

8 Introdução ao Método Epidemiológico

Naomar de Almeida Filho, Maurício L. Barreto e Maria Zélia Rouquayrol

Denomina-se método científico (ou, por metonímia, metodologia científica) ao conjunto de lógicas aplicadas, protocolos e práticas empregadas pelas diversas ciências para a produção de conhecimento sistemático, válido e legitimado pelas instituições das ciências. O método científico assume, portanto, as peculiaridades dos respectivos objetos de conhecimento em cada campo disciplinar da ciência. Somente desse ponto de vista podemos aceitar com reservas a denominação "método epidemiológico" por referência às estratégias, técnicas e procedimentos estruturados de pesquisa no campo da Epidemiologia.

Neste capítulo, em primeiro lugar, definiremos a noção de Metodologia, no sentido de uma teoria do método na pesquisa científica em geral. Em segundo lugar, pretendemos analisar o processo de problematização (construção do problema científico) no campo epidemiológico, com ênfase em suas peculiaridades e particularmente sua contribuição para a metodologia científica em geral. Em terceiro lugar, vamos expor os conceitos metodológicos de variável e indicador, base das metodologias quantitativas na ciência contemporânea, apresentando algumas tipologias úteis para sua operacionalização na pesquisa em saúde e, particularmente, na Epidemiologia. Finalmente, discutiremos o conceito de hipótese como eixo articulador entre teoria e método, com especial ênfase àquelas que têm como objetivo testar relações de associação ou determinação de riscos, ou seja, hipóteses epidemiológicas *stricto sensu*.

O modo de produção de saber característico da ciência ocidental baseia-se no domínio do homem sobre a natureza, mediante contínua anexação de novos territórios de conhecimento. Estes territórios, ou campos de ciência, são demarcados por meio da formulação de questões e solução de problemas. Questões e pro-

► Conceito de metodologia

A concepção da ciência como prática de construção e análise de problemas foi bastante desenvolvida por Mário Bunge (1980), físico e filósofo argentino radicado no Canadá, que propõe que uma investigação terá atingido seus objetivos científicos ao cumprir as seguintes etapas, cujo conjunto constitui uma metodologia, sem necessariamente restringir-se a elas, ou esgotá-las, ou mesmo obedecer à ordem em que estão propostas:

1. *Descobrimiento do problema* ou lacuna em um conjunto de conhecimentos. Se o problema não estiver enunciado com clareza, deve-se passar à etapa seguinte; se estiver, passa-se à subsequente.
2. *Colocação do problema* com precisão, se possível, em termos matemáticos, mesmo que não necessariamente quantitativos. Ou, ainda, recolocação de um velho problema à luz de novos conhecimentos (empíricos ou teóricos, substantivos ou metodológicos).
3. *Procura de conhecimentos* ou instrumentos relevantes para a solução do problema (p. ex., dados empíricos, teóricos, aparelhos de medição, técnicas de cálculo ou de medição). Ou seja, exame do conhecimento para tentar resolver o problema.
4. *Solução* do problema com auxílio dos meios identificados. Se a tentativa resultar inútil, passa-se para a etapa seguinte; em caso contrário, à subsequente.

5. *Invento de novas ideias* (hipóteses, teorias ou técnicas) ou produto de novos dados empíricos que prometam resolver o problema.
6. *Obtenção de uma solução* (exata ou aproximada) do problema, com auxílio do instrumental conceitual ou empírico disponível.
7. *Investigação das consequências* da solução obtida. No caso de uma teoria, identificação de predições e prognósticos. Em se tratando de novos dados, exame das suas consequências para teorias relevantes.
8. *Prova da solução*: confronto da solução com as teorias e informação empírica pertinente. Se o resultado for satisfatório, a pesquisa é dada por concluída. Caso contrário, passa-se à etapa seguinte.
9. *Correção* das hipóteses, teoria, procedimentos ou dados empregados na obtenção da solução incorreta. Trata-se, naturalmente, do começo de um novo ciclo de investigação.

Na concepção de Samaja (1994), eminente epistemólogo argentino contemporâneo, a ciência não constitui uma entidade em si mesma, mas de fato se estrutura como modo de produção realizado por seres concretos e singulares, a quem denominamos “cientistas”. O produto desse processo produtivo peculiar é uma modalidade de saber sistemático e integrador chamada “conhecimento científico”, composta de elementos teóricos e componentes empíricos. É nesse sentido que se pode entender a afirmação de Samaja (1994) de que a ciência produz duas coisas: fatos e teorias, e ainda, analisa também “duas modalidades contrapostas na noção de método da ciência”:

1. Conjunto de ações destinadas à revelação ou aquisição de novas informações (que Samaja denomina “modo de descobrimento”);
2. Sequência de passos para a verificação da cientificidade do conhecimento (que ele chama de “modo de validação”).

O processo de produção do conhecimento humano, dentro do padrão sistemático de aplicação racional considerado como científico (ou baseado em evidências, para usar uma expressão corrente), tem várias fases, articuladas na noção de “cadeia do

conhecimento”. Eminentíssimos epistemólogos (Bhaskar, 1978; Chalmers, 1982; Goldmann, 1988; Samaja, 1994) defendem distintas abordagens que correspondem a diferentes denominações para cada uma dessas etapas e seus produtos intermediários.

Consideremos o processo de produção científico-tecnológico como uma cadeia produtiva, composta por etapas de transformação do objeto científico e seus respectivos produtos intermediários (Figura 8.1).

Este processo produtivo peculiar se inicia pela observação. Tomadas como “matéria-prima”, as *observações* serão transformadas em *dados* que, processados para produzir *informação*, no final do processo produtivo, emergem como *conhecimento científico e tecnológico* (Almeida Filho, 2003).

Na linguagem corrente da metodologia científica, o termo *observação* designa o processo de identificação, seleção, coleção e registro sistemático de signos referentes a propriedades ou atributos relevantes de objetos naturais, culturais ou sociais. Uma célula, um corpo humano, uma doença, uma população, uma sociedade, uma situação de saúde, são exemplos de objetos; permeabilidade de membrana, massa corporal, patogenicidade, fertilidade, desigualdade, iniquidade, são exemplos de propriedades desses objetos. Por metonímia, o produto desse primeiro elo na cadeia produtiva do conhecimento é também chamado de ‘observação’.

Uma observação pode ser produzida diretamente por meio do nosso aparato sensorial (principalmente mediante o sentido da visão) articulado à rede neural cognitiva ou indiretamente via algum tipo de instrumento, dispositivo ou aparelho destinado a ampliar ou substituir a percepção humana. A observação produzida pode ser gravada mecanicamente ou eletronicamente, ou representada em qualquer das línguas naturais como uma descrição escrita, ou ainda registrada usando-se alguma notação matemática. Atributos dimensionais normalmente são mensuráveis e atributos discretos são computáveis, enquanto situações, traços, processos, opiniões, narrativas e eventos, ou observações de natureza similar, não são mensuráveis nem computáveis, mas sim descritíveis. No primeiro caso, a observação assume a forma de medida, enquanto no segundo caso esta se expressa como descrição ou registro.

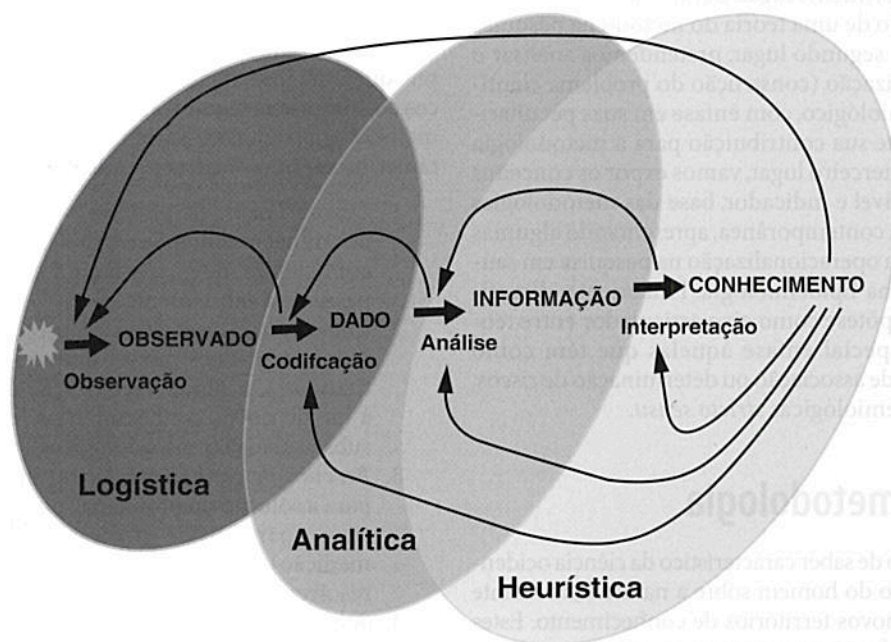


Figura 8.1 O ciclo de produção de conhecimento.

Boxe 8.1 Curiosidades etimológicas

O vocábulo "informação" origina-se diretamente do latim *informatio*, que significava justificativa ou explicação de uma palavra, concepção ou ideia. Objeto da ação do verbo *informare*, composição resultante do prefixo *in-* (dentro, sob etc.) e da raiz *formare* (dar forma), informar originalmente equivalia a "imaginar" ou "dar uma forma (imagem) na mente". O termo *enformation* foi primeiro incorporado no francês medieval com uma conotação jurídica equivalente a "inquerito criminal". Posteriormente, em meados do século XV, o termo passou a ser empregado no plural, designando o conjunto de conhecimentos de alguém, no caso um "informante". Na segunda metade do século XIX, a forma moderna *information* já era empregada especificamente no sentido de "ação de obtenção de dados ou relatos" e "relatório, documento de registro de dados", conotação com que se difundiu para outros idiomas, inclusive o português. Vale registrar o termo derivado *informática*, tradução direta do francês *informatique*, vocábulo criado por Philippe Dreyfus a partir do modelo "mathe-matique" especialmente para designar a ciência e a prática de organização e tratamento da informação.

Fonte: Rey, 1993.

de certa criança desnutrida, ou a contagem de células de uma dada amostra de sangue), essencial para se estabelecer a validade operacional concreta dos dados, deixa de ser relevante depois da transformação do dado em informação, sendo substituída pela categoria "variável" enquanto "definidor chave" no âmbito metodológico.

Informação, portanto, refere-se sempre a variáveis (resultantes do processamento de dados produzidos com as observações dos atributos ou propriedades de um dado objeto, que variam para cada caso).

A informação tem valor limitado além do escopo de um problema prático ou tecnológico. A fim de transcender a mera generalização e assim alcançar um grau útil de universalidade, a informação deve ascender ao nível do *conhecimento* (Almeida Filho, 2003). A informação torna-se conhecimento científico e tecnológico somente após articulada em algum marco de referência conceitual hierarquizado. Isso implica que informações científicas devem necessariamente ser elaboradas para subsidiar a construção de um objeto conceitual ou de um objeto de intervenção, ou seja, para formular uma teoria científica ou enquadrar um modelo de aplicação tecnológica. Dessa forma, não é válido falar de "informação científica" no mesmo sentido em que se diz "conhecimento científico e tecnológico". Em suma, a tecnologia resulta do conhecimento (mas não apenas deste) e não diretamente da informação científica (ou da "evidência").

A transformação de informação em conhecimento científico e tecnológico é regulada por processos sintéticos. Nesse sentido, a síntese constitui um procedimento especial de interpretação de informações, destinado a colocar a informação em um nível supracontextual. Conhecimento, por conseguinte, implica informação posta fora do seu próprio contexto e situada em um contexto mais geral, pronta para auxiliar pesquisadores, profissionais técnicos e tomadores de decisão a compreender outros contextos ou novas situações. Em comparação com aplicações baseadas em informação, abordagens baseadas no conhecimento são mais versáteis e flexíveis e, por conseguinte, mais úteis para lidar com novos problemas tanto no campo da ciência quanto no âmbito da técnica.

As observações são feitas por referência a casos ou situações singulares, e não se comprometem com outras esferas de generalização. Para que uma generalização (ou inferência) ocorra, é necessário que uma observação seja submetida a um processo complexo de transformação em outras categorias cognitivas hierarquicamente superiores (Almeida Filho, 2003).

A primeira transformação da cadeia produtiva do conhecimento científico e tecnológico opera da observação para o dado. Um *dado* é um signo. Mais especificamente, trata-se de um signo constituído a partir de um atributo observado em um objeto qualquer, que recebe um significado. Portanto, os dados podem ser definidos como observações com significado. Abordagens linguísticas e filosóficas dessa ordem poderiam subsidiar avaliações da pertinência (ou impertinência) teórica do uso geral dos conceitos de dado, informação, conhecimento e tecnologia, que extrapolam o escopo deste texto. Voltamos portanto ao tema da cadeia do conhecimento.

Os dados podem ser classificados, de acordo com o seu nível de "estruturção", como estruturados, semiestruturados e não estruturados. Dados estruturados são aqueles para os quais um sistema de codificação fixa já se encontra predeterminado. Dados semiestruturados não seguem um padrão de código prévio porém da própria produção de observações deveria resultar um sistema de codificação. Dados estruturados e semiestruturados *grossos modo* correspondem a abordagens de pesquisa quantitativa ou numérica. Nesse caso, os dados constituem produto do trabalho de traduzir observações para a linguagem codificada de uma dada pesquisa. Dados não estruturados são o produto de estratégias de pesquisa que não se baseiam em nenhum tipo de codificação.

Em termos estritos da sistematização científica, os dados são expressos como indicadores no sentido de que indicam parâmetros ou propriedades. Um parâmetro compreende um valor ideal de uma dada dimensão (ou propriedade quantificável) do objeto concernido sob investigação. No processo de transformação da medida em indicador, estimativas e parâmetros merecem atenção especial porque funcionam como produtos de etapas intermediárias. A partir de um conjunto de medidas, produzem-se estimativas expandindo-se o âmbito de generalização por meio de técnicas de extrapolação que se justificam, na maior parte das vezes, por regras inferenciais. Quando se faz uso de amostragem probabilística, a validade da conexão entre uma estimativa e um parâmetro é justificada estatisticamente. A validade de um indicador em relação a um parâmetro somente pode ser estabelecida no contexto de uma teoria científica.

Entretanto, os dados não fazem sentido em isolamento. Para que tenham algum valor científico e possibilitem apoiar processos de tomada de decisão sobre a validade de alguma posição sobre o mundo (em nosso caso, sobre a saúde), os dados precisam ser transformados em *informação*.

A passagem do dado para informação é determinada por processos de transformação analítica. Informação se produz a partir de dados analisados de modo adequado, no sentido de que estes devem ser processados com o objetivo de resolver um problema, responder uma questão ou testar uma hipótese. Nesse sentido, análise implica um processo de organização, indexação, classificação, condensação e interpretação de dados, com o objetivo de identificar comunalidades de dimensões, atributos, predicados e propriedades entre casos individuais.

Dessa maneira, buscam-se em cada caso os elementos individuais de universalidade, por isso o potencial de generalização torna-se aí o foco do processo de produção de conhecimento. Nesse nível, distinções, singularidades e idiosincrasias dos casos individuais não são importantes. O atributo ou propriedade individual (p. ex., o peso de um feto em particular, ou a estatura

O “definidor chave” nesta etapa do processo de produção de conhecimento é a categoria de “conceito”. Unidade elementar do modelo teórico, base do conhecimento científico e tecnológico, o “conceito” situa-se também como produto final de uma cadeia própria de produção cognitiva. Especificamente em relação aos dados estruturados e semiestruturados típicos da pesquisa epidemiológica, podemos identificar uma cadeia metodológica paralela à cadeia produtiva do conhecimento: medida-estimativa-parâmetro-indicador-variável-conceito. No que se refere a dados não estruturados (típicos da pesquisa antropológica, por exemplo), esta cadeia paralela terá somente dois elos: indicador-conceito.

Para processos de tomada de decisão nos níveis científico, tecnológico e pragmático, o conhecimento é hierarquicamente superior às informações e aos dados. O uso de dispositivos cognitivos como conceitos, modelos, teorias e protocolos, característicos da esfera do conhecimento, mostra-se mais eficiente para lidar com a complexidade e a emergência dos novos objetos científicos e tecnológicos, porque as sínteses do conhecimento não são limitadas pelos laços estreitos que fazem a informação depender de contextos, populações, aplicações ou situações de referência.

Observe-se ainda na Figura 8.1 uma indicação importante sobre os âmbitos de produção do conhecimento. As etapas de produção de observações e dados compreendem uma logística metodológica; etapas de produção de dados e informações implicam uma analítica; etapas de produção do conhecimento a partir da informação conformam uma heurística da pesquisa. Podemos assim propor que a metodologia científica, enfim, compõe-se de uma logística, uma analítica e uma heurística, articuladas de modo dinâmico e processual.

A prática da ciência resulta, enfim, de uma dialética fundamental entre o conhecimento assentado e os problemas gerados pela interação com o real (Goldmann, 1988). Estratégias de problematização efetivamente propiciam o crescimento da capacidade de o homem conhecer e dominar as realidades factuais do mundo circunstante, transcendendo a sua alienação, essencial na construção histórica de sua emancipação (Bhaskar, 1989). Enfim, criar problemas; é disso que se trata na pesquisa científica.

Vejamos agora como a ideia de problematização se aplica à ciência epidemiológica.

► Problematização na pesquisa epidemiológica

Na Epidemiologia, o problema científico aparece quando doenças (ou agravos à saúde, de qualquer natureza) acometem grupos humanos. A necessidade social de reconhecer, controlar e remover fatores ambientais, culturais, biológicos ou físico-químicos nocivos à saúde, implicando a criação de condições que a promovam, determina a problemática própria da Epidemiologia. A situação enigmática e intrigante com que se defronta o epidemiologista-pesquisador é geralmente de natureza diversa daquela posta perante o investigador de outros ramos do saber. A solução do problema epidemiológico muitas vezes representa a diferença entre vida e morte para muitos membros de uma dada comunidade.

Muitas vezes, um problema epidemiológico é demasiado evidente, saltando aos olhos por assim dizer, como, por exemplo, uma grave epidemia. A dramaticidade de um problema científico dessa natureza pode ser constatada no próprio evento considerado como fundante da Epidemiologia, ocorrido na Londres de 1854 com a publicação do relatório de pesquisa do Dr. John

Snow “*Sobre a maneira de transmissão do cólera*” (Snow, 1994 [1854]). Como vimos no Capítulo 2, trata-se de um marco histórico porque pela primeira vez se procedia de forma sistemática a uma investigação epidemiológica, buscando-se determinar a causa de um surto epidêmico.

Outras vezes, ocorrem problemas epidemiológicos latentes, posto que não se apresentavam abertamente como tais. Vejamos um exemplo. Durante anos, sem maiores preocupações além do atendimento clínico curativo ou paliativo, a medicina conviveu com surdez, catarata, retardamento mental e anormalidades cardíacas, afetando recém-nascidos e crianças. No início da dé-

Boxe 8.2 Dietilestilbestrol e câncer de vagina

Herbst e Usfelder, cancerologistas, e Poskanzer, epidemiologista, relatam a descoberta de uma associação causal entre adenocarcinoma de vagina em jovens e terapia por dietilestilbestrol nas mães durante a gestação. Em uma primeira publicação, seus autores relatam 68 casos de câncer vaginal primário tratados em dois hospitais, em um período de 36 anos, de 1927 a 1963. Na totalidade dos casos, a doença apresentou-se em mulheres com mais de 20 anos de idade: 62 casos (91%) nos grupos etários cuja idade era igual ou superior aos 40 anos. Em um segundo relatório, apresentaram uma casuística de oito mulheres jovens (idade abaixo de 25 anos) com adenocarcinoma vaginal diagnosticado, nos últimos 4 anos. Cancerologistas clínicos haviam levantado um problema: pela concentração de casos, tipos de tumor e idade das mulheres, tratava-se de um evento de alguma maneira inusitado. Foi proposto um estudo epidemiológico para a sua solução. O grupo de pesquisadores partiu da hipótese de que os oito casos tinham uma causa comum, desconhecida. Utilizou-se o desenho de caso-controle (ver adiante, Capítulo 8). Procederam à varredura de um amplo espectro de fatores, para a detecção daqueles que, pela reiterada aparição nas histórias dos casos, pudessem ser isolados como possíveis fatores causais. Para cada caso foram selecionados quatro controles dentre as nascidas no mesmo hospital que a paciente.

Houve a precaução de escolher, para controles, jovens nascidas com diferença de apenas 5 dias, no máximo, de cada caso, no mesmo tipo de serviço (enfermaria ou apartamento). Ficou evidenciado que nenhuma das oito pacientes havia feito uso de irritante intravaginal, duchas ou tampão. A relação a seguir mostra fatores adicionais cujas diferenças de ocorrência nos pacientes e nos controles não foram significativas: peso ao nascer; idade da menarca; complicações durante a vida intrauterina; medicamentos usados pela mãe durante a gravidez, à exceção de estrógenos; doenças da infância; história de amigdalectomia; tipo de alimentação durante a infância; presença de animais domésticos; uso de cosméticos; hábito de fumar; consumo de álcool, bem como ocupação e nível de instrução dos pais. Dentre os fatores maternos cobertos pelo inquérito, os seguintes não apresentaram diferenças significantes entre os grupos: idade da mãe; hábito de fumar da mãe; aleitamento no seio e exposição a raios X. Somente uma dentre as pacientes deixou de ser exposta ao dietilestilbestrol durante a vida fetal, enquanto nenhuma do grupo-controle havia sido exposta. A droga fora ministrada às mães devido à perda fetal em gestações anteriores ou por perda de sangue na atual. Estudos posteriores confirmaram a associação entre adenocarcinoma de vagina e o uso de certos estrógenos durante a gravidez. A partir da divulgação destes resultados, o dietilestilbestrol passou a ser medicação proibida para gestantes.

Fonte: Herbst, Usfelder & Poskanzer, 1971.

determinam a maneira pela qual os elementos de qualquer conjunto são diferentes entre si. Tomando como exemplo a população brasileira, pode-se dizer que as pessoas aí incluídas serão diferenciadas entre si por atributos, tais como sexo, religião, peso ou estatura. As variáveis, quanto a sua natureza, podem ser categorizadas como *qualitativas* e *quantitativas*.

Variáveis qualitativas são as que implicam diferenças radicais ou essenciais. A variável sexo, por exemplo, inclui as categorias masculino e feminino, as quais mantêm entre si diferenças não apenas de nível, volume ou quantidades, mas também de natureza. Exemplos de outras variáveis qualitativas que eventualmente podem despertar interesse epidemiológico são: local de residência, local de trabalho, ocupação, procedência, situação conjugal etc.

Variáveis quantitativas, por sua vez, envolvem distinções não substantiais, no sentido de diferenças traduzíveis em desigualdades de grau, frequência, intensidade, volume. Retenham-se as propriedades que mantêm a mesma natureza em toda sua extensão ou dimensão, que se mostram com maior ou menor expressão, podendo ser manifestadas em termos numéricos; temperatura, pressão sanguínea, peso e estatura são bons exemplos.

As variáveis quantitativas são *descontínuas* ou *discretas* quando, entre dois valores consecutivos expressos por números inteiros, não é possível a inclusão de valores fracionários: número de casos de uma doença ou frequência de batimentos cardíacos, por exemplo. As variáveis podem ser também *contínuas*, quando admitem valores fracionários entre quaisquer valores consecutivos (pressão barométrica ou temperatura corporal, por exemplo).

Em estudos epidemiológicos, as doenças específicas são consideradas ora como variáveis, ora como constantes. Quando tomadas como variáveis, seus valores normalmente podem ser ausência e presença. Assim, ao analisar-se a distribuição de dada doença em um grupo populacional homogêneo quanto ao critério "local de moradia", este será dividido em dois subgrupos: portadores e não portadores da doença.

A fim de possibilitar a codificação, análise e compreensão de conjuntos de fatos, processos e fenômenos, as diferenças entre seus elementos – ou variáveis – devem estar formalmente explicitadas. Isto pode ser realizado mediante operações de classificação, contagem ou mensuração da propriedade ou variável considerada. Se tratarmos da variável "sexo", as pessoas serão classificadas em duas categorias: masculino ou feminino; se "peso" ou "altura", serão medidas em quilogramas ou em centímetros, respectivamente; se "número de pessoas acometidas", será feita a contagem daqueles que possam ser aí classificados.

Na prática epidemiológica, quando se acompanha de descritivamente a evolução de fatos de interesse científico, ou quando se procede à investigação de fenômenos instáveis relacionados com a saúde ou a doença, analisando e enunciando problemas ou propondo hipóteses explicativas, busca-se no fim das contas identificar relações entre variáveis. Ao se estabelecerem tais relações, opera-se, na maioria dos casos, com valores atribuídos às variáveis por classificação, mensuração ou contagem. Em termos metodológicos, a mais importante e útil relação entre variáveis é a que as categoriza como *independentes* e *dependentes*.

Os termos "variável independente" e "variável dependente" foram emprestados da matemática. As variáveis representadas no eixo dos x, das abscissas, são as variáveis independentes, e aquelas representadas no eixo dos y, eixos ordenadas, são as dependentes. Formular uma relação entre variáveis significa assumir que a variável dependente deve variar concomitantemente com as mudanças ocorridas na variável independente.

cada de 1940 um oftalmologista australiano, chamado Norman Gregg, teve sua atenção despertada por uma cliente, mãe de uma criança com catarata, para o fato de que outra mãe, na sala de espera, trazia para consulta um filho também com catarata e que ambas tinham sido acometidas de rubéola durante a gravidez. Com a informação que espontaneamente lhe era dada, o Dr. Gregg foi capaz de intuir e depois formular um problema científico. Da pesquisa daí originada, e que não se restringiu apenas a catarata em recém-nascidos, resultou o conhecimento atual sobre os efeitos da rubéola em filhos de gestantes, expostas nos primeiros meses da gravidez (Gregg, 1941).

Fatos encobertos às vezes emergem pela utilização de estratégias de investigação mais ou menos elaboradas, resultado da intermediação da sensibilidade e percepção do investigador, juntamente com algum lampejo de intuição ou por alguma circunstância fortuita, como foi o caso das pesquisas empreendidas por Gregg. Somente dessa forma passam a constituir problemas científicos genuínos. Outras vezes, para a configuração de um problema epidemiológico original, é preciso mais que intuição, sorte, criatividade ou oportunidade, tornando-se necessário um investimento logístico e institucional de alta monta, como no exemplo recente da síndrome de imunodeficiência adquirida (AIDS) (ver Capítulo 43).

Entretanto, o desafio maior para a metodologia científica em geral, e para a ciência epidemiológica em particular, consiste na correta produção de hipóteses e no rigoroso processo de validação destas na busca de solução para os problemas identificados. Trata-se da própria essência do processo de produção do conhecimento científico, na medida em que criar problemas pode não ser tão difícil quanto resolvê-los.

Quando se trata de um problema complexo, pode-se partir de um enunciado bastante geral e, na medida em que o problema vai sendo analisado, este será decomposto em problemas mais simples, solucionados por meio de hipóteses claras, concisas e refutáveis. Tais hipóteses, ferramenta fundamental do raciocínio científico, são constituídas por elementos de tradução dos conceitos epidemiológicos sob a forma de variáveis. A seção seguinte discute, em maior profundidade, a natureza das variáveis epidemiológicas e o seu relacionamento com o processo de formulação de hipóteses neste campo específico da ciência contemporânea.

► Variáveis e indicadores epidemiológicos

Retornemos por um momento à cadeia produtiva do conhecimento (Figura 8.1), mais especificamente na passagem da observação ao dado. Considerando-se um conjunto qualquer de processos, fatos ou fenômenos observados, duas categorias de propriedades estarão ali presentes. Em primeiro plano, ressaltam as propriedades constantes, denominadas simplesmente *constantes*. Estas são exibidas por todos os elementos do conjunto de igual forma e, por isso, podem ser tomadas como critério para delimitar conjuntos homogêneos a partir de elementos esparsos. Exemplificando: tomando-se como critério o "país de nascimento" para se proceder à inclusão de pessoas em um conjunto homogêneo, pode-se definir uma população de "brasileiros". Este conjunto é complemento de outros conjuntos diferentes, porém homogêneos quanto ao mesmo critério: "americanos", "africanos", "argentinos".

Em nível mais profundo de análise dos dados, são discerníveis propriedades variáveis ou simplesmente *variáveis*. Estas

Esta é talvez a forma mais simples de se pensar os dois tipos de variáveis colocadas em relação, por não implicar o uso de termos discutíveis ou hipotéticos, tais como “causa” e “efeito”. Porém, quando se trabalha em um referencial de causalidade, a variável independente será o fator causal, ou seja, a causa presumida da variável dependente, sendo esta o efeito resultante da primeira. De todo modo, sempre se define variável independente como antecedente e variável dependente como consequente.

Em estudos experimentais, a variável independente é aquela que tem seus valores escolhidos e determinados pelo pesquisador. Quando, por exemplo, se testa um novo planorbicida, as concentrações da droga a que serão expostos os caramujos são determinadas pelo experimentador – esta é a variável independente da pesquisa. Neste exemplo, o número de caramujos mortos será a variável dependente – aquela que, em estudos experimentais, escapa ao controle do investigador e cuja variação se pretende mensurar.

Na pesquisa não experimental, que corresponde à maioria das investigações epidemiológicas, não é possível a manipulação de variáveis. Geralmente, a escolha de qual será a variável dependente e de qual será a independente é determinada pela suposição de que certa condição variável produz uma mudança no estado de saúde ou de doença; esta condição variável será tomada como variável independente, e o efeito, doença ou não doença, como variável dependente. É comum encontrar relações nas quais a variável independente foi escolhida pelo fato de que os eventos a ela associados apresentam-se com anterioridade aos eventos que a partir daí são tomados como dependentes.

A variável dependente é normalmente a condição cuja explicação está sendo tentada, ou seja, cujo conhecimento sobre sua determinação está sendo construído. Em muitos manuais de metodologia epidemiológica, a variável dependente tem sido chamada de “variável resposta”, efeito presumido ou desenlace (do inglês *outcome*). Seus valores dependem dos valores assumidos pela variável independente, denominada na pesquisa epidemiológica de “variável preditora”. Assim, para se explicar doença ou não doença como variáveis dependentes, pode-se pensar em uma multiplicidade de fatores corresponsáveis por sua determinação, ou seja, as variáveis preditoras. Não existem restrições teóricas quanto ao número de variáveis, sejam dependentes ou independentes, em um modelo explicativo epidemiológico.

Em uma perspectiva mais geral, as variáveis epidemiológicas são expressas como dados por meio de indicadores. Indicadores epidemiológicos sintetizam a relação entre o subconjunto de doentes (ou óbitos por uma dada doença, ou sujeitos portadores de uma condição relacionada com a saúde) e o conjunto de membros da população. Conforme vimos no Capítulo 5, tal relação equivale ao cálculo da probabilidade de adoecer, ou seja, constitui a expressão mais geral e simplificada do *risco*.

Para compreender a lógica básica dos indicadores epidemiológicos, precisamos retomar a Figura 5.1, enriquecendo-a com delimitações alternativas de denominadores. Dela podemos derivar a Figura 8.2, onde a base populacional do Risco continua sendo representada através do conjunto-mestre **P**.

Observem que agora outros subconjuntos além de **D** são definidos por referência à condição de saúde em questão. O mais importante destes é o subconjunto **O**, que representa os óbitos resultantes da doença **d**. Por isso **O** está contido em **D** (conjunto de doentes ou portadores da enfermidade **d**) que, por sua vez, é um subconjunto de **P**.

Por outro lado, nem todos os membros de **D** encontram-se sob o mesmo risco de morrer, porque a doença **d** produz casos

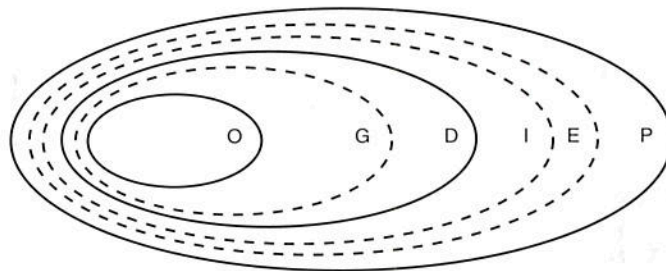


Figura 8.2 Subconjuntos componentes dos indicadores epidemiológicos.

com distintos níveis de gravidade. Chamemos **G** a este subconjunto de casos graves de **Dd**. Portanto, **O** está contido em **G**, que está contido em **D**, que por sua vez é um subconjunto de **P**.

Como nem todos os membros de **P** são suscetíveis ao efeito do agente de **d** e, além disso, nem todos os infectados tornam-se doentes, podemos incorporar o subconjunto **I**, de infectados (ou vulneráveis, no caso de doenças não infecciosas). Agora, **O** está contido em **G**, que está contido em **D**, que está contido em **I**, que por sua vez é um subconjunto de **P**.

Enfim, considerando que nem todos os membros da população **P** encontram-se expostos aos agentes ou fatores de risco de **d** e que destes nem todos tornam-se doentes, resta indicar o subconjunto **E**, de exposição. Então **O** está contido em **G**, que está contido em **D**, que está contido em **I**, que está contido em **E**, que enfim é um subconjunto de **P**.

Conforme podemos verificar de modo intuitivo, relações entre subconjuntos e grupos populacionais definidos por referência à condição patológica em questão propiciam o cálculo de indicadores epidemiológicos diversos.

Na perspectiva da série de conjuntos hierárquicos da Figura 8.2, podemos classificar os indicadores epidemiológicos de acordo com nível de referência, como:

- Macroindicadores – aqueles cujos denominadores se referem à base populacional plena **P**;
- Microindicadores – aqueles que tomam como denominador qualquer dos subconjuntos indicados, hierarquicamente inferiores a **P**.

Nos textos de epidemiologia escritos em português, estas modalidades especiais de razão ou proporção têm sido chamadas às vezes de taxa, às vezes de coeficiente. Ora, se respeitarmos a terminologia emprestada da matemática e da física, de acordo com a crítica clássica de Elandt-Johnson (1975), não se trata nem de uma coisa nem de outra. Rigorosamente, uma taxa (que tem sido tomada como equivalente ao termo inglês *rate*) denota uma medida de variação, como, por exemplo, a velocidade em física cinética e as taxas de crescimento ou de inflação na econometria. O termo coeficiente, por sua vez, designa funções derivadas do cálculo, como, por exemplo, o coeficiente de resistência estrutural na física de materiais ou o coeficiente de penetração na aerodinâmica. A propósito, o termo *coefficient* praticamente não é usado no jargão epidemiológico em inglês, idioma de origem da Epidemiologia.

Não obstante, considerando a necessidade de padronização terminológica, porém respeitando um modo de uso já dominante no campo da saúde coletiva no Brasil, propomos adotar os termos “taxa” para designar macroindicadores epidemiológicos e “coeficiente” para microindicadores, parte por mera convenção e parte pelo caráter eminentemente pragmático dos indicadores epidemiológicos.

Assim, tomando os subconjuntos inclusivos da Figura 8.2,

- O/P = Taxas de mortalidade
- D/P = Taxas de incidência (e prevalência) de doença ou simplesmente prevalência (e incidência)
- I/P = Taxas de incidência (e prevalência) de infecção ou simplesmente prevalência (e incidência)
- D/I = Coeficientes de patogenicidade
- G/D = Coeficientes de virulência
- O/D = Coeficientes de letalidade

Os indicadores epidemiológicos típicos abordados neste capítulo, como por exemplo as taxas de prevalência, incidência e mortalidade, têm como denominador o conjunto mais abrangente, referido à população P. Indicadores especiais de morbidade, tais como os coeficientes de patogenicidade, virulência e letalidade, têm como denominadores subconjuntos de P que incluem infectados I (ou prediagnósticos, em se tratando de doenças não infecciosas) e doentes D.

Os indicadores epidemiológicos podem também ser classificados descritivamente, de acordo com o tipo de evento referido, como medidas de morbidade e medidas de mortalidade. Referem-se respectivamente aos subconjuntos da população formados por indivíduos que adquiriram doenças ou faleceram em um dado intervalo de tempo. Por motivos óbvios – o óbito é um evento pontual e absoluto, só é possível se estimar taxas de mortalidade tipo incidência.

Morbidade e mortalidade de fato constituem os principais indicadores empregados na Epidemiologia para abordar o estado de saúde das comunidades. Por esse motivo, merecem um tratamento mais aprofundado no Capítulo 9. Nesse capítulo, aborda-se a forma empírica dos indicadores epidemiológicos, detalhando fontes, características e instrumentos de coleta como meio de identificação de exemplares ou casos das variáveis dependentes.

Apesar da hegemonia de medidas de ocorrência de doenças, agravos e óbitos como sucedâneo (*proxy*) de indicadores de saúde, cresce na Epidemiologia uma vertente metodológica que busca quantificar a saúde e seus correlatos de modo positivo. Este tema será analisado e discutido no Capítulo 10, convergindo para uma apresentação sumária de indicadores compostos de saúde como qualidade de vida. O Capítulo 11 avalia, em uma perspectiva prática, o processo de produção de dados e de informação, mediante observação e registro de fenômenos epidemiológicos, ampliados com uma abordagem dos determinantes da saúde-doença. Fechando essa sequência de textos metodológicos, o Capítulo 12 pretende oferecer uma sistemática operacional para a montagem ou adaptação de instrumentos e técnicas de produção de dado epidemiológico, especialmente dispositivos literários (inventários, formulários, questionários etc.), já que a produção de dados clínicos será objeto do Capítulo 32, Parte 4.

▶ Hipóteses epidemiológicas

Em um certo sentido, a pesquisa em Epidemiologia busca sempre o teste de um tipo formal de hipótese: a de que uma dada variável de exposição constitui ou não fator de risco para certa patologia. Chamemos a essa forma geral de “hipótese epidemiológica”.

Na pesquisa populacional em geral, as técnicas utilizadas para a atribuição de valores numéricos ao grau ou nível com que variáveis se associam ou correlacionam são eminentemente estatísticas. No campo da Epidemiologia, porém, a estatística

- se afirmarem relações entre variáveis;
- se forem abertas à validação (ou refutação).

Além da função que as define, que é a de adiantar respostas tentativas a problemas novos ou revisitados, e como decorrencia dessa mesma função, as hipóteses, de algum modo, orientam e determinam a natureza dos dados a serem coletados e, portanto, a metodologia da pesquisa. Dados são produzidos ou colhidos para satisfazer a um objetivo, qual seja avaliar a validade da hipótese, buscando refutá-la.

A validação de hipóteses se dá mediante refutação das premissões ou consequências enunciadas por dedução. Tais premissões constituem enunciados menos gerais que, mesmo antes da sua explicitação, preexistiam virtualmente no enunciado da hipótese. Uma vez corroborada em alguma de suas consequências, a hipótese fica beneficiada por uma validação parcial. Nem mesmo a confirmação de muitas consequências fará da hipótese um conhecimento totalmente certo. A rejeição de uma só de suas premissões pode invalidar a hipótese por completo. De acordo com o eminente filósofo da ciência Sir Karl Popper (1985), o conhecimento científico avança quando uma hipótese mais abrangente substitui uma hipótese anterior, rejeitada através da investigação experimental ou observacional de suas consequências lógicas.

Ocorrem generalizações de pequeno alcance que, como de costuma, produzem um número bastante restrito de premissões. Podemos considerá-las como hipóteses pobres. Outras hipóteses, cientificamente mais fereis, trazem implicações em seu enunciado um número grande de consequências diretas ou indiretamente verificáveis. No extremo da generalização de curto alcance, encontram-se hipóteses cuja única verificação possível é definida pelos seus próprios termos.

No outro extremo, encontram-se hipóteses com *status* de teoria. Seu maior mérito é o de propiciar a produção de uma multiplicidade de premissões refutáveis, não só no campo do conhecimento em que foram pronunciadas, como também em disciplinas correlatas. São as consequências das hipóteses que fazem progredir o conhecimento, por sua capacidade de ajustar as teorias e orientar a prática. Estas constituem origens e finalidades dos experimentos e observações orientadas, que buscam validar ou refutar enunciados hipotéticos.

A formulação de hipóteses é etapa indispensável em qualquer pesquisa que se pretenda científica dentro dos paradigmas dominantes na ciência contemporânea. Por um lado, frequente-

mente encontramos “pesquisas” cujos dados foram coletados sem um objetivo-diretor do processo de produção de dados, isto é, sem hipóteses. Por outro lado, existem hipóteses engenhosas que passam por conhecimento estabelecido, faltando-lhes, no entanto, validação, seja de natureza experimental ou de ordem observacional. Em ambos os casos, o ciclo produtivo do conhecimento científico não foi devidamente completado.

As hipóteses podem ser originais, substitutivas ou dedutivas (Buck, 1975). Ao se buscar explicação para um fenômeno novo, inusitado, uma hipótese formulada e a seguir testada é original no sentido de ser a primeira a tentar esclarecer o problema. Além disso, a partir de teorias ou de hipóteses abrangentes, deduzem-se hipóteses menos gerais que são suas consequências. A medida do valor de uma hipótese são a qualidade e o número de consequências preditivas que dela podem ser deduzidas.

Quando o poder explicativo de uma hipótese já não é suficiente para esclarecer fenômenos novos, aparece a necessidade de se encontrarem hipóteses substitutivas. Uma hipótese substitutiva deve satisfazer a um dos seguintes critérios: (a) permitir predições mais precisas; (b) explicar maior volume de observações anteriores; (c) explicar com mais detalhes observações feitas previamente; (d) ser aplicável nas situações em que a hipótese anterior falhou; (e) indicar novas predições não sugeridas pela hipótese original; (f) relacionar ou unificar fenômenos que antes não estavam conectados.

Uma hipótese epidemiológica compreende um enunciado que propõe uma explicação para algum fenômeno relativo à distribuição ou determinação do surgimento de doentes em populações, através do relacionamento de variáveis que representam risco e fatores de risco. Ao ser formulada, a hipótese epidemiológica deve levar em consideração os aspectos da doença na população e as variações perceptíveis nos componentes ambientais (físicos, químicos, biológicos, sociais etc.) associados à exposição aos fatores de risco.

A hipótese epidemiológica, dependendo de seu alcance e generalidade, pode fazer emergir transformações em outras disciplinas, nos campos biológico, social e de outras ciências da saúde. Os estudos epidemiológicos referentes à distribuição das doenças são fundamentais na elucidação de mecanismos causais. As hipóteses, geradas com base nesses estudos, têm como meta última a explicação dos padrões de distribuição e, para isso, devem ter como objetivo imediato desvelar fatores de risco. Este é o contexto da exploração (ou modo de descobrimento, na terminologia samajiana).

Na análise epidemiológica, as hipóteses causais são primeiramente introduzidas para avaliar informações descritivas já recolhidas. Se, nessa etapa, a hipótese for considerada coerente com conhecimentos acumulados *a priori*, deverá então ser validada por meio de estudos planejados especialmente para esta finalidade. Todavia, a grande maioria dos estudos em Epidemiologia é observacional por motivos éticos (lidamos com pessoas humanas), lógicos (essas pessoas vivem em um contexto concreto) e operacionais (estudamos, em geral, grandes amostras e pequenos efeitos). Difícilmente são encontrados estudos de campo em Epidemiologia realizados com inclusão e controle do fator suspeito em um grupo experimental. Alguns estudos

podem ser executados com base na remoção do fator suspeito. Dessa forma, podemos dizer que o critério de verdade (ou de prova, como preferem os epistemólogos) em Epidemiologia é, em última análise, a eficácia e a efetividade em prevenção e controle. Este é o contexto da verificação ou o modo de validação, conforme Samaja (1994).

Para o teste das hipóteses epidemiológicas, deve-se necessariamente destacar os elementos empíricos do problema epidemiológico, fundamentados nos seus respectivos componentes teóricos. Por essa razão, o trabalho científico na Epidemiologia depende, por um lado, de modelos conceituais de saúde-doença capazes de orientar o processo de problematização na disciplina e, por outro lado, de intenso esforço de validação empírica das questões e hipóteses epidemiológicas através de um amplo repertório de estratégias metodológicas de pesquisa de campo, em laboratórios e em observatórios.

A metodologia epidemiológica será realizada mediante a operação de conjuntos complexos de planos, procedimentos, etapas, instrumentos e técnicas de produção e análise de dados, genericamente designados como “estratégias de pesquisa”, conforme apresentamos no Capítulo 13. Os textos seguintes desta Parte 3 (Capítulos 14 a 18) tratam em maior detalhe dos desenhos de pesquisa mais utilizados na investigação epidemiológica, respectivamente: Estudos Ecológicos (ou de Dados Agregados); Estudos Transversais; Estudos de Caso-Control; Estudos Longitudinais (ou de Coorte); Estudos de Intervenção ou Experimentais.

Enfim, aplicando de modo pertinente esse conjunto de lógicas, marcos conceituais, estratégias de pesquisa, instrumentos e técnicas de produção de dados, modos de análise e de interpretação, o método epidemiológico passa a ter existência específica como prática concreta de investigação científica sobre os fenômenos da saúde-doença-cuidado.

► Referências bibliográficas

- Almeida Filho N. Integração metodológica na pesquisa em saúde: nota crítica sobre a dicotomia quantitativo-qualitativo. In: Goldenberg P, Marsiglia RMG, Gomes MHA. *O clássico e o novo: tendências, objetos e abordagens em ciências sociais e saúde*. Rio de Janeiro: FIOCRUZ, 2003. p. 143-156.
- Bhaskar R. *A realist theory of science*. Hassocks: Harvester Press, 1978.
- Bhaskar R. *Scientific realism & human emancipation*. London: Verso, 1989.
- Buck C. Popper's philosophy for epidemiologists. *International Journal of Epidemiology* 4(3):159-68, 1975.
- Bunge M. *Epistemologia: curso de atualização*. São Paulo: TAQ/EDUSP, 1980.
- Chalmers I. *O que é ciência afinal?* São Paulo: Brasiliense, 1982.
- Elandt-Johnson R.C. Definition of rates: some remarks on their use and misuse. *Amer J Epidemiology* 102:267-71, 1975.
- Goldmann L. *Sciences Humaines et Philosophie*. Paris: PUF, 1988.
- Gregg N. Congenital cataract following German measles in the mother. *Trans Austr Ophthalmol* 3:35-6, 1941.
- Herbst A, Ufelfelder H, Poskanzer D. Adenocarcinoma of the vagina: association of maternal stilbestrol therapy with tumor appearance in young women. *New England Journal of Medicine* 284:878-81, 1971.
- Popper K. *Lógica da pesquisa científica*. São Paulo: EDUSP, 1985.
- Rey A. *Dictionnaire Historique de la Langue Française*. Paris, Dictionnaires Le Robert, 1993.
- Samaja J. *Epistemología y metodología*. Buenos Aires: Eudeba, 1994.
- Snow J. *Sobre a maneira de transmissão do cólera*. (Tradução do texto de 1854.) São Paulo: Hucitec, 1994.