ESTIMABILIDADE DE MEDIDAS DE ASSOCIAÇÃO E DE RISCO EM ESTUDOS CASO-CONTROLE ESPACIAIS

Ricardo Cordeiro[1] (coordenador)

Laecio Carvalho de Barros[2] (pesquisador principal)

Cláudia Torres Codeço[3] (pesquisador)

Paulo Justiniano Ribeiro Junior[4] (pesquisador)

Trevor Charles Bailey[5] (pesquisador)

[1] Professor Associado do Departamento de Medicina Preventiva e Social da Faculdade de Ciências Médicas da Universidade Estadual de Campinas – UNICAMP

[2] Professor Associado do Departamento de Matemática Aplicada do Instituto de Matemática e Ciências da Computação da Universidade Estadual de Campinas – UNICAMP

[3] Professora Doutora do Programa de Computação Científica da Fundação Oswaldo Cruz – FIOCRUZ

[4] Professor Adjunto do Departamento de Estatística da Universidade Federal do Paraná – UFPr

[5] Professor Titular do Department of Mathematical Sciences, School of Engineering, Computer Sciences & Mathematics, University of Exeter, UK.

Campinas
Outubro de 2006

1. INTRODUÇÃO: A EVOLUÇÃO DOS ESTUDOS CASO-CONTROLE

O surgimento de uma série de casos de uma doença em quantidade maior do que a esperada geralmente atrai a atenção. Na caracterização dessa situação, três medidas de incidência são freqüentemente utilizadas. Talvez, a mais comum delas seja a taxa de incidência[43], entendida como a expressão da mudança do estado *saudável* para *doente* dos indivíduos de uma população. Outra medida bastante usada é a proporção de incidência[43], definida como a proporção de indivíduos que adoecem em uma população fechada sob risco durante um dado período de tempo. Ainda um outro modo de medir a incidência pode ser obtido por meio do *odds* de incidência[43], entendido como a razão entre o número de indivíduos que adoecem e o número de indivíduos que não adoecem em uma população fechada sob risco durante um dado período de tempo.

Na tentativa de entender as origens de uma ocorrência não usual de casos (uma epidemia), profissionais da saúde comumente buscam examinar a história individual de cada um dos acometidos procurando exposições compartilhadas por eles que não sejam freqüentes na população que deu origem a esses casos.

Na busca do entendimento dos mecanismos que governam o fenômeno em questão, o próximo passo é verificar se há relação entre o surgimento dos casos e as exposições eventualmente identificadas. Este é o campo da epidemiologia, área do conhecimento que estuda relações entre exposições e fenômenos do processo saúde/doença ocorrendo em populações humanas. Uma medida usual dessa relação é o risco relativo (RR), definido como a razão entre a incidência (medida de qualquer uma das maneiras acima mencionadas) de um agravo em uma população exposta e em outra não exposta a um dado conjunto de exposições de interesse, em um determinado intervalo de tempo[36].

A relação entre exposição e doença em populações humanas também pode ser avaliada por meio da medida fração atribuível (FA), definida como a porcentagem

de casos que deixariam de ocorrer, em uma população num dado intervalo de tempo, caso fosse eliminado um determinado fator causal dessa doença[28].

O entendimento de relações causais que poderiam explicar parcial ou totalmente o surgimento de séries de casos, como o acima exemplificado, é obtido comumente por meio da execução de estudos caso-controle. O desenvolvimento desses estudos constitui a maior contribuição metodológica da epidemiologia[9]. Este tipo de investigação propicia uma maneira eficiente de obter estimativas populacionais de RR e FA. A razão desta eficiência advém da capacidade que o estudo caso-controle tem, a um custo relativamente baixo, de estimar a distribuição populacional de exposições a partir do seu grupo controle. Esta propriedade dispensa o consumo de tempo, bem como recursos financeiros e administrativos, na enumeração e seguimento de um grande conjunto de indivíduos, classificados de acordo com níveis de exposição.

A epidemiologia trabalha com relações causais entre exposições e doenças, que ocorrem em populações humanas. Os casos surgem em uma população, nem sempre facilmente identificável, chamada população fonte. A pedra de toque da eficiência do estudo caso-controle é a seleção dos controles. Para garantir a estimabilidade das medidas RR e FA, à parte flutuações randômicas, a distribuição de exposições entre controles deve ser a mesma da população fonte de casos.

Os primeiros estudos caso-controle começaram a ser realizados na década de 1920, nos Estados Unidos[48]. Cole refere que até o final dos anos 1950 eles eram "*raramente executados, pobremente entendidos e pouco conceituados*"[8]. A essa época, Mantel e Haenszel referiram que "*estatísticos têm se mostrado um pouco relutantes em analisar dados gerados em estudos caso-controle, possivelmente porque seu treinamento enfatiza a importância em se definir um universo e especificar regras de contagem de eventos ou de obtenção de amostras probabilísticas nesse universo. Para os estatísticos, caminhar do efeito para a causa, com sua conseqüente perda de especificidade da população sob risco, parece uma aproximação pouco natural*"[31].

A década de 1950 foi decisiva para o desenvolvimento metodológico dos estudos caso-controle[40]. Em 1951, Cornfield publica o primeiro estudo sobre este método[10], demonstrando que as freqüências de exposição entre casos e controles, medidas de fácil obtenção, podem estimar um parâmetro de grande interesse epidemiológico: "*razão da freqüência de doença entre indivíduos expostos relativa àquela entre indivíduos não expostos*"[10]. Este parâmetro ganhou vários nomes e diferentes interpretações nos anos subseqüentes, sendo hoje amplamente identificado como risco relativo. Neste texto, Cornfield argumenta que a razão do *odds* de exposição entre casos e controles é um bom estimador do risco relativo, contanto que a freqüência da doença estudada seja rara. Em 1954, Cochran fornece as bases da análise estratificada e mostra como é possível combinar duas ou mais tabelas 2x2 em um único indicador sumariante de associação[7]. Em 1955, Wolf propõe as bases da análise do *odds ratio* (OR) utilizando a transformação logarítmica[55]. Em 1956, Cornfield deduz um método de obtenção de intervalos de confiança para OR[11], procedimento rotineiramente executado nos dias de hoje por pacotes estatísticos. Em 1959, Mantel e Haenszel sistematizam alguns dos pressupostos dos estudos caso-controle e indicam dois métodos para a estimativa de OR em estudos estratificados[31], largamente utilizados até o presente.

Estas análises, que conferiram credibilidade aos estudos caso-controle, tinham como referência um desenho com seleção de casos prevalentes. Daí o termo "razão de freqüência de doença", cunhado por Cornfield, ao invés de "razão de incidência de doença", que seria mais adequado à idéia de risco relativo. Sob este desenho, onde os controles são alocados após a ocorrência de todos os casos (isto é, entre os "sobreviventes" da doença), e sem considerações a respeito da população fonte de casos, a raridade da doença estudada é um pressuposto necessário para que o OR obtido estime a razão de proporções de incidência buscada na população base do estudo[36]. Isto porque na população fonte de casos, em condições estacionárias, a proporção de expostos a um fator causal que não adoecem decresce durante o estudo. Desse modo, controles amostrados apenas no final do período de estudo subestimam a freqüência da exposição na população fonte, produzindo, conseqüentemente, uma estimativa de associação

exposição/doença superestimada[42], que pode ser negligenciada apenas se a doença for rara. Uma outra maneira de entender esse viés é lembrar que um dos princípios dos estudos caso-controle estabelece que os controles devem ser amostrados de modo independente da exposição em estudo[43]. Entretanto, as pessoas que sobrevivem à doença têm maior probabilidade de não terem sido expostas, o que torna o número de controles não expostos maior do que o esperado, inflacionando o OR estimado[42]. A necessidade de pressuposição da raridade da doença para que as estimativas OR obtidas em estudo caso-controle estimem de modo adequado o RR se difundiu amplamente na literatura epidemiológica mundial, sendo incorporado de modo erroneamente restritivo a livros texto clássicos, como o de Macmahon e Pugh[30], ecoando inclusive até hoje na literatura nacional, como pode ser visto, por exemplo, no livro texto de Rouquayrol e Almeida Filho[44], amplamente difundido no Brasil.

Dando prosseguimento ao refinamento metodológico dos estudos caso-controle, além de considerações importantes a respeito da identificação e do controle de vieses, novas abordagens amostrais foram incorporadas na década de 1970. A essa época, Thomas[51] e Kupper & colaboradores[27] sugeriram a conveniência de, em estudos caso-controle, amostrar os controles a partir de toda a população base no início do estudo, antes de os casos incidirem. Esse procedimento amostral passou a ser conhecido como “case-base sampling”[34] ou “case-cohort design”[39]. Miettinen em 1976, no trabalho “Estimability and estimation in case-referent studies”, argumenta que em estudos caso-controle com esse desenho amostral, que abre a possibilidade de um mesmo indivíduo ser amostrado como controle no início do estudo e como caso no seu decorrer, o OR obtido é um estimador não enviesado da razão de proporção de incidência na população base do estudo, sem que seja necessária a pressuposição de raridade da doença estudada. Esta propriedade pode ser vista pormenorizadamente em Cordeiro[9] (2005).

Alguns anos mais tarde, já na década de 1980, consolida-se um outro modo de amostrar controles em estudos caso-controle de base populacional. Neste formato – conhecido como “risk-set sampling”[41], “sampling from the study base”[33] e

"density sampling"[26] – os controles vão sendo aleatoriamente amostrados da população fonte de casos, na medida em que os casos estudados incidem, o que torna a probabilidade de ser amostrado proporcional ao tempo em que cada indivíduo esteve sob risco nesta população. Em estudos caso-controle com este desenho amostral, que de modo semelhante possibilita a um mesmo indivíduo participar do estudo tanto como caso, quanto como controle, o OR obtido estima sem viés a razão de taxa de incidência na população base do estudo, também sem que seja necessária nenhuma pressuposição sobre a raridade da doença estudada. Esta propriedade pode ser vista pormenorizadamente em Cordeiro[9] (2005).

Dando um fecho à discussão metodológica que esses novos desenhos amostrais ensejaram, no início da década de 1990 Pearce[36] afirma que o significado das estimativas de OR obtidas em estudos caso-controle de base populacional difere se os controles são amostrados a partir das pessoas em risco no início do seguimento, da experiência pessoa-tempo da população fonte de casos ou da população não doente ao final do seguimento. Estudos caso-controle utilizando esses três métodos de seleção de controles estimam razão de proporção de incidência (RPI), razão de taxa de incidência (RTI) e OR, respectivamente. Nenhuma dessas estimativas depende de qualquer pressuposição sobre a raridade da doença estudada. A pressuposição de raridade é necessária apenas para que a estimativa de OR seja uma aproximação do risco relativo em estudos com o desenho tradicional de Cornfield, Mantel e Haenszel, isto é, com amostragem de controles entre não doentes, após já identificados todos os casos[9].

A estimação da distribuição de exposições na população fonte, que em estudos caso-controle é feita a partir do grupo controle, geralmente envolve considerável esforço. Por essa razão, à época da síntese acima mencionada, já apareciam na literatura epidemiológica propostas de desenhos de estudo envolvendo apenas casos de doença. São exemplos dessas propostas os desenhos "case-genotype studies"[15;18;25;46;47;49;50;52], "case-cross-over study"[16;17;29;32], "case-specular study"[56],

entre outros[4;35;38]. Estes estudos foram genericamente chamados por Greenland[20] de “case-distribution studies” ou “case-only studies”. O que eles trazem de avanço são tentativas, em situações particularmente favoráveis, de construir hipotéticas distribuições de exposições em populações fonte, ou, equivalentemente, hipotéticas séries de controles, a partir de informação externa ao estudo[20]. A validade das estimativas de estudos “case-only” depende da validade das pressuposições utilizadas para a construção da distribuição de exposições em populações fonte. Por exemplo, para os “case-genotype studies”, depende do modelo de prevalência genotípica adotado; para os “case-cross-over studies”, depende de condições estacionárias das exposições dos casos ao longo do tempo; para os “case-specular studies”, depende da suposta intercambialidade entre exposições na residência do caso e na casa à frente no outro lado da rua[20].

Também a essa época, na medida em que o desenvolvimento da microinformática torna viável o desenvolvimento e a popularização do uso de sistemas de informação geográfica (SIG) e de ferramentas de análise espacial de dados, a dimensão espacial começa a ser incorporada ao método caso-controle.

Em certos lugares o risco de adoecer é maior do que em outros. A incidência de um determinado agravo também pode ser função do espaço. Nos anos 1990 começa a se desenvolver uma especialização dos estudos caso-controle que incorpora explicitamente o espaço ao conjunto de covariáveis que modulam o risco. São os chamados estudos caso-controle espaciais, que então passam a contribuir com a incorporação de métodos de análise espacial de dados ao instrumental analítico da epidemiologia. Este é o campo onde se aninha este projeto de pesquisa.

2. METODOLOGIA DOS ESTUDOS CASO-CONTROLE ESPACIAIS

A preocupação com a distribuição espacial de casos de uma doença parece ser uma constante na epidemiologia desde o seu nascimento enquanto disciplina.

Entretanto, apenas nas duas últimas décadas a análise espacial de pontos começou a se consolidar enquanto método em estudos epidemiológicos, sendo ainda um campo do conhecimento novo e em consolidação[54]. Formalmente, um padrão espacial de pontos pode ser entendido como um conjunto de localizações numa determinada região de estudo representando o registro de eventos de interesse[19]. Tais localizações podem estar acompanhadas de informações adicionais relativas aos eventos registrados, entre elas, de particular importância, o tempo associado a cada ocorrência.

De um ponto de vista estatístico, um padrão observado de pontos pode ser modelado como a realização de um processo estocástico. O modelo mais simples de uma distribuição de pontos é o da aleatoriedade espacial completa, no qual os eventos se distribuem de forma independente entre si sobre uma região de interesse. Esse modelo tem pouca aplicação epidemiológica (exceção feita a estudos em escalas grandes) uma vez que as populações fontes de caso distribuem-se elas próprias em aglomerados espaciais determinados por fatores ambientais e sociais. Na ausência de associação entre espaço e doença, hipótese nula, a ocorrência de casos espelha a heterogeneidade espacial da população fonte. A incorporação dos métodos de análise espacial de dados ao instrumental epidemiológico se deu a partir da indagação sobre o quanto um aglomerado observado de casos se deve à aglomeração de base da população fonte.

Um modo intuitivo de incorporar a dimensão espacial aos estudos caso-controle pode ser encontrado em Bithell[5] e em Gatrell e colaboradores[19]. Estes autores utilizaram a razão entre a densidade espacial de casos e de controles, obtidas por estimação kernel[2], área a área, como um estimador da distribuição espacial de risco relativo.

Cuzick & Edwards[12] propuseram um modo de investigar a existência de aglomerados espaciais de doença, comparando a distribuição observada de casos e controles em um dado estudo com distribuições hipotéticas geradas a partir da rotulação aleatória (casos ou controles) dos pontos observados.

Usando uma variação do método do "vizinho mais próximo"[2], Vieira e colaboradores[53] propuseram um algoritmo para estimação da distribuição espacial de OR em estudos caso-controle que, basicamente, dado um conjunto de localizações de casos e controles, consiste em calcular o odds caso/controle em pequenas áreas arbitrariamente estabelecidas cobrindo inteiramente uma região de interesse e fazer a razão desses valores pelo odds caso/controle da região como um todo.

Os métodos acima citados apresentam limitações importantes, tais como a dependência de escolha arbitrária de largura de banda a ser utilizada, e a dificuldade em incorporar covariáveis à análise. Uma alternativa promissora é o desenvolvimento de estudos caso-controle espaciais, que parecem ser úteis para verificação de variabilidade espacial do risco, identificação de confundimento espacial, avaliação da influência no mapa de risco da utilização de tempo de latência no estudo, bem como sugerir hipóteses causais para futuras investigações. Abaixo, são descritos os pressupostos teóricos e metodológicos que embasam os estudos caso-controle espaciais.

Para uma determinada doença e uma delimitada região $\Re$ de interesse (uma cidade, um distrito, um bairro), seja a função risco relativo espacial, conforme definida por Bithell[5], uma função bivariada descrita da seguinte forma:

$$\theta(\mathbf{x}) = \frac{\lambda(\mathbf{x})}{\pi(\mathbf{x})} \quad (1)$$

onde **x** é um vetor de coordenadas geográficas em $\Re$, $\lambda(\mathbf{x})$ é a função densidade espacial de ocorrência de casos da doença em $\Re$ (número de ocorrências por unidade de área na localização **x**) e $\pi(\mathbf{x})$ é a função densidade espacial da população fonte de casos em $\Re$. A função $\theta(\mathbf{x})$ integra 1 sobre $\Re$, se usada a densidade da população fonte como função ponderadora[6], isto é:

$$\iint_{\Re} \theta(\mathbf{x})\pi(\mathbf{x})d\Re = 1 \quad (2)$$

O risco relativo assim definido em uma dada localização $\mathbf{x} \in \Re$ representa o risco de ocorrência de um caso em **x** relativo ao risco médio de ocorrência em $\Re$. Se este for conhecido, pode ser usado para se obter o risco absoluto em **x**, conforme será visto adiante.

A partir de um conjunto observado de casos e controles, Kelsall & Diggle[24] e Diggle[14] propuseram um modo interessante de estimar $\theta(\mathbf{x})$, abaixo descrito.

Seja C e P os conjuntos de localizações (coordenadas geográficas) de casos e população fonte em estudo, respectivamente. Assume-se que as localizações em C e P sejam realizações de dois processos planares de Poisson[13;22], com intensidade $\lambda(\mathbf{x})$ e $\pi(\mathbf{x})$, respectivamente.

Seja $\mathbf{x_i}$, i = 1, 2, ..., n_1, n_1 localizações observadas do conjunto C.

Seja $\mathbf{x_i}$, i = n_1+1, n_1+2, ..., n_1+n_2, n_2 localizações observadas do conjunto P, onde $n = n_1 + n_2$.

Seja y_i um indicador associado ao ponto $\mathbf{x_i}$, tal que $y_i = 1$ se $\mathbf{x_i} \in$ C, $y_i = 0$ se $\mathbf{x_i} \in$ P.

Assume-se que os casos e os controles incluídos no estudo são amostras aleatórias de todos os casos ocorridos (C) em $\Re$ e da população fonte (P) existente em $\Re$ no período de estudo, nas proporções q_1 e q_2, respectivamente. Desse modo, condicional aos pontos $\mathbf{x_i}$, y_i são realizações mutuamente independentes de uma variável aleatória de Bernoulli[13], onde a probabilidade de qualquer ponto, observado no conjunto de casos e controles, ser um caso [isto é, $p(\mathbf{x}) = Pr(y_i = 1 \setminus \mathbf{x_i} = \mathbf{x})$] é dada por:

$$p(\mathbf{x}) = \frac{q_1 \lambda(\mathbf{x})}{q_1 \lambda(\mathbf{x}) + q_2 \pi(\mathbf{x})} \quad (3)$$

A partir das equações 1 e 3, com álgebra simples demonstra-se que $p(\mathbf{x})$ está relacionado ao risco espacial $\theta(\mathbf{x})$ conforme a expressão abaixo:

$$\ln[\theta(\mathbf{x})] = \ln\left[\frac{p(\mathbf{x})}{1-p(\mathbf{x})}\right] + \ln\left(\frac{q_2}{q_1}\right) = \text{logit}[p(\mathbf{x})] + c = \rho(\mathbf{x}) \quad (4)$$

Desse modo, o logaritmo da função risco relativo espacial, ρ(**x**), aparte uma constante aditiva c, depende apenas do logito da função p(**x**), definido na equação 3.

Assim, a tarefa de modelar θ(**x**) se reduz efetivamente à modelagem do logit[p(x)] a partir de um conjunto observado de desfechos Bernoulli caso/controle, y_i, associados às suas localizações espaciais, $\mathbf{x}_i$. Em essência, este é um problema trivial de regressão logística[23], embora aqui, dado o provável comportamento não linear da superfície espacial de risco a ser estimada, o uso de um método não paramétrico seja mais apropriado do que a usual abordagem paramétrica. Para tanto, Kelsall & Diggle[24] propuseram o uso do modelo aditivo generalizado[21] (GAM) como ferramenta básica, onde:

$$\rho(\mathbf{x}) = c + \alpha + g(\mathbf{x}) \quad (5)$$

sendo c uma constante conhecida, α um intercepto a ser estimado e g(**x**) uma função bivariada suave do vetor de localização **x** que pode ser não parametricamente estimada por vários algoritmos iterativos bem estabelecidos[21], tais como regressão de kernel[45], regressão polinomial local[45] ou regressão spline penalizada[45].

Uma extensão bastante conveniente deste método é a incorporação de covariáveis não espaciais, o que resulta no modelo semiparamétrico:

$$\rho(\mathbf{x},\mathbf{z}) = c + \alpha + \boldsymbol{\beta}\mathbf{z} + g(\mathbf{x}) \quad (6)$$

onde **z** é um vetor de covariáveis de interesse e **β** representa seu vetor de efeitos. O log-risco $\rho(\mathbf{x},\mathbf{z})$ é assim modelado como linearmente dependente de um componente não espacial paramétrico representando o efeito das covariáveis incorporadas ao modelo, associado ao componente espacial não paramétrico e possivelmente não linear $g(\mathbf{x})$. Estimativas de α, **β** e $g(\mathbf{x})$, bem como seus erros

padrão associados, são obtidas segundo a aplicação do modelo aditivo generalizado[21] acima mencionado. Se o risco estimado for invariante no espaço, g(**x**) é constante e o modelo (equação 6) se reduz a uma regressão logística trivial[23].

$$\rho(\mathbf{z}) = c + \alpha + \boldsymbol{\beta}\mathbf{z} \qquad (7)$$

Finalmente, a estimativa da função risco relativo espacial (equação 1) em $\Re$ é obtida como:

$$\hat{\theta}(\mathbf{x},\mathbf{z}) = e^{\hat{\rho}(\mathbf{x},\mathbf{z})} \quad (8)$$

onde $\hat{\rho}(\mathbf{x},\mathbf{z})$ representa a estimativa obtida no ajuste da equação 6. Se desejado, a estimativa da distribuição espacial da incidência de casos em $\Re$ pode ser obtida como:

$$\hat{I}(\mathbf{x},\mathbf{z}) = \hat{i}\hat{\theta}(\mathbf{x},\mathbf{z}) = \hat{i}\, e^{\hat{\rho}(\mathbf{x},\mathbf{z})} \quad (9)$$

onde î representa a estimativa da proporção média de incidência de casos em $\Re$.

A significância dos componentes paramétricos estimados na equação 6 são obtidas de modo usual, como nos modelos lineares generalizados, com base na magnitude da razão entre estimativas e seus respectivos erros padrão. Para testar a significância da variação espacial do risco estimada na equação 6, não foram ainda descritos métodos analíticos. Para tanto, Kelsall & Diggle[24] propuseram um procedimento de Monte Carlo, abaixo descrito:

a) O modelo completo expresso pela equação 6 é ajustado aos dados via GAM. Desse modo, predições de respostas são obtidas para cada ponto de uma grade previamente definida (200 x 200, por exemplo) cobrindo $\Re$.

b) O modelo reduzido (hipótese nula) expresso pela equação 7, que omite o componente espacial g(**x**), é ajustado aos dados por meio de uma regressão logística usual. Desse modo, predições da probabilidade de ser caso são

geradas para cada um dos sujeitos estudados (casos e controles), de acordo com o componente não espacial ($\boldsymbol{\beta z}$) individual.

c) Para cada um dos sujeitos estudados é determinado um novo status de caso ou controle de modo aleatório usando-se as probabilidades preditas no passo *b* acima.

d) O modelo completo expresso pela equação 6 é ajustado aos dados modificados no passo *c* via GAM, e assim novas predições de respostas são obtidas para cada ponto da mesma grade previamente definida no passo *a*.

e) Os passos *c* e *d* são repetidos um grande número de vezes (por exemplo, 500 vezes) produzindo simulações de superfícies resposta sob H_0 (isto é, sob a hipótese de homogeneidade espacial do risco).

f) Ponto a ponto sobre a grade definida em *a*, a resposta predita em *a* é comparada com a distribuição de simulações sob H_0 produzidas em *e*. Se aquela se situa entre os percentis $(\alpha/2)100\%$ e $(1 - \alpha/2)100\%$ desta, a correspondente localização na grade é rotulada "não significante". Caso contrário, é rotulada "significante".

g) Finalmente, um conjunto de mapas é produzido a partir da distribuição espacial da incidência de casos predita na equação 9, de acordo com valores pré-definidos de interesse do componente não espacial (**z**), excluindo-se todas as posições na grade rotuladas "não significantes" no passo *f*. Esses mapas, portanto, mostram apenas as estimativas que diferem significativamente da incidência média em $\Re$, para a configuração escolhida de z ($p<\alpha$).

3. OBJETIVOS

Face o grande desenvolvimento metodológico dos estudos caso-controle nos últimos 50 anos, bem como a contribuição analítica que os sistemas de

informação geográfica e os métodos de análise espacial de dados trazem para a epidemiologia, este projeto tem como objetivos, em estudos caso-controle espaciais:

a) verificar o comportamento e o significado epidemiológico da função risco-relativo espacial[5] em função dos desenhos amostrais "case-base sampling"[34], "risk-set sampling"[41] e amostragem a partir de "sobreviventes" da doença estudada[36].

b) desenvolver modelos computacionais que simulem, para um conjunto de parâmetros demográficos, geográficos e epidemiológicos pré-definidos, cenários de dinâmica espaço-temporal de ocorrência de casos e o processo de execução de estudos caso-controle espacial nesta população. Estas simulações permitirão a avaliação do comportamento e da precisão do estimador da função risco relativo espacial obtido a partir dos desenhos amostrais acima citados, em situações que simulem a complexidade do processo gerador de casos.

c) aplicar o modelo de regressão logística multinomial[23] para a estimação da distribuição espacial do risco de ocorrências quando estas são classificadas de acordo com sua "gravidade", isto é, em contraposição ao modelo binário acima descrito, que classifica os indivíduos estudados em casos ou controles, desenvolver um tratamento multinomial, que possibilite a estimação da distribuição espacial do risco de ocorrência de casos, em função de sua gravidade.

d) desenvolver testes de significância para os componentes não-lineares da equação 6 que possam ser utilizados paralelamente ou em substituição ao procedimento Monte-Carlo acima descrito.

e) testar os resultados metodológicos obtidos, particularmente o explicitado no item c acima, na reanálise de dois estudos caso-controle espaciais: "Distribuição espacial do risco de acidentes do trabalho no mercado

informal de Piracicaba (FAPESP 05/03491-9)" e "Distribuição espacial do risco de dengue na região sul do município de Campinas (FAPESP 06/01224-6).

4. MÉTODO

Os objetivos enunciados nos itens 3.a, 3.b, 3.c e 3.d acima serão obtidos por meio de:

i) constituição de um grupo de trabalho composto por todos os pesquisadores do projeto, quatro alunos bolsistas de pós-doutorado e quatro alunos bolsistas de doutorado, quatro alunos bolsistas de mestrado e quatro alunos bolsistas de iniciação científica.

ii) leitura de toda a bibliografia publicada nos últimos 16 anos sobre o tema, por parte do grupo de trabalho constituído, tomando como marco inicial o texto de Hastie & Tibshirani[21]: Generalized additive models.

iii) realização de seminários semanais descentralizados (isto é, nas cidades onde trabalham os pesquisadores do projeto) para a discussão e integração dos conceitos apreendidos na leitura realizada, bem como a discussão e o desenvolvimento do método analítico para a obtenção de estimativas de significância mencionadas no item 3.d acima e o desenvolvimento da abordagem multinomial mencionada no item 3.c acima.

iv) realização de seminários trimestrais reunindo todo o grupo de trabalho constituído para a apresentação, discussão e consolidação dos resultados obtidos nos seminários semanais (item 4.iii), bem como a programação do trabalho durante o próximo trimestre. O pesquisador da instituição estrangeira (Dr. Bailey) deverá participar de um desses seminários a cada ano.

v) anualmente, o pesquisador da instituição estrangeira deverá trabalhar pessoalmente com o restante do grupo de pesquisadores durante um período de um mês no Brasil, na sede do projeto.

vi) realização de seminários especiais sobre tópicos relacionados aos objetivos do projeto com os professores visitantes convidados de instituições nacionais e internacionais.

vii) visitas de trabalho a departamentos universitários nacionais e internacionais onde se desenvolve pesquisa sobre tópicos relacionados aos objetivos deste projeto.

viii) com relação aos modelos de simulação mencionados no objetivo 3.b acima, a modelagem computacional passará por três etapas.

1. Definição das condições iniciais: o modelo requer como entrada, um conjunto de parâmetros que represente a distribuição espacial da população em risco (e seus atributos individuais associados ao risco), e a distribuição espacial de fatores de risco ecológicos (associados ao lugar). Estes parâmetros deverão ser representativos do conjunto de problemas aos quais os estudos de caso-controle pretendem ser aplicados. Estas distribuições espaciais serão utilizadas para a criação de 'populações de risco sintética', isto é, um conjunto de indivíduos com atributos pessoais e associados a um local no espaço.

2. Definição do processo gerador de eventos: dois processos geradores de eventos serão considerados. O primeiro é um processo contagioso, onde a probabilidade de um indivíduo sofrer um evento é função do contato com indivíduos que o sofreram. Este processo gera estruturas de correlação espaciais e temporais. Eventos que se encaixam nesta

categoria são, por exemplo, doenças transmissíveis e são modelados por processos estocásticos do tipo SIR[1]. A forma da força de infecção pode ser alterada para contemplar diferentes padrões de transmissão, incluindo vias de transmissão diretas e indiretas, além de variações temporais e espaciais na força de infecção, etc. O segundo processo gerador é não contagioso, onde a probabilidade de um indivíduo sofrer o evento é função de características do próprio indivíduo ou do lugar onde ele está. Neste caso, a ocorrência de um evento não interfere na probabilidade de ocorrência de outro. Independente do caso, o processo gerador será realizado a partir da associação de probabilidades de transição 'não caso' - 'caso' a cada indivíduo da população de risco sintética em função das características do indivíduo e de seu ambiente. Séries temporais serão assim simuladas.

3. Simulação do processo amostral: Uma vez o processo de ocorrência de eventos seja simulado, pode-se aplicar procedimentos amostrais que obterão informações de uma amostra dos indivíduos sintéticos, em determinados momentos no tempo. A partir destas amostras, calcula-se os estimadores que poderão ser comparados com os 'valores verdadeiros' obtidos da população como um todo, ou a partir do conhecimento do próprio processo gerador.

Os modelos de simulação serão implementados em linguagem Python e operados em linux.

Ainda como método de obtenção do objetivo 3.b acima mencionado, será desenvolvido um modelo matemático, sob uma estrutura de incerteza, a partir de técnicas de autômatos celulares[37] combinadas com lógica fuzzy[3], simulando o espalhamento espacial de doenças, num ambiente criado para

a execução simulada de estudos caso-controle espaciais, como mais uma abordagem para o estudo do comportamento e da precisão do estimador da função risco-relativo espacial em função dos desenhos amostrais acima citados.

Na fase final da realização do projeto, como “teste de campo” e ilustração dos resultados obtidos, serão reanalizados os dados produzidos nas duas pesquisas referidas no item 3.e acima. Tratam-se de dois grandes estudos caso-controle espaciais atualmente em andamento (com previsão de término em 2008), financiados pela FAPESP e coordenados pelo coordenador deste projeto. Os dois estudos têm método amostral diferentes entre si. Na coleta dos dados, os casos de ambos foram também classificados quanto a gravidade. Isto torna bastante interessante e informativo a reanálise desses resultados com as ferramentas e métodos produzidos neste projeto.

5. RESULTADOS ESPERADOS

O produto esperado deste projeto é desenvolvimento metodológico. Espera-se ao final do estudo:

a) saber como as propriedades dos estimadores de risco nos estudos caso-controle espacial variam em função da estratégia amostral;

b) saber como levar em consideração a gravidade dos casos na estimação da distribuição espacial do risco em estudos caso-controle utilizando-se um modelo multinomial;

c) saber como dimensionar as flutuações amostrais dos estimadores de efeito nos estudos caso-controle espaciais (questão vital pois diretamente relacionada à verificação da significância estatística das estimativas obtidas) de um modo melhor do que o atualmente disponível na literatura;

d) produzir um ambiente computacional de simulação de ocorrências epidêmicas, útil, entre outras coisas, para a simulação de realizações de estudos caso-controle espaciais visando estudar empiricamente a eficiência de diferentes estratégias de planejamento do estudo;

Os objetivos propostos neste projeto são inéditos na literatura. Ao atingi-los, em suma, o projeto contribuirá para que avanços metodológicos obtidos no desenvolvimento dos estudos caso-controle sejam estendidos e incorporados à análise espacial de dados epidemiológicos.

Espera-se também que o desenvolvimento deste projeto contribua para a consolidação de um grupo de pesquisa e desenvolvimento de método epidemiológico em nosso meio, voltado para a análise espacial de dados, bem como para o fortalecimento do Laboratório de Análises Espacial de Dados Epidemiológicos (EpiGeo) do Departamento de Medicina Preventiva e Social da Faculdade de Ciências Médicas da Unicamp.

6. CRONOGRAMA DE ATIVIDADES

Este é um projeto temático cujo tema é a incorporação de avanços metodológicos dos estudos caso-controle à análise espacial de dados epidemiológicos. Diferentemente da maioria dos projetos de pesquisa epidemiológica (que envolvem amostragem, recrutamento e treinamento de equipe de campo, coleta de dados, controle de qualidade, digitação, e dezenas de outras etapas), as atividades serão leitura, reflexão, discussão, seminários internos e seminários com pesquisadores visitantes, desenvolvimento computacional e visitas científicas a centros de pesquisa. Dada a complexidade e a originalidade dos objetivos propostos e resultados previstos, prevê-se que o projeto se estenda por quatro anos, da seguinte maneira:

ATIVIDADE	SEMESTRE

	1º	2º	3º	4º	5º	6º	7º	8º
Levantamento bibliográfico extensivo	X	X						
Seminários semanais descentralizados	X	X	X	X	X	X	X	X
Seminários trimestrais	X	X	X	X	X	X	X	X
Seminários especiais com professor(es) convidado(s)		X		X		X		X
Visitas a centros internacionais de pesquisa		X		X		X		X
Desenvolvimento de método analítico para estimativas de significância e aplicação do modelo multinomial		X	X	X	X	X	X	
Desenvolvimento de modelos de simulação		X	X	X	X	X	X	X
Reanálise de estudos caso-controle espaciais					X	X	X	X
Redação de artigos científicos					X	X	X	X

7. ASPECTOS ÉTICOS

Este projeto não envolve experimentação com humanos ou animais, bem como a geração de resíduos químicos. Os estudos que serão reanalizados durante a execução deste projeto receberam pareceres favoráveis da Comissão de Ética em Pesquisa da Faculdade de Ciências Médicas da Unicamp.

8. APRESENTAÇÃO DA EQUIPE

Para o desenvolvimento do projeto, formou-se uma equipe multidisciplinar de pesquisadores com experiência nas áreas de epidemiologia, matemática aplicada, estatística espacial e modelagem da dinâmica do espalhamento de doenças. Abaixo, são apresentados os pesquisadores do projeto.

8.1. Ricardo Cordeiro (Coordenador)

Possui graduação em Medicina pela Universidade de São Paulo (1983), mestrado em Medicina Concentração Saúde Coletiva pela Universidade Estadual de Campinas (1991), doutorado em Saúde Coletiva pela Universidade Estadual de Campinas (1995), Pós-Doutorado em Epidemiologia na University of California at Los Angeles (1999-2000) e Livre-Docência em Epidemiologia na Universidade Estadual Paulista (2001). Atualmente é Professor Associado (MS-5) da Universidade Estadual de Campinas. Tem experiência em Epidemiologia, com ênfase em métodos quantitativos, atuando principalmente nos seguintes temas: análise espacial de dados, acidente do trabalho, saúde do trabalhador, vigilância em saúde.

Coordenou a realização de quatro Auxílios Individuais à Pesquisa (FAPESP 95/4342-3, 96/7583-4, 97/12782-9 e 00/09105-0) e um Auxílio à Pesquisa Políticas Públicas (FAPESP 00/13719-3). Atualmente coordena a realização de dois estudos caso-controle espaciais financiados na forma de Auxílios Individuais à Pesquisa (FAPESP 05/03491-9 e 06/01224-6).

Nos últimos cinco anos tem se dedicado a projetos de investigação incorporando a análise espacial de dados epidemiológicos. Em 2005, como Bolsista do Programa de Aperfeiçoamento Científico no Exterior da FAPESP (processo 04/09859-5), trabalhou no projeto “"Semiparametric modelling of the spatial distribution of occupational accident risk in the casual labor market, Piracicaba, Southeast Brazil" conjuntamente com o Professor Trevor Bailey, um dos participantes da equipe deste projeto. Atualmente, coordena o Laboratório de Análise Espacial de Dados

Epidemiológicos – EpiGeo, do DMPS/FCM/Unicamp.
Responsabilidades no Projeto: Coordenação, objetivos 3.a e 3.e.

8.2. Laecio Carvalho de Barros (Pesquisador Principal)

Possui graduação em Matemática pela Universidade de São Paulo (1979), mestrado em Matemática Aplicada pela Universidade Estadual de Campinas (1992) e doutorado em Matemática Aplicada pela Universidade Estadual de Campinas (1997). Atualmente é professor adjunto da Universidade Estadual de Campinas, atuando principalmente nos seguintes temas: biomatemática, sistemas dinâmicos, conjuntos fuzzy, controladores fuzzy, sistemas dinâmicos fuzzy. Desde 2001 é revisor dos periódicos Fuzzy Sets and Systems (0165-0114) e IEEE Transactions on Fuzzy Systems (1063-6706).
Responsabilidades no Projeto: objetivos 3.b. (determinista) e 3.e.

8.3. Cláudia Torres Codeço (Pesquisadora)

Possui graduação em Ciências Biológicas pela Universidade Federal do Rio de Janeiro (1992), mestrado em Engenharia Biomédica pela Universidade Federal do Rio de Janeiro (1995) e doutorado em Quantitative Biology - University Of Texas At Arlington (1998). Atualmente é pesquisadora associada da Fundação Oswaldo Cruz - Programa de Computação Científica. Tem experiência na área de Dinâmica de Populações, com ênfase em Modelagem Matemática de Dinâmica de Doenças Infecciosas. Atuando principalmente nos seguintes temas: biomatemática, epidemiologia, dinâmica de processos infecciosos, ecologia populacional e análise de sobrevida.
Responsabilidade no Projeto: objetivos 3.b. (estocástico) e 3.e.

8.4. Paulo Justiniano Ribeiro Junior (Pesquisador)

Possui graduação em Agronomia pela Universidade Federal de Lavras (1989) , mestrado em Agronomia pela Universidade de São Paulo (1992) e doutorado em Statistics pela University of Lancaster (2002) . Atualmente é Professor adjunto da Universidade Federal do Paraná. Tem experiência na área de Probabilidade e

Estatística , com ênfase em Estatística. Atuando principalmente nos seguintes temas: geostatistics, spatial statistics, statistical software, Bayesian inference. Atualmente é o Presidente da Região Brasileira da Sociedade Internacional de Biometria (RBRAS), membro do Conselho da Associação Brasileira de Estatística (ABE) e membro do Conselho da International Biometric Society (IBS).
Responsabilidades no Projeto: objetivos 3.c, 3.d. e 3.e.

8.5. Trevor Charles Bailey (Pesquisador)

Academic Qualifications: BSc., Mathematics, Hons., 1st Class, (Lond.), 1975; MSc., Mathematical Statistics, (Lond.), 1976; PhD., (Exon.), 1981. Joined the University of Exeter in 1986, having formerly lectured for four years at the Australian Graduate School of Management, University of New South Wales, Sydney, Australia. Research interests: computational statistics, applied statistical modelling, spatial statistics, spatial epidemiology. Current administrative responsibilities: Associate Dean - Faculty of Undergraduate Studies; Director of Undergraduate Studies - School Engineering, Computer Science and Mathematics; Programme Coordinator (Mathematics Programmes) - School Engineering, Computer Science and Mathematics.
Obs: O professor Bailey não tem currículo cadastrado na base Lattes. Seu currículo pode ser consultado em http://www.secamlocal.ex.ac.uk/~tcb/HomePage.html
Responsabilidades no Projeto: objetivos 3.c e 3.d.

Além dos pesquisadores acima, também fazem parte da equipe de pesquisa, como pesquisadores colaboradores, os professores Marília Sá Carvalho (epidemiologista, Professora Associada da FIOCRUZ), Carlos Alberto de Bragança Pereira (estatístico, Professor Titular do IME/USP) e Miguel Antônio Vieira Monteiro (engenheiro, Pesquisador Senior do INPE).

9. REFERÊNCIAS

1. Anderson RM, May RM. Infectious diseases of humans. Dynamics and control. Oxford: Oxford University Press, 1991.

2. Bailey TC, Gatrell AC. Interactive spatial data analysis. Harlow, UK: Longman, 1995.

3. Barros LC, Bassanezi RC. Tópicos de lógica fuzzy e biomatemática. Campinas: Coleção IMECC - Textos didáticos 5., 2006.

4. Begg CB, Zhang Z. Statistical analysis of molecular epidemiology studies employing case series. *Cancer Epidemiology, Biomarkers, and Prevention* 1994;**3**:173-5.

5. Bithell J. An application of density estimation to geographical epidemiology. *Statistics in Medicine* 1990;**9**:691-701.

6. Bithell JF. Disease mapping using the relative risk function estimates from areal data. In Lawson A, Biggeri A, Bohning D, Lesaffre E, Viel JF, Bertillini R, eds. *Disease mapping and risk assessment for public health.*, pp 248-55. Chichester: John Wiley & Sons, 1999.

7. Cochran WG. Some methods for strengthening the common X^2 tests. *Biometrics* 1954;**10**:417-51.

8. Cole P. Introduction. In Breslow NE, Day NE, eds. *Statistical methods in cancer research.*, pp 14-40. Geneve: IARC, 1980.

9. Cordeiro R. O mito da doença rara. *Revista Brasileira de Epidemiologia* 2005;**8**:111-6.

10. Cornfield J. A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix. *Journal of the National Cancer Institute* 1951;**11**:1269-75.

11. Cornfield, Jerome. A statistical problem arising from retrospective studies. Neyman, J. Proceedings of The Third Berkeley Symposium. IV. 56. Berkeley, University of California.

12. Cuzick J,.Edwards R. Spatial clustering for inhomogeneous populations (with discussion). *Journal of the Royal Statistical Society B* 1990;**52**:73-104.

13. DeGroot MH. Probability and statistics, 2nd ed. Reading, MA: Addison-Wesley Publishing Company, 1986.

14. Diggle PJ. Statistical analysis of spatial point patterns, 2nd ed. London: Arnold, 2003.

15. Falk CY, Rubinstein P. Haplotype relative risk: an easy reliable way to construct a proper controls sample for risk calculations. *Annals of Human Genetics* 1987;**51**:227-33.

16. Feldman U. Design and analysis of drug safety studies. *Journal of Clinical Epidemiology* 1993;**46**:237-44.

17. Feldman U. Epidemiologic assessment of adverse reactions associated with intermittent exposure. *Biometrics* 1993;**49**:419-28.

18. Flanders WD, Khoury MJ. Analysis of case-parental control studies: method for the study of associations between disease and genetic markers. *American Journal of Epidemiology* 1996;**144**:696-703.

19. Gatrell AC, Bailey TC, Diggle PJ, Rowlingson BS. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions - Institute of British Geographers* 1996;**NS 21**:256-74.

20. Greenland S. A unified approach to the analysis of case-distribution (case-only) studies. *Statistics in Medicine* 1999;**18**:1-15.

21. Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman and

Hall, 1990.

22. Hogg RV, Craig AT. Introduction to mathematical statistics, 4th ed. New York: Macmillan Publishing Co, 1978.

23. Hosmer Jr, David W and Lemeshow, Stanley. Applied logistic regression. 2000. New York, John Wiley & Sons.

24. Kelsall JE, Diggle PJ. Spatial variation in risk of disease: a nonparametric binary regression approach. *Applied Statistics* 1998;**47**:559-73.

25. Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene-environment interactions: case-control studies with no controls! *American Journal of Epidemiology* 1996;**144**:207-13.

26. Kleinbaum DG, Kupper LI, Morgenstern H. Epidemiologic research - principles and quantitative methods. Belmont, CA: Lifetime Learning Publications, 1982.

27. Kupper LL, McMichael AJ, Spirtas R. A hybrid epidemiologic design useful in estimating relative risk. *Journal of the American Statistical Association* 1975;**70**:524-8.

28. Levin ML. The occurrence of lung cancer in man. *Acta Unio Internationalis Contra Cancrum* 1953;**9**:531-41.

29. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* 1991;**133**:144-53.

30. MacMahon B, Pugh JF. Epidemiology. Principles and methods. Boston: Little, Brown and Co., 1970.

31. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 1959;**22**:719-48.

32. Marshall RJ, Jackson RT. Analysis of case-crossover designs. *Statistics in Medicine* 1993;**12**:2333-41.

33. Miettinen OS. Theoretical epidemiology - Principles of occurrence research in medicine. New York: John Wiley, 1985.

34. Miettinen OS. Designs options in epidemiologic research: un up-date. *Scandinavian Journal of Work, Environment and Health* 1982;**9**:7-14.

35. Mittleman MA, Maclure M, Robins JM. Controls sampling strategies for case-crossover studies: an assessment of relative efficiency. *American Journal of Epidemiology* 1995;**142**:91-8.

36. Pearce N. What does the odds ratio estimate in a case-control study? *International Journal of Epidemiology* 1993;**22**:1189-92.

37. Peixoto, M S. Sistemas dinâmicos e controladores fuzzy. Um estudo da dispersao da morte súbita dos citrus no estado de São Paulo. Tese de doutorado. 2005. Campinas, UNICAMP.

38. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* 1994;**13**:153-62.

39. Prentice RL. A case-cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika* 1986;**73**:1-11.

40. Rego MAV. Aspectos históricos dos estudos caso-controle. *Cadernos de Saúde Pública* 2001;**17**:1017-24.

41. Robins JM, Gail MH, Lubin JH. More on biased selection of controls for case-control analysis of cohort studies. *Biometrics* 1986;**42**:1293-9.

42. Rothman KJ. Epidemiology. An introduction. Oxford: Oxford University Press, 2002.

43. Rothman KJ, Greenland S. Modern epidemiology. Philadelphia, PA: Lippincott-Raven, 1998.

44. Rouquayrol MZ, Almeida Filho N. Introdução a epidemiologia moderna, 6a edição. Rio de Janeiro: MEDSI, 2003.

45. Ruppert D, Wand MP, Carroll RJ. Semiparametric regression. Cambridge: Cambridge University Press, 2003.

46. Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *American Journal of Human Genetics* 1993;**53**:1114-26.

47. Schaid DJ, Sommer SS. Comparison of statistics for candidate-gene associations studies using case and parents. *American Journal of Human Genetics* 1994;**55**:402-9.

48. Schlesselman JJ. Case-control studies - Design, conduct, analysis. New York: Oxford University Press, 1982.

49. Self SG, Longton G, Kopecky KJ, Liang KY. On estimating HLA/disease association with application to a study of aplastic anaemia. *Biometrics* 1991;**47**:53-61.

50. Thomas D, Pitkäniemi J, Langholz B, Tuomilehto-Wolf E, Tuomilehto J. Variation in HLA-associated risks of childhood insulin-dependent diabetes in the Finnish population: II. Haplotype effects. *Genetic Epidemiology* 1995;**12**:455-66.

51. Thomas DB. The relationship of oral contraceptives to cervical carcinogenesis. *Obstetrics and Gynecology* 1972;**40**:508-18.

52. Umbach DM, Weinberg CR. Designing and analysing case-control studies to exploit independence of genotype and exposure. *Statistics in Medicine* 1997;**16**:1731-43.

53. Vieira V, Webster T, Aschengrau A, Ozonoff D. A method for spatial analysis of risk in a population-based case-control study. *International Journal of Hygiene and Environmental Health* 2002;**205**:115-20.

54. Vieira V, Webster T, Weinberg J, Aschengrau A, Ozonoff D. Spatial analysis of lung, colorectal, and breast cancer on Cape Cod: an application of generalized additive models to case-control data. Environmental health: a global access science source. 2005. Biomed Central. 2006. [http://www.ehjournal.net/content/4/1/11, accessed on 4 Sep 2006]

55. Woolf B. On estimating the relatiion between blood group and disease. *Annals of Human Genetics* 1955; **19**:251-3.

56. Zaffanella LE, Savitz DA, Greenland S, Ebi KL. The residential case-specular method to study wire codes, magnetic fields, and disease. *Epidemiology* 1998;**9** :16-20.