

Revista de Saúde Pública

JOURNAL OF PUBLIC HEALTH

Atualização

Current Comments

Metodologia caso-controle em epidemiologia de doenças cardiovasculares. II - Análise de dados

Case-control design in cardiovascular disease epidemiology. II - Analysis of data

Suzana Alves de Moraes e José Maria Pacheco de Souza

*Departamento de Enfermagem Materno-Infantil e Saúde Pública da Escola de Enfermagem de Ribeirão Preto da Universidade de São Paulo. Ribeirão Preto, SP - Brasil (S.A.M.),
Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, SP - Brasil (J.M.P.S.)*

Moraes, Suzana Alves de, Metodologia caso-controle em epidemiologia de doenças cardiovasculares. II - Análise de dados. *Rev. Saúde Pública*, **32** (1): 82-8, 1998.

Metodologia caso-controle em epidemiologia de doenças cardiovasculares. II - Análise de dados

Case-control design in cardiovascular disease epidemiology. II - Analysis of data

Suzana Alves de Moraes e José Maria Pacheco de Souza

Departamento de Enfermagem Materno-Infantil e Saúde Pública da Escola de Enfermagem de Ribeirão Preto da Universidade de São Paulo. Ribeirão Preto, SP - Brasil (S.A.M.), Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, SP - Brasil (J.M.P.S.)

Resumo

A larga difusão, nos últimos anos, de softwares estatísticos indicados para análises multivariadas, aliada à facilidade de seu manuseio, poderá conduzir os usuários à obtenção de estimativas enganosas caso não se conheçam os pressupostos teóricos necessários para o emprego destas técnicas. Com o propósito de se destacar alguns desses pressupostos, apresentam-se as etapas executadas para a análise dos dados de um estudo caso-controle, conduzido no Município de São Paulo, entre março de 1993 e fevereiro de 1994, cujo objetivo foi testar a associação do diabetes mellitus com a doença isquêmica do coração após o ajustamento para possíveis variáveis de confusão e/ou modificadoras de efeito. Para uma melhor visualização didática desses pressupostos metodológicos são destacadas quatro etapas fundamentais durante a fase de análise dos dados: a) construção do banco de dados; b) cálculo do poder estatístico; c) categorização e codificação das variáveis e d) a escolha do modelo de regressão logística multivariada.

Estudos de casos e controles. Diabetes mellitus. Isquemia miocárdica.

Abstract

The widespread distribution of statistical software recommended for multivariate analysis as well as the ease in handling it can lead the users into adopting wrong measures if they do not pay attention to the theoretical principles behind those methods. With a view to bringing out some of these principles some steps for the data analysis of a case-control study undertaken in the city of S. Paulo-Brazil from March, 1993 to February, 1994 in order to test the association between diabetes mellitus and ischaemic heart disease after adjusting for potential confounders and/or modifiers of effect are presented. Methodologic issues are emphasized in the development of four steps: a) the data bank structure; b) the calculation of statistical power; c) the definition of variables strata and codification and d) the choice of the logistic regression method.

Case-control studies. Diabetes mellitus. Myocardial ischemia.

Correspondência para/Correspondence to: Suzana Alves de Moraes - Av. Santa Luzia, 440/81 Sumaré - 14025-090 Ribeirão Preto, SP - Brasil. E-mail: samoraes@usp.br

Edição subvencionada pela FAPESP (Processo nº 97/09815-2).

Recebido em 30.10.1996. Reapresentado em 30.6.1997. Aprovado em 19.8.1997.

INTRODUÇÃO

Entre as técnicas estatísticas disponíveis para a análise de dados provenientes de estudos caso-controle, destaca-se a regressão logística multivariada, recomendada quando a variável resposta é um evento binário, ou seja, os indivíduos da amostra são classificados como “casos” ou “controles”.⁹

A possibilidade de cálculo de medidas de associação (odds ratios) ajustadas simultaneamente para o efeito de múltiplas variáveis de confusão e/ou modificadoras de efeito, sem perda de poder estatístico, coloca em destaque as vantagens da regressão logística multivariada sobre técnicas tradicionais como a análise estratificada.

Entretanto, a larga difusão de softwares estatísticos de fácil manuseio tem conduzido ao uso, por vezes indiscriminado, dessas técnicas sem que sejam levados em consideração os pressupostos recomendados para a sua utilização⁵.

Com o objetivo de destacar alguns desses pressupostos, ao se optar pela técnica estatística de regressão logística, os autores apresentam, no presente trabalho, algumas etapas executadas para a análise dos dados de um estudo caso-controle, conduzido no Município de São Paulo, entre 1993 e 1994, onde a variável resposta foi representada pela doença isquêmica do coração (DIC)¹⁴.

CONSTRUÇÃO DOS BANCOS DE DADOS PARA O ESTUDO

O estudo teve por objetivo testar a associação do diabetes mellitus com a doença isquêmica do coração, procedendo-se a ajustamentos para possíveis variáveis de confusão e/ou modificadoras de efeito.

Ao todo, foram entrevistados 900 indivíduos: 300 casos, 300 controles de vizinhança e 300 controles hospitalares. Os registros que contiveram perdas de informações foram excluídos. Após a exclusão desses registros, a amostra remanescente foi composta por 833 indivíduos: 280 casos e 553 controles. Foram compostos dois bancos de dados para que os casos pudessem ser comparados com dois diferentes conjuntos de controles. Na composição “A” 266 casos foram comparados com 266 controles de vizinhança e na composição “B” 280 casos foram comparados com 268 controles hospitalares. Cada banco de dados, representado por um determinado tipo de composição, está apresentado no

Tabela 1. A escolha de mais de um grupo-controle tem sido recomendada com o propósito de se obter medidas de efeito (“odds ratios”) em diferentes composições de controles servindo estas medidas de efeito para avaliar a existência de possíveis vícios de seleção como o “viés” ou “falácia” de Berkson^{6, 10, 16}.

Os bancos de dados foram estruturados no programa FOXPRO vs.2.0 para microcomputador⁷.

Tabela 1 - Caracterização dos bancos de dados do estudo.
Table 1 - Characterization of the data banks of the study.

Composição	Elementos	Nº Casos	Nº Controles	Total
A	Casos e controles de vizinhança	266	266	532
B	Casos e controles hospitalares	280	268	548

CÁLCULO DO PODER ESTATÍSTICO

O poder estatístico para cada uma das composições foi calculado, levando-se em consideração o número de controles por caso e o modelo logístico utilizado, e tendo-se em vista a detecção de “odds ratios” $\geq 2,0$ em nível de 5% monocausal de significância. A escolha de um teste monocausal justificase, neste caso, visto que todo o conhecimento a priori acerca dos mecanismos fisiopatológicos do diabetes mellitus apontam-no como um fator de risco para a doença isquêmica do coração e não como um fator de proteção.

Os valores obtidos para o poder estatístico foram calculados segundo os modelos propostos por Schlesselman¹⁸, considerando-se a prevalência de diabetes (exposição principal), como sendo 10%. Tais valores, em cada composição, encontram-se na Tabela 2 onde também são apresentadas as fórmulas utilizadas para o cálculo.

CATEGORIZAÇÃO E CODIFICAÇÃO DAS VARIÁVEIS

Variável Dependente

A variável dependente (evento-resposta) foi representada pela doença isquêmica do coração (DIC). Aos casos foi atribuído o código “1”. Aos controles foi atribuído o código “0”.

Tabela 2 - Valores de poder estatístico* em cada composição do estudo.**Table 2** - Values of statistical power in each study composition.

Composição	Tipo de análise	Nº Controles por caso	Poder estatístico	Fórmula
A	Condicional	1::1	81%	$Z_{\beta} = \frac{\left[-Z_{\frac{\alpha}{2}} + \sqrt{m \left(P - \frac{1}{2} \right)^2} \right]}{\sqrt{P(1-P)}}$ <p> $m = \text{n}^{\circ}$ pares discordantes $P = R/1+R$ $R = \text{odds ratio}$ que se quer detectar no estudo </p>
B	Não condicional	0,96::1	86%	$Z_{\beta} = [n(p_1 - p_0)^2 \div (1 + \frac{1}{c}) \bar{p}\bar{q}]^{0.5} - Z_{\alpha}$ <p> $n = \text{n}^{\circ}$ de casos do estudo $p_1 = \text{proporção}$ de exposição entre os casos $p_0 = \text{proporção}$ de exposição entre os controles $c = \text{relação:}$ controles/casos $\bar{p} = (p_1 + cp_0) \div (1 + c)$ $\bar{q} = 1 - \bar{p}$ </p>

* O poder estatístico é expresso como $1 - \beta = P(Z \leq Z_{\beta})$

Variáveis Independentes (Variáveis de Exposição)

Variáveis Dicotômicas (expostos x não expostos)

Elenco: *sexo; história de diabetes mellitus* (variável principal); antecedentes familiares de cardiopatia; história de hipercolesterolemia e história de hipertensão arterial sistêmica. A categoria de referência para a variável sexo foi representada pelos indivíduos do sexo feminino. As respectivas categorias de referência para as demais variáveis dicotômicas foram representadas pelos “não expostos”.

Variáveis com mais de Duas Categorias

Elenco: *idade; hábito de fumar e histórico do índice de massa corporal*. A variável idade foi classificada em 4 categorias, com intervalos de 10 anos: “30 a 39”; “40 a 49”; “50 a 59” e “60 a 69” anos. A primeira categoria foi considerada a de referência. O hábito de fumar foi classificado em três categorias: “não fumantes” (categoria de referência) “ex-fumantes” e “fumantes”.

Durante a entrevista foram tomadas medidas relativas ao peso e altura para o cálculo do índice de massa corporal atual (peso/altura²). Os indivíduos foram também questionados sobre seu peso, 10 anos antes da data da entrevista, servindo essa informação para o cálculo do índice de massa corporal no passado. Os índices de massa corporal (IMC) foram

primeiramente classificados em 3 categorias: “normal” (IMC < 25); “sobrepeso” (IMC: 25 A 29,9) e “obeso” (IMC ≥ 30), segundo os critérios recomendados por Bray². A partir desta classificação foi criada a variável BEV, codificada em 4 níveis e indicativa da evolução do índice de massa corporal no decorrer dos 10 anos que antecederam a entrevista. Os indivíduos com IMC normal há 10 anos, atualmente se constituíram na categoria de referência “nível 1”. Os obesos ou com sobrepeso no passado e IMC normal, no presente, representaram o “nível 2”. Os que apresentaram IMC normal no passado e estiveram obesos ou com sobrepeso, no presente, compuseram o “nível 3”. Os indivíduos obesos ou com sobrepeso nos dois momentos representaram o “nível 4” de exposição.

TÉCNICAS ESTATÍSTICAS DE ANÁLISE DOS DADOS: A REGRESSÃO LOGÍSTICA MULTIVARIADA

Seleção do Modelo

A regressão logística multivariada foi a técnica de eleição para a análise dos dados, em virtude de se adequar à necessidade de controle de múltiplas variáveis de confusão, de servir para a identificação de fatores modificadores de efeito e de se ter utilizado, como resposta, um evento binário, a doença isquêmica do coração (DIC).⁹

A existência ou não de relação dose-resposta entre as variáveis independentes com mais de duas categorias e a variável resposta foi testada indistintamente, através da observação dos respectivos “logits” (logarítmos dos “odds ratios” ou coeficientes de regressão) ou das respectivas medidas de efeito (“odds ratios”). Para tanto, optou-se por transformar todas as variáveis com mais de duas categorias em variáveis do tipo indicadora “dummy”, o que também se prestou à obtenção de “odds ratios” específicos para determinadas categorias de exposição, de interesse para o estudo. Pode-se citar, como exemplo, o caso de variáveis como o hábito de fumar, em que a transformação em categorias indicadoras permitiu visualizar a não linearidade dos respectivos “logits”, o que não teria sido possível caso fosse mantida a codificação inicial, em categorias ordinais. As variáveis com mais de duas categorias foram inicialmente codificadas na forma ordinal: 0,1, 2, 3..., conforme os níveis de exposição. Com tal forma de categorização obter-se-ia, através do modelo logístico, um coeficiente médio de regressão, para tais categorias, tendo-se ainda que efetuar um cálculo (e^{bx}) para obtenção dos “odds ratios” em cada nível de exposição da variável. Ao optar-se pela forma “dummy” de categorização, criou-se, na verdade, a possibilidade de obtenção direta de um coeficiente de regressão específico para cada categoria de exposição pois, com tal transforma-

ção, cada estrato foi tratado como se fosse uma nova variável. O número de variáveis “dummy” correspondeu ao número de categorias ordinais de cada variável menos um⁹. Na Tabela 3 encontra-se um exemplo do processo de formação de variáveis “dummy”.

A criação das duas variáveis “dummy” para *hábito de fumar* permitiu a obtenção de “odds ratios” específicos para cada categoria de exposição que foram comparados com a categoria de referência (não fumantes) cujo “odds ratio” =1.

No estudo caso-controle, objeto destas considerações¹⁴, havia interesse em se comparar as medidas de efeito correspondentes a categorias específicas de algumas variáveis, nas duas composições do banco de dados. Embora a transformação em variáveis indicadoras possa ter sacrificado o ajuste dos modelos, pôde-se, através deste procedimento, garantir a informação subjacente em categorias específicas de algumas variáveis, além de se tornar possível a comparabilidade das medidas de efeito nas duas composições do estudo.

No Tabela 4 são apresentados, a título de ilustração, os coeficientes de regressão (β) e os respectivos “odds ratios” ajustados, equivalentes às categorias de exposição da variável *hábito de fumar*, nas composições “A”(casos + controles de vizinhança) e “B”(casos + controles hospitalares). A obtenção de “odds ratios” específicos para cada categoria de fumo permitiu não apenas a comparação das respectivas magnitudes em cada composição, como também entre as duas composições do estudo. A observação de “odds ratios” de maior magnitude para a categoria de ex-fumantes que para a de fumantes, em relação aos não fumantes, foi indicativa da presença de casos prevalentes no estudo, com possíveis modificações do hábito de fumar após um diagnóstico prévio da doença.

Considera-se oportuno salientar que mesmo utilizando-se a forma indicadora (“dummy”) para a codificação de variáveis com mais de duas categorias, a relação entre o número de casos e o número de variáveis independentes foi mantida em todas as

Tabela 3 - Codificação de 2 variáveis dummy* a partir das 3 categorias ordinais da variável hábito de fumar.

Table 3 - Codification of 2 dummy variables on the basis of 3 ordinal categories of the smoking habit variable.

Variável	Categorias ordinais	Variáveis dummy (D)	
		(D1)	(D2)
Háb. fumar			
Não fumante	0	0	0
Ex-fumante	1	1	0
Fumante	2	0	1

* O programa MULTLR¹⁰ oferece uma opção para a formação automática de variáveis “dummy”.

Tabela 4 - “Odds Ratios” ajustados de cada uma das categorias da variável hábito de fumar nas duas composições do estudo.

Table 4 - Adjusted Odds Ratios of each of the categories of the smoking habit variable in each of the study compositions.

Variável	Composição A (casos + controles de vizinhança)		Composição B (casos + controles hospitalares)	
	Coefficiente (β)*	“Odds Ratio” Ajustado*	Coefficiente (β)	“Odds Ratio” Ajustado
Háb. fumar				
Não fumante	0	1	0	1
Ex-fumante	1,0355	2,82	0,8343	2,30
Fumante	0,2975	1,35	0,7395	2,09

*A transformação em categorias “dummy” permitiu o cálculo de coeficientes de regressão (β) e “odds ratios” específicos para cada uma das categorias de exposição da variável hábito de fumar.

composições do estudo como sendo $> 10/1$. A relação mínima de 10 eventos resposta para cada variável independente tem sido recomendada com o propósito de se evitarem distorções nas medidas de efeito provenientes de um superajustamento⁵.

Seleção das Variáveis

A escolha inicial do elenco das possíveis variáveis de confusão esteve fundamentada, em princípio, pelo seu reconhecimento em estudos epidemiológicos anteriores, sendo procuradas, tanto quanto possível, as bases biológicas que explicassem suas associações com a variável resposta e com as demais variáveis de confusão, evitando-se, assim, a inclusão de variáveis intervenientes. Miettinen¹³ cita que “a consideração de determinada variável como de confusão não deve estar baseada apenas em procedimentos mecanicistas, como mudanças nas estimativas de efeito, decorrentes do controle para tal variável, devendo tal seleção, antes de tudo, estar baseada no entendimento de suas relações com as demais variáveis, o que só poderia advir do conhecimento *a priori*”.

Greenland⁸ recomenda que “para maximizar a validade de estimativas baseadas em modelos, o pesquisador deveria usar conjuntamente os próprios dados e as informações prévias, de forma a poder criticamente avaliar as pressuposições epidemiológicas e estatísticas inerentes a estes modelos”.

Após tais considerações, que nortearam a escolha do elenco inicial de possíveis variáveis de confusão, os procedimentos estatísticos adotados para a inclusão ou exclusão definitiva de variáveis nos modelos foram aqueles recomendados por Hosmer e Lemeshow⁹:

- a) Para cada composição foram inicialmente realizadas análises brutas (univariadas), obtendo-se os valores “p” (testes de Wald) correspondentes aos níveis de significância dos coeficientes de regressão. O critério adotado para a seleção das variáveis que deveriam compor os modelos multivariados, a partir das análises brutas, foi a obtenção de valores “p” para os testes de Wald $\leq 0,25$.
- b) O modelo inicial foi composto pelas variáveis que apresentaram valor $p \leq 0,25$ nos testes de Wald e também pelos termos de interação (termos produto), obtidos pelas combinações de interesse entre a variável principal (diabetes mellitus) e as demais variáveis.
- c) A estatística utilizada para a exclusão definitiva de variáveis, a partir dos modelos multivariados, foi a razão de verossimilhança (RV),

utilizando-se níveis de significância $\leq 0,05$, por ser a RV menos conservadora que o teste de Wald, que algumas vezes pode falhar em rejeitar H_0 ^{9,11}. Isto significa que os coeficientes de regressão (β) de algumas variáveis podem apresentar valores “p” descritivos nos testes de Wald $> 0,05$ (β não significantes) sinalizando para a possibilidade de exclusão dessas variáveis dos modelos, enquanto tal exclusão não será permitida quando utilizado o teste da razão de verossimilhança. Esta constatação indica que a estatística de Wald se constitui num bom teste durante a triagem inicial das variáveis (análises univariadas), servindo para apontar, nesta etapa, quais as variáveis que deverão compor os modelos multivariados. Uma vez composto o elenco de variáveis para os modelos multivariados, o critério de exclusão a partir de então deverá estar baseado no valor obtido para a razão de verossimilhança.

Considera-se oportuno destacar a importância de se realizar inicialmente os testes de interação (modelo inicial), antes da exclusão, *a priori*, de algumas variáveis, pois, segundo Hosmer e Lemeshow⁹, algumas delas, consideradas não significantes para compor os modelos multivariados, podem apresentar interação ao serem combinadas com outras variáveis do estudo. Esta situação encontra suporte, sobretudo, em situações onde a interação é plausível do ponto de vista biológico. Por outro lado destaca-se que a modificação de efeito (interação) foi testada por modelos estatísticos multiplicativos que não permitem conclusões sobre interações biológicas de natureza aditiva e que poderiam existir entre tais variáveis¹⁹.

A análise de regressão logística multivariada, em cada composição, foi, portanto, realizada em três etapas, após as análises univariadas: 1) modelo inicial, contendo os termos de interação; 2) modelo sem os termos de interação e 3) modelo final. O software estatístico utilizado foi o “MULTLR”⁴.

Para informações mais detalhadas acerca dos resultados do estudo recomenda-se consultar Moraes e Souza^{14, 15}.

Avaliação de Multicolinearidade

Diversos autores têm chamado a atenção para a distorção das medidas de efeito quando as variáveis independentes mantêm entre si uma relação linear (multicolinearidade)^{1,5,9,11,12}. Kleinbaum-Kupper-Muller¹¹, a propósito de modelos de regres-

são múltipla, destacam que as estimativas dos coeficientes de regressão em tais modelos são equivalentes a:

$$q_j = c_j \left[\frac{1}{1 - r^2 (X_j, X_j)} \right]$$

q_j = coeficiente de regressão

c_j = constante calculada a partir dos dados

r = valor da correlação

X = variáveis independentes

Sob a condição de multicolinearidade, $r^2 = 1$, donde a estimativa q_j do coeficiente de regressão de tal variável seria indeterminada (denominador = 0). Embora o cálculo apresentado acima seja relativo a estimativas de coeficientes em análise de regressão múltipla, Hosmer e Lemeshow⁹ apontam a existência de cálculos similares para modelos de regressão logística, destacando que, nesses modelos, o emprego de testes para multicolinearidade apenas seriam indicados quando a degradação dos modelos ficasse evidente, o que poderia ser detectado a partir da observação de valores aberrantes de " β " e, principalmente, na grande variabilidade dos respectivos erros-padrão. Concato e col.⁵ reconhecem, por outro lado, que, em algumas situações, as variáveis podem passar por tais testes, mesmo na presença de multicolinearidade.

No estudo caso-controle, ao qual se referem estas considerações, não foram realizados testes estatísticos para multicolinearidade. Entretanto, a observação das estimativas obtidas para os coeficientes de regressão, respectivos erros-padrão e intervalos de confiança não foram indicativos da existência de multicolinearidade entre as variáveis independentes¹⁴.

Regressão Logística Condicional e Não Condicional

Os casos do estudo foram emparelhados por sexo e faixa etária com dois tipos de controles: de vizinhança e hospitalares. A composição "A", conforme citado anteriormente, foi constituída pelos casos e respectivos controles de vizinhança e a composição "B" pelos casos e controles hospitalares. O emparelhamento por sexo e faixa etária (emparelhamento por frequência) é considerado, segundo Schlesselman¹⁸, um emparelhamento "grosseiro", visto que "qualquer controle" desde que do mesmo sexo e faixa etária, está apto a compor um par com

o respectivo caso. Esta situação, segundo o autor, difere do emparelhamento "fino", como quando consideradas variáveis como "parentesco" ou "vizinhança" que restringem a "indivíduos específicos" a aptidão para compor os pares com os respectivos casos. Sob esta ótica, poder-se-ia dizer que na composição "A" (casos + controles de vizinhança) procedeu-se a um emparelhamento "fino" entre casos e controles.

O emparelhamento por sexo e faixa etária foi procedido, no desenho do estudo, para efeito de agrupamento (obtenção de equilíbrio, entre casos e controles, para sexo e faixa etária). A variável "matching" de interesse foi a "vizinhança", sendo recomendado, apenas sob estas condições, o emprego da análise de regressão logística condicional^{10,18}.

Com base nestes pressupostos, procedeu-se à análise de regressão logística condicional para a composição "A" e à análise não condicional para a composição "B".

Por outro lado, o emprego inadvertido da análise não condicional de dados submetidos a um emparelhamento "fino" pode gerar medidas de efeito superestimadas sobretudo quando o número de controles por caso é pequeno³.

Tabela 5 - "Odds Ratios" ajustados da variável diabetes, segundo os modelos não condicional e condicional de regressão logística multivariada para os dados da composição "A".

Table 5 - Adjusted Odds Ratios of the diabetes variable, by nonconditional and conditional models of multivariate logistic regression for the data of composition "A".

Variável	Análise não condicional		Análise condicional	
	OR	IC (95%)	OR	IC (95%)
Diabetes mellitus				
Não diabéticos	1		1	
Diabéticos	2,61	1,18-5,77	1,25	0,51-3,07

OR = "Odds Ratio"

IC = Intervalo de Confiança

A título de ilustração, na Tabela 5 estão apresentadas as medidas de efeito ajustadas da variável diabetes, obtidas por análise não condicional e condicional dos dados relativos à composição "A" (casos + controles de vizinhança). Torna-se oportuno destacar que o uso inapropriado da análise não condicional para uma composição onde os casos e os controles foram submetidos a um emparelhamento "fino" poderia conduzir à conclusão de que o diabetes é um fator de risco independente para a doença isquêmica do coração, o que foi claramente refutado a partir dos resultados da análise condicional.

CONSIDERAÇÕES FINAIS

Pretendeu-se enfatizar a importância de se levar em conta os pressupostos metodológicos recomendados para o emprego da regressão logística multivariada. Não tendo por objetivo uma exposição detalhada das técnicas estatísticas relacionadas aos métodos multivariados, insistiu-se em chamar a atenção dos potenciais usuários de softwares estatísticos para o fato de que a relativa simplicidade quanto ao manuseio desses programas poderá conduzir a estimativas enganosas, caso não se conheçam os princípios metodológicos relacionados à eleição e condução adequada dessas técnicas.

Além de recomendações relacionadas à correta condução da fase de delineamento de estudos caso-controle^{16,17,20}, o emprego adequado de técnicas estatísticas para a análise dos dados é de importância fundamental para que se possa garantir a validade das estimativas.

AGRADECIMENTOS

Aos relatores, pelas sugestões e comentários que muito contribuíram para a versão final do presente trabalho.

REFERÊNCIAS

1. ARMITAGE, P. & BERRY, G. *Statistical methods in medical research*. Oxford, Blackwell Scientific Publications, 1987.
2. BRAY, G.A. Obesity: basic considerations and clinical approaches. *Dis. mon.*, **35**:454-537, 1989.
3. BRESLOW, N.E. & DAY, N.E. *Statistical methods in cancer research*. Lion, Interation Agency of Research on Cancer, 1980. v. 1. (IARC Scientific Publication, 32).
4. CAMPOS-FILHO, N. & FRANCO, E.L. MULTLR-A microcomputer program for multiple logistic regression by unconditional and conditional maximum likelihood methods. *Am. J. Epidemiol.*, **129**:439-44, 1989.
5. CONCATO, J.; FEINSTEIN, A.R.; HOLFORD, T.R. The risk of determining risk with multivariable models. *Ann. Intern. Med.*, **118**:201-10, 1993.
6. FLANDERS, W.D.; BOYLE, C.A.; BORING, J.R. Bias associated with differential hospitalization rates in incident case-controls studies. *J. Clin. Epidemiol.*, **42**:395-401, 1989.
7. FOX HOLDINGS INC. FOXPRO vs 2.0. C/S. Pat. Pend. American Edition, USA, 1991.
8. GREENLAND, S. Modeling and variable selection in epidemiologic analysis. *Am. J. Public Health*, **79**:340-9, 1989.
9. HOSMER, D.W. & LEMESHOW, S. *Applied logistic regression*, New York, John Wiley and Sons, 1989.
10. KLEINBAUM, D.G.; KUPPER, L.L.; MORGENSTERN, H. *Epidemiologic research: principles and quantitative methods*. London, Lifetime Learning, 1982.
11. KLEINBAUM, D.G.; KUPPER L.L.; MÜLLER, K.E. *Applied regression analysis and other multivariable methods*. Boston. PWS-KENT Publishing Company, 1988.
12. Mc GEE, D.; REED D.; YANO, K. The results of logistic analysis when the variables are highly correlated: an empirical example using diet and CHD incidence. *J. Chron. Dis.*, **37**:713-9, 1984.
13. MIETTINEN, O.S. & COOK, E.F. Confounding: essence and detection. *Am. J. Epidemiol.*, **114**:593-603, 1981.
14. MORAES, S.A. & SOUZA, J.M.P. Diabetes mellitus e doença isquêmica do coração: um estudo tipo caso-controle. *Rev. Saúde Pública*, **30**:364-71, 1996.
15. MORAES, S.A. & SOUZA, J.M.P. Efeito dose-resposta de fatores de risco para a doença isquêmica do coração. *Rev. Saúde Pública*, **30**:471-8, 1996.
16. MORAES, S.A. & SOUZA, J.M.P. Metodologia caso-controle em epidemiologia de doenças cardiovasculares. I- Delineamento. *Rev. Med. Ribeirão Preto*, **29**:460-7, 1996.
17. ROTHMAN, K.J. *Modern epidemiology*. Boston, Little Brown and Company, 1986.
18. SCHLESSELMAN, J.J. *Case control studies-design, conduct and analysis*. New York, Oxford, University Press, 1982.
19. SIEMIATYCKI, J. & THOMAS, D.C. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int. J. Epidemiol.*, **10**:383-7, 1981.
20. WACHOLDER, S.; McLAUGHLIN, J.K.; SILVERMAN, D.T.; MANDEL, J.S. Selection of controls in case-control studies. I-Principles. *Am. J. Epidemiol.*, **135**:1019-28, 1992.