Este conceito é distinto do de agregação espacial e temporal, e também tem sido denominado interação espaço-temporal (Knox, 1991; Jacquez et al., 1996). De fato, agregação espaço-temporal pode ocorrer na ausência de agregação espacial e temporal, ou mesmo estar ausente quando existe agregação nas duas dimensões (Estève et al., 1994).

Estas situações podem ser ilustradas através de um cenário hipotético no qual nove regiões do mesmo tamanho e população são observadas ao longo de dez anos para o estudo de uma doença X. Suponha que casos que ocorrem em uma mesma região são considerados próximos no espaço, e quando ocorrem em um mesmo ano são considerados próximos no tempo. As Figuras 1 e 2 mostram dois padrões hipotéticos de distribuição espacial e temporal dos casos de doença. Agregação espacial e temporal é evidente nas Figuras 1(a) e 1(b), enquanto nas Figuras 2(a) e 2(b) ambas parecem estar ausentes. Como agregação espaço-temporal ocorre apenas quando há proximidade espacial e temporal simultânea entre casos, as informações disponíveis nestas figuras são insuficientes para captar este fenômeno.

Figuras 1(c) e 1(d) mostram duas possíveis alternativas para a distribuição espacial e temporal conjunta dos casos apresentados em 1(a) e 1(b). Casos são numerados segundo região de ocorrência (1, 2, ..., 9) e distribuídos segundo ano de ocorrência. Figura 1(c) mostra agregação espaço-temporal quando também há agregação espacial e temporal isoladamente. Figura 1(d) mostra ausência de agregação espaço-temporal na mesma situação. Figura 2(c) mos-

Figura 1

Distribuição espacial e temporal hipotética de casos de doença. (a) Distribuição espacial de casos em nove regiões. Agregação espacial é evidente (região 8). (b) Distribuição temporal dos casos ao longo de dez anos. Agregação temporal é evidente (ano 6). (c) Distribuição espaço-temporal dos casos. Cada caso está numerado segundo região de ocorrência (1, 2, ..., 9) e distribuídos segundo ano de ocorrência. Há agregação espaço-temporal pois os casos que ocorrem próximos no espaço (região 8) também ocorrem simultaneamente próximos no tempo (ano 6). (d) Mesma situação de (c), mas agregação espaço-temporal não existe. Casos que ocorrem próximos no espaço estão proporcionalmente distribuídos ao longo de todos os anos.

Figura 1a

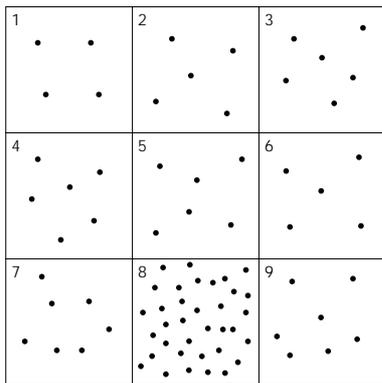


Figura 1b

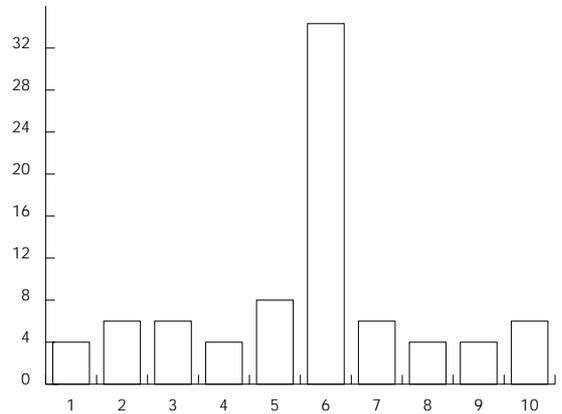


Figura 1c

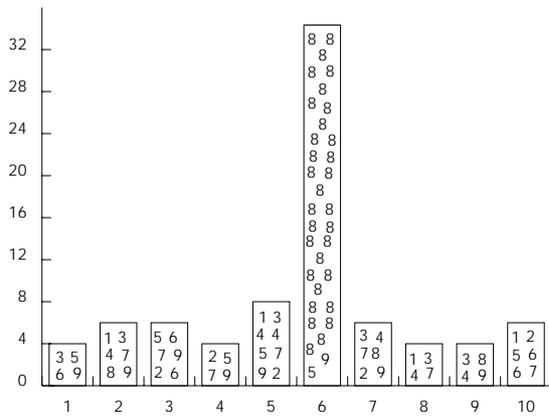


Figura 1d

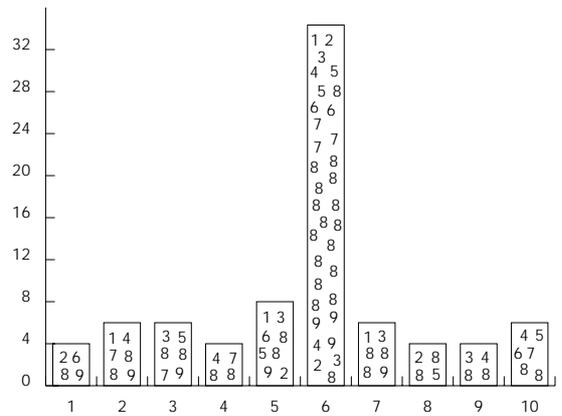


Figura 2

Distribuição espacial e temporal hipotética de casos de doença. (a) Distribuição espacial de casos em nove regiões. Aparentemente não há agregação espacial. (b) Distribuição temporal dos casos ao longo de dez anos. Agregação temporal também parece estar ausente. (c) Distribuição espaço-temporal dos casos. Cada caso está numerado segundo região de ocorrência (1, 2, ..., 9) e distribuídos segundo ano de ocorrência. Há agregação espaço-temporal pois os casos que ocorrem próximos no espaço também ocorrem simultaneamente próximos no tempo. (d) Mesma situação de (c), mas agregação espaço-temporal não existe. Casos que ocorrem próximos no espaço estão distribuídos ao longo de todos os anos.

Figura 2a

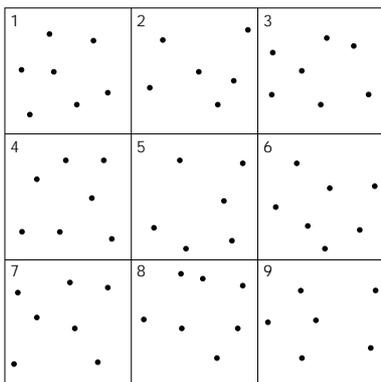


Figura 2b

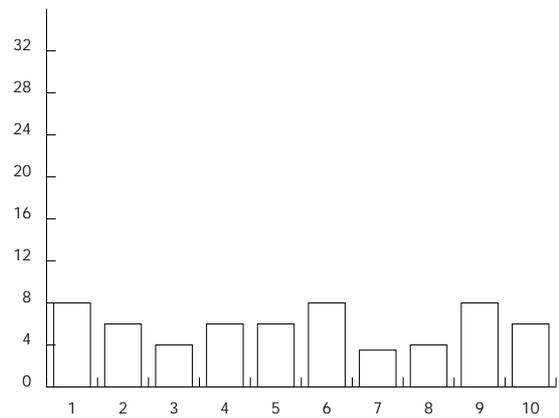


Figura 2c

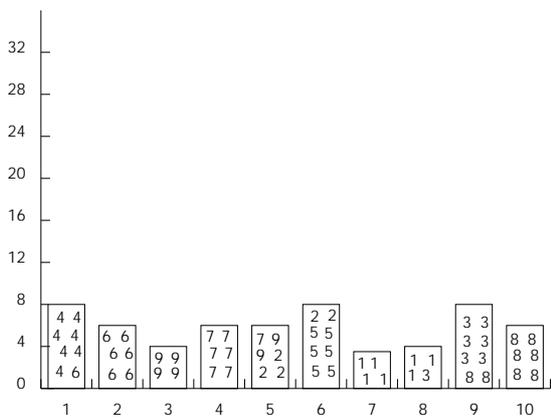
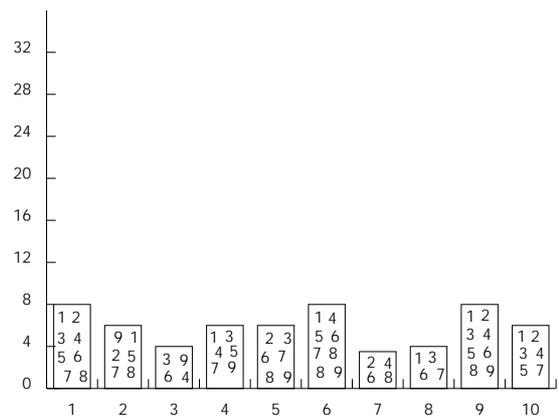


Figura 2d



tra agregação espaço-temporal quando não há agregação espacial e temporal isoladamente. Figura 2(d) mostra ausência de agregação espaço-temporal na mesma situação.

A idéia de interação espaço-temporal também pode ser concebida como sendo uma versão restrita do conceito de modificação de efeito. De modo geral, há modificação de efeito ou interação entre duas variáveis dicotômicas quando o efeito de uma varia na dependência da presença/ausência da outra. No entanto, haverá interação espaço-temporal apenas na situação específica de modificação de efeito na qual o efeito da proximidade espacial na ocorrência de doença é potencializado na presença de um efeito positivo da proximidade temporal, e vice-versa. Desta maneira, situações em que o efeito da proximidade espacial é potencializado justamente na ausência de efeito de proximidade temporal não correspondem à existência de agregação espaço-temporal, ainda que constitua uma situação evidente de modificação de efeito. Em razão desta inconsistência conceitual, sugerimos que o termo interação espaço-temporal seja descontinuado em detrimento do conceito clássico e mais intuitivo de agregação espaço-temporal.

De um modo geral, o objetivo central da investigação de agregados espaço-temporais é o esclarecimento do mecanismo subjacente responsável pela formação do agregado. São dois os mecanismos explicativos de interesse epidemiológico que podem contribuir para a formação de agregados espaço-temporais de doença. No primeiro, fatores de risco transitórios (por exemplo, exposição ambiental a poluentes) expõem indivíduos, independentemente, a um maior risco de adoecer em espaço e tempo delimitados. No segundo, processos infecciosos ou heranças genéticas são responsáveis pela formação dos agregados, na qual a independência entre eventos não é observada (Marshall, 1991).

Agregação espaço-temporal de doença tem sido compreendida como expressão de processos contagiosos (Knox, 1964a; Bailey & Gattrell, 1995). Por conseguinte, as técnicas para a identificação de agregados espaço-temporais foram utilizadas prioritariamente na investigação de infectividade de doenças de etiologia desconhecida (Mantel, 1967; Marshall, 1991). Doenças raras, particularmente as neoplasias, têm sido as mais comumente estudadas (Jacquez et al., 1996). Entretanto, outros usos potenciais têm sido propostos, como: avaliar impacto de programa de prevenção (McAuliffe & Afifi, 1984), sugerir padrões de disseminação de doenças transmissíveis (Klauber & Angulo, 1974; Cliff & Ord, 1981), monitorar a ocorrên-

cia de doenças infecciosas em ambiente hospitalar e subsidiar a especificação de modelos analíticos formais para processos espaço-temporais (Jacquez et al., 1996).

Descrição das técnicas analisadas

Características gerais

Segundo Besag & Newell (1991), as técnicas para investigação de agregados espaciais e temporais podem ser divididas em dois subgrupos: testes de agregação (T1) e testes para detecção de agregados (T2). Testes de agregação (T1) podem ser de dois tipos: gerais (T1-G) ou focais (T1-F). Os primeiros (T1-G) objetivam identificar um padrão geral de distribuição da doença em uma grande região. Os segundos (T1-F) utilizam dados de pequenas regiões, geralmente definidas em função de uma fonte de exposição (por exemplo, instalação nuclear). Testes para detecção de agregados (T2), por outro lado, são utilizados na vigilância rotineira de grandes quantidades de pequenas áreas, buscando evidenciar agregados pontuais de doença, porém sem preconcepção acerca de suas localizações. Em geral, funcionam como testes de rastreamento para posterior investigação detalhada das áreas selecionadas.

Neste trabalho, foram escolhidas as seguintes técnicas para avaliação: Knox (Knox, 1964a), Mantel (Mantel, 1967), McAuliffe (McAuliffe & Afifi, 1984), EMM (Ederer et al., 1964) e Moran (Moran, 1948; Moran, 1950; Cliff & Ord, 1981). De acordo com a proposta de Besag & Newell (1991), as três primeiras podem ser classificadas como técnicas para detecção de agregados (T2) e as duas últimas como técnicas para a identificação de tendências gerais à formação de agregados (T1-G). Técnicas utilizadas para investigação de agregados em torno de uma potencial fonte de risco (T1-F) foram excluídas desta revisão. Ao leitor interessado sugere-se a revisão feita por Hills & Alexander (1989).

A escolha destas técnicas pretendeu abarcar três diferentes maneiras de abordar o problema. A primeira abordagem, representada pelas técnicas de Knox, Mantel e McAuliffe, utiliza distâncias espaciais e temporais entre casos de doença. Estas técnicas são usualmente denominadas métodos de distância. Utilizam dados do tipo localização/data do evento, isto é, necessitam da especificação precisa das coordenadas geográfica e temporal de cada evento. O segundo tipo, exemplificado por EMM, examina o padrão de ocupação de caselas, ou seja, a distribuição de casos em regiões durante pe-

riodos especificados (unidades de espaço-tempo). A técnica de Moran representa uma terceira abordagem baseada em modelos de autocorrelação. EMM e Moran usam dados do tipo área/intervalo de tempo. Aqui, contagens ou medidas de frequência de doença são definidas para regiões geográficas e intervalos temporais.

Procurou-se enfatizar técnicas com maior tradição de uso em epidemiologia ou que foram concebidas especificamente para solucionar fragilidades de outras técnicas. Knox, Mantel e EMM são das mais populares na literatura epidemiológica (Smith, 1982; Wartenberg & Greenberg, 1994). Do ponto de vista histórico-evolutivo, Knox e Mantel são as técnicas precursoras, ao passo que McAuliffe foi proposta especificamente para a solução de problemas detectados nestas técnicas. A técnica de Moran é uma das mais comumente aplicadas no estudo de agregados espaciais, em geral acompanhando mapas de distribuição de doenças (Cliff & Ord, 1981; Walter, 1993). A Tabela 1 apresenta algumas características das técnicas analisadas.

Knox

O conjunto de dados utilizados na técnica descrita por Knox (1964a) é composto pelas localizações geográfica e temporal de n casos de doença. A técnica baseia-se na mensuração da distância espacial e temporal entre todos os $N = n(n-1)/2$ diferentes pares de casos. A cada par de casos está associada uma medida de distância espacial e outra temporal. As medidas de distância espacial podem ser a Euclideana ou qualquer outra regra que descreva a proximidade entre casos (por exemplo, casos ocorrendo em uma mesma região são considerados próximos).

Arbitra-se um valor crítico para cada distância (espacial e temporal), que são utilizados para classificar todos os possíveis pares segundo dois critérios: 1) se a distância espacial observada entre casos i e j (d_{ij}) é menor ou igual/maior do que o valor crítico definido previamente (d_0); 2) se a diferença temporal observada (t_{ij}) é menor ou igual/maior do que o valor crítico definido *a priori* (t_0). Assim, quatro categorias podem ser identificadas: pares próximos no espaço e no tempo, próximos apenas no espaço ou apenas no tempo e pares distantes, conforme apresentado na Tabela 2.

A casela (A) compreende os pares de casos que estão próximos no espaço e no tempo. O teste da hipótese nula, de ausência de agregação no espaço-tempo, toma (A) como uma

variável aleatória com distribuição Poisson. Quanto maior for o valor de (A), maior será o indicativo da existência de agregação. A aplicação da distribuição de Poisson não é estritamente correta, pois a formação de N pares induz a uma certa interdependência, na medida em que o mesmo evento é utilizado diversas vezes (Williams, 1984). No entanto, quando as distâncias críticas escolhidas tornam a expectativa de (A) pequena, o seu uso é apropriado (Selvin, 1991).

Uma das principais vantagens desta técnica é que ela avalia a presença de associação entre as medidas de distância temporal e espacial sem requerer informação sobre a densidade populacional ou força de morbidade. É uma técnica de alta especificidade, ou seja, indiferente à existência de agregados somente no espaço ou no tempo. No entanto, agregação espaço-temporal pode ser detectada apenas por fenômenos demográficos, como a variação geográfica da distribuição populacional ao longo do tempo (Roberson, 1990). Assim, sua utilidade torna-se restrita a avaliações em períodos relativamente curtos e em regiões com população estável. Outro potencial problema é a escolha dos pontos de corte, o que leva à perda de informação e pode afetar seu poder para a detecção de diferentes padrões de agregação.

Esta técnica tem sido prioritariamente utilizada para investigar a existência (ou não) de mecanismos de contágio em doenças de etiologia desconhecida, como, por exemplo: leucemia (Knox, 1964b; Gilman & Knox, 1995), doença de Hodgkin (Chen et al., 1984), anomalias congênitas (Siemiatycki & McDonald, 1972), diabetes mellitus (Samuelsson et al., 1994), síndrome da morte súbita infantil (Rodrigues et al., 1992) e suicídio (Gould et al., 1994). Klauber & Angulo (1974) ampliaram o seu potencial de aplicação, utilizando-a para descrever aspectos do mecanismo de difusão de uma doença infecciosa (varíola menor) entre escolares.

Mantel

Mantel (1967) desenvolveu uma técnica para identificação de agregados espaço-temporais que utiliza a mesma estrutura de dados da técnica de Knox (1964a), e também baseia-se em medidas de distâncias entre pares de casos. Mantel (1967) propõe que, independentemente da escala de proximidade utilizada, pode-se conceber uma estatística geral de teste do tipo: Equação 1 $Z = \sum_i \sum_j X_{ij} Y_{ij} (i \neq j)$ onde X_{ij} é uma função da distância espacial entre dois pontos i e j , e Y_{ij} uma função da distância temporal.

Tabela 1

Caracterização geral de técnicas para investigação de agregados espaço-temporais.

	Knox	Mantel	McAuliffe	EMM	Moran
Objetivo principal	Detecção de agregados (T2)	Detecção de agregados (T2)	Detecção de agregados (T2)	Identificação de padrão geral (T1-G)	Identificação de padrão geral (T1-G)
Estrutura dos dados	Localização espacial e temporal do caso (<i>point processes</i>)	Localização espacial e temporal do caso (<i>point processes</i>)	Localização espacial e temporal do caso (<i>point processes</i>)	Número de casos por intervalo de tempo e região (<i>lattice data</i>)	Taxas de adoecimento por região e intervalo de tempo (<i>lattice data</i>)
Medida de conectividade básica	Distância entre pares de casos (dicotomizada)	Recíproca da distância entre pares de casos	Distância entre pares de casos vizinhos mais próximos	Ocorrência conjunta em unidades espaço-temporais	Distância entre centróides ou indicador de vizinhança geográfica
Estatística de teste	Número de pares próximos no espaço e tempo	Correlação entre distâncias espaciais e temporais	Média ponderada das distâncias de vizinhos próximos para cada intervalo de tempo	Soma do nº máximo de casos em unidades espaço-temporais	Medida ponderada de autocorrelação entre taxas
Informação populacional	não	não	não	não	sim

Tabela 2

Classificação dos pares de casos de doença segundo distância espacial e temporal.

	$t_{ij} < t_0$	$t_{ij} \geq t_0$	Total
$d_{ij} < d_0$	A	B	A + B
$d_{ij} \geq d_0$	C	D	C + D
Total	A + C	B + D	N

d_{ij} – distância espacial observada entre casos i e j
 d_0 – distância espacial crítica
 t_{ij} – diferença temporal observada entre casos i e j
 t_0 – diferença temporal crítica

Um importante problema na utilização deste teste é o excesso de variabilidade imposto pelas grandes distâncias. Como estratégia de restrição desta variabilidade, Mantel (1967) propôs o uso das recíprocas das distâncias espaciais e temporais, enfatizando a proximidade em detrimento da distância entre casos. Esta formulação permite a incorporação de relações de distâncias assimétricas, onde $X_{ij} \neq X_{ji}$ (por exemplo, quando as distâncias são definidas relativamente às posições dos outros casos). Pode-se suprimir a restrição $i \neq j$, especificando uma constante (usualmente zero) para X_{ii} e Y_{ii} .

A distribuição de Z sob a hipótese nula pode ser obtida através de abordagens permutacionais (Good, 1994). Tendo n localizações de casos no espaço e n localizações no tempo, a

hipótese de ausência de aglomeração é equivalente à de que as localizações no espaço são aleatoriamente pareadas com localizações no tempo. Obtém-se, assim, $n!$ conjuntos equiprováveis de pares. Listam-se as $n!$ possíveis permutações, computa-se Z para cada uma delas e obtém-se a distribuição nula de Z contra a qual o valor observado de Z pode ser julgado. Quando n é muito grande, a abordagem permutacional completa torna-se impraticável e devem-se utilizar métodos de Monte Carlo para obter uma distribuição de Z para o teste de significância (Williams, 1984). Quando a amostra é suficientemente grande, a correção de Z por suas média e variância estimadas pelo procedimento acima resulta em uma variável aleatória com distribuição normal padronizada. Valores altos de $(Z - E(Z)) / (Var(Z))^{1/2}$ são indicativos de aglomeração espaço-temporal.

Um dos principais problemas desta técnica é a ocorrência de distâncias espaciais ou temporais iguais a zero. Nestes casos, o uso das recíprocas das distâncias necessita de adição de constantes para evitar a divisão por zero. Assim como na técnica de Knox (1964a), a escolha de constantes é um problema crítico. Constantes pequenas levam a distribuição de Z a afastar-se da normalidade, e a adição de valores altos elimina o efeito de pares próximos, o que acarreta perda de potência (McAuliffe & Afifi, 1984).

A técnica de Mantel (1967) também tem sido predominantemente utilizada no esclarecimento da existência (ou não) de mecanismos

de contágio em determinadas doenças, como: leucemia (Glass & Mantel, 1969; Klauber & Mustachii, 1970), doença de Hodgkin (Chen et al., 1984) e anomalias congênitas (Siemiatycki & McDonald, 1972). Não se identificou nenhum estudo que propusesse novos usos potenciais para esta técnica. Entretanto, dado suas similitudes com a técnica de Knox (1964a), a técnica de Mantel (1967) também pode ser utilizada para descrever aspectos da difusão de doenças infecciosas (Klauber & Angulo, 1974).

McAuliffe

A técnica desenvolvida por McAuliffe & Afifi (1984) é baseada na distância entre cada caso e sua vizinhança mais próxima (*nearest neighbor*), e utiliza a mesma estrutura de dados das técnicas anteriormente descritas. Baseia-se na distância espacial entre um evento i e o evento geograficamente mais próximo (vizinho mais próximo) dentre aqueles localizados em um intervalo l de tempo anterior. Assim, pode-se dividir o tempo anterior ao período em que i ocorreu em 'janelas' onde outros eventos podem ter acontecido. Para cada janela l procura-se a ocorrência geograficamente mais próxima de i e mede-se a distância espacial entre elas. Esta distância espacial entre a ocorrência de referência e seu vizinho mais próximo é $d_i(l)$, onde $l = 0, 1, 2, \dots, t_i - 1$ (t_i é o número de períodos temporais definidos a partir de cada i). Se nenhuma ocorrência foi observada l períodos antes, então $d_i(l)$ não é definida. Para a comparação entre distâncias, cada $d_i(l)$ é padronizada por sua distribuição sob a hipótese nula de ausência de agregação espaço-temporal. Assim:

$$\text{Equação 2} \quad V_i(l) = \frac{d_i(l) - \mu_0(i,l)}{\sigma_0(i,l)}$$

onde $\sigma_0(i,l)$ e $\mu_0(i,l)$ são, respectivamente, o desvio padrão e a média de $d_i(l)$ sob a hipótese nula de não-agregação. Estes parâmetros são estimados via abordagem permutacional análoga àquela utilizada em Mantel (1967). A concepção básica é a de que, se existe agregação espaço-temporal para alguma diferença fixa de tempo, então existem ocorrências entre aquelas com esta diferença de tempo que estão mais próximas no espaço do que o esperado, daí que $V_i(l)$ tenderá a ser negativo na vigência de agregação.

Para um intervalo de tempo fixo l , $B(l)$ é a coleção dos índices i para os quais a distância $d_i(l)$ pode ser definida e $N(l)$ o tamanho do conjunto $B(l)$. A média ponderada

$$\text{Equação 3} \quad Z(l) = \frac{\sum_{i \in B(l)} V_i(l)}{\sqrt{N(l)}}$$

é uma medida que resume a proximidade entre todos os pares de ocorrências que têm uma diferença temporal l . Se as distâncias padronizadas são independentes, a média e variância de $Z(l)$ sob a hipótese nula serão 0 e 1, respectivamente. Na presença de agregação, a média de $Z(l)$ é menor que zero.

Além de não necessitar do conhecimento da distribuição espacial da população subjacente, este método tem a vantagem de evitar a definição de constantes arbitrárias como ocorre em Knox (1964a) e Mantel (1967). Na dimensão espacial, isto efetivamente ocorre, pois a variabilidade é restringida pelo uso das distâncias entre vizinhos próximos. No entanto, o poder deste teste é afetado pela definição do número de períodos de tempo entre as ocorrências inicial e final. Da mesma forma, o poder é afetado pelo número de ocorrências em cada período. Tanto um como outro problema reflete a influência do tamanho amostral em $Z(l)$.

Ainda que adequada para a investigação de mecanismos de contágio de doenças, nenhum estudo neste âmbito pode ser identificado na literatura biomédica. Os formuladores desta técnica utilizaram-na, em comparação com a de Knox (1964a) e Mantel (1967), na avaliação do impacto de um programa educativo antitabagismo (McAuliffe & Afifi, 1984).

Ederer-Myers-Mantel (EMM)

A estrutura dos dados usada pelo teste proposto por Ederer et al. (1964) é formada por unidades de espaço-tempo (l), definidas como regiões geográficas associadas a um período de tempo definido. Considere uma região e um período de tempo suficientemente curto para não haver grandes mudanças populacionais (cinco anos). Nesta unidade espaço-temporal (i) ocorrem n_i casos de doença distribuídos por j menores intervalos de tempo (por exemplo, um ano calendário). A estatística de teste pode ser definida da seguinte forma:

$$\text{Equação 4} \quad EMM = \frac{\sum_{i=1}^I (m_i - E(m_i|n_i))}{\sqrt{\sum_{i=1}^I \text{Var}(m_i|n_i)}}$$

onde m_i é o número máximo de casos ocorridos em um dos intervalos de tempo em cada região estudada. Assumindo que o número de casos de doença em uma região durante um período de tempo especificado tem distribuição Poisson, então, condicionalmente ao número total de casos desta região (n_i), a distribuição de casos ao longo dos intervalos de tempo segue uma distribuição multinomial. Os valores dos momentos $E(m_i|n_i)$ e $\text{Var}(m_i|n_i)$ para $n_i \leq 500$ e $j \leq 3$ ou para $n_i \leq 200$ e $j = 4$ e 5 , assim como uma fórmula assintótica, são apre-

sentados por Mantel et al. (1976). Se o número de unidades de espaço-tempo é suficientemente grande e nenhuma das unidades geográficas é muito maior que as outras, a estatística EMM tem aproximadamente distribuição normal sob a hipótese nula de ausência de agregação espaço-temporal.

Esta estatística não usa explicitamente distâncias intereventos e, portanto, não se enquadra na mesma categoria das técnicas descritas anteriormente (métodos de distância). Seu uso está indicado nas situações em que não existe exato conhecimento da localização dos eventos. Como Knox e Mantel, também é uma técnica que prescinde do conhecimento da distribuição espacial da população subjacente. Uma das principais vantagens é a possibilidade de ser estimada uma medida do tipo risco atribuível (Wallenstein et al., 1989).

Dois principais limites são: sensibilidade à agregação apenas na dimensão temporal e à utilização de janelas temporais fixas. O primeiro caso obriga a definição de janelas temporais suficientemente curtas, para evitar o efeito de tendência, e suficientemente grandes, para evitar efeito de sazonalidade. O segundo problema implica diluição de agregações que se dão nas zonas limítrofes de janelas temporais adjacentes. Para períodos de um ano, por exemplo, pode-se não detectar eventos agregados em torno dos últimos dias de um ano e dos primeiros do ano seguinte. Naus (1965) propôs a estatística *scan* como estratégia para a solução deste problema. A estatística *scan* usa o número de máximo de eventos em qualquer período de tamanho fixo (por exemplo, 365 dias).

Tanto a técnica EMM, quanto sua similar, a técnica *scan* (Naus, 1965), foram utilizadas primordialmente para a investigação de potenciais mecanismos de transmissão de doenças, como, por exemplo, leucemia (Ederer et al., 1964), suicídio (Gould et al., 1994; Wallenstein et al., 1989) e doença de Hodgkin (Chen et al., 1984). Raubertas (1989) propôs um novo uso potencial desta técnica no monitoramento da qualidade de dados gerados por sistemas de vigilância epidemiológica.

Moran

A técnica *I* de Moran (Moran, 1948; Moran, 1950; Cliff & Ord, 1981) foi pensada originalmente para explorar padrões espaciais de doenças, mas também pode ser utilizada para investigar o componente espaço-temporal (Cliff & Ord, 1981). O conjunto de dados utilizado é composto por taxas de incidência de doença (ou outra medida de ocorrência) por regiões se-

gundo períodos de tempo delimitados. A estatística de Moran, adaptada para o estudo de agregação espaço-temporal, é:

$$\text{Equação 5 } I = \frac{n \sum_{i \neq j} w_{ij} (y_{i,t} - y_t)(y_{j,t-k} - y_{t-k})}{S_0 \sum_{i=1}^n (y_{i,t} - y_t)^2}$$

onde n é o número de regiões, w_{ij} mede a proximidade da região j em relação à região i , $y_{i,t}$ é a taxa da doença na região i no tempo t , y_t é a taxa média da doença no tempo t , $y_{j,t-k}$ é a taxa da doença na região j no tempo $t-k$, y_{t-k} é a taxa média da doença no tempo $t-k$, e $S_0 = \sum_{i \neq j} w_{ij}$. Os valores de w_{ij} podem ser binários (1 se regiões fazem fronteira e 0 se não) ou configurar uma matriz geral de ponderações W , permitindo utilizar funções da distância Euclideana ou outras métricas que expressem uma interação entre i e j . Uma alternativa é considerar apenas influências de vizinhos próximos, usando $w_{ij} = 0$ se i e j não são vizinhos imediatos (autocorrelação espacial de ordem um). No entanto, nada impede de que sejam investigadas relações entre segundos-vizinhos ou outros ainda mais distantes. Segundos-vizinhos seriam aquelas regiões j que não são vizinhas de i , mas de uma certa região h , que por sua vez é vizinha de i . Da mesma maneira, o índice $t - k$ pode ser restrito a $k = 1$ (autocorrelação temporal de ordem um) ou incorporar períodos temporais anteriores ($k \geq 1$).

O índice de Moran (I) é descrito como um coeficiente de autocorrelação espaço-temporal, onde no numerador está uma medida de covariância de y 's, e no denominador está uma medida de variância. Pode-se interpretá-lo de maneira similar ao coeficiente de correlação, ou seja, restrito ao intervalo $[-1, +1]$ com valores próximos à unidade, significando forte relação (Cliff & Ord, 1981). Agregação espaço-temporal é indicada quando I assume valores próximos de $+1$ e significa que regiões próximas no espaço tendem a ter taxas de adoecimento similares (maiores ou menores que a taxa média) em períodos de tempo próximos. Se não existir correlação, I é próximo de 0 , e se a correlação for inversa, isto é, se regiões próximas no espaço tendem a ter taxas de adoecimento discrepantes em períodos próximos no tempo, I é próximo de -1 . Porém, os limites de I em geral não alcançam ± 1 , e em situações onde regiões com $(y_{i,t} - y_t)$ muito grandes são excessivamente ponderadas estes valores podem até ser superados (Jong et al., 1984).

Admite-se que, com aproximadamente vinte observações (áreas), I é assintoticamente normalmente distribuído, permitindo testes de significância baseados em médias e variâncias teóricas (Walter, 1993). Podem-se também utilizar os procedimentos permutacionais aleató-

rios descritos para as outras técnicas. Neste caso, considera-se o valor observado de I em relação ao conjunto de todos valores que I pode tomar se $\{y_j\}$ é repetidamente permutado aleatoriamente no sistema geográfico (Cliff & Ord, 1981).

Uma vantagem desta técnica é que ela não se restringe à identificação de agregados, mas fornece informações sobre a estrutura de dependência espaço-temporal da doença (Cliff & Ord, 1981; Lam et al., 1996). Raubertas (1988) descreve uma medida geral de autocorrelação espaço-temporal para um conjunto de dados semelhante à utilizada por Moran. Propõe um modelo aditivo que descreve a discrepância entre probabilidades observadas e esperadas em cada unidade espaço-temporal. Os parâmetros do modelo representam efeitos principais associados ao espaço, ao tempo e ao efeito de interação entre espaço e tempo. Sob a hipótese nula de ausência de agregação espaço-temporal, este último parâmetro é igual a zero. A maior vantagem deste método é permitir a indicação de cada região e período de tempo que contribuem para a formação do agregado.

A técnica de Moran tem sido prioritariamente utilizada no âmbito espacial como um teste para detecção de autocorrelação espacial nas taxas de mortalidade de câncer (Glick, 1979; Smans & Esteve, 1992; Esteve et al., 1994). Oden (1993) propôs uma técnica análoga (IDIFF) para a investigação da direcionalidade da autocorrelação espacial, que pode ser utilizada para distinguir processos epidêmicos isotrópicos daqueles que se difundem com uma direção preferencial. O seu maior potencial de utilização está no estudo do processo de difusão espaço-temporal de doenças infecciosas como, por exemplo, cólera (Cliff & Ord, 1981; Adesina, 1984) e Sida/Aids (Lam et al., 1996).

Estrutura estatística geral das técnicas estudadas

A estatística Z definida em Equação 1 pode ser vista como uma medida geral de autocorrelação entre funções de distância espacial e temporal (Marshall, 1991). De fato, todas as outras estatísticas aqui descritas podem ser derivadas desta.

Mantel (1967) aponta que a estatística de Knox (1964a) pode ser vista como um caso especial da estatística Z , onde X_{ij} e Y_{ij} tomam valores 0 (pares distantes) ou 1 (próximos no espaço e tempo) e $i < j$. Neste caso, a estatística é igual ao número de pares próximos tanto no tempo como no espaço. A categorização das

distâncias proposta por Knox (1964a) serve como uma outra estratégia de combater o excesso de variabilidade imposta pelas grandes distâncias (Mantel, 1967).

A estatística $Z(l)$ descrita em Equação 3 (McAuliffe & Afifi, 1984) também pode ser escrita como função de uma estatística Z . A distância $d_j(l)$ pode ser definida para cada i e l como o valor obtido para uma estatística geral Z na qual X_{ij} é igual a 1 quando j é a vizinhança mais próxima de i e igual a zero em qualquer outra situação, e Y_{ij} são os valores de distância espacial padronizada entre todos os pares ij . A estatística $Z(l)$ seria então a soma de cada um destes valores padronizados e divididos por $N(l)^{1/2}$.

Na técnica EMM, a estatística de teste (m_j) é somada para cada unidade espaço-temporal i , onde X_{ij} é igual a 1 para o período j , onde ocorre o maior número de casos dentre os períodos observados e zero em caso contrário, e Y_{ij} o número de casos observados corrigidos por sua média e variância em cada período j .

Por fim, a técnica I de Moran pode ser expressa como uma transformação da estatística geral Z , quando X_{ij} é w_{ij} e Y_{ij} é $(y_{i,t} - y_t) / (y_{i,t-k} - y_{t-k})$ (Hubert et al., 1981).

A confluência destas técnicas para funções da estatística Z tem importância teórica e prática na medida em que, na dependência da especificação de X_{ij} e Y_{ij} e do tamanho amostral, sua padronização leva à obtenção de uma estatística com distribuição normal (Hubert et al., 1981).

Por outro lado, a estatística descrita por Tango (1984), que também tem estrutura geral de uma estatística Z , pode ser escrita como uma função linear de uma estatística U bidimensional (Rayens & Kryscio, 1993). Se este fato puder ser demonstrado também para aquelas estatísticas aqui descritas, resultados assintóticos para os momentos destas variáveis poderão ser obtidos com base na teoria de estatísticas U (Lehman & D'Abbrera, 1975).

Perspectivas de investigação e considerações finais

Muito se tem discutido sobre os limites e possibilidades dos estudos de agregados em epidemiologia. Para Rothman (1990), a investigação de um único potencial agregado de doença tem limitado valor científico. O autor aponta cinco justificativas para tal avaliação: 1) os agregados são usualmente pequenos para constituir um estudo epidemiológico com o adequado controle de variáveis de confundimento; 2) ha-

bitualmente os agregados relatados utilizam definição vaga da doença; 3) sua delimitação pode dificultar a definição de uma população base adequada para o cálculo das taxas de incidência; 4) as exposições sob investigação são geralmente mal caracterizadas e em baixa concentração e 5) uma vez que existe excessiva publicidade em torno da descoberta de um possível agregado, a obtenção de dados válidos torna-se difícil.

Por outro lado, estratégias para a detecção de padrões de agregação espaço-temporal poderiam ter maior utilidade na medida em que as definições utilizadas no processo de investigação pudessem ser baseadas em critérios mais objetivos. Além do mais, as possíveis hipóteses explanatórias podem ser restringidas a exposições que variam em um padrão similar ao da doença sob investigação. Entretanto, Rothman (1990) aponta que a identificação de agregação espaço-temporal pode servir apenas como estratégia auxiliar na avaliação de hipóteses mais específicas. Para ele, é preferível dispendir mais esforços conduzindo estudos grandes e bem delineados que testem hipóteses específicas em populações com grande variabilidade na exposição de interesse. Rothman (1990) conclui que estudos de agregados teriam utilidade apenas quando a doença é extremamente rara ou previamente desconhecida.

No entanto, mesmo reconhecendo seus limites, estudos bem conduzidos que identifiquem a concomitância espaço-temporal de doença e alta prevalência de fatores de risco podem ser uma fonte geradora de hipóteses causais (Kheifets, 1993). Jacquez et al. (1996), por exemplo, preferem conceber os estudos de agregados sob o espírito da análise exploratória de dados, ou seja, objetivam a identificação de padrões nos dados e geração de hipóteses referentes à origem destes padrões. Esta abordagem seria a preferida quando o conhecimento incompleto sobre o processo de difusão de uma doença não permite a construção de modelos mais formais. Neste sentido, estes estudos podem ser realizados antes dos mais tradicionais, demorados e caros delineamentos epidemiológicos, ajudando a definir prioridades para investigação complementar e implementação de ações preventivas (Wartenberg & Greenberg, 1993). Ainda assim, para que a utilização destas técnicas em epidemiologia seja otimizada, é essencial o enfrentamento de uma série de desafios metodológicos.

Um problema crucial é a definição adequada dos critérios de vizinhança e das escalas de espaço e tempo para a delimitação dos agrega-

dos. Para que o processo inferencial dedutivo seja privilegiado, todas estas definições devem ser pautadas em teorias sobre os mecanismos de disseminação da doença sob investigação. Pike & Smith (1968), por exemplo, foram pioneiros na adaptação da técnica de Knox (1964a) para investigação de patologias com longo período de latência incorporando os períodos de susceptibilidade, infectividade e latência de cada caso. A solução usualmente proposta baseia-se na utilização de múltiplos testes para diferentes critérios, favorecendo, porém, a ocorrência de erro tipo I, ou seja, afirmar a existência de agregação espaço-temporal quando esta não existe. A correção de Bonferroni ou Tukey para múltiplos testes são recomendadas para que estes erros sejam reduzidos (Kleinbaum et al., 1988).

A exploração de novos usos para estas técnicas em epidemiologia é ainda restrita. Isto decorre em parte do não-reconhecimento dos diferentes modelos para ocorrência de doença que estão subjacentes às diferentes técnicas utilizadas (Waller & Jacquez, 1995). Por exemplo, a típica hipótese nula utilizada pelos métodos aqui estudados é a de completa aleatoriedade na distribuição espaço-temporal da doença, ou seja, um modelo de ocorrência de doença com taxa constante no espaço-tempo (Waller & Jacquez, 1995). No entanto, tal modelo é uma simplificação grosseira da maioria dos processos estudados em epidemiologia. Neste sentido, mais do que testes de significância estatística, são desejáveis técnicas que visem também à descrição de facetas do processo de ocorrência de doença. Complementarmente, estudos de potência são uma necessidade premente, na medida em que os processos de difusão de doenças são responsáveis pelas mais variadas configurações espaço-temporais e as técnicas podem obter melhor *performance* em situações bem distintas (Wartenberg & Greenberg, 1993). Modelos teóricos de difusão de doenças podem ser simulados com diferentes parâmetros especificados, o que permite avaliar como estas diferenças influenciam a potência destes métodos (Wartenberg & Greenberg, 1990a e b).

As possibilidades de utilização destas metodologias no âmbito da vigilância epidemiológica ainda não foram exploradas em toda a sua extensão. A delimitação e identificação de quem pertence aos agregados (Raubertas, 1988), assim como a detecção da direcionalidade da máxima associação (Lawson & Viel, 1995) podem sugerir e facilitar a investigação de fatores etiológicos comuns. O uso do correlograma para identificar a direção e variação espaço-

temporal da força da associação (como uma medida indireta do grau de contágio), bem como avaliar se o processo de disseminação da doença é dominado por fatores intrínsecos à região e/ou por difusão espacial também tem sido pouco desenvolvido (Cliff & Ord, 1981).

São amplas as dificuldades operacionais para aplicação e interpretação destes métodos no estudo de doenças infecciosas no Brasil. O impacto dos intensos fluxos migratórios, por exemplo, pode resultar em uma dinâmica de adoecimento de difícil apreensão. Por um lado, implicando modificações de hábitos e estilos de vida importantes para a disseminação da enfermidade. Por outro, modificando a estrutura epidemiológica incorporando novos infectantes, diluindo o compartimento de susceptíveis e/ou aumento da heterogeneidade em decorrência de novos padrões individuais e coletivos de adoecimento. Também as modificações nos limites geopolíticos e dos critérios de agregação de informação em saúde dificultam a análise dos dados, especialmente a definição de padrões de proximidade geográfica. A incorporação destes fenômenos no desenho e/ou

análise de estudos de agregados é mais um desafio ainda inexplorado.

Este artigo pretendeu focar algumas questões centrais para a compreensão e melhor utilização destas técnicas. Neste sentido, é feita uma caracterização das técnicas enfatizando aspectos que permitem identificar similitudes e diferenças, assim como uma descrição dos procedimentos necessários para utilizar tais técnicas na prática epidemiológica. São apontadas suas principais fragilidades e usos potenciais. Muitos dos problemas apresentados só poderão ser solucionados se os instrumentais estatístico e epidemiológico puderem ser compatibilizados sob um mesmo referencial teórico-conceitual. Foi com este espírito que apresentamos uma proposta de generalização conceitual e estatística das técnicas analisadas. Ainda assim, é evidente que muitas questões permanecem abertas para investigação, em particular a incorporação de concepções, medidas e delineamentos epidemiológicos como estratégia de aprimorar a interpretação e utilidade de estudos de agregados espaço-temporais.

Agradecimentos

Gostaríamos de agradecer a Célia L. Szwarcwald, Flávio F. Nobre e Ronir R. Luiz pelas contribuições durante a elaboração deste artigo, período em que GLW era bolsista da Capes e CJS foi parcialmente financiado pelo CNPq.

Referências

- ADESINA, H. O., 1984. Identification of the colera diffusion process in Ibadan, 1971. *Social Science and Medicine*, 18:429-440.
- BAILEY, N. T. J. & GATRELL, A., 1995. *Interactive Spatial Data Analysis*. Essex: Longman
- BESAG, J. & NEWELL, J., 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A*, 154:143-155.
- CDC (Center for Disease Control), 1990. Guidelines for investigating clusters of health events. *Morbidity and Mortality Weekly Reprot*, 39(RR-11):1-23.
- CHEN, R.; MANTEL, N. & KLINGBERG, M. A., 1984. A study of three techniques for time-space clustering in Hodgkin's disease. *Statistics in Medicine*, 3:173-184.
- CLIFF, A. D. & ORD, J. K., 1981. *Spatial Processes*. London: Pion.
- EDERER, F.; MYERS, M. H. & MANTEL, N., 1964. A statistical problem in space and time: Do leukemia cases come in clusters? *Biometrics*, 20:626-638.

- ESTÈVE, J.; BENHAMOU, E. & RAYMOND, L., 1994. *Statistical Methods in Cancer Research, Vol. IV, Descriptive Epidemiology (IARC Scientific Publications nº 128)*. Lyon: International Agency for Research on Cancer.
- GESLER, W., 1986. The uses of spatial analysis in medical geography: a review. *Social Science and Medicine*, 23:963-973.
- GILMAN, E. A. & KNOX, E. G., 1995. Childhood cancers: space-time distribution in Britain. *Journal of Epidemiology and Community Health*, 49:158-163.
- GLASS, A. G. & MANTEL, N., 1969. Lack of time-space clustering of childhood leukemia in Los Angeles County, 1960-1964. *Cancer Research*, 29:1995-2001.
- GLICK, B., 1979. The spatial autocorrelation of cancer mortality. *Social Science and Medicine*, 13D:123-130.
- GOOD, P., 1994. *Permutation Tests*. New York: Springer-Verlag.
- GOULD, M. S.; PETRIE, K.; KLEINMAN, M. H. & WALLENSTEIN, S., 1994. Clustering of attempted suicide: New Zealand national data. *International Journal of Epidemiology*, 23:1185-1189.
- HILLS, M. & ALEXANDER, F., 1989. Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review. *Journal of the Royal Statistical Society A*, 152:353-363.
- HUBERT, L. J.; GOLLEDGE, R. G. & CONSTANZO, C. M., 1981. Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis*, 13:224-233.
- JACQUEZ, G. M.; WALLER, L. A.; GRIMSON, R. & WARTENBERG, D., 1996. The analysis of disease clusters, part I: state of the art. *Infection Control and Hospital Epidemiology*, 17:319-327.
- JONG, P.; SPRENGER, C. & VEEN, F. V., 1984. On extreme values of Moran's I and Geary's c. *Geographical Analysis*, 16:17-24.
- KHEIFETS, L. I., 1993. Cluster analysis: a perspective. *Statistics in Medicine*, 12:1755-1756.
- KLAUBER, M. R. & ANGULO, J. J., 1974. Variola Minor in Bragança Paulista county, 1956: space-time interactions among Variola Minor cases in two elementary schools. *American Journal of Epidemiology*, 99:65-74.
- KLAUBER, M. R. & MUSTACCHI, P., 1970. Space-time clustering of childhood leukemia in San Francisco. *Cancer Research*, 30:1.969-1.973.
- KLEINBAUM, D. G.; KUPPER, L. L. & MORGENSTERN, H., 1982. *Epidemiologic Research: Principles and Quantitative Methods*. New York: Lifetime Learning Publications.
- KLEINBAUM, D. G.; KUPPER, L. L. & MULLER, K. E., 1988. *Applied Regression Analysis and Other Multivariable Methods*. 2a ed., Belmont: Duxbury.
- KNOX, E. G., 1964a. The detection of space-time interactions. *Applied Statistics*, 13:25-29.
- KNOX, E. G., 1964b. Epidemiology of childhood leukaemia in Northumberland and Durham. *British Journal of Preventive and Social Medicine*, 18:17-24.
- KNOX, E. G., 1991. Spatial and temporal studies in epidemiology. In: *Oxford Textbook of Public Health: Methods of Public Health* (W. W. Holland, R. Detels & G. E. Knox, eds.), vol. 2, pp. 95-105, Oxford: Oxford University Press.
- LAM, N. S.; FAN, M. & LIU, K., 1996. Spatial-temporal spread of the AIDS epidemic, 1982-1990: a correlogram analysis of four regions of the United States. *Geographical Analysis*, 28:93-107.
- LAWSON, A. B. & VIEL, J. F., 1995. Tests for directional space-time interactions in epidemiological data. *Statistics in Medicine*, 14:2.383-2.391.
- LEHMANN, E. L. & D'ABRERA, H. J. M., 1975. *Non-parametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.
- LILIENFELD, A. M. & LILIENFELD, D. E., 1980. *Foundations of Epidemiology*. 2a ed, New York: Oxford Press.
- MacMAHON, B. & PUGH, T., 1970. *Epidemiology: Principles and Methods*. Boston: Little, Brown & Co.
- MANTEL, N., 1967. The detection of disease clustering and a generalised regression approach. *Cancer Research*, 27:209-220.
- MANTEL, N.; KRYSZCIO, R. J. & MYERS, M. H., 1976. Tables and formulas for extended use of the Ederer-Myers-Mantel disease-clustering procedure. *American Journal of Epidemiology*, 104:576-584.
- MARSHALL, R. J., 1991. A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society A*, 154:421-441.
- McAULIFFE, T. L. & AFIFI, A. A., 1984. Comparison of a nearest neighbor and other approaches to the detection of space-time clustering. *Computational Statistics & Data Analysis*, 2:125-142.
- MORAN, P. A. P., 1948. The interpretation of statistical maps. *Journal of the Royal Statistical Society B*, 10:243-251.
- MORAN, P. A. P., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37:17-23.
- NAUS, J., 1965. The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*, 60:532-538.
- ODEN, N. L., 1993. Assessing directional effects in spatial data. *Statistics in Medicine*, 12:1.795-1.805.
- PIKE, M. C. & SMITH, P. G., 1968. Disease clustering: a generalization of Knox's approach to the detection of space-time interactions. *Biometrics*, 24:541-556.
- PINKEL, D.; DOWD, J. E. & BROSS, I. D. J., 1963. Some epidemiological features of malignant solid tumors of children in the Buffalo, N.Y., area. *Cancer*, 16:28-33.
- RAUBERTAS, R. F., 1988. Spatial and temporal analysis of disease occurrence for detection of clustering. *Biometrics*, 44:1.121-1.129.
- RAUBERTAS, R. F., 1989. An analysis of disease surveillance data that uses the geographic locations of the reporting units. *Statistics in Medicine*, 8:267-271.
- RAYSON, M. K. & KRYSZCIO, R. J., 1993. Properties of Tango's index for detecting clustering in time. *Statistics in Medicine*, 12:1.813-1.827.

- RICHARDSON, S., 1992. Modélisation statistique des variations géographiques en épidémiologie. *Revue d'Epidémiologie et Santé Publique*, 40:33-45.
- ROBERSON, K. P., 1990. Controlling for time-varying population distributions in disease clustering studies. *American Journal of Epidemiology*, 132:S131-S135.
- RODRIGUES, L. C.; MARSHALL, T.; MURPHY, M. & OSMOND, C., 1992. Space time clustering of births in AIDS: do perinatal infections play a role? *International Journal of Epidemiology*, 21:714-719.
- ROTHMAN, K. J., 1990. A sobering start for the cluster busters' conference. *American Journal of Epidemiology*, 132:S6-S13.
- SAMUELSSON, U.; JOHANSSON, C.; CARSTENSEN, J. & LUDOVIGSSON, J., 1994. Space-time clustering in insulin-dependent Diabetes Mellitus (IDDM) in south-east Sweden. *International Journal of Epidemiology*, 23:138-142.
- SELVIN, S., 1991. *Statistical Analysis of Epidemiologic Data*. New York: Oxford.
- SIEMIATYCKI, J. & McDONALD, A. D., 1972. Neural tube defects in Quebec: a search for evidence of 'clustering' in time and place. *British Journal of Preventive and Social Medicine*, 26:10-14.
- SMANS, M. & ESTÈVE, J., 1992. Practical approaches to disease mapping. In: *Geographical & Environmental Epidemiology: Methods for Small-Area Studies* (P. Elliot, J. Cuzick, D. English & R. Stern, eds.), pp.141-150, Oxford: Oxford University Press.
- SMITH, P. G., 1982. Spatial and temporal clustering. In: *Cancer Epidemiology and Prevention* (D. Schottenfeld & J. F. Fraumeni, eds.), pp. 391-407, Philadelphia: Saunders.
- TANGO, T., 1984. The detection of disease clustering in time. *Biometrics*, 40:15-26.
- WALLER, L. A. & JACQUEZ, G. M., 1995. Disease models implicit in statistical tests of disease clustering. *Epidemiology*, 6:584-590.
- WALLENSTEIN, S.; GOULD, M. S. & KLEINMAN, M., 1989. Use of the scan statistic to detect time-space clustering. *American Journal of Epidemiology*, 130:1.057-1.064.
- WALTER, S. D., 1993. Assessing spatial patterns in disease rates. *Statistics in Medicine*, 12:1.885-1.894.
- WARTENBERG, D. & GREENBERG, M., 1990a. Detecting disease clusters: the importance of statistical power. *American Journal of Epidemiology*, 132: S156-S166.
- WARTENBERG, D. & GREENBERG, M., 1990b. Space-time models for the detection of clusters of disease. In: *Spatial Epidemiology* (R.W. Thomas, ed.), pp. 17-34, London: Pion.
- WARTENBERG, D. & GREENBERG, M., 1993. Solving the cluster puzzle: clues to follow and pitfalls to avoid. *Statistics in Medicine*, 12:1.763-1.770.
- WARTENBERG, D. & GREENBERG, M., 1994. Characterizing cluster studies: a review of the literature. *Conference on Statistics and Computing in Disease Clustering*. Vancouver, British Columbia, Canada.
- WILLIAMS, G. W., 1984. Time-space clustering of disease. In: *Statistical Methods for Cancer Studies* (R. G. Cornell, ed.), pp. 167-227, New York: Dekker.