

Validade Interna em Estudos de Corte Transversal: Reflexões a Partir de uma Investigação sobre Esquistossomose Mansônica e Condições Socioeconômicas

Internal Validity in Cross-Sectional Studies: Comments Based on an Investigation on the Association Between Socioeconomic Factors and Schistosomiasis

Ricardo A. de A. Ximenes¹

Thalia V. B. de Araújo²

XIMENES, R. A. A. & ARAÚJO, T. V. B. *Internal Validity in Cross-Sectional Studies: Comments Based on an Investigation on the Association Between Socioeconomic Factors and Schistosomiasis*. *Cad. Saúde Públ.*, Rio de Janeiro, 11 (1): 118-127, Jan/Mar, 1995.

The purpose of this paper was to make some comments on validity based on an investigation developed in an urban area in the Greater Metropolitan Area of Recife, Brazil, the objective of which was to study the association between socioeconomic factors and schistosomiasis. It discusses the distortion in the odds ratio that may be introduced either by the way individuals are selected for the study or by the way the information is collected. The article describes the mechanisms used by the authors to avoid bias when planning or conducting the investigation and considerations were made on how these types of bias could have influenced the results.

Key words: Schistosomiasis; Epidemiology; Research Methodology; Internal Validity; Cross-sectional Studies

INTRODUÇÃO

A interpretação dos resultados de uma pesquisa é uma preocupação importante da rotina de trabalho dos pesquisadores. Uma questão central é evitar conclusões espúrias que podem derivar de problemas metodológicos ligados, seja ao desenho do estudo, seja a análise. Costuma-se considerar duas fontes principais de erro, erro randômico e erro sistemático. O primeiro aparece como uma imprecisão da estimativa obtida a partir do estudo em relação ao parâmetro que está sendo avaliado,

sendo resultante da variação amostral. É, portanto, fundamentalmente dependente do tamanho da amostra e de características específicas do parâmetro a ser estimado, ou seja, sua variância. O erro sistemático, viés, ou bias como referido pelos autores de língua inglesa, difere em sua natureza do erro randômico, por ter sua origem na existência de uma diferença entre o parâmetro a ser estimado e o verdadeiro efeito que se quer medir. É um problema de validade e não de precisão da estimativa usada em uma investigação. O seu surgimento está relacionado à forma pela qual os indivíduos são selecionados ou a maneira pela qual as informações são obtidas, registradas ou interpretadas em um estudo.

Sendo assim, erro sistemático pode resultar, nos estudos analíticos, em uma estimativa incorreta da associação entre a exposição e a doença, constituindo então, uma possível explicação alternativa a uma associação encontrada (Kleinbaum et al., 1982).

¹ Departamento de Medicina Tropical, Universidade Federal de Pernambuco. Hospital das Clínicas. Avenida Professor Moraes Rego, s/n, Bloco A, térreo. Cidade Universitária. Recife, PE, 50670-420, Brasil.

² Faculdade de Ciências Médicas de Pernambuco, Universidade de Pernambuco. Rua Arnóbio Marques, 310, Recife, PE, 53100-130, Brasil.

A analogia que habitualmente é feita com o “tiro ao alvo” parece bastante ilustrativa. A validade diz respeito ao fato do indivíduo estar ou não atirando no alvo correto, enquanto que a precisão refere-se a variação individual de tiro para tiro quando o alvo correto está sendo considerado.

Diferentes sistemas de classificação já foram utilizados por diversos autores; por sua simplicidade, utilizar-se-á a nomenclatura proposta por Rothman (1986), na qual os erros sistemáticos são categorizados em: bias de seleção, quando referente a identificação dos indivíduos que comporão a população de estudo e o bias de informação, que inclui qualquer erro sistemático relacionado a obtenção de informação sobre a exposição ou o efeito (doença). O terceiro tipo de erro sistemático, o confundimento, que consiste no efeito de uma terceira variável, denominada variável de confusão, sobre a associação entre a doença e o fator de risco estudado, por suas especificidades não será abordado neste texto. Pretende-se aqui concentrar a discussão na questão de validade, tomando como referencial um fragmento de um trabalho desenvolvido na área urbana de São Lourenço da Mata, na Região Metropolitana do Recife. Para responder aos diferentes objetivos do trabalho, realizou-se dois tipos de estudo, um corte transversal e um caso-controle de base populacional (Ximenes, 1991).

As reflexões sobre validade neste texto serão restritas ao estudo de corte transversal e focalizadas na apreciação, quanto a possíveis distorções, em uma medida particular de associação, ou efeito, a Razão dos Produtos Cruzados. Limitam-se, também, os comentários ao problema de validade interna, ou seja, a validade das inferências para a população alvo, usando informações obtidas de uma amostra. A validade externa, que diz respeito a inferência para além da população alvo, e que pressupõe a generalização dos achados, a partir de diferentes premissas, não será trabalhada nesse momento.

O estudo transversal foi realizado para medir a prevalência de esquistossomose entre as famílias com indivíduos no grupo etário de 10 a 25 anos e para investigar a associação entre fatores socioeconômicos e esquistossomose. A hipótese de trabalho foi de que a esquistossomose naquela população não ocorria de forma aleató-

ria, mas que a estrutura econômica e social produzia e condicionava a sua distribuição. A determinação socioeconômica foi trabalhada em duas diferentes formas. Primeiro, usando indicadores socioeconômicos. Esse tipo de abordagem, embora reduza uma realidade social complexa a um número de fatores que por si só não expressam a complexidade do processo social que está ocorrendo, pode mostrar a associação entre esses indicadores e esquistossomose e identificar subgrupos com maior risco de infecção. Segundo, apresentando uma visão mais ampla do processo social, caracterizando a posição que o chefe da família tem no processo de produção. Essa abordagem permitiu fazer uma aproximação de uma estratificação social na área de estudo; considerou-se que a inserção de cada grupo no processo produtivo originaria diferenciais em termos de esquistossomose.

Na análise, quando os dados dos grupos sociais ou classes foram relacionados com o risco de esquistossomose, a informação utilizada para caracterizar a estratificação social era relativa ao chefe da família; considerando esse aspecto e o fato de que havia uma agregação familiar dos casos (testada através da técnica proposta por Smith & Pike (1976)], optou-se por assumir a família como unidade de análise. A regressão logística incondicional foi usada na análise dos dados. Apesar de que as aplicações iniciais do modelo logístico envolviam a análise de dados dos estudos de coorte, ele pode também ser aplicado diretamente para análise de dados dos estudos de caso-controle e dos cortes transversais. Os *odds ratios* foram estimados com relação aos níveis de referência e foram calculados os intervalos de confiança e os valores de *p* para os *odds ratios*, além da razão de verossimilhança.

Quatro entrevistadores foram responsáveis pela aplicação dos questionários e o diagnóstico da esquistossomose foi feito, utilizando-se o método de Kato (Kato & Miura, 1954), modificado por Katz et al. (1972).

Os estudos de prevalência, ou corte transversal, visam estimar a frequência de evento em uma população num determinado período de tempo e caracterizar grupos de risco. Nesses a população de estudo consiste em uma amostra representativa selecionada de uma população alvo, a partir da qual os indivíduos são examinados,

observados e/ou inquiridos sobre a sua doença, sobre características e exposições, atuais ou anteriores, e outras variáveis de interesse. A exposição atual pode ser utilizada como uma aproximação da exposição relevante quando se tem uma fundamentação para acreditar que existe uma boa correlação entre esta e a exposição prévia, assumindo-se que a memória referente à exposição no passado não é confiável (Rothman, 1986).

Esses estudos são freqüentemente interpretados como diferentes dos estudos de coorte e de caso-controle, mas podem ser compreendidos como o caso-controle análogo ao coorte de base geográfica, onde a série de casos é constituída, a partir dos casos prevalentes identificados no estudo e os controles dos demais indivíduos.

Nos estudos de corte-transversal, a taxa de prevalência é a medida utilizada para estimar a freqüência da doença e fazer comparações desta entre subgrupos da população. Pode ser analisada através da construção de razões entre essas taxas nas diferentes categorias de exposição, o que denomina-se de Razão de Prevalência. Essa medida, à semelhança da medida obtida nos estudos de coorte: o risco relativo, razão entre ta-

xas de incidência, expressaria o risco associado a exposição. Outra alternativa seria a utilização da Razão dos Produtos Cruzados (RPC), *odds Ratio*, razão entre as taxas de exposição dos casos e dos controles, medida utilizada como estimativa do risco relativo nos estudos de caso-controle. A RPC, como estimativa do risco, tem a vantagem de não se alterar mesmo quando da existência de uma variação no número de não-casos, desde que não haja alteração na proporção de exposição entre os controles (Hennekens & Buring, 1987).

A utilização da RPC na estimativa do risco relativo ficou restrita, até recentemente, a doenças com baixa prevalência, ou seja, raras, o que poderia sugerir a sua inaplicabilidade aos estudos de corte transversal, voltados a investigação de doenças não-raras, contudo, essa limitação imposta ao seu uso tem sido questionada por diversos autores (Greenland & Thomas, 1982; Rodrigues & Kirkwood, 1990).

A Tabela 1 apresenta o cálculo da RPC e RP, em uma situação simplificada onde a variável independente, potencial fator de risco, e a variável dependente, a doença, são representadas como variáveis dicotômicas, isto é, com apenas duas categorias (presente, ausente).

TABELA 1. Distribuição dos Indivíduos com Respeito a Presença ou Ausência de Exposição e Doença

	Doença		
	Presente	Ausente	
Fator de Risco Presente	a	b	a+b
Fator de Risco Ausente	c	d	c+d
	a+c	b+d	N (a+b+c+d)

$$RPC = a/c / b/d = axd / cxb$$

$$RP = a/a+b / c/c+d$$

BIAS DE SELEÇÃO

O bias de seleção ocorre quando a população acessível, a partir da qual é retirada uma amostra representativa, difere da população alvo, a qual se deseja inferir os resultados, quanto à freqüência relativa dos parâmetros (exposição ou efeito) em estudo; sendo a RPC calculada como ad/bc , como apresentado na Tabela 1, se a população acessível mantém, em cada uma das células A',

B', C' e D', a mesma freqüência relativa que existe para os quatro grupos na população alvo, nenhum bias será introduzido na estimativa dessa medida. No entanto, caso aconteça, por exemplo, que na população acessível, haja, relativamente, uma maior representação dos "casos expostos e dos não-casos não-expostos" do que "dos não-casos expostos e os casos não-expostos", haverá um superdimensionamento da magnitude do *odds ratio*.

O fundamental para a compreensão do bias de seleção é, conseqüentemente, o entendimento dos possíveis mecanismos através dos quais as células de uma tabela 2X2 podem ser alteradas, em sua proporcionalidade, quando se passa da população alvo para a população acessível.

Kleinbaum et al. (1981) denomina como alfa a probabilidade (ou freqüência relativa) dos indivíduos doentes e expostos existentes na população alvo (A) pertencerem a este mesmo grupo na população acessível A' ($\alpha = A'/A$); beta, gama e delta corresponderiam a essas mesmas probabilidades para as células B, C e D, respectivamente. A existência de bias de seleção na estimativa da RPC dependeria, então, da possibilidade do produto alfa x gama / beta x delta diferir de 1, ou seja, de o *odds* de seleção para os doentes (α/β), diferir do *odds* de seleção para os não-doentes (gama/delta), ou ainda, quando os indivíduos são selecionados, a partir da exposição, do *odds* de seleção para os expostos (α/gama) diferir do *odds* de seleção para os não-expostos (β/delta). Com relação a razão de prevalência poder-se-ia afirmar, de forma genérica, que não haverá distorção na sua estimativa quando o *odds* de seleção para os expostos (α/gama) e para os não-expostos (β/delta) forem ambos iguais a 1.

Entre os problemas que podem produzir essa distorção estão imperfeições no desenho do estudo, especialmente relativos a escolha dos grupos a serem comparados (em todos os tipos de estudo) e a escolha do *sampling frame* (notadamente nos estudos de caso-controle e nos estudos transversais); perda de indivíduos durante o acompanhamento nos estudos de coorte ou falha na obtenção de informações devido a não-resposta ou sobrevivência seletiva em estudos de caso-controle ou transversais. Existe, ainda, a possibilidade de que o procedimento empregado para identificar a doença varie de acordo com a situação da exposição (Kleibaum et al., 1982; Abramson, 1990).

A identificação de doentes, casos, e não-doentes, controles, a partir de uma amostra randômica obtida da população a qual se pretende inferir os resultados, implica na construção ou identificação de uma listagem ou mapa de referência, *sampling frame*, que a presente de forma adequada e permita opera-

cionalizar o processo de amostragem. O cumprimento deste requisito se faz importante não só para evitar erros na estimativa de taxas de prevalência mas é crucial para evitar bias de seleção na estimativa dos fatores de riscos a ela associados.

Considerando-se as observações acima é importante revisar o processo de amostragem adotado no estudo. A cidade de São Lourenço da Mata está dividida em doze microáreas, Unidades Espaciais de Referência (UERs), compostas a partir da fusão de setores censitários, para as quais, existia uma listagem dos domicílios disponível no Cadastro da Prefeitura Municipal. O cadastro de domicílios por UER excluía as áreas de invasão e um conjunto habitacional, o Parque Capibaribe.

A estratificação foi realizada para garantir que as diferentes áreas da cidade estariam adequadamente representadas na mostra. Aceita-se que uma amostra randômica estratificada com fração constante de amostragem tende a ter mais precisão do que uma amostra randômica simples, além de ser geralmente mais conveniente sob o ponto de vista operacional. O ganho em precisão deve-se ao fato de que a variação entre os estratos da população é exatamente refletida na amostra (Moser & Kalton, 1985).

O ponto a ser discutido aqui seria se a exclusão dessas áreas na amostra produziria um viés de seleção em alguma direção. O primeiro aspecto a ser considerado refere-se ao contingente populacional ali residente. Essas áreas correspondiam a uma fração importante das habitações e, conseqüentemente, da população de São Lourenço da Mata: 8,28% para as áreas de invasão e 21,16% para o Parque Capibaribe. Acrescente-se que os indivíduos ali residentes representavam diferentes situações em termos de condições de vida, aqueles que moravam nas áreas de invasão sendo particularmente desfavorecidos, enquanto que, os residentes na outra área apresentavam uma condição de vida relativamente privilegiada, tendo-se como referência o padrão local. Considerando o exposto, dificilmente seria obtida uma amostra representativa se essas áreas fossem deixadas a parte.

Uma vez que uma das propostas do trabalho era estudar a associação entre a esquistossomose e alguns indicadores socioeconômicos, a não-inclusão dessas áreas poderia

levar a uma sub-representação de uma ou mais categorias das variáveis independentes. Apenas para facilitar o raciocínio, restringir-se-á os comentários a uma situação hipotética em que tem-se uma única variável independente e dicotômica, por exemplo, renda familiar e uma variável dependente, também dicotômica, esquistossomose. Renda, expressa em salários mínimos, poderia ser categorizada em ≤ 2 salários mínimos e > 2 sm; esquistossomose categorizada em positivo (indivíduos com parasitológico positivo) e negativo. Considere-se então duas possibilidades:

1. A exclusão do Parque Capibaribe e das “áreas de invasão” implicaria na subrepresentação dos dois níveis de renda na amostra, porém, em cada um dos níveis, a proporção de indivíduos positivos seria a mesma das respectivas categorias na população de origem. Isso equivale a dizer que em uma tabela 2X2, a exemplo da Tabela 1, os totais das linhas (a+b e c+d correspondendo, respectivamente, aos dois níveis da variável independente – renda) estariam desigualmente representados, mas, as duas células que compõem cada uma das linhas guardariam com os totais das linhas a mesma proporcionalidade existente na população alvo. Se essa proporcionalidade é mantida isso significa que $\alpha \times \delta / \beta \times \gamma$ seria igual a 1.

Na situação acima descrita a estimativa da Razão dos Produtos Cruzados, ad/bc , não seria distorcida, porém o número total de indivíduos infectados na população seria subestimado, admitindo-se a existência de uma associação entre as duas variáveis. Caso fosse utilizada para a estimativa do risco a Razão de Prevalência esta, também, não seria alterada.

2. A inclusão do Parque Capibaribe e da “área de invasão” poderia implicar na super-representação dos indivíduos nas duas categorias de renda na amostra, porém, como na situação anterior, sem que a proporção de indivíduos positivos em cada um dos níveis de renda fosse diferente daquela existente na população de origem. Isso poderia ocorrer pelo fato de que na formação do estrato “Parque Ca-

pibaribe” e do estrato “área de invasão” empregou-se um critério diferente daquele utilizado na composição das UERs. Isso poderia implicar torná-los mais homogêneo com relação as condições de vida do que as demais áreas e, ao proceder-se a estratificação, resultar na super-representação dos níveis de renda predominantes em uma ou outra área. Não é provável que isso tenha acontecido, as casas pertencentes ao Parque Capibaribe, a semelhança das UERs, estão situadas em uma mesma área geográfica. Não é, porém, esta a situação das casas construídas nas áreas de invasão que estão distribuídas por toda a cidade. Tomando-as como uma única área, assume-se que existem similaridades nas condições socioeconômicas dos indivíduos, morando nas diversas áreas de invasão como no interior das demais UERs selecionadas, para as quais foi utilizado o critério de homogeneidade socioeconômica na sua composição. Embora não houvessem disponíveis dados específicos que respaldassem esse pressuposto, baseou-se o pesquisador em sua própria impressão, na discussão da questão com os residentes e com a administração local. Não sendo esta a realidade, ao assumir os domicílios construídos nas áreas de invasão como correspondendo a uma única área, haveria uma representação excessiva da população de menor poder aquisitivo na amostra, mas esta não distorceria a associação entre pobreza, medida por meio da renda familiar, e esquistossomose. A situação, na realidade corresponderia ao inverso da descrita no item anterior no que se refere a estimativa do número total de indivíduos infectados que estaria superestimado; mas, seria semelhante ao item anterior, na medida em que as duas células que compõem cada uma das linhas guardariam com o total da linha a mesma proporcionalidade existente na população alvo. Caso essa proporcionalidade não fosse mantida, aí então, poderia vir a ocorrer uma distorção na estimativa da RPC, ou mesmo da Razão de Prevalência.

Os problemas descritos como resultantes da supressão de parcelas da população alvo do processo amostral podem emergir em um estudo epidemiológico como decorrência da não-participação, ou recusa, de indivíduos previamente selecionados, o que denomina-se de bias de não-resposta (Criqui, 1979).

A recusa a participação ou a não-resposta pode ser outra fonte de bias de seleção. O percentual de recusas com relação ao questionário ou com relação ao parasitológico de fezes, no estudo, foi bastante reduzido. Focaremos a atenção, então, na falta de informação concernente a um item específico do questionário, com respeito ao salário do chefe da família.

A informação relativa ao salário do chefe da família foi um dos atributos utilizados para uma operacionalização de classe social que permitisse apreender as características mais importantes da estrutura social de São Lourenço da Mata, a partir de uma perspectiva econômica, sendo um dos critérios usados para fracionar os **empregados** em três diferentes grupos: burguesia gerencial, proletariado e subproletariado; e para distinguir, entre os **autônomos** o subproletariado do proletariado e dos trabalhadores não-proletários. Naturalmente esses indivíduos não puderam ser enquadrados nessas categorias e como representavam uma fração importante da amostra (24%), a exclusão dos mesmos dessa etapa da análise mereceu uma maior discussão.

O perigo de bias devido a não-resposta recai na possibilidade de que os não-respondentes difiram significativamente dos respondentes. A medição da magnitude do bias devido a não-resposta depende da possibilidade de obtenção da informação sobre os não-respondentes que viabilizem a comparação entre respondentes e não-respondentes. Nesse estudo, como a omissão de resposta referia-se a um item específico: o salário do chefe da família, estando disponível a informação relativa às demais variáveis, incluindo algumas com forte conexão com o salário, comparou-se os empregados e autônomos, para os quais dispunha-se da informação sobre o salário com aqueles para quem a informação não foi obtida, com relação aos seguintes fatores: posição na produção, setor econômico, nível de educação, posses no interior da casa, tipo de habitação e abastecimento d'água. Um teste de qui-quadrado foi utilizado para testar se a distribuição dos indivíduos entre os diferentes níveis das variáveis socioeconômicas ocorria independentemente da inclusão, ou exclusão, dos não-respondentes da classificação de classe social, devido a falta de informação quanto ao salário. A título de exemplo apresentaremos apenas as tabelas relativas a setor econômico e posses no interior da casa (Tabelas 2 e 3). Os resultados indicam que não existe associação entre as variáveis: setor econômico e posses no interior da casa e a não-resposta.

TABELA 2. Comparação entre os Chefes de Famílias Incluídos e Excluídos da Classificação de Classe Social, por Setor Econômico

Grupos	Setor Econômico							
	Indústria		Terciário		Agricultura		Total	
	Frequência	Percentual	Frequência	Percentual	Frequência	Percentual	Frequência	Percentual
Incluídos	271	36,97	440	60,03	22	3,00	733	100,00
Excluídos	151	37,19	243	59,85	12	2,96	406	100,00

X=0,007, 2 g1, p>0,05

TABELA 3. Comparação entre os Chefes de Família Incluídos e Excluídos da Classificação de Classe Social, por Posse de Bens de Consumo Duráveis (F = Frequência; % = Percentual)

	Posse de Bens Duráveis									
	Nível 1		Nível 2		Nível 3		Nível 4		Total	
	F	%	F	%	F	%	F	%	F	%
Incluídos	44	6,00	183	24,97	438	59,75	68	9,28	733	100,00
Excluídos	12	2,89	108	26,02	250	60,24	45	10,84	415	100,00

X = 6,05, 3 gl, p > 0,05

Resultados semelhantes (não-estatisticamente significantes) foram obtidos para as demais variáveis, sugerindo que os empregados e autônomos que foram excluídos da classificação de classe social, devido a falta de informação sobre o salário, não diferiam significativamente daqueles que foram incluídos. As variáveis testadas foram aquelas utilizadas para definir classe social, ou variáveis estreitamente vinculadas ao salário ou, ainda, variáveis que haviam demonstrado uma forte associação com esquistossomose.

Buscando-se ter maior segurança, a exclusão dos não-respondentes não introduziu um viés importante; testou-se se a associação entre esquistossomose e as variáveis socioeconômicas

era semelhante nos dois grupos. Com essa finalidade foram selecionadas duas variáveis relacionadas com salário: posses de bens de consumo duráveis e tipo de habitação, e a informação sobre a família foi transformada em dados individuais de forma a permitir testar a prevalência de esquistossomose em cada um dos níveis das ditas variáveis escolhidas. A apresentação dos resultados encontrados limita-se, aqui, aos resultados relativos à primeira variável (Tabela 4).

O teste de qui-quadrado utilizado para comparar a prevalência, em cada um dos níveis, entre os dois grupos (incluídos e excluídos) mostra que para nenhum dos níveis a diferença foi estatisticamente significativa.

TABELA 4. Prevalência da Esquistossomose por Categoria de Posses de Bens de Consumo Duráveis para Aqueles Incluídos e Excluídos da Classificação de Classe Social

	Incluídos		Excluídos	
	Frequência	Prevalência	Frequência	Prevalência
Nível 1*				
Positivo	47	55,3%	12	48,0%
Negativo	38		13	
Total	85		25	
Nível 2*				
Positivo	141	35,9%	71	30,0%
Negativo	252		166	
Total	393		237	
Nível 3*				
Positivo	215	21,8%	130	21,6%
Negativo	772		472	
Total	987		602	
Nível 4*				
Positivo	22	14,7%	13	0,8%
Negativo	128		81	
Total	150		94	

*p>0,05

Em resumo, como constatou-se que o grupo de não-respondentes (com relação a salário) era similar ao grupo de respondentes com respeito aos outros indicadores socioeconômicos e como a prevalência em cada categoria das variáveis socioeconômicas testadas era semelhante em ambos os grupos, não há razão para acreditar que a não-resposta quanto ao salário tenha distorcido os resultados em qualquer direção.

BIAS DE INFORMAÇÃO

Bias de informação resultam de diferenças sistemáticas na forma pela qual os dados sobre a exposição ou doença são obtidos. Podem ocorrer, entre outros motivos, devido a um registro inadequado das informações ou a alguma forma de influência do entrevistador sobre o respondente, quer por características da personalidade do primeiro, quer por inadequações na formulação das questões. Pode também estar relacionada a fatores ligados a memória dos fatos por parte do entrevistado; assim indivíduos doentes podem lembrar-se ou referir-se a fatos passados de forma diferente dos não-doentes. Se, por algum desses motivos, os dados são inaccurados ou incompletos, isso pode implicar no aparecimento de associações espúrias; na realidade elas ocorrerão na medida em que tais problemas afetem de forma diferente os dois grupos, isto é, expostos e não-expostos ou doentes e não-doentes (Hennekens & Buring, 1987).

Os problemas ligados a entrevista podem não ser facilmente identificáveis durante o desenvolvimento da investigação e devem portanto merecer uma atenção especial durante a fase de planejamento. Com o objetivo de minimizar a possibilidade de que viessem a interferir nos resultados do estudo em questão foi-se particularmente cuidadoso quanto a escolha dos entrevistadores; selecionou-se indivíduos de sexos opostos (dois homens e duas mulheres) e residindo na área, conseqüentemente familiarizados com os códigos sociais daquela população. Dispensou-se um grande cuidado ao treinamento dos mesmos, a elaboração do questionário e a adequação das palavras as especificidades da linguagem local. A

codificação do questionário, embora sob a responsabilidade do entrevistador, era posteriormente revisada pelo investigador principal ou seu assistente; algumas respostas, face a sua complexidade, eram codificadas apenas pelo investigador principal. Em complemento a esses cuidados, a coleta de dados foi planejada de tal maneira que para cada uma das áreas cada entrevistador aplicasse a mesma proporção de questionários.

Qualquer que seja o problema relativo a informação; o resultado final será que o indivíduo (ou grupo) poderá ser erroneamente classificado com relação à exposição e/ou com relação ao efeito; no entanto, as conseqüências serão diferentes dependendo de se o erro de classificação em um eixo (seja exposição ou doença) é independente (ou não) da classificação no outro eixo. A existência de erros de classificação em 1 eixo que não são independentes do outro eixo é conhecida como erro de classificação diferencial, enquanto a existência de erros de classificação para a exposição ou para a doença que são independentes do outro eixo é referida como erro de classificação não-diferencial (Rothman, 1986).

Kleinbaum et al. (1982) propõe que se trabalhe com os conceitos de sensibilidade e especificidade como referenciais para a compreensão do bias de informação. A sensibilidade para uma determinada condição é definida como a probabilidade de que os indivíduos com aquela condição sejam classificados, no estudo, como tendo a condição. Especificidade é definida como a probabilidade de que as pessoas sem a condição sejam classificadas, no estudo, como não tendo a condição. Os erros de classificação diferencial e não-diferencial poderiam então ser compreendidos da seguinte maneira:

- não-diferencial: quando ao classificar os indivíduos como doentes ou não-doentes o investigador obtivesse a mesma sensibilidade e a mesma especificidade no grupo de expostos e no grupo de não-expostos. Similarmente, quando estivesse classificando a exposição, a sensibilidade e a especificidade seriam a mesma para o grupo de doentes e não-doentes.
- diferencial: seria caracterizado por diferentes sensibilidades e diferentes especificidades na identificação da exposição nas duas ca-

tegorias de doença (doente, não-doente) ou, ainda, diferentes sensibilidades e especificidades na identificação da doença nas diferentes categorias de exposição (exposto, não-exposto).

O erro de classificação diferencial pode resultar em um bias de informação que pode exagerar ou subestimar um efeito. O outro tipo de erro de classificação, o erro de classificação não-diferencial, que pode ocorrer quando uma classificação de exposição ou doença é incorreta para uma proporção semelhante de indivíduos nos grupos comparados, tem sido compreendido como uma ameaça menor a validade do que o erro de classificação diferencial, uma vez que o bias introduzido por um erro de classificação não-diferencial geralmente ocorre em direção a hipótese nula (Bross, 1954; Copeland et al., 1977; Rothman, 1986). Esse tipo de bias é uma preocupação maior na interpretação de estudos que parecem indicar a ausência de um efeito (Rothman, 1986).

No estudo em questão, não há evidência de que tenha havido um erro de classificação diferencial. Nem o entrevistador, nem o respondente sabiam, no momento da entrevista, quem estava e quem não estava infectado. Por outro lado, os dois técnicos de laboratório desconheciam as características socioeconômicas dos indivíduos para os quais estavam realizando o exame; além disso, os exames de cada área eram alocados para cada um dos técnicos independentemente da família a qual os indivíduos pertenciam. Não existe, portanto, nenhuma razão para que o diagnóstico da esquistossomose fosse influenciado pela condição socioeconômica ou vice-versa. Em outras palavras, a sensibilidade e a especificidade para classificação dos indivíduos como doentes e não-doentes em cada uma das duas categorias de exposição seria a mesma, ou a sensibilidade e especificidade para identificação dos indivíduos como expostos ou não-expostos nas duas categorias de doença seria semelhante.

Considerando as limitações dos instrumentos e das técnicas usadas para caracterização das exposições e da doença no trabalho em questão, o erro de classificação não-diferencial dificilmente seria evitado. Vale ressaltar que, como a distorção se faz em uma direção previsí-

vel, em direção a hipótese nula, o seu efeito, se presente, seria o de subestimar a magnitude do valor, das associações encontradas. A sua discussão seria de fundamental importância nos estudos com achados negativos, o que não é o caso do referido trabalho.

RESUMO

XIMENES, R. R. A. & ARAÚJO, T. V. B.
Validade Interna em Estudos de Corte Transversal: Reflexões a Partir de uma Investigação sobre Esquistossomose Mansônica e Condições Socioeconômicas.
Cad. Saúde Públ., Rio de Janeiro, 11 (1): 118-127, jan/mar, 1995.

O objetivo desse trabalho foi discutir a questão de validade, tomando como referencial um fragmento de uma pesquisa desenvolvida em área urbana na Região Metropolitana do Recife, com a finalidade de investigar a associação entre fatores socioeconômicos e esquistossomose. Discute-se as possíveis distorções na razão dos produtos cruzados (*odds ratio*) que podem ser introduzidas, quer pela forma como os indivíduos são selecionados para o estudo, quer por problemas ligados a obtenção das informações. Ressalta-se alguns dos mecanismos utilizados, em diferentes momentos da investigação, no sentido de minimizar a possibilidade do aparecimento desses tipos de bias ou de avaliar a sua influência na interpretação dos resultados.

Palavras-Chave: Esquistossomose; Epidemiologia; Metodologia de Pesquisa; Validade Interna; Estudos de Prevalência

REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAMSON, J. H., 1990. *Survey Methods in Community Medicine*. 4th ed., London: Churchill Livingstone.
- BROSS, I., 1954. Misclassification in 2x2 tables. *Biometrics*, 10: 478-486.
- COPELAND, K. T.; CHECKOWAY, H. & McMICHAEL, A. J., 1977. Bias due to misclassification in the estimation of relative risk. *American Journal of Epidemiology*, 105: 488-495.

- CRIQUI, M. H., 1979. Response bias and risk ratios in epidemiologic studies. *American Journal of Epidemiology*, 109: 394-399.
- GREENLAND, S. & THOMAS, D. C., 1982. On the need for the rare disease assumption in case-control studies. *American Journal of Epidemiology*, 115: 547-553.
- HENNEKENS, C. H. & BURING, J. E., 1987. *Epidemiology in Medicine*. Boston/Toronto: Little, Brown and Company.
- KATO, K. & MIURA, M., 1954. Comparative examinations. *Japanese Journal of Parasitology*, 2: 35.
- KATZ, N.; CHAVES, A. & PELLEGRINO, J., 1972. A simple device for quantitative stool thicksmear technique in Schistosomiasis Mansoni. *Revista do Instituto de Medicina Tropical de São Paulo*, 14: 397-400.
- KLEINBAUM, D. G.; MORGENSTERN, H. & KUPPER, L. L., 1981. Selection bias in epidemiologic studies. *American Journal of Epidemiology*, 113: 452-463.
- , 1982. *Epidemiologic Research, Principles and Quantitative Methods*. New York: Van Nostrand Reinhold Company.
- MOSER, C. & KALTON, G., 1985. *Survey Methods in Social Investigation*. 2nd ed., Aldershot: Gower Publishing Company.
- RODRIGUES, L. & KIRKWOOD, B., 1990. Case control design in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *International Journal of Epidemiology*, 19: 205-212.
- ROTHMAN, K., 1986. *Modern Epidemiology*. Boston/Toronto: Little, Brown and Company.
- SMITH, P. G. & PIKE, M. C., 1976. Generalization of two teste for the detection of household aggregation of disease. *Biometrics*, 32: 817-828.
- XIMENES, R. A. A., 1991. Urbanization of Schistosomiasis in the Metropolitan Region of Recife (Social causation of Schistosomiasis in the age group 10-25 years in the urban area of São Lourenço da Mata). Tese de Doutorado, London: London School of Hygiene and Tropical Medicine, London University.