

Alguns pilares para a apreciação da validade de estudos epidemiológicos

Pillars for assessing validity in epidemiological studies

Michael E. Reichenheim

**Departamento de Epidemiologia
Instituto de Medicina Social
Universidade do Estado do Rio de Janeiro**
Rua São Francisco Xavier, 524 – 7º andar
20559-090 Rio de Janeiro, RJ – Brasil
E-mail: michaelr@ims.uerj.br

Claudia Leite Moraes

**Departamento de Epidemiologia
Instituto de Medicina Social
Universidade do Estado do Rio de Janeiro**
Rua São Francisco Xavier, 524
20559-090 Rio de Janeiro, RJ – Brasil

Este trabalho foi parcialmente realizado no Medical Research Council - Biostatistics Unit, University of Cambridge, quando MER foi apoiado pelo CNPq (nº 200122/82-7).

Resumo

Abordagem da questão da validade dos estudos epidemiológicos. Distinguem-se seis tipos de validade. Na validade conceitual enfocam-se os elementos que sustentam a construção de um modelo teórico explicativo do fenômeno de interesse, articulado no plano do discurso e conceitos abstratos. Na validade operacional discutem-se os processos nos quais os conceitos que compõem as diferentes dimensões do modelo teórico são reduzidos a variáveis e indicadores, viabilizando sua instrumentação no nível empírico. À apresentação da validade de domínio, levanta-se a questão da intercambialidade dos indivíduos da população alvo e da população de estudo como sustentação para a generalização dos resultados. A validade de comparação dos grupos de estudo gira em torno da identificação das circunstâncias que rompem a estrutura de comparabilidade, a saber, a situação de confusão e a exclusão seletiva de elementos. À abordagem da validade de informação, aprofundam-se os problemas referentes à aferição/mensuração e os conseqüentes vieses. Visando algumas diretrizes para avaliação da validade de informação, explora-se os fundamentos das relações temporais envolvendo as ocorrências e os processos de coleta. Por fim, apresentam-se algumas importantes questões relacionadas à escolha e validade de especificação do modelo estatístico, bem como comentários sobre as estratégias mais utilizadas na seleção de variáveis e diagnóstico de modelo. Na parte final, sustenta-se que um enfrentamento sistemático das questões de validade pode servir como um dos pilares para a construção do conhecimento em epidemiologia.

Palavras-chave: Epidemiologia. Modelos teóricos. Fatores de confusão. Vies. Modelos estatísticos.

Abstract

The article outlines several validity issues in epidemiological studies. Six types of validity are covered. Conceptual validity concerns the elements sustaining a theoretical explanatory model, which, by definition, exists only on an abstract and discursive level. Operational validity concerns the processes by which concept-pertaining dimensions at a theoretical level are reduced to variables and indicators, enabling empirical testing. In validity of domain, focus is shed on the exchangeability between subjects of the target population and those of the study population as a basis for generalizing results. Comparison validity concerns the identification of circumstances that rupture the comparability structure between groups, namely, confounding and selective exclusion of subjects. Information validity relates to data collection and ensuing biases. Attention is drawn to the problems of timing and order of measurement related to events and data collection procedures. Finally, focus is placed on some important issues related to specification validity of statistical models, identifying guidelines for selecting those models. In the final part, systematic handling of validity issues is contended as a possible building block for attaining knowledge in epidemiology.

Keywords: Epidemiology. Models, theoretical. Confounding factors. Models, statistical. Research design. Bias. Data collection, methods.

Introdução

A validade dos estudos epidemiológicos é um tema de preocupação e em permanente debate entre pesquisadores. Muitos autores apontam para a necessidade de detalhamento sobre as possíveis fontes de erros sistemáticos e aleatórios, na tentativa de evitar ou minimizar vieses¹⁻³. Este artigo visa oferecer alguns alicerces para a crítica, enfocando a trajetória de elaboração, execução e análise de estudos epidemiológicos sob o prisma das questões de validade. Aceita-se que o escopo da epidemiologia não se restringe aos estudos de causalidade e/ou determinação e que, portanto, nem todos os componentes de validade abordados aqui são universalmente pertinentes. Todavia, tomam-se os estudos "analíticos" como base das apreciações pois, nestes, todas as etapas são permeadas por questões de validade, especialmente nos estudos observacionais onde o completo controle sobre os grupos de comparação é impossível.

As possibilidades de organizar as categorias de validade não são exaustivas. Portanto, o artigo é orientado por outros estudos teórico-metodológicos³⁻⁵ e incorpora algumas reflexões desenvolvidas ao longo dos anos nos cursos de pós-graduação. O recorte dado parte da proposta de Greenland que enfoca os possíveis erros sob quatro aspectos de validade: comparação, seguimento, mensuração e especificação do modelo de análise³. Propõe-se uma reordenação e aprofundamento destas categorias, além da expansão da classificação para incluir outros aspectos, a saber, validade conceitual, operacional e de domínio. Reconhecendo que uma abordagem minuciosa sobre as questões de validade de estudos epidemiológicos desenbocaria em um tratado sobre o próprio método epidemiológico, este artigo visa somente abrir portas, justificando, assim, o seu título. Seis tipos de validade são discutidos a seguir.

Validade dos estudos epidemiológicos

Validade conceitual

A epidemiologia opera nos campos teórico e empírico. A conexão entre ambos é mediada através da formulação de hipóteses que expressam as relações terminais de um modelo (quadro) teórico, servindo como ponte entre este e a realidade⁶. Um modelo teórico é o primeiro recorte da realidade e tem o objetivo de representar o que se pretende dela extrair. A partir da sua construção, o pesquisador organiza suas idéias em relação ao processo de causação ou determinação da doença/agravo. Além disso, o modelo torna coerentes as indagações de pesquisa, interpretação dos dados e, em última instância, o direcionamento dos achados, identificando os interlocutores precípuos que deles usufruirão⁷.

É boa prática que os estudos visando inferência causal partam de modelos teórico-conceituais construídos a partir de uma ampla e rigorosa revisão de literatura e/ou de intuições, experiências e observações do próprio pesquisador. Independentemente da origem, desses elementos edifica-se um composto de conceitos, nexos e relações. Com um quadro teórico bem definido é possível identificar as hipóteses de pesquisa e as dimensões teóricas putativamente envolvidas no processo em estudo e, a partir daí, estabelecer uma estratégia de captação empírica das relações de ocorrência.

Interessa destacar que, tendo esta construção um forte elemento subjetivo, é difícil, ou mesmo impossível, certificar-se que o modelo concebido é de fato válido. Os resultados de uma dada pesquisa são apenas instâncias de corroboração ou refutação de hipóteses propostas. No entanto, a validade conceitual pode ser aprimorada, desde que o investigador dê atenção aos quadros teóricos que subjazem os componentes do modelo proposto. Este refinamento é ainda mais importante quando o modelo a ser investigado é composto de intricados constructos conectados no plano abstrato do discurso, sem representa-

ção factual imediata (e.g., classe social ou violência intradomiciliar).

Vale alertar que nem sempre os estudos epidemiológicos têm se preocupado com este componente de validade, assumindo os elementos reduzidos a partir dos conceitos no plano teórico como os elementos teóricos em si. Esta deficiência compromete não somente a validade conceitual, mas também prejudica a validade operacional ao delegar um conceito a um indicador de forma equivocada.

Validade operacional

As questões relacionadas à validade operacional dizem respeito à construção do modelo a ser testado empiricamente, composto pelas variáveis de desfecho, exposição, confusão e modificação de efeito. A escolha de variáveis e indicadores para representar os conceitos abstratos do modelo conceitual requerem atenção e aprofundamento, pois o traslado de um componente teórico, já um recorte da realidade, para o plano empírico inevitavelmente produz ainda mais simplificações. O pouco cuidado no processo de redução pode fazer com que um “falso” representante do conceito seja erradamente incorporado no sistema. Chama-se a atenção sobre a necessidade de explicitação desta etapa. A abertura dos processos que envolvem estas representações tende a aumentar o potencial de crítica de uma investigação, permitindo, conseqüentemente, o seu aprimoramento.

É importante diferenciar entre conceitos e escalas no plano teórico, e suas respectivas definições operacionais. Miittinen⁴ sugere que estas definições sejam estabelecidas na fase de desenho de estudo, evitando que o pesquisador demarque as escalas empíricas com a finalidade de chegar a resultados esperados. Outros autores também aceitam que a definição das escalas teóricas e de sistemas de codificação devam ser realizadas antes da análise de dados, porém, argumentam que a categoria-base, o número de categorias e os sistemas de escores devem ser sempre definidos após a coleta de dados, permitindo uma maior eficiência do estudo⁸.

Na prática, a possibilidade de uma definição antes da coleta de dados depende fundamentalmente do estágio em que se encontra o programa de investigação. Na medida que o programa amadurece, cresce o conhecimento prévio, aumenta o refinamento teórico e operacional, e aumenta a possibilidade de especificação *a priori* das escalas. Obviamente, em programas de investigação com quadros teóricos rudimentares e representantes empíricos ainda nebulosos é necessário, senão inevitável, que se proceda de maneira retrospectiva.

Algumas variáveis requerem vários indicadores para que suas representações sejam válidas. Isto se aplica tanto para a especificação do desfecho, como para a definição das exposições, fatores de confusão e dos modificadores de efeito. Nos estudos epidemiológicos, a escala empírica utilizada tem influência direta nos resultados da investigação e, logo, necessita ser explicitada. Seu estabelecimento deve ser norteado por questões referentes ao processo de determinação em estudo, almejando-se alocar sempre um número suficiente de indivíduos a todas as categorias de exposição.

Este último ponto pode trazer problemas pois é tentador criar categorias que maximizem a precisão*, i.e., que aloquem o maior número de sujeitos possíveis aos estratos da covariada, em detrimento do significado das categorias ligadas ao conceito que a variável representa. As questões de conteúdo devem ter precedência sobre as questões de eficiência, mesmo porque a última pode ser manipulada no desenho de estudo (e.g., superamostrando-se um estrato e cor-rigindo-se na análise).

Vale lembrar que questões de cunho eminentemente técnico imporão limites ao número de variáveis e suas categorias. Sabe-se que para um modelo estatístico produzir estimadores precisos, o número de parâmetros não deve exceder à razão entre o total de eventos de desfecho e o total de variáveis introduzidas na modelagem. Quando for grande o número de parâmetros a esti-

mar, processos de redução de dados podem ser usados¹⁰. Estes procedimentos, por definição, levam à desfiguração da significação nominal das variáveis originais, ainda que a eficiência na captação da informatividade dos elementos de estudo esteja aumentada. No entanto, se envolvem somente potenciais fatores de confusão e não a variável de exposição de interesse central, estas reduções não infringem a máxima que propõe o respeito à teoria subjacente no processo de redução.

Uma situação análoga a esta concerne o uso de uma variável *proxi* de um *constructo* na tarefa de representar o conceito subjacente¹¹. A validade conceitual não estará comprometida se for possível estabelecer uma estreita correspondência entre a *proxi* e o *constructo* ao se alocar adequadamente os elementos de análise às respectivas categorias. Por outro lado, uma precária correspondência de conteúdo leva à má-classificação. Vale notar que existe uma distinção entre o processo de categorização/classificação e a mensuração propriamente dita. É possível conceber uma situação em que ocorra um problema de classificação mas a aferição seja adequada. Por outro lado, mesmo que a categorização seja perfeita, ainda assim pode ocorrer um problema no processo de medição. Ambas as situações levam a uma má-classificação dos elementos, ponto aprofundado na seção sobre validade de informação.

Validade de domínio

A validade de um estudo epidemiológico está intimamente atrelada ao seu escopo inferencial. Neste sentido, é fundamental que haja uma clara definição do escopo das relações de ocorrência captadas pelo estudo e às quais os resultados se referem, ou seja, seu domínio⁴. O seu estabelecimento obedece a diversos critérios e depende do tipo de investigação que está sendo realizada. Em algumas situações, a limitação do domínio pode ser inerente ao próprio problema investigado. Nos estudos "particularistas", as relações de ocorrência

* Ao longo do texto entende-se precisão e validade como definido em Last⁹.

se restringem a uma população específica, ou seja, os resultados da pesquisa não se prestam a generalizações. O domínio do estudo é restrito *a priori* e a base é formada levando isto em consideração^{4,12,13}. No desenho de estudos "abstratos", a população fonte das informações está sendo utilizada apenas para captar as relações de ocorrência em geral e não as relações específicas àqueles indivíduos do estudo ou à população que estes representam. Neste tipo de investigação, o domínio do estudo transcende a população fonte.

Em estudos onde, por questões logísticas, é mais fácil a captação de indivíduos oriundos de um certo subgrupo, é comum restringir a população de estudo a este estrato. Há aumento de eficiência devido à redução de custo operacional, já que a informação é mais facilmente captada. Por exemplo, ao se investigar a relação do tabagismo com câncer de pulmão é mais eficiente utilizar indivíduos do sexo masculino, pois neste estrato, a proporção de expostos ao fumo é maior. Mesmo que de forma provisória, esta estratégia não invalida a generalização para mulheres se for aceita a premissa de que ser homem é irrelevante para a ação carcinogênica do tabaco no tecido pulmonar. Tal assunção é baseada no conhecimento da homogeneidade da ação carcinogênica do fumo e na similaridade biológica entre os pulmões de homens e mulheres².

Uma situação oposta também merece destaque. Victora e col.¹⁴, estudando o efeito do aleitamento materno na ocorrência de diarreia na infância, aponta para a influência do diferencial social na magnitude da proteção. Os autores mostram que o efeito somente ocorre nos estratos sócio-econômicos mais baixos e está ausente nos mais abastados, caracterizando uma situação de modificação de efeito. Imaginando que um pesquisador confinasse a sua população de estudo aos indivíduos pertencentes a apenas um destes estratos, somente caracterizaria uma validade de domínio se a inferência se restringisse a este grupo, enquanto que uma extensão à toda a população a violaria flagrantemente.

Em suma, é necessário, de saída, identificar-se a população externa à qual a população de estudo se refere. A capacidade de generalização dos resultados é baseada no julgamento dos pesquisadores e no esclarecimento de quais condições são relevantes para a extrapolação dos resultados. Deve ser enfatizado que a capacidade de generalização dos resultados de um estudo não é uma questão de representatividade amostral, e sim, de representatividade inferencial.

Validade de comparação

Nas pesquisas que visam a inferência causal, pressupõe-se que haja validade de comparação quando o grupo de não expostos tiver a capacidade de prever o que teria ocorrido com os expostos, caso estes não fossem expostos. Esta visão assume explicitamente o preceito contrafactual^{15,16} de intercambialidade dos grupos de exposição¹⁷. Estritamente, esta situação é apenas teórica porque há a pressuposição de se estar comparando o elemento de análise quando exposto com ele mesmo quando não exposto. Na impossibilidade factual desta comparação, substitui-se o "exposto quando não exposto" por um elemento referente não exposto. Para haver comparabilidade e, portanto, uma estimação não enviesada, é necessário que todos os outros determinantes estejam atuando igualmente no elemento índice e seu referente. Na prática, observa-se uma falta de comparabilidade sempre que os indivíduos do grupo de referência não forem equivalentes aos do grupo índice no que diz respeito à distribuição dos demais fatores de risco para o desfecho de interesse^{3,18}.

Para uma classificação satisfatória que permita uma apreciação adequada de possíveis problemas é necessário a identificação das circunstâncias que rompem esta estrutura de comparabilidade. O corte dado aqui assemelha-se à proposta de Steineck e Ahlbom¹⁹. Separam-se explicitamente estas circunstâncias que rompem a comparabilidade em situação de confusão (*confounding*) e exclusão seletiva de elementos.

Estes são fenômenos bem distintos. A situação de confusão decorre de uma distorção da homogeneidade interna dos estratos de exposição e é algo imanente ao fenômeno sob investigação. Ocorre quando certos eventos estão interligados e se distribuem de forma agrupada nos elementos da base. É crucial notar que a heterogeneidade dos grupos de comparação ocorreria mesmo que todos os elementos da hipotética hiperpopulação fizessem parte do estudo. Não se trata de uma distorção induzida por um recorte na captação da base.

Esta deficiência, por sua vez, rompe a estrutura de comparabilidade pela exclusão seletiva de certos elementos. Mesmo que dois grupos populacionais sejam absolutamente semelhantes (salvo a exposição) – descaracterizando, portanto, uma situação de confusão – uma amostra desta população terá estratos de exposição completamente heterogêneos se por algum motivo houver seletividade no processo de incorporação dos elementos.

Postas estas considerações de fundo, o restante da seção aprofunda algumas questões sobre as estratégias para abordar a situação de confusão e, subsequentemente, oferece um panorama sobre o problema de seletividade da população de estudo em relação à base.

Situação de confusão (*confounding*)

É preciso que se faça uma distinção entre fator de confusão e situação de confusão, assinalando que o primeiro é o indicador nos dados do imanente fenômeno que está ocorrendo na base. Kleinbaum e col.¹ descrevem três propriedades fundamentais para que uma variável seja considerada um fator de confusão: (a) ser um fator de risco para a doença em questão entre os não-expostos; (b) estar associada à exposição na população em estudo, especificamente, na coorte em estudos de seguimento e nos controles em estudos de caso-controle; e (c) não ser intermediária na relação entre a exposição e a doença.

De uma maneira geral, recomenda-se que os três critérios sejam utilizados na

avaliação do papel de uma covariada. No entanto, retomando e aprofundando Miettinen e Cook²⁰, Robins e Morgenstern²¹ questionam a necessidade imperiosa de obediência a estes critérios. No contexto de vários delineamentos epidemiológicos, os autores examinam uma gama de situações em que o conhecimento prévio sobre as associações entre exposição, covariada e desfecho deve prevalecer sobre as informações contidas nos dados. Por exemplo, em um estudo caso-controle em que a informação *a priori* não indique que a variável seja fator de confusão, os autores sugerem que a estimativa mais válida de efeito da exposição seja a medida bruta, mesmo que ocorra nos dados uma associação da covariada com a exposição e o desfecho.

O conhecimento *a priori* também é importante quando, na base, uma covariada é intermediária no processo de causação/determinação. O controle equivocado de variáveis intervenientes leva a estimadores espúrios, e neste caso, a medida de efeito menos viesada pode ser a não ajustada. A decisão de introduzir ou não uma covariada no modelo de análise requer um quadro teórico que defina a posição hierárquica da exposição no processo^{22,23}.

É interessante notar que, nesta perspectiva, a discussão sobre a situação de confusão é baseada nas informações provenientes de diversas fontes: o processo de determinação apreendido em investigações anteriores; considerações sobre as associações entre as variáveis nos dados coletados; e informações sobre o tipo de desenho epidemiológico em questão. Esta posição se opõe àquela que se guia pela regra meramente estatística de colapsabilidade das medidas de efeito²⁴. Esta considera que existe confusão somente quando ocorre uma diferença estatisticamente significativa entre as medidas de efeito estratificadas (condicionais) e a medida bruta correspondente. Miettinen e Cook²⁰ apontam que a identificação da situação de confusão desta forma faz com que as conclusões sejam divergentes, dependendo do parâmetro escolhido para medir o efeito da exposição. Importa salientar que, do ponto de vista prático, a escolha deste ou

do outro posicionamento está atrelada ao estágio em que se encontra o programa de investigação no qual se insere a pesquisa. Ainda que recomendável em situações de programas maduros, na ausência de conhecimento prévio sobre as relações de ocorrência, não se tem muitas opções a não ser aceitar as informações empíricas como guia.

Seletividade na inclusão/exclusão das ocorrências da base

Como visto na introdução desta seção, além da situação de confusão, a outra razão que pode romper a estrutura de comparabilidade é a seletiva exclusão de elementos representativos da base de referência, levando a uma estrutura de dados ausentes. A grosso modo, existem três mecanismos básicos de ausência²⁵⁻²⁷. Na *ausência completamente aleatória*, a exclusão de elementos é incondicional às covariadas e ao próprio desfecho. O modelo de ausência contempla somente um parâmetro de seleção, i.e., $p(A|X, Y, \Phi) = p(A|\Phi)$, tomando-se a ausência do desfecho Y como exemplo. O coeficiente Φ regula o processo amostral do indicador de ausência no modelo de seleção. Na chamada *ausência aleatória* existe o condicionamento a uma ou mais covariadas e ao desfecho observados nos dados. Formalmente tem-se $p(A|X, Y, \Phi) = p(A|X, Y_{obs}, \Phi)$. Se além destas condições existir também uma dependência aos valores dos elementos não observados, tem-se uma condição de não ignorabilidade (*ausência não ignorável*) e formalmente, $p(A|X, Y, \Phi) = p(A|X, Y_{obs}, Y_{aus}, \Phi)$.

A repercussão destes mecanismos na estrutura de comparabilidade é diversa e de importância distinta. Na *ausência completamente aleatória* há uma intercambiabilidade irrestrita dos elementos encontrados na população de estudo com os da população fonte (base) de onde são amostrados. É o caso de uma amostra aleatória simples da base, que Gelman e col.²⁷ denominam de dados completos ou potenciais neste contexto. É digno de nota que, condicional à inexistência de confusão (*confounding*), a comparabilidade continua

valendo nos dados, sem necessidade de qualquer procedimento de correção.

No caso da *ausência aleatória*, ainda que haja seletividade na amostra, o conhecimento por observação empírica direta dos mecanismos condicionantes permite o resgate da estrutura de comparabilidade. Esta situação pode ocorrer tanto em estudos de seguimento como em estudos de caso-controle. A seletividade ocorre quando um fator estranho modifica a probabilidade de seleção de indivíduos. Se as covariadas são identificadas e mensuradas é possível remover o erro através de uma análise apropriada que contemple o controle destas.

De maior relevância para esta discussão é a *ausência não-ignorável*, pois qualquer tentativa de resgatar a informação dependerá do que acontece (ou melhor, teria acontecido) aos elementos excluídos que, por definição, são de informatividade distinta dos elementos da população de estudo e, principalmente, desconhecidos.

Visando uma sistematização das condições de exclusão de dados podem-se propor mais dois eixos de classificação, além dos mecanismos de ausência descritos acima. Um, discerne se a exclusão é *intencional* (voluntária) ou *não-intencional* (involuntária)²⁷. Quando voluntária, o investigador tem condições de obter as informações necessárias para resgatar a estrutura de comparabilidade perdida pela seletividade forçada. É o caso das amostras estratificadas²⁸ e multi-fásicas²⁹ em que se usa o método de reponderação na análise para recompor a estrutura original.

O outro eixo classifica a exclusão de acordo com o momento no percurso da investigação. Neste enfoque, a exclusão pode ocorrer antes do momento potencial de entrada na população de estudo - uma *censura à esquerda* ou *truncamento*, empregando-se a terminologia dos estudos de sobrevivência, ou após a captação dos elementos de estudo, caracterizando uma *censura à direita*. Ambas as censuras podem acontecer no mesmo estudo.

É interessante notar que este eixo de classificação não caracteriza implícita-

mente o tipo de mecanismo de ausência. Nas duas pontas este pode ser *completamente aleatório*, *aleatório* ou *não-ignorável*. A repercussão e a maneira de enfrentamento é que diferem. A Figura 1 resume os três eixos. Como as combinações são muitas, oferecem-se, a seguir, apenas algumas situações, a troco de ilustração e contextualização da tipologia.

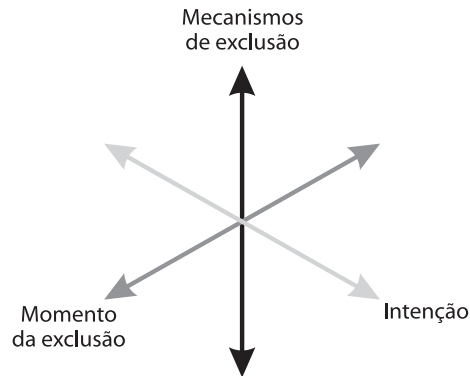


Figura 1 - Eixos de classificação relativos à exclusão de elementos da base de estudo.

Figure 1 - Classification axis concerning the exclusion of subjects from the study base.

Nos estudos caso-controle há, por definição, uma exclusão *intencional* pois os controles são sempre uma subamostra da base. Quando não há distúrbio de captação tem-se uma ausência não informativa, razão pela qual o mecanismo de seleção é ignorado na análise, como ocorre tradicionalmente. De acordo com os eixos propostos, no estudo caso-controle simples há exclusão *intencional* com *truncamento* e um mecanismo de *ausência completamente aleatória*.

Classificação distinta têm os estudos de caso-controle que se utilizam da estratégia de pareamento para a seleção de controles. É amplamente reconhecido que o pareamento introduz uma forma de viés de seleção mas, desde que o fator de pareamento tenha sido apropriadamente registrado e utilizado na análise, é possível reconstituir a estrutura de comparabilidade perdida inicialmente^{2,30,31}. A possibilidade de correção caracteriza o tipo de *ausência aleatória*.

Na perspectiva da identificação de rupturas de comparabilidade, no entanto, interessam mais as situações de não ignora-

bilidade, decorrentes de problemas involuntários de captação, levando a viéses de seleção. Vários são conhecidos¹. Para situar o problema, um bom exemplo é o viés de *Berkson* que resulta da diferença de probabilidade de hospitalização dos casos e controles, condicional ao *status* de exposição. Trata-se de um problema de encaminhamento (drenagem) diferenciado que pode ser mitigado com a escolha de patologias-controle com o mesmo sistema de referenciamento dos casos. Merece ser lembrado que qualquer tentativa de lidar com o problema deve constar da estratégia de captação de elementos, pois, uma vez instalado o viés, a possibilidade de recuperação da informação excluída e, mais grave, não ignorável, é extremamente complicada, senão impossível³².

Nos estudos longitudinais um dos principais problemas é a perda de elementos ao longo do período de acompanhamento, caracterizando uma *censura à direita*. Alguns são perdidos antes do final da pesquisa, quer porque desistiram de participar do estudo, quer porque se mudaram da localidade da investigação, ou mesmo porque faleceram por outras causas. Independente do motivo que tenha levado a estas perdas, a conseqüente falta de informação pode ser um obstáculo à adequada estimação de medidas de efeito. A validade de seguimento estará assegurada quando o risco de perda durante o período de acompanhamento não está associado ao desfecho³.

Chama-se a atenção para o fato de que as tradicionais análises de sobrevivência tacitamente assumem a ignorabilidade das perdas. O modelo supõe que a informação perdida pela exclusão de um elemento é reconstituída através da informação proveniente de um outro que permanece sob observação naquele momento. No entanto, pode-se dizer que esta situação é a exceção e não a regra. Em muitas ocasiões as perdas são dependentes do virtual desenlace, caracterizando uma *ausência não-ignorável*. A correção é delicada, uma vez que é necessário se ter alguma informação sobre o mecanismo de seletividade da perda.

Neste caso, a informação adicional pode ser incorporada à análise como exemplificam Best e col.³³

Os estudos transversais apresentam particulares problemas de seletividade. O mais comum e importante é o viés de sobrevivência seletiva que ocorre quando a perda de um elemento selecionado para o estudo acontece antes da coleta das informações, por exemplo, por morte ou migração³⁴. Se a exclusão destes elementos na população de estudo depender do potencial e ausente *status* do desfecho, de acordo com a tipologia proposta, ter-se-á uma *ausência não-intencional e não-ignorável*. Merece ser apontado que esta *ausência não-ignorável* somente tem relevância se o estudo em questão tiver pretensões de inferência causal. Nos estudos transversais que pretendem descrever pontualmente o estado de uma população, uma exclusão de elementos não constitui uma violação da comparabilidade e sim, por oposição aos inclusos, o foco precípua de descrição.

Vale comentar que a ausência de informações, conseqüente a intermitentes recusas de respostas ou omissões de registro, pode também comprometer a validade de comparação e merece toda a atenção. Problemas podem ocorrer em qualquer tipo de delineamento. Uma situação crítica sucede quando todas as informações de um elemento são excluídas da análise, como requer a implementação de modelos de regressão, ainda que a ausência se dê em apenas algumas variáveis. A redução do número de elementos pode ser substantiva quando se consideram todas as combinações de ausência nas muitas variáveis contempladas numa análise. Este fato, em princípio, leva a um declínio da eficiência e poder do estudo. Existe abundante literatura sobre os procedimentos de imputação para lidar com o problema³⁵.

Todavia, há situações onde as perdas decorrentes do tratamento dos dados levam a uma quebra da estrutura de comparabilidade. Se, por exemplo, em um inquérito de vacinação, o padrão de extravio de cartões de registro for dependente do "verdadeiro" mas desconhecido estado vacinal

– caracterizando uma *ausência não-ignorável*, retirar da análise os indivíduos sem informação sobre o desfecho enviesará os resultados. Na situação de não-ignorabilidade, o resgate da comparabilidade é de difícil solução mas factível mediante métodos de modelagem complexa^{25,26,36,37}.

Validade de informação

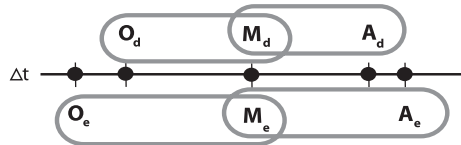
Problemas no processo de mensuração e o enfrentamento destes têm tido uma atenção especial no meio epidemiológico e bioestatístico, como testemunha uma crescente e recente literatura sobre o assunto³⁸⁻⁴¹. A abordagem aqui parte da classificação tradicional que separa confiabilidade de validade^{1,39}. No entanto, é central precisar o que se tem em mente ao se confrontarem estes dois conceitos. Confiabilidade diz respeito ao processo de aferição específico de um estudo, não sendo, portanto, uma característica imanente do instrumento de aferição. Pode-se pensar a confiabilidade como conotando a "robustez" da aferição (ou sua falta) em um certo estudo, apreendendo as "pressões" exercidas pelo aferidor e o aferido sobre o instrumento. Pelo caráter particular desta interação, a confiabilidade necessita ser investigada em cada estudo e seus resultados são intransferíveis.

A validade do instrumento pode ser considerada uma propriedade deste, havendo, pois, transponibilidade para uma população externa àquela onde o estudo de validação foi realizado. Porém, é mister distingüir entre a validade própria de um instrumento de aferição e a de todo o processo de aferição. A última depende também da confiabilidade. Por exemplo, um instrumento aceito como válido (*a priori* ou estudado anteriormente) pode ter, circunstancialmente, uma precária estabilidade e replicabilidade, dois componentes centrais da confiabilidade³⁹. Esta deficiência implicará em uma má qualidade da informação captada, comprometendo conseqüentemente a validade do processo de aferição como um todo.

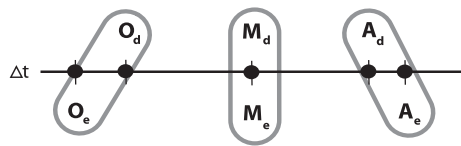
Uma gama de vieses de aferição tem como fundamento as relações temporais

envolvidas na investigação, merecendo atenção precoce na etapa de construção do desenho de estudo. Resgatando as observações de Greenland e Morgenstern^{42,43}, propõem-se dois eixos de classificação. A Figura 2 ilustra estes eixos, focalizando a relação entre exposição e desfecho.

(a) Primeiro eixo: temporalidade intra-evento



(b) Segundo eixo: temporalidade inter-evento



O Ocorrência Índices:
M Mensuração **e** Exposição
A Apuração **d** Desfecho

Figura 2 - Eixos temporais para a caracterização da relação entre ocorrência, mensuração e apuração dos eventos de interesse.

Figure 2 - Time axis concerning the relationship between occurrence, measurement and information recovery of events of interest.

O primeiro – a ser chamado de *temporalidade intra-evento* – envolve as relações entre a ocorrência, mensuração e apuração* de um evento de interesse (desfecho ou exposição ou fator de confusão) e compreende três elos. Um se refere à relação do momento da ocorrência e da mensuração de cada evento (desfecho, exposição ou fator de confusão); o segundo relaciona a mensuração com a apuração; e, por extensão aos outros dois, o terceiro elo se refere à relação entre a ocorrência e a apuração. Em todas estas situações a *temporalidade intra-evento* pode ser retrospectiva quando a ocorrência do evento é anterior ao momento de sua mensuração e/ou apuração, ou concorrente, se houver simultaneidade.

No planejamento de uma investigação é fundamental a especificação de quais os eventos que estão sendo abordados e suas relações. Por exemplo, em certos estudos usando dados primários, a temporalidade (*intra-evento*) da relação mensuração-apuração referente à exposição é, por desenho, concorrente, mas retrospectiva na relação ocorrência-mensuração. Dependendo da defasagem entre estes dois momentos, o viés de memória pode ser prominente, gerando um severo erro de informação. Em um estudo de coorte histórico, por outro lado, as três relações se dão em momentos diferentes, na seqüência ocorrência-mensuração-apuração. Esta é a situação mostrada na Figura 2.

O segundo eixo – a ser chamado de *temporalidade inter-eventos* – se baseia na ordem das ocorrências ou mensurações ou apurações entre os diversos eventos de interesse (exposição, fatores de confusão e desfecho). A relação pode ser denominada anterior ou produtiva, sempre que a ocorrência da exposição ou fator de confusão anteceder a ocorrência do desfecho (idem para a mensuração/apuração). Do contrário, a relação pode ser denominada posterior ou reversa. Problemas referentes à ordem de aferição podem acontecer em qualquer momento. Não é difícil conceber que um conhecimento prévio por parte do mensurador sobre certas características do observado introduza distorções no registro de um evento (mensuração) e, anos depois, à execução de um estudo de coorte histórico, uma apuração não mascarada possa, pelos mesmos motivos, agravar estas distorções.

Na seção referente a validade operacional indicou-se que uma má-classificação pode resultar tanto de problemas operacionais no processo de redução de conceitos a variáveis, quanto de problemas na mensuração em si. A despeito da origem, a má-classificação pode ser diferencial ou não-diferencial, quer independente ou não^{1,2}. Grande parte da literatura sobre as estraté-

* Distingue-se explicitamente mensuração de apuração. Mensuração se refere a um procedimento diretamente efetuado com o elemento de estudo. A apuração se refere a um procedimento de aferição retrospectivo através de registros, como ocorre, por exemplo, nos estudos de coorte histórica.

gias de correção de erros não-diferenciais e independentes na aferição de variáveis dicotômicas e contínuas assume que a magnitude e direção destes erros não está relacionada ao verdadeiro valor da variável. Nestas circunstâncias, a aferição equivocada de uma variável de exposição leva a uma subestimação de seu efeito^{3,30}. Recentemente, no entanto, alguns autores têm apontado que, na presença de dependência entre os erros de mensuração e os verdadeiros valores das variáveis, a má-classificação pode até aumentar ou mesmo reverter a medida de efeito de interesse⁴⁴⁻⁴⁶.

Quando se está diante de variáveis que tenham mais de dois níveis, apenas o conhecimento de que os erros são do tipo não diferencial e independentes também não garante que a direção do viés seja em direção ao nulo⁴⁷. Quando a má-classificação ocorre na aferição de variáveis de confusão, a situação de confusão é subestimada. Lamentavelmente, a tentativa de tratamento por estratificação ou modelagem não previne a confusão por esta covariada. A situação de confusão residual ocorre devido à impossibilidade de construção de estratos internamente homogêneos em relação à real variável de confusão. Pequenos erros na aferição de importantes fatores de confusão podem levar a estimativas enviesadas do efeito da exposição e de outras covariadas⁴⁸⁻⁵⁰.

Resultados e inferências válidas demandam, portanto, a utilização de condutas que impeçam, minimizem ou tentem corrigir os erros de aferição. É corriqueiro o enfrentamento destes somente nas fases pré-analíticas, mediante condutas preventivas concentradas na fase de desenho do estudo. Quando realizados, estudos de replicação ou validação de instrumentos são usados mais para colocar em perspectiva os resultados do estudo principal. Algumas vezes, a média das medidas replicadas é utilizada na análise de fundo. O aparecimento de técnicas estatísticas e abordagens mais sofisticadas tem alterado este panorama. Trata-se de uma literatura em constante crescimento e cujo aprofundamento transcende o escopo do artigo. O

leitor poderá encontrar alguns exemplos interessantes em Clayton⁵¹ (modelos lineares generalizados com efeitos aleatórios); Wacholder⁴⁵ (modelos que incorporam a correlação entre os erros e o verdadeiro valor das variáveis); Richardson⁵² e Spiegelhalter e col.⁵³ (abordagem Bayesiana); e Wacholder e col.⁴⁴, Brenner⁵⁴ e Spiegelman e col.⁵⁵ (estratégias de correção para estudos que utilizam um padrão "dourado" - *alloyed gold standard*).

Validade de especificação de modelo

A questão central na apreciação da validade do processo de modelagem é avaliar a especificação do modelo estatístico. Ajuda reconhecer que o processo de especificação é norteado por alguns pontos que merecem ser explorados. Esquemáticamente, numa fase inicial importa identificar o propósito do estudo, o parâmetro de ocorrência desejado e o tipo e estrutura dos dados coletados. Fechando o processo, avalia-se o ajuste do modelo (*goodness-of-fit*). A Figura 3 procura captar as inter-relações destas etapas.

Antes de apreciar os pontos norteadores do processo de especificação do modelo estatístico convém apresentar seus dois

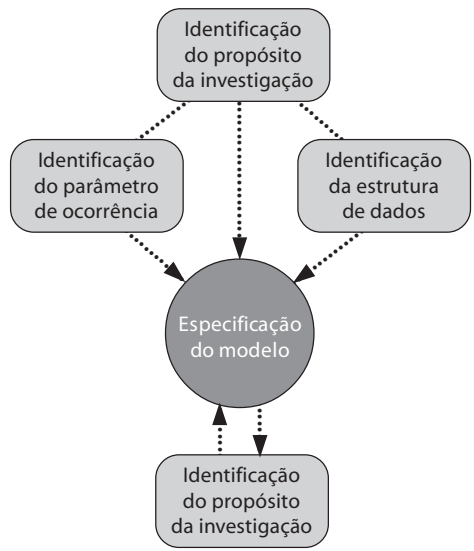


Figura 3 - Pontos norteadores para uma adequada especificação do modelo estatístico.

Figure 3 - Cornerstones for an adequate specification of a statistical model.

componentes - o modelo amostral e o estrutural. Todas as técnicas estatísticas, incluindo as não paramétricas, assumem um modelo para o processo de geração dos dados. Nos métodos paramétricos, este modelo amostral consiste de uma relação matemática que expressa a probabilidade de observar os dados como uma função de certas constantes ou parâmetros. A forma matemática deste componente incorpora apenas os aspectos aleatórios do processo de geração dos dados, deixando de lado os aspectos sistemáticos³. Sucintamente, o modelo amostral contempla o aspecto distribucional dos parâmetros do modelo estatístico.

O modelo estrutural expressa os parâmetros do componente amostral (distribucional) como uma função das variáveis de estudo. A especificação do modelo estrutural depende, pois, das variáveis selecionadas/usadas e da função de ligação que as conecta ao(s) parâmetro(s) do modelo amostral. A conjunção dos dois modelos pode ser melhor entendida com um exemplo descrevendo uma simples análise logística, tão comum em epidemiologia:

$$y_i \sim \text{Binomial}(\pi, n) \\ \text{logito}(\pi) = \beta_0 + \beta_k x_{ki}$$

Esta especificação informa que o desfecho y_i tem uma distribuição binomial com parâmetro π (e denominador n), cujo logito é uma função linear dependente das variáveis e os respectivos coeficientes. O modelo assume que π cresce ou decresce de forma logística com x_{ki} . A violação dessa premissa implica numa inadequação de especificação do modelo estatístico, quer pelo lado do componente amostral, quer pelo estrutural⁵⁶.

Um ponto de apoio importante na especificação de um modelo estatístico e, por conseguinte, na sua apreciação, passa pelo reconhecimento do propósito do estudo. É um equívoco comum contemplar a análise de um estudo epidemiológico presumindo, irrefletidamente, uma investigação de fatores de risco ou prognóstico e cuja modelagem demanda o reconhecimento explícito da situação de confusão e os demais quesitos de validade já abordados nas seções anteriores.

De fato, muitos estudos em epidemiologia não têm este enfoque. Estudos epidemiológicos podem, por exemplo, ser usados para previsões temporais, espaciais ou de sub-grupos de risco⁵⁷. Aqui importa muito mais a maximização de ajuste do modelo frente aos preditores escolhidos do que a boa especificação do modelo estatístico no sentido amplo do termo. Não interessa tanto o constructo teórico subjacente mas somente uma especificidade do modelo que atenda à proposta de projetar valores no tempo e/ou espaço. Neste sentido, cabe perfeitamente a implementação de um processo automático de seleção de variáveis do tipo passo-a-passo^{58,59}, centrada na inclusão/exclusão de variáveis de acordo com sua significância estatística. O processo de seleção gira em torno de eficiência, prescindindo do apelo à teoria. Aqui, é preciso registrar que os procedimentos automáticos baseados em significância estatística têm sido usados, senão abusados, na perspectiva de análises causais. Neste contexto, cabem severas críticas quando se reconhece que o controle da situação de confusão não é uma questão de precisão, mas de validade (comparabilidade). Fica o reconhecimento de que modelagem não é panacéia e seu uso merece sempre uma cautelosa reflexão^{60,4}.

Se o propósito da investigação é a inferência causal, a apreciação da validade do modelo estatístico vai além dos quesitos de eficiência. Como visto adiante, o componente amostral é mais dependente do tipo de variável de desfecho. A especificação do componente estrutural, no entanto, depende diretamente da validade das variáveis de exposição, confusão e modificação de efeito incluídas. O ideal é que haja um modelo teórico subjacente orientando o processo de seleção destas. Uma estrutura teórica hierarquizada contendo diferentes dimensões do processo de causação, bem como a relação entre as variáveis que as compõem, são definidas já no início do estudo. Aceita-se a premissa de preexistência de validade conceitual.

Todavia, em programas de investigação em que pouco ou mesmo nada é conheci-

do em relação ao fenômeno em estudo, uma orientação teórica fica prejudicada. Em estudos exploratórios ou rastreadores*, as dimensões que compõem o quadro teórico e suas respectivas variáveis podem não estar bem especificadas no início do estudo. Nestes casos, seria legítimo usar alguma estratégia de seleção de variáveis baseada na mudança da magnitude do coeficiente de regressão, apesar desta ser criticável por não levar em conta a variabilidade estatística das estimações⁵⁹. Uma variante que contempla de maneira mais adequada o problema da colinearidade entre as covariadas se calca na mudança no intervalo de confiança do estimador. Uma opção ainda melhor seria usar os dois critérios simultaneamente⁶⁵.

É preciso notar que estas estratégias não fogem a escolhas balizadas por decisões funcionais. Uma alternativa ao processo inteiramente automático é forçar, no modelo final, algumas variáveis de interesse especial, independentemente de qualquer critério de mudança (p-valor, estimador de ponto ou intervalo de confiança). Esta proposta subscreve a realização de uma análise de dados interativa, na qual o pesquisador não se redime da responsabilidade de assumir certas questões subjetivas respaldadas por informações existentes *a priori*, por mais parcas que estas sejam.

O mesmo princípio deve dirigir uma possível inclusão de termos de interação⁶⁶. Nesta perspectiva, priorizam-se as interações epidemiológicas e biológicas em detrimento das interações estatísticas**. No entanto, vale novamente lembrar como é comum a situação em que pouco se sabe sobre o processo em estudo. Em programas de pesquisa jovens, muitas vezes não se tem opção senão partir para uma exploração estatística de termos de interação. Problemas de subdetecção podem ocorrer

pois a identificação de interações, mesmo as de baixa ordem, dependem de tamanhos amostrais consideráveis (poder do estudo).

Ao se optar por procedimentos automáticos ou semi-automáticos de especificação de modelos – quer os de orientação estatística ou não – encontra-se uma dificuldade adicional, pois são muitas as combinações de "bem ajustados" modelos amostrais e estruturais especificáveis a partir de um mesmo conjunto de dados. A questão que se impõe é como apreciar e discernir modelos estatísticos com diferentes componentes amostrais e estruturais e que se ajustem igualmente bem. Como podem estar todos captando o "verdadeiro" fenômeno? Este impasse não é de fácil resolução. Uma sugestão é se ater à perspectiva de um programa de investigação continuado. Testa-se a presença de interação através de técnicas estatísticas formais*** e, ao se identificar uma, aceita-se provisoriamente a condição de modificação de efeito. Evidências subseqüentes deste achado no curso do programa de pesquisa passam a servir, então, como instâncias de corroboração do fenômeno, permitindo elevar-se o achado inicial ao *status* de conhecimento em algum momento.

Ainda tangente à identificação do propósito do estudo, vale apontar que, se a intenção é caracterizar sub-grupos e/ou investigar a magnitude de certo agravo em diferentes estratos populacionais, uma simples análise estratificada será totalmente satisfatória. Por exemplo, em estudos "particularistas" no âmbito dos serviços de saúde, a identificação de um agravo e sua distribuição na clientela adscrita é uma informação importante para os profissionais de saúde diretamente envolvidos no atendimento. Claramente, uma sofisticada análise multivariada de nada contribuiria neste contexto, a não ser na res-

* Usam-se os termos exploratório ou rastreador de forma ampla. Considera-se que, mesmo em estudos com pretensões à inferência causal e delineamentos apropriados para tal, a inexistência de conhecimento prévio leva o estudo a ser, no máximo, uma exploração inicial da configuração teórica e um rastreador desta possível configuração a partir de conceitos/variáveis putativas.

** As definições de interação estatística, epidemiológica e biológica podem ser encontradas em Rothman e col.⁶⁷ Siemiatycki e Duncan⁶⁸ e Morgenstern³⁰. Sucintamente, o primeiro tipo atende apenas a requisitos matemáticos. Já, a interação biológica é inerente às relações qualitativas existentes no nível do indivíduo. Dois fatores são considerados biologicamente dependentes - sinérgicos ou antagonísticos - quando o mecanismo de ação de um é afetado pelo outro. A interação epidemiológica (modificação de efeito) resulta do impacto bruto de vários tipos de interações biológicas ocorrendo nos indivíduos que compõem a população. Pode ser definida como uma heterogeneidade de efeito da exposição em substratos de uma covariada.

*** São exemplos, os testes de homogeneidade das Razões de Chances⁶⁹ ou testes de interação em modelos paramétricos⁷⁰.

trita situação em que se pretendesse encontrar um modelo preditivo como uma ferramenta de rastreamento (*screening*) para o agravo em questão.

Para a identificação do parâmetro de ocorrência – o segundo ponto norteador aqui proposto – importam as características do fenômeno estudado e o respectivo delineamento adotado para a captação da ocorrência empírica. No caso de estudos transversais, é desejável usar a razão de prevalência ou a razão de chance de prevalência? Em se tratando de um estudo prospectivo, quer-se usar a razão de risco, a razão de chances ou uma razão de taxas? Em certas situações o próprio delineamento impõe a escolha. Por exemplo, em um estudo caso-controle com amostragem por conjunto de risco (pareado no tempo) é recomendável usar-se a razão de produtos cruzados como um estimador não enviesado da razão de taxas.

Deve ser mencionado que a disponibilidade de aplicativos que, no entanto, implementam limitados modelos e só permitem a modelagem de restritos parâmetros de ocorrência, não pode ser mais aceita como uma camisa de força. A ampliação dos métodos que permitem usar modelos mais complexos e, portanto, outros parâmetros de ocorrência, tende a aumentar o escopo dos parâmetros estudados. A diferença de risco, risco atribuível, mediana, centis, ranques e outros podem agora ser adequadamente modelados⁵³.

No processo de especificação do modelo estatístico interessa também identificar o tipo de estrutura dos dados, discernindo-se, inicialmente, o tipo de variável de desfecho em questão. A distribuição presumida no modelo amostral depende disto. Se a variável for contínua, pode-se optar por uma distribuição gaussiana; em se tratando de uma proporção ou risco, opta-se por uma distribuição binomial; caso os dados se originem de um estudo prospectivo em que o desfecho seja a contagem de eventos no tempo, pode-se optar por uma distribuição Poisson⁷¹.

Estas são apenas algumas distribuições corriqueiramente utilizadas em epidemiolo-

gia. Todavia, diante da gama de outras distribuições conhecidas, nada impede que sejam utilizadas, desde que existam tecnologias para isto. Problemas decorrentes de estruturas conglomeradas de dados⁷², dados ausentes não-ignoráveis^{26,35} ou de erros de mensuração⁴⁰ também implicam em problemas de especificação da estrutura de modelos e necessitam ser considerados à análise.

Por fim, o quarto ponto norteador para a especificação do modelo estatístico concerne a checagem de ajuste de modelo. Esta etapa coroa as outras três mas, ao indicar problemas, pode reorientar as decisões tomadas anteriormente. A Figura 3 procura ressaltar esta interatividade.

São muitas as variantes e critérios para a apreciação de ajuste¹⁰ e um aprofundamento foge muito ao escopo deste texto. A troca de reflexão, no entanto, merece ser apontado que as abordagens clássicas, baseadas na comparação de valores observados e preditos por modelos aninhados, não estão isentas de crítica. Gelfand⁷³ considera de pouca utilidade uma comparação de dois modelos distintos completamente especificados e que só funciona na restritíssima situação de aninhamento, i.e., um sendo a extensão do outro mediante a adição de termos no componente estrutural. Uma comparação entre modelos estatísticos de componentes amostrais e estruturais diferentes não pode ser realizada usando-se os métodos habituais. Todavia, são precisamente estes os procedimentos de averiguação de ajuste de modelo que mais interessam, indo além da mera checagem em modelos contíguos. Novas técnicas têm sido desenvolvidas neste sentido⁷³⁻⁷⁷.

Considerações finais

Seria surpreendente se, diante de tantos e tão severos obstáculos à validade, restasse ainda alguma ingenuidade quanto à possibilidade de verdade objetiva. No entanto, tal conclusão não deve conduzir à paralisia. É possível evocar a noção de “veracidade construída” – fundamentada na apreciação crítica de estudos pontuais,

bem como de todo um programa de investigação consistente – como pista para contornar o problema da incerteza.

A idéia de uma “veracidade construída” fica bem clara quando se percebe, por exemplo, o quanto a validade conceitual (teórica) está interligada à validade de especificação do modelo estatístico. A última necessita de um quadro teórico para sua instalação, por mais precário e ainda turvo que este seja. O modelo teórico, por sua vez, cresce e se consolida lentamente no bojo de um programa de investigação mediante seqüências de testes de conjecturas baseadas em procedimento de análises que são inteiramente dependentes de adequadas, ainda que nunca estritamente corretas, especificações de modelos estatísticos. A circularidade do processo é transparente.

Contudo, conceder que o conhecimento é sempre construído e negociado, não

implica em subestimar os rigores necessários à construção e execução de um estudo epidemiológico. Pelo contrário, esta opção epistemológica, ao reconhecer a busca do conhecimento como algo plenamente disputável, demanda, incondicionalmente, a plena abertura à crítica dos conteúdos teóricos e empíricos. O enfrentamento sistemático das questões de validade, se já não importante enquanto um requisito técnico no âmbito estrito de estudos singulares, pode ser encarado de forma mais ampla, como o sustentáculo do exercício da crítica e contra-crítica, servindo para diminuir, pelo menos parcialmente, as dúvidas e indeterminações inerentes aos programas de investigação epidemiológica. E é nesta articulada seqüência de estudos, que erros, tais quais teorias, podem ser sistematicamente descartados, possibilitando, no limite, o estabelecimento do próprio conhecimento.

Summary

The article outlines several validity issues in epidemiological studies. Six types of validity are covered. Conceptual validity concerns the elements sustaining a theoretical explanatory model, which, by definition, exists only on an abstract and discursive level. This model represents the ideas the researcher has on the causation/determination process being studied. It is contended as an important aspect of any (epidemiological) study, although often downplayed and bypassed. Operational validity concerns the processes by which concept-pertaining dimensions at a theoretical level are reduced to variables and indicators, enabling empirical testing. Distinction is made between concepts and scales at a theoretical level and their respective operational definitions. The timing for these issues to be specified in a study is also brought up, along with the place of proxy variables as markers for theoretical concepts. In validity of domain, focus is placed upon the exchangeability between subjects of the target population and those of the study population as a basis for

generalizing results. It is argued that appropriate inference must acknowledge the applicability of the study population with respect to the source population to whom the results apply. An example is given where the characteristics of the study population are irrelevant so far as generalization is concerned, as opposed to another where the study population is adequate in one situation but misrepresents the base, if the domain of inference is not properly specified. Comparison validity concerns the identification of circumstances that rupture the comparability structure between groups, namely, confounding and selective exclusion of subjects. The view is taken that the former must be distinguished from the latter since it is an intrinsic part of the phenomenon under investigation. Confounding occurs when certain characteristics are clustered within study subjects, and the problem persists even if all elements of the hyperpopulation are accrued. Selective exclusion of subjects, on the other hand, entails missing data structures and ensuing

problems. Three axes are presented whereby exclusions can be classified, namely, missing data, moment of exclusion, and whether this is voluntary or not. Examples of epidemiological study designs and how they can be classified accordingly are provided. The section also discusses whether and how these problems may be mitigated or corrected. Information validity relates to data collection and resulting biases. Attention is drawn to the problems of timing and order of measurements related to events and data collection procedures. An explicit distinction is made between a time axis which concerns the relationship between occurrence, measurement, and information recovery (from logs) of a particular event, and a time axis

which relates different events in terms of their occurrence or measurement or information recovery. The section briefly discusses how measurement errors may be handled. Finally, light is shed on some important issues related to specification validity of statistical models, identifying guidelines for selecting those models: purpose of the study, type of occurrence parameters, type of data structure and model adjustment. How those criteria lead to different sampling and structural models and the basic components of a statistical model are shown. In the final part of the article, systematic handling of validity issues is contended as a possible building block for attaining knowledge in epidemiology.

Referências

1. Kleinbaum D, Kupper L, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: Van Nostrand Reinhold Company; 1982.
2. Rothman K, Greenland S. Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven Publishers, 1998.
3. Greenland S. Validity concepts in epidemiologic research, draft 3c. Los Angeles: UCLA School of Public Health; Department of Epidemiology; 1993.
4. Miettinen O. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley & Sons; 1985.
5. Moraes C. A contribuição à epidemiologia aos estudos de avaliação da performance clínica através de um evento traçador: o caso das prescrições médicas às crianças com infecção respiratória aguda. [Tese de Mestrado]. Rio de Janeiro: Centro Biomédico da Universidade do Estado do Rio de Janeiro; 1995.
6. Almeida Filho N. Epidemiologia sem números: uma introdução crítica à ciência epidemiológica. Rio de Janeiro: Campus; 1989.
7. Krieger N, Zierler S. The need for epidemiologic theory. Epidemiology 1997;8:212-3.
8. Blettner M, Sauerbrei W. Influence of model-building strategies on the results of a case-control study. Stat Med 1993;12:1325-38.
9. Last J. A dictionary of epidemiology. 2nd ed. New York: Oxford University Press; 1988.
10. Harrel F. Predicting outcomes: applied survival analysis and logistic regression. Charlottesville: Division of Biostatistics and Epidemiology, Department of Health Evaluation Sciences, School of Medicine, University of Virginia; 1997.
11. Nelson L, Longstreth Jr W, Koepsell T, VanBelle G. Proxy respondents in epidemiologic research. Epidemiol Rev 1990;12:71-86.
12. Miettinen O. The "case-control" study: valid selection of subjects. J Chron Dis 1985;38:543-8.
13. Wacholder S, McLaughlin J, Silverman D, Mandel J. Selection of controls in case-control studies. I. Principles. Am J Epidemiol 1992;135:1019-28.
14. Victora C, Hutly S, Barros F, Martines J, Vaughan J. Prolonged breast-feeding and malnutrition: confounding and effect modification in a Brazilian cohort study. Epidemiology 1991;2:175-81.
15. Rubin D. Estimating the causal effects of treatment in randomized and nonrandomized studies. J Educ Psychol 1974;66:688-701.
16. Rubin D. Assignment of treatment group on basis of a covariate. J Educ Stat 1976;2:1-26.
17. Greenland S, Robins J. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15:412-8.
18. Greenland S. Randomization, statistics and causal inference. Epidemiology 1990;1:421-9.
19. Steineck G, Ahlbom A. A definition of bias founded on the concept of the study base. Epidemiology 1992;3:477-82.
20. Miettinen O, Cook E. Confounding: essence and detection. Am J Epidemiol 1981;114:593-603.
21. Robins J, Morgenstern H. The foundations of confounding in epidemiology. Comput Math Applic 1987;14:869-916.
22. Reichenheim M, Hasselmann M. Estudo sobre violência doméstica e desnutrição severa na infância: construção de um modelo teórico. [Comunicação apresentada no 3º Congresso Brasileiro de Epidemiologia, Salvador, 1995].

23. Victora C, Hutly S, Fuchs S, Olinto M. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol* 1997;26:224-7.
24. Boivin J-F, Wacholder S. Conditions for confounding of the risk ratio and of the odds ratio. *Am J Epidemiol* 1985; 121:152-8.
25. Diggle P, Kenward M. Informative dropout in longitudinal data analysis (with discussion). *Appl Stat* 1994;43:49-94.
26. Little R. Modeling drop-out mechanism in repeated-measures studies. *J Am Stat Assoc* 1995;90:1112-21.
27. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian data analysis*. London: Chapman and Hall; 1995.
28. Moser C, Kalton G. *Survey methods in social investigation*. London: Heinemann Educational Books; 1985.
29. Pickles A, Dunn G, Vazquez-Barquero J. Screening for stratification in two-phase ("two-stage") epidemiological surveys. *Stat Methods Med Res* 1995;4:73-89.
30. Morgenstern H. *Epidemiologic methods I e II - class notes*. Los Angeles: UCLA School of Public Health, Division of Epidemiology; 1995.
31. Greenland S. The effect of misclassification in matched-pair case-control studies. *Am J Epidemiol* 1982; 116:402-6.
32. Walter S. Berkson's bias and its control in epidemiology. *J Chron Dis* 1980;33:721-5.
33. Best N, Spiegelhalter D, Thomas A, Brayne C. Bayesian analysis of realistically complex models. *J R Stat Soc A* 1996;159:323-42.
34. Kleinbaum D, Morgenstern H, Kupper L. Selection bias in epidemiologic studies. *Am J Epidemiol* 1981;113:452-63.
35. Little R, Rubin D. *Statistical analysis with missing data*. New York: John Wiley; 1987.
36. Gilks W, Richardson S, Spiegelhalter D. Introducing Markov chain Monte Carlo. In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov chain Monte Carlo in practice*. London: Chapman and Hall; 1995. p. 1.
37. Gamerman D. Simulação estocástica via cadeias de Markov. In: 2º SINAPE; 1996 jul 29 a ago 2; Caxambu. São Paulo: Associação Brasileira de Estatística; 1996. p.196.
38. Dunn G. *Design and analysis of reliability studies: the statistical evaluation of measurement errors*. New York: Oxford University Press; 1989.
39. Streiner D, Norman G. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press; 1989.
40. Carroll R, Ruppert D, Stefanski L. *Measurement errors in nonlinear models*. London: Chapman and Hall; 1995.
41. Kupper L. Measurement and misclassification errors. [A 2-day Short Course; 1996 May 31-June 1; Brussels].
42. Greenland S, Morgenstern H. What is directionality? *J Clin Epidemiol* 1989;42:821-4.
43. Greenland S, Morgenstern H. Classification schemes for epidemiologic research designs. *J Clin Epidemiol* 1988;41: 715-6.
44. Wacholder S, Armstrong B, Hartge P. Validation studies using an allowed gold standard. *Am J Epidemiol* 1993; 137:1251-58.
45. Wacholder S. When measurements errors correlate with truth: surprising effects of nondifferential misclassification. *Epidemiology* 1995;6:157-61.
46. Kipnis V, Freedman L, Brown CC, Hartman A.M, Schatzkin A.W S. Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol* 1997; 146:842-55.
47. Dosemeci M, Wacholder S, Lubin J. Does non-differential misclassification always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746-8.
48. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112:546-69.
49. Brenner H. Bias due to non-differential misclassification of polytomous confounders. *J Clin Epidemiol* 1993;46:57-63.
50. Marshall J, Hastrup J. Mismeasurement and the resonance of strong confounders: uncorrelated errors. *Am J Epidemiol* 1996;143:1069-78.
51. Clayton D. Models for the logitudinal analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer J, Feinleib M, Lippert P, Hoffmeister H, editors. *Statistical models for longitudinal studies of health*. New York: Oxford University Press; 1992. p. 301.
52. Richardson S. Measurement error. In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov chain Monte Carlo in practice*. London: Chapman and Hall; 1995. p.401.
53. Spiegelhalter D, Thomas A, Best N, Gilks W. *BUGS 0.5 examples*. Cambridge: Medical Research Council Biostatistics Unit; 1996. 2 v.
54. Brenner H. Correcting for exposure missclassification using allowed gold standard. *Epidemiology* 1996;7:406-10.
55. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an "allowed gold standard." *Am J Epidemiol* 1997;145:184-96.
56. McCullagh P, Nelder J. *Generalized linear models*. London: Chapman & Hall; 1989.
57. Lehmann E. Model specification: the views of Fisher and Neyman, and later developments. *Stat Sci* 1990;6:160-8.
58. Mickey R, Greenland S. A study of the impact of confounder - selection criteria on effect estimation. *Am J Epidemiol* 1989; 129:125-37.
59. Maldonado G, Greenland S. Simulation study of confounder - selection strategies. *Am J Epidemiol* 1993;38:923-36.

60. Vandembroucke J. Should we abandon statistical modeling all together? *Am J Epidemiol* 1987;126:10-3.
61. Evans S. Uses and abuses of multivariate methods in epidemiology. *J Epidemiol Community Health* 1988;42:311-5.
62. Greenland S. Modeling and variable selection in epidemiologic. *Am J Public Health* 1989;79:340-9.
63. James F, McCulloch C. Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Rev Ecol Syst* 1990;21: 129-66.
64. Efron B, Tibshirani R. Statistical data analysis in the computer age. *Science* 1991;253:390-5.
65. Kleinbaum D. Logistic regression: a self learning text. New York: Springer-Verlag; 1994.
66. Hlatky W, Whittemore A. The importance of models in the assessment of synergy. *J Clin Epidemiol* 1991;44:1287-8.
67. Rothman K, Greenland S, Walker A. Concepts of interaction. *Am J Epidemiol* 1980;112:467-70.
68. Siemiatycki J, Duncan C. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol* 1989;10:383-7.
69. Breslow N, Day N. Statistical methods in cancer research. Lyon: IARC; 1980. v.1; The analysis of case-control studies.
70. Hosmer D, Lemeshow S. Applied logistic regression. New York: Wiley & Sons; 1989.
71. Lindsey J. Models for repeated measures. Oxford: Clarendon Press; 1993.
72. Clayton D. Some approaches to the analysis of recurrent event data. *Stat Methods Med Res* 1994;3:244-62.
73. Gelfand A. Model determination using sampling-based methods. In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov chain Monte Carlo in practice*. London: Chapman and Hall; 1995. p. 145.
74. Aitkin M. Posterior Bayes factor. *J R Stat Soc B* 1991;154: 111-42.
75. Kass R, Raftery A. Bayes factor. *J Am Stat Soc* 1995;90:773-95.
76. Raftery A. Hypothesis testing and model selection. In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov chain Monte Carlo in practice*. London: Chapman and Hall; 1995. p. 163.
77. Gelman A, Meng X-L. Model checking and model improvement. In: Gilks W, Richardson S, Spiegelhalter D, editors. *Markov chain Monte Carlo in practice*. London: Chapman and Hall; 1995. p. 189.