

Como **unidade de análise** do dado entende-se o objeto da descrição, classificação, enumeração ou medida. Isto é, o que ou quem se pretende descrever, classificar, contar ou medir: os indivíduos ou as pessoas, as instituições ou organizações, os países, territórios ou ambientes, os recursos financeiros, as subunidades simples com as quais se trabalha frequentemente em Epidemiologia. Por exemplo, o dado que representa o número dos que foram acometidos de dengue pode ser obtido pela enumeração de pessoas atendidas em um serviço de emergência clínica que tiverem o diagnóstico da doença.

Em uma pesquisa científica ou nas práticas da derivadas, a estrutura do dado em geral compreende pelo menos três elementos: a *unidade de análise*, a *variável* e o *valor ou quantidade* atribuído à variável como resultado da quantificação ou da medida (Alazraqi, Mota & Spinelli, 2006). O dado epidemiológico de um estudo epidemiológico, tais elementos são cuidadosamente examinados para todos os dados necessários à obtenção dos indicadores de condições de saúde, de morbidade ou de mortalidade, entre outros, que atendam ao cumprimento dos objetivos. A unidade de análise, a variável e o valor do dado formam um todo coerente.

## ► A estrutura dos dados

do sob a forma de informação qualificada. A análise e a interpretação agregam valor, contextualizam e referenciam os dados, progressivamente, na extensão em que avançam, e deles se extrai uma representação mais completa e significativa da realidade. É por essa razão que a *qualidade* dos dados é tão essencial à produção de informações. O que se pretende, por exemplo, ao registrar o registro de todos os casos de uma doença atendidos em um serviço de saúde ou identificados por um inquérito domiciliar é poder aplicar o dado apurado ao número de casos – ao conhecimento e entendimento da ocorrência do problema de saúde na população, isto é, saber quantos, onde, quando, como e por que ocorreram os casos da doença. Por isso, é importante conhecer como os dados se estruturam.

dos. Uma vez que sejam analisados, têm seu significado ampliado.

*Eduardo Mota, Marcia Furgum de Almeida e Francisco Viacava*

# O Dado Epidemiológico: Estrutura, Fontes, Propriedades e Instrumentos

## 9

A informação epidemiológica é produzida a partir de dados. Os dados epidemiológicos são produzidos com a finalidade de descrever, acompanhar e comparar características de populações, grupos de indivíduos e coletividades humanas no que ateta a saúde, o bem-estar e a qualidade de vida, bem como de permanecer da ocorrência e distribuição dos eventos de saúde. A estrutura do dado em Epidemiologia constitui a base sobre a qual se formulam adequadamente objetivos, hipóteses e metodologias dos estudos. Há inúmeros dados que interessam ao raciocínio epidemiológico, dirigidos à descrição e entendimento de problemas e situações de saúde, que resultam da aplicação de amplo espectro de métodos das ciências da saúde, das ciências humanas e sociais e das ciências biológicas, entre outras.

Neste capítulo, serão estudadas as características dos dados produzidos pela aplicação dos métodos epidemiológicos. Veremos como se estruturam e como se produzem dados a partir de observações a partir das principais fontes e instrumentos de coleta aplicados à informação em saúde.

Dados epidemiológicos podem ser enumerados (contagens) ou quantidades: uma medida, uma frequência ou uma distribuição numérica de atributos em categorias previamente definidas. Entretanto, o conceito corrente de dado como a expressão de um valor quantitativo não trabalhado (Mota & Carvalho, 2003), isto é, sem ter sido submetido à análise e interpretação, não traduz por completo a extensão do seu papel na construção da informação epidemiológica. O dado representa, ainda que parcialmente, a realidade que se pretende revelar. O número de óbitos infantis ou a frequência absoluta de casos de dengue, por exemplo, podem ser considerados dados tanto quanto simplesmente enumerem esses eventos. Mas esses, ao mesmo tempo, apresentam aspectos da mortalidade e da morbidade que trazem significados: a enumeração de óbitos infantis indica quantas crianças morrem antes de completar 1 ano de idade, e a frequência de casos notificados de dengue oferece igualmente uma noção da magnitude desse problema de saúde, que fazem parte da experiência humana em coletividade, mesmo que estejam fora de um contexto explicativo. Esses dados podem ser aplicados ao cálculo de coeficientes, indicadores e índices, referidos a uma população, em um território e tempo determinada.

Características da unidade de análise quanto à natureza daquilo que expressa possibilitam classificar os dados epidemiológicos em dois grandes grupos: *individuado* e *agregado*. O dado é classificado como *individuado* quando a contagem ou a medida a ser aplicada à análise é registrada para cada indivíduo singularmente, como quando se obtém a frequência de pessoas que têm um atributo, e, nesse caso, cada dado representa a realidade de cada pessoa e esta será a unidade de análise do dado. O dado é dito *agregado* quando se obtém valores que expressam coletivos (população de um município, residências em um distrito sanitário, famílias). Por exemplo, em um estudo que objetive comparar vários municípios que notificaram casos de dengue em 1 ano, o total de casos notificados será um dado agregado para cada município. Aqui a unidade de análise do dado é município e o resultado da comparação será válido somente para os municípios participantes do estudo. Realizam-se estudos epidemiológicos com propósitos e metodologias distintos com cada tipo de unidade de análise do dado.

A **variável ou dimensão** constitui o elemento do dado que qualifica e especifica o objeto da contagem (a unidade de análise). Aqui se definem atributos e características do que se pretende descrever, classificar, quantificar ou medir. No exemplo anterior, a presença da doença “dengue” se refere aos sinais, sintomas e exames que possibilitam diagnosticá-la e que sejam considerados “típicos” dessa doença (e de nenhuma outra); é a variável daquele dado. Para orientar o diagnóstico do problema de saúde, uma *definição de caso* é necessária. Idade, sexo, cor da pele ou etnia, escolaridade, ocupação, nascimentos e mortes, tipos de profissionais de saúde e de serviços (p. ex., hospitais universitários, creches públicas), tipos de poluentes ambientais etc. são exemplos de variáveis que qualificam diferentes unidades de análise de dados. A variável pode ser classificada em *quantitativa* (p. ex., glicemia), que é essencialmente numérica (ordinal, discreta ou contínua), ou *qualitativa*, conhecida também como categórica ou categorizada (p. ex., sexo, ocupação, escolaridade, nível pressórico alto, tabagismo, alcoolismo). Tendo natureza simples ou complexa, na dependência de como seja definida, a variável representa sempre um atributo da unidade de análise. O conceito de variável epidemiológica e seus correlatos foram apresentados no capítulo anterior.

O **valor** atribuído à variável é a própria essência da medida, o seu resultado numérico ou quantificável, e a finalidade do registro do dado. Dito de outra maneira, é o elemento do dado que se aplica diretamente aos cálculos dos parâmetros, indicadores e índices. A simples enumeração requer que o valor quantitativo obtido expresse o número de unidades de análise que apresentaram um dado atributo. Por exemplo, em um estudo agregado, o “número” de casos acometidos de dengue representará o “valor” atribuído à variável “de dengue” no município ou bairro considerado. Isto é, corresponderá à frequência absoluta de “indivíduos” que se julgou estarem afetados pela doença. Por outro lado, em um estudo *individuado*, o “diagnóstico” de dengue representará o “valor” atribuído à variável “dengue” aos indivíduos atendidos. Em resumo:

Dado	Unidade de análise = Indivíduo	Variável ou dimensão = Doença dengue	Valor ou quantidade = Número de casos da doença
------	-----------------------------------	---	--

Cada variável requer um valor típico ou que lhe é próprio: a glicemia, por exemplo, poderá ser expressa em *miligramas por 100 ml de glicose no sangue*. Com os valores das dosagens de

glicemia será possível classificar um grupo de indivíduos de acordo com a presença ou ausência de valores “elevados” (segundo a adoção de um ponto de corte ou valor máximo de tolerância) e, dessa maneira, saber quantos indivíduos diabéticos existem no grupo estudado (*i. e.*, quantos têm valores elevados de glicemia), dado esse que terá diferentes aplicações no cálculo de coeficientes de morbidade e mortalidade. Os valores são definidos em uma dada *unidade de medida*, isto é, aquilo que especifica o valor, e segundo uma *escala de medida* correspondente, isto é, a variação admitida aos valores (p. ex., limites máximo e mínimo). Tratando-se do peso ao nascer, um dado importante que possibilita calcular indicadores das condições de saúde ao nascimento, se poderá medi-lo em gramas e os valores poderão variar em certo estudo na escala de 1.000 a 6.000 gramas para crianças que nasceram vivas.

A unidade de medida e a escala são por sua vez atributos do *instrumento* utilizado para obter ou coletar os dados; no caso do peso ao nascer será uma balança antropométrica, e para gerar e registrar o valor da medida é necessário conhecer sob que formato, e em que condições e limites esse instrumento é capaz de fazê-lo. Instrumentos de coleta devem ser, nos limites do conhecimento e da tecnologia conhecidos, perfeitamente adequados aos dados que se deseja obter. Em outro exemplo, pretende-se realizar a contagem de “indivíduos adultos que fizeram pelo menos uma consulta odontológica nos últimos 12 meses”. Isto poderá ser feito por inquérito domiciliar, aplicando-se um questionário (instrumento com a pergunta específica) em entrevista aos participantes de um estudo. Nesse caso, a unidade de medida será “realizou/não realizou consulta”, em geral codificada, respectivamente, como 1 = sim, 0 = não, e a escala será numérica simples com variação de zero a “n” (tamanho do grupo de estudo), expressando a frequência de respostas “sim”. Convém lembrar que, nesse exemplo, a unidade de análise é “indivíduo participante” e a variável é “consulta odontológica”, além de outras variáveis (idade dos participantes, período de referência etc.).

Para entender melhor a estrutura dos dados, suponhamos que se pretenda realizar um “estudo epidemiológico sobre as principais causas de hospitalização de indivíduos idosos em uma região em um dado ano”. No planejamento do estudo será necessário incluir a obtenção de “dados de hospitalizações” registradas naquele ano, em consulta a um sistema de informação dos serviços de saúde da região. Certamente serão adotadas algumas definições prévias à coleta de dados: a unidade de análise poderá ser “indivíduo hospitalizado” ou “internação realizada” (considerando que uma pessoa poderá ser hospitalizada mais de uma vez no ano); a variável de interesse é a “causa” da hospitalização, uma vez que o dado a ser obtido será consolidado segundo a causa ou diagnóstico, e aqui uma classificação padronizada de diagnósticos deverá ser utilizada; o valor ou quantidade será a frequência com que cada causa foi registrada entre as internações hospitalares ocorridas no período e na área definida. Outros atributos poderão ser levados em consideração nesse caso: a idade dos indivíduos hospitalizados será igual ou maior que 60 anos, o sexo, o mês da internação etc.

Retornando à questão da qualidade dos dados, note-se que para a unidade de análise a limitação e o desafio são de ordem conceitual (Alazraqui, Mota & Spinelli, 2006). Definir apropriadamente, ou melhor, coerentemente, as unidades de análise dos dados é algo que se realiza desde a escolha do tema do estudo, se continua na definição dos objetivos e hipóteses e se completa nas opções metodológicas. No exemplo acima, se o objetivo for “descrever o perfil de morbidade de indivíduos idosos” pelo estudo das causas de hospitalização, certamente a unidade de análise não será representada por “crianças atendidas na emer-

seu uso representa um grande desafio à análise e à interpretação dos dados secundários. Dado em saúde podem ser coletados ou obtidos de maneira contínua ou episódica e, nesse último caso, com ou sem periodicidade definida. Dados episódicos podem ser obtidos em uma pesquisa transversal para a descrição da situação de saúde, em um estudo retrospectivo ou em pesquisa longitudinal para o acompanhamento de um grupo populacional com o objetivo de estudar fatores de risco e determinantes de doenças e agravos à saúde. Por coleta contínua entende-se o processo de registro sequencial e universal de dados estabelecido institucionalmente e, em geral, de cobertura nacional ou integral para um território. Fazem parte do conjunto de dados de registro contínuo as estatísticas do registro civil (como nascimentos, óbitos, casamentos, divórcios e separações), as notificações e registros de casos de doenças e agravos à saúde e os assim chamados registros administrativos da assistência à saúde.

## Dados de registro contínuo

Agências e órgãos governamentais são responsáveis pela geração de estatísticas oficiais sobre nascimentos e óbitos, cujos dados provêm de atestados e certidões emitidas pelos cartórios de registro civil. O uso desses dados para fins epidemiológicos é limitado ao tipo de informação disponível relacionada com os eventos, em geral restrita às variáveis: sexo, idade, local de residência e de ocorrência do evento. No caso dos nascimentos, dada a frequência importante de registros feitos em anos posteriores ao evento, é preciso considerar o registro tardio. As estatísticas vitais baseadas em registros de nascimentos e óbitos são de responsabilidade nacional do Instituto Brasileiro de Geografia e Estatística (IBGE).<sup>1</sup> Essa instituição, que realiza o censo populacionais, responde também por todos os dados de registros oficiais brasileiros e por outras informações socioeconômicas e ambientais de interesse, sendo responsável pela coordenação do Sistema Estatístico Nacional de acordo com a Lei n.º 6.183 de 1974. Entretanto, com a ampliação dos sistemas de informação do Ministério da Saúde e a realização de inquéritos, há dados de interesse em saúde obtidos ainda por outros órgãos governamentais, não se constituindo ainda no país um sistema nacional de informações em saúde que possibilite uma visão abrangente e integrada de todos os dados do setor (Viacava, Dachs & Travassos, 2006).

O registro contínuo de dados sobre doenças ou agravos de notificação compulsória, no Brasil e na maioria dos países, é de responsabilidade do Ministério da Saúde. Entretanto, a seleção das doenças e agravos que são considerados de notificação obrigatória varia de país para país. No caso brasileiro fazem parte dessa seleção quase 40 tipos de doenças transmissíveis que são

gências. Para as variáveis e os seus valores, outros desafios da qualidade dos dados se apresentam: a disponibilidade ou o acesso aos dados, o preenchimento adequado dos instrumentos de coleta, o registro correto das alternativas de resposta etc., aspectos que serão abordados mais adiante.

## Características e fontes dos dados

Com propriedade diz-se que a Epidemiologia interessam todos os dados que direta ou indiretamente contribuam para revelar o quadro sanitário da população, possibilitando o entendimento dos processos saúde-doença-cuidado. Com efeito, são aplicados ao cálculo de indicadores epidemiológicos distintos dados que se apresentam com amplo espectro de características, como se verá adiante, e que se originam de fontes diversas. Quanto às fontes primárias, destacam-se os próprios indivíduos participantes das investigações que são examinados ou entrevistados, como é o caso dos inquéritos e dos estudos observacionais em Epidemiologia. Ademais, o registro sistemático das atividades de profissionais de saúde em serviços diversos e quando da condução de ações e intervenções constituem fontes especiais de dados, como no caso de dados de registro contínuo em saúde que alimentam sistemas de informação. Nesse último tipo se incluem também os dados de cartórios do registro civil e os que são registrados em prontuários clínicos, nos arquivos de laboratórios e nos serviços de medicina legal, entre outros. Formulários que captam dados que são registrados nos sistemas de informação em saúde são fontes de dados epidemiológicos de uso constante, com destaque para a Declaração de Óbito, a Declaração de Nascimento e a Ficha de Notificação de casos de doenças e agravos à saúde.

Uma importante fonte de conhecimentos sobre os dados de interesse em saúde no Brasil pode ser encontrada nos produtos da Rede Interagencial de Informações para a Saúde (RIPSA), formada por iniciativa do Ministério da Saúde (MS) e da Organização Pan-Americana da Saúde. A RIPSA disponibiliza, via internet, um grande conjunto de dados e informações com a publicação *Indicadores e Dados Básicos para a Saúde no Brasil* (Risi Junior, 2006; RIPSA, 2008). Nesta se destacam as fichas de qualificação dos indicadores com informações úteis sobre as características dos dados, fontes e limitações de uso.

Além dos aspectos já examinados sobre a estrutura dos dados dos que os caracterizam como epidemiológicos, outros devem ser conhecidos, e serão apresentados aqui de acordo com os tipos básicos de procedimentos de coleta, isto é, da própria organização das atividades de registro de dados. Esses processos, relativos à finalidade básica da obtenção dos dados, devem ser conhecidos porque dependem do uso que se pode fazer, e de como lidar com suas limitações e potencialidades nos estudos epidemiológicos.

De maneira geral, os dados podem ser classificados como primários ou secundários, segundo a relação desses com quem investiga, os obtêm ou os aplica. Dados primários são aqueles obtidos diretamente por quem formulou e escolheu os métodos, planejou a coleta e/ou participou de sua análise; os secundários – em estudos observacionais – possibilita o controle da qualidade dos dados e sua melhor adequação ao cumprimento dos objetivos do estudo. Como dados secundários denominam-se aqueles coletados por outros pesquisadores e que estão à disposição para os estudos epidemiológicos. Nesse caso, os métodos utilizados na coleta dos dados foram definidos em processos já estabelecidos, e com finalidade própria, e não podem ser modificados por nenhuma pessoa que deseje utilizá-los. Assim,

<sup>1</sup> Consultar o endereço eletrônico <http://www.ibge.gov.br/home>.

objeto das atividades de vigilância epidemiológica. Do Sistema Nacional de Vigilância Epidemiológica participam órgãos do Ministério da Saúde (Secretaria de Vigilância em Saúde) e das secretarias de saúde de todos os estados e municípios. O registro é feito pelo Sistema de Informação de Agravos de Notificação (SINAN) a partir de dois documentos com uma variável identificadora única:

- a) Ficha Individual de Notificação (FIN), que contém um conjunto padronizado de dados para cada doença (identificação do caso, dados sobre a unidade assistencial notificadora, informações mínimas sobre o agravo de notificação etc.) e que é encaminhada pelas unidades de saúde a partir da suspeita clínica da ocorrência de algum dos agravos notificáveis;
- b) Ficha Individual de Investigação (FII), que se configura, na maior parte das vezes, como um roteiro de investigação epidemiológica, distinto por tipo de agravo.<sup>2</sup>

Informações valiosas para os epidemiologistas, como a causa básica dos óbitos ou as circunstâncias em que ocorreram os nascimentos, podem ser obtidas, no Brasil, a partir de consultas aos dados do Sistema de Informações sobre Mortalidade (SIM) e do Sistema de Informações sobre Nascidos Vivos (SINASC), ambos sob responsabilidade do Ministério da Saúde e que também são de registro contínuo. No SIM, os dados provêm de cópia das Declarações de Óbito (DO) emitidas por médicos e captadas nos cartórios de registro civil e pelas secretarias de saúde em grande parte dos municípios (há municípios com outros fluxos de dados). Além da especificação da causa básica e das causas associadas do óbito, e das características de óbitos fetais, é possível encontrar também dados de idade, sexo, raça/cor da pele, ocupação, local de ocorrência e de residência dos indivíduos que faleceram e, mais recentemente, informações sobre sua escolaridade. No SINASC, os dados originais constam da Declaração de Nascido Vivo (DN), feita na grande maioria dos casos pelas maternidades e encaminhadas às secretarias de saúde. Dentre as informações coletadas na DN constam dados da mãe (idade, raça/cor da pele, estado civil, escolaridade, ocupação, município de residência, número de filhos vivos e mortos); dados da gestação (semanas de gestação, tipo de gravidez, número de consultas de pré-natal; data do nascimento); e dados da criança (Apgar, peso, raça/cor da pele e presença de malformação congênita). Além disso, são coletados dados do estabelecimento de saúde onde foi realizado o parto, incluindo o endereço completo e o código do estabelecimento, bem como sobre o responsável pelo preenchimento da declaração (nome, função, identidade, órgão emissor e data de emissão) (Mello-Jorge, Laurenti & Gottlieb, 2007).

Além desses, são também de registro contínuo dados sobre outros agravos e doenças como: neoplasias,<sup>3</sup> acidentes de trabalho e benefícios da previdência social por motivos de saúde,<sup>4</sup> acidentes de trânsito<sup>5</sup> e intoxicações e envenenamentos, entre outros. Em cada um desses existem registros mantidos por órgãos governamentais com graus diferenciados de cobertura nacional, disponibilização e acesso. No caso das neoplasias há duas bases importantes: o Registro de Câncer de Base Populacional e o Registro Hospitalar de Câncer, sob a coordenação do

Instituto Nacional do Câncer (INCA) do Ministério da Saúde. Enquanto o primeiro é orientado para uma abordagem epidemiológica considerando os casos existentes em uma determinada área geográfica, o registro hospitalar contém informações relacionadas com a qualidade da atenção hospitalar.

Os dados sobre os acidentes de trabalho são comunicados pelas empresas ao INSS e registrados na DATAPREV no formulário de Comunicação de Acidentes de Trabalho (CAT). Esses devem conter a qualificação do segurado (nome, endereço, data de nascimento e filiação materna), identificação do empregador, causa do acidente (registrada com o código da Classificação Internacional de Doenças [CID] 10ª Revisão), tipo e data do acidente e ocorrência de óbito. Outros dados sobre benefícios e aposentadorias da previdência social, inclusive os concedidos por motivos de saúde, são registrados pelo Ministério da Previdência Social.

Por sua vez, o DENATRAN mantém um Sistema Nacional de Estatísticas de Trânsito (SINET) onde estão registrados os dados sobre acidentes de trânsito segundo várias características, como: tipo de colisão, com ou sem vítimas, tipo de veículo, qualificação das vítimas (condutor, passageiro, pedestre, ciclista, motociclista), horário do acidente, sexo e idade das vítimas, idade e habilitação dos condutores. Esses e os dados estaduais sobre acidentes de trânsito não têm sido usados com frequência em pesquisas em saúde, mas apresentam bom potencial de uso, a exemplo da análise realizada recentemente por Oliveira *et al.* (2008).

Dados de registro contínuo sobre intoxicações e reações adversas ao uso de produtos para a saúde podem ser obtidos, respectivamente, do SINITOX,<sup>6</sup> sob gestão da FIOCRUZ e do NOTIVISA, da Agência Nacional de Vigilância Sanitária (ANVISA),<sup>7</sup> que são órgãos do Ministério da Saúde.

Os dados da assistência médico-odontológica no Brasil são registrados em sistemas de informação próprios que por sua natureza relacionada com o pagamento de prestadores de serviços de saúde se tornaram conhecidos como *dados administrativos*, de uso voltado às atividades gerenciais em saúde. Porém, se constituem em um importante conjunto de dados para análises epidemiológicas e também são, por definição, de coleta contínua. Nesse grupo se incluem as bases de dados sobre internações hospitalares e atendimentos ambulatoriais,<sup>8</sup> oferta de serviços e disponibilidade de recursos humanos<sup>8</sup> sob a responsabilidade nacional do Ministério da Saúde [Departamento de Informática do SUS (DATASUS)].

As hospitalizações no setor público e no setor privado contratado pelo SUS têm seus dados originados da Autorização de Internação Hospitalar (AIH), documento usado para faturamento/registro das internações onde constam informações demográficas (sexo e idade), dados clínicos (diagnóstico principal e secundário, tempo de internação, número de diárias em unidade de terapia intensiva, procedimentos solicitados e realizados, condição de saída do paciente – inclusive óbitos hospitalares) e outras informações, como, por exemplo, especialidade médica, motivo da hospitalização, instituição e profissional responsável pela assistência ao paciente, valores pagos, município de internação e de residência. Os dados das AIH alimentam o Sistema de Internações Hospitalares do SUS (SIH/SUS). Algumas limitações de uso desses dados devem ser conhecidas para sua adequada aplicação aos estudos de morbimortalidade: existem

<sup>2</sup> Consultar "Sistema de Informação" em [www.saude.gov.br/svs](http://www.saude.gov.br/svs).

<sup>3</sup> Ver [www.inca.gov.br](http://www.inca.gov.br).

<sup>4</sup> Consultar [www.dataprev.gov.br](http://www.dataprev.gov.br).

<sup>5</sup> Consultar RENAEST/DENATRAN para estatísticas do SINET em [www.denatran.gov.br](http://www.denatran.gov.br).

<sup>6</sup> Consultar [www.fiocruz.br/sinitox](http://www.fiocruz.br/sinitox).

<sup>7</sup> Consultar [www.anvisa.gov.br](http://www.anvisa.gov.br).

<sup>8</sup> Consultar [www.datasus.gov.br](http://www.datasus.gov.br).

populacionais fornecem indicadores como leitos hospitalares por mil habitantes etc.

O uso dos dados dos registros contínuos de saúde em estudos epidemiológicos tem crescido em importância. Avaliações dos sistemas nacionais de informação em saúde têm apontado a utilidade desses dados em inúmeras aplicações de interesse na vigilância epidemiológica, nos estudos transversais sobre desigualdades em saúde, na descrição dos perfis de morbidade e mortalidade, entre outros, com resultados válidos, mesmo considerando as limitações de cobertura e confiabilidade dos dados. Completa revisão desses estudos para o SIH/SUS foi publicada recentemente e há outros estudos e publicações que se referem igualmente aos dados do SIM, do SINASC e de outros sistemas de informações em saúde (Mendes *et al.*, 2000; Silva *et al.*, 2001; Bittencourt, Camacho & Leal, 2006; Mello-Jorge, Laurenti & Gottlieb, 2007; Rede, 2008; Laurenti, Mello-Jorge & Gottlieb, 2008).

## Dados de registro periódico

Há um conjunto de bases de dados compostas por dados coletados, episódica ou periodicamente, em pesquisas amostrais, como os inquéritos domiciliares sobre situação de saúde ou nos levantamentos que cobrem todo o universo populacional, como os censos demográficos. Os inquéritos domiciliares se distinguem das bases de dados administrativos em alguns aspectos. Em primeiro lugar coletam dados sobre morbidade percebida, referida pelos entrevistados e, portanto, não são baseados em diagnósticos médicos. Um segundo aspecto que caracteriza essas bases é que os dados geram informações de acordo com contextos familiares e sociais diferentemente da informação médica gerada pelos registros administrativos ou contínuos em saúde. Por último, mas não menos importante, é preciso ter em conta que enquanto nos registros administrativos os dados referem-se à população usuária dos serviços públicos de atenção à saúde, nesse caso a informação é de base populacional e considera todas as pessoas, sejam elas usuárias ou não desses serviços.

No Brasil, os inquéritos domiciliares constituem o formato mais comumente utilizado na área da saúde, mas existem também pesquisas amostrais realizadas em instituições como escolas, a exemplo dos Inquéritos Nacionais sobre Saúde Bucal, realizados em 1986, 1996 e 2002-2003. Os dados desses inquéritos e de pesquisas amostrais citadas adiante podem ser consultados nos endereços eletrônicos do IBGE e dos órgãos do Ministério da Saúde.

O inquérito de saúde com maior abrangência geográfica e periodicidade é o suplemento-saúde da Pesquisa Nacional por Amostra de Domicílios (PNAD),<sup>13</sup> que foi a campo pela primeira vez em 1981. O suplemento continha um amplo questionário que abordou diversos temas, incluindo morbidade, uso de serviços, cobertura vacinal, cuidado à mãe e à criança, hospitalização, assistência odontológica, portadores de deficiências e de incapacidade física, gastos privados em saúde e fontes de financiamento do consumo de serviços de saúde.

Uma versão menor e modificada do suplemento de 1981 acompanhou a PNAD de 1986, e naquela ocasião foram coletados dados sobre morbidade referida, utilização de serviços, suplementação alimentar e anti-concepção. Na PNAD de 1988, algumas questões sobre uso de medicamentos e uso de serviços públicos de saúde foram incorporadas ao suplemento Partici-

limites de financiamento do SUS com recursos federais para os estados e municípios e disto resulta um número de AIH apre-sentadas e outro número das faturas, isto é, efetivamente pagas; a AIH é essencialmente um documento relativo ao principal procedimento realizado na internação e ainda que existam campos destinados ao diagnóstico principal e secundário, condições de acordo com a CID, esses podem não ter a confiabilidade necessária para estudos de morbidade; outro cuidado que deve ser tomado, especialmente nos estudos de mortalidade de hospitalar, refere-se ao fato de que as admissões não geram AIH quando o período de internação for menor que 24 h. As internações hospitalares realizadas no sistema privado e financiadas pelas operadoras de planos de saúde ou diretamente pelos usuários não fazem parte desse sistema. A Agência Nacional de Saúde Suplementar (ANS/MMS) tem obtido dados consolidados dos atendimentos clínicos neste setor e mantém sistemas de informação sobre a população beneficiária e operadoras dos planos privados de saúde.<sup>9</sup>

Os atendimentos ambulatoriais realizados no âmbito do SUS são atualmente registrados no Sistema de Informações Ambulatoriais do SUS (SIA/SUS), e os procedimentos relacionados com a diálise renal, transplantes, radio e quimioterapia, entre outros, fazem parte da Autorização para Procedimentos de Alta Complexidade (APAC). Outros dados desse tipo são registrados em outros sistemas de informação, como os da atenção básica no Sistema de Informação sobre Atenção Básica (SIAB).<sup>10</sup> Dados de registro contínuo de interesse em saúde incluem também aqueles produzidos no âmbito do Programa Nacional de Imunizações. Os registros de doses aplicadas de todas as vacinas em todos os serviços públicos de saúde convergem para o Sistema de Informações do Programa Nacional de Imunizações (SI-PNI),<sup>11</sup> um sistema de informação gerido pela Secretaria de Vigilância em Saúde do Ministério da Saúde em conjunto com as secretarias estaduais e municipais de saúde de todo o país. Os serviços municipais e estaduais coletam os dados no Boletim Diário de Doses Aplicadas de Vacinas, com os quantitativos de indivíduos vacinados por faixa etária e por tipo de vacina, consolidando essas informações para registro nesse sistema de base nacional.

Nos últimos anos, alguns sistemas de informação foram criados pelo Ministério da Saúde para o registro contínuo de dados sobre problemas de saúde específicos e de dados gerenciais, seja por sua importância epidemiológica, seja pelas atividades de controle e gestão associadas. Nesse grupo se incluem, entre outros: dados sobre o câncer do colo de útero (SISCOCO), os atendimentos a pacientes com hipertensão e diabetes (HIPERDIA), do controle do câncer de mama (SISMAA), o controle das doenças sexualmente transmissíveis e AIDS, a vigilância nutricional (SISVAN), a atenção pré-natal (SISPRENATAL) e os orçamentos públicos em saúde (SIOPS).<sup>12</sup>

O Cadastro Nacional de Estabelecimentos de Saúde (CNE/S) também constitui uma importante base de registro contínuo e capta dados sobre a oferta de serviços e recursos humanos envolvidos, permitindo caracterizar a rede de serviços de saúde disponíveis e sua complexidade ao fornecer o número e tipo de estabelecimentos. Podem-se obter o número de unidades básicas de saúde, número de leitos, leitos de UTI e outros dados que ao serem trabalhados em conjunto com dados po-

<sup>9</sup> Consultar [www.ans.gov.br](http://www.ans.gov.br).

<sup>10</sup> Consultar <http://siab.datasus.gov.br/SIAB/default.php?area=03>.

<sup>11</sup> Consultar <http://pni.datasus.gov.br>.

<sup>12</sup> Consultar <http://w3.datasus.gov.br/datasus/datasus.php>.

<sup>13</sup> Ver em [www.ibge.gov.br](http://www.ibge.gov.br).

pação Político-Social. Após um intervalo de 10 anos, um novo suplemento sobre acesso e utilização de serviços foi aplicado. Com esse suplemento, inaugurou-se uma série histórica quinzenal (1998, 2003 e 2008) de informação nacional de base populacional sobre saúde e consumo de serviços, contemplando os seguintes aspectos: morbidade, cobertura por plano de saúde, acesso e utilização de serviços, limitação de atividades físicas para maiores de 13 anos e gastos privados com saúde. A partir de 2008 o suplemento passou a contar com uma seção sobre tabagismo e sedentarismo. A PNAD, como outros inquéritos, coleta dados individuados e agregados (módulo domiciliar).

A Pesquisa Nacional sobre Demografia e Saúde (PNDS) é outro inquérito domiciliar realizado periodicamente. Sua primeira versão, elaborada pela Sociedade Civil Bem-estar Familiar no Brasil (BEMFAM) em 1986, obteve em nível nacional dados sobre o comportamento reprodutivo das mulheres em idade fértil, incluindo o planejamento familiar. Também incluiu informações sobre serviços de saúde materno-infantil, mortalidade, amamentação, nupcialidade, estado nutricional e antropometria de crianças menores de 5 anos (Arruda *et al.*, 1986). Em 1991, a mesma pesquisa foi realizada apenas na região Nordeste, incluindo, além dos aspectos cobertos pela pesquisa de 1986, conhecimento sobre doenças sexualmente transmissíveis e AIDS (BEMFAM, 1992). A Pesquisa Nacional sobre Demografia e Saúde de 1996 (PNDS-96) deu continuidade às que foram realizadas no marco do programa de Pesquisas de Demografia e Saúde, conhecidas pela sigla DHS (*Demographic and Health Survey*). Na versão daquele último ano, além da inclusão da população feminina, foi também considerada uma subamostra de 25% dos domicílios selecionados para registrar dados da população masculina sobre conhecimento, atitudes e práticas relacionadas com o planejamento familiar, intenções reprodutivas, comportamento sexual em face da AIDS, entre outras. Em 2006, uma nova pesquisa foi financiada pelo Ministério da Saúde e realizada por um consórcio de instituições coordenado pelo Centro Brasileiro de Análise e Planejamento (CEBRAP).<sup>14</sup>

Os gastos com saúde também são periodicamente coletados pelo IBGE através da Pesquisa de Orçamento Familiar (POF). Os dados domiciliares são anotados em cadernetas e os entrevistadores visitam os domicílios durante 1 semana, permanecendo 12 meses em campo para controlar o efeito sazonal do padrão de consumo. No caso específico do gasto em saúde, há detalhamento para medicamentos, plano de saúde, consultas, exames e internações, e aquisição ou aluguel de aparelhos médicos. A pesquisa já foi realizada em 1974/75 (ENDEF), 1987/88, 1995/96, 2002/2003 e 2007/2008.

Mais recentemente, inquéritos vêm sendo realizados no âmbito do Ministério da Saúde que geram estimativas populacionais sobre doenças e agravos não transmissíveis e fatores comportamentais de risco associados. Em 2003, o Instituto Nacional do Câncer (INCA) do Ministério da Saúde conduziu o Inquérito Domiciliar sobre Comportamentos de Risco e Morbidade Referida de Doenças e Agravos Não Transmissíveis, realizado nas capitais. Outra iniciativa importante do Ministério da Saúde é o VIGITEL, inquérito domiciliar com entrevistas feitas por telefone em todos os municípios das capitais. Em cada domicílio é sorteado um adulto com 18 anos ou mais que responde sobre características sociodemográficas, alimentação, atividade física, tabagismo, consumo de álcool e outras drogas, imagem corporal, saúde bucal, comportamento sexual, violência e rede de proteção, e medidas antropométricas. Além dessas pesquisas

também foram realizados no Brasil outros inquéritos domiciliares episódicos, como a Pesquisa Nacional sobre Saúde e Nutrição (PNSN), a Pesquisa sobre Padrões de Vida (PPV)<sup>15</sup> e a Pesquisa Mundial de Saúde – Brasil (PMS),<sup>16</sup> entre outras de abrangência nacional e as que são realizadas em grandes centros urbanos por iniciativas de alguns municípios.

Uma extensa revisão publicada por Viacava (Viacava, 2002) sobre os inquéritos populacionais no Brasil e em outros países dá a dimensão da importância dos dados de registro periódico para o conhecimento da situação de saúde. Em outro trabalho (Viacava, Dachs & Travassos, 2006), discute-se a necessidade de integrar as iniciativas do registro periódico de dados em saúde e dos inquéritos de base populacional com os sistemas de informação para registro contínuo de dados, tal que conformassem um sistema nacional de estatísticas de saúde. No Quadro 9.1, apresentam-se informações sobre os principais inquéritos domiciliares de abrangência nacional no Brasil.

## ► Instrumentos de coleta e qualidade dos dados

Existe uma diversidade de instrumentos de coleta dependendo do objetivo do estudo epidemiológico e da fonte de dados a ser utilizada. Para os dados de registro contínuo, os sistemas de informação em saúde empregam instrumentos de coleta padronizados, dos quais bons exemplos são a Declaração de Óbito, a Declaração de Nascido Vivo e a Ficha de Notificação de Casos ao SINAN. Essa padronização é fundamental para assegurar a comparabilidade dos dados entre grupos populacionais, períodos de tempo e entre áreas. Na formulação dos instrumentos de coleta é necessário assegurar que contenham as variáveis suficientes à produção da informação: insuficiência de variáveis trará prejuízos aos estudos epidemiológicos e, em contrapartida, instrumentos muito extensos e complexos têm o seu uso dificultado durante as atividades dos serviços e a pesquisa, resultando em erros e falhas de preenchimento. Com isto pretende-se assinalar que a qualidade dos dados depende diretamente da adequação e qualidade dos instrumentos de coleta, entre outros fatores, tais como: o treinamento dos profissionais, a adequada inserção do uso dos instrumentos nos processos de trabalho, a organização do registro de dados etc. Dos diversos tipos de instrumentos de coleta, citam-se: formulários, protocolos e roteiros.

Os dados obtidos em inquéritos ou em pesquisas amostrais periódicas são registrados em *formulários* padronizados. Esses são utilizados em entrevistas individuais, como no caso do censo demográfico e da PNAD. Quando se trata de extrair dados de prontuários clínicos, de serviços de saúde ou laboratórios, utilizam-se os chamados *protocolos*, onde se indicam de maneira padronizada que dados devem ser transcritos das fontes primárias e sob que formato. *Roteiros* para a observação de situações ou de entrevistas podem resultar em registro não padronizado de dados. Porém, prestam-se bem para a coleta de dados que descrevem características de processos de trabalho, atitudes, comportamentos e práticas, opiniões e percepções dos participantes, condições ambientais etc., que poderão ser consolidados em frequências simples ou utilizados para contextualizar outros tipos de dados.

<sup>14</sup> Consultar [www.ibge.gov.br](http://bvsms.saude.gov.br/bvs/pnds/index.php).

<sup>15</sup> Consultar [www.ensp.fiocruz.br](http://bvsms.saude.gov.br/bvs/pnds/index.php).

<sup>16</sup> Ver: <http://bvsms.saude.gov.br/bvs/pnds/index.php>.

■ Quadro 9.1 Inquéritos domiciliares de abrangência nacional (Brasil, 2008)

Inquérito	Ano de coleta	Instituição responsável	Metodologia	Temas cobertos	Representatividade e abrangência	Endereço eletrônico de referência
Pesquisa sobre Organização Familiar (POF)	1974/75, 1987/88, 1995/96, 2002/2003 e 2007/2008	IBGE	Inquérito domiciliar	Composição dos gastos das famílias, com detalhamento para alimentação, plano de saúde, medicamentos, serviços de saúde e aparelhos médicos	Brasil, grandes regiões, e situação urbana e rural. Regiões metropolitanas e Unidades da Federação: total e situação urbana	www.ibge.gov.br
Suplemento-Saúde da PNAD	1981, 1986, 1998, 2003 e 2008	IBGE	Inquérito domiciliar com módulos domiciliares e individual	Morbidade referida, acesso, uso e financiamento de serviços de saúde, tabagismo e sedentarismo	Área rural da região norte excluída de 1981 a 2003	www.ibge.gov.br
Pesquisa Nacional sobre Demografia e Saúde	1986 (PNSMIPF), 1991 (PSFNE), 1996 e 2006	BEMFAM, IBGE, MS	Inquérito domiciliar com módulos para mulheres de 15 a 49 anos, crianças de 0 a 5 anos e homens	Uso de métodos contraceptivos, saúde da mulher e da criança, conhecimentos sobre DST/AIDS	Brasil, grandes regiões e estados (São Paulo, Rio de Janeiro, Minas Gerais, Rio Grande do Norte, Bahia, Pernambuco, Ceará e Rio Grande do Sul)	www.bemfam.org.br, www.saude.gov.br
Pesquisa Nacional sobre Saúde e Nutrição	1989	IBGE	Inquérito domiciliar com módulos domiciliares e individual	Apurar os indicadores da situação nutricional das famílias, população	Brasil, estados e nove regiões metropolitanas, área urbana e rural	www.ibge.gov.br
Pesquisa sobre Padrão de Vida	1996/97	IBGE	Inquérito domiciliar com módulos domiciliares e individual	Domicílios, famílias, anticoncepção, saúde, antropometria, educação, trabalho, empreendimentos, mobilidade ocupacional e uso do tempo	Região metropolitana de Fortaleza, Recife, Salvador, Rio de Janeiro, Belo Horizonte e São Paulo: urbano e rural	www.ibge.gov.br
Inquérito Domiciliar sobre Comportamentos de Risco e Morbidade Referida de Doenças e Agravos não Transmissíveis	2002-2003	INCA/MS	Inquérito domiciliar com módulos: adulto de 25 anos ou mais e jovem de 15 a 24 anos	Tabagismo, atividade física, peso, dieta, álcool, detecção precoce de câncer, hipertensão arterial, diabetes, doenças cardiovasculares, percepção da saúde e condição funcional	Municípios das capitais (15) e Distrito Federal	www.inca.gov.br
Pesquisa Mundial da Saúde – Brasil	2003	Fiocruz/MS	Inquérito domiciliar com módulos domiciliares e individual (18 anos ou mais)	Morbidade referida, uso e financiamento de serviços de saúde, avaliação do desempenho dos serviços	Brasil	www.ct.fiocruz.br
VIGITEL – Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico	2006-2007	SVS/MS	Inquérito domiciliar por telefone em pessoas de 18 anos ou mais	Alimentação, atividade física, tabagismo, consumo de álcool e outras drogas, imagem corporal, saúde bucal, comportamento sexual, violência e rede de antropometria	Municípios das capitais e Distrito Federal	www.saude.gov.br

Na elaboração de um instrumento de coleta de dados é habitual que se inicie com campos para o registro da identificação do participante, com ou sem nome e endereço – em função da necessidade ou não de guardar anonimato, seguindo-se as demais variáveis de interesse. Nos procedimentos de coleta, se por entrevista em que os dados são registrados por entrevistador ou se do tipo autoaplicado (em geral, denominado “questionário” e utilizado para questões “sensíveis”) em que os dados são registrados pelo participante, o instrumento será formulado de maneira adequada ao grau de conhecimento e experiência de quem deve preenchê-lo. Atenção especial deve ser dada a coleta de dados sobre pessoas ausentes valendo-se de informantes secundários, isto é, os que compartilham a residência, o local de trabalho ou que tenham relação de parentesco com o indivíduo de quem se deseja obter dados. Por exemplo, entrevistas com as mães para obter dados sobre crianças e adolescentes. Nesse caso, as possibilidades de erro aumentam com o grau de subjetividade dos temas pesquisados (Viacava, 2002).

Em um instrumento de coleta, perguntas equilibradas, com alternativas de respostas que não introduzam tendência ou não induzam uma escolha, serão formuladas de maneira clara, concisa, objetiva e inequívoca. Variáveis contínuas terão campos apropriados para valores numéricos das unidades de medida e respectivas escalas (p. ex., idade, em anos completos: 0-99 anos). Variáveis categóricas terão campos para o registro da(s) alternativa(s) de resposta (p. ex., ocupação do chefe da família: código de ocupação ou ocupação informada). Nesse último caso, será apropriado adotar uma classificação já existente e que pode ser referida. Para ocupação, por exemplo, a Classificação Brasileira de Ocupações está disponível via internet; para a codificação de doenças e agravos à saúde se utiliza a Classificação Internacional de Doenças. O cuidado na formulação do instrumento de coleta, com relação às variáveis e valores a serem registrados, irá assegurar a qualidade dos dados, maior facilidade na formação do banco de dados e conseqüentemente qualidade na análise dos resultados. O IBGE disponibiliza em seu sítio na internet os instrumentos aplicados na PNAD e em outras pesquisas domiciliares que podem ser usados como exemplos de formulários.

A coleta de dados deve ser feita com procedimentos uniformizados e padronizados em qualquer situação, tanto nos registros contínuos quanto em pesquisas amostrais e de base populacional do tipo episódica. Isto possibilita comparabilidade dos dados resultantes da aplicação dos instrumentos de coleta entre todos que dela participam. Para assegurar uniformidade, os instrumentos são testados (pré-teste) em um ou mais grupos de indivíduos, avaliando-se a propriedade e a compreensão das perguntas formuladas, dos campos de registro das alternativas de resposta, da seqüência das variáveis etc. Após a realização do pré-teste o instrumento é revisto e se necessário novamente testado até que se esteja seguro de que cumpre satisfatoriamente os objetivos da produção de dados. Há situações em que o instrumento de coleta já existe e foi publicado e o investigador pretende utilizá-lo em outro contexto de pesquisa. Ao lado da adequação desse instrumento aos objetivos do estudo, atenção especial deve ser dada à sua *validação*, em especial no caso de questionários estruturados com fim diagnóstico.

Acompanham os instrumentos de coleta os respectivos manuais de instrução ou manuais do entrevistador. Neles estão contidas todas as informações necessárias ao correto preenchimento de formulários e protocolos na pesquisa epidemiológica. Especifica-se inclusive o propósito de cada pergunta ou campo de preenchimento do formulário. A definição de caso ou desfecho, dos fatores de exposição, das variáveis de pessoa, tempo e lugar, das condições do acompanhamento dos participantes, das

instituições, da consulta a documentos (pessoais ou institucionais) quando apropriado, entre outras, deve constar dos manuais, que serão consultados sempre que for necessário por quem coleta os dados. Além disso, com os manuais e os instrumentos se realizam treinamentos e testes de consistência do preenchimento para assegurar o melhor nível possível de qualidade dos dados. Manuais e toda a documentação dos sistemas de informação em saúde de base nacional estão disponíveis em meio eletrônico nos endereços na internet dos órgãos do Ministério da Saúde, do IBGE e de outras instituições de interesse.

Bons instrumentos de coleta, treinamentos e manuais de instrução não asseguram isoladamente a qualidade dos dados. Atividades de supervisão dos processos de produção e registro de dados em saúde, tanto nas pesquisas amostrais e inquéritos quanto para os dados de registro contínuo, são necessárias para a manutenção da qualidade. A supervisão da coleta, realizada contínua ou periodicamente por profissionais especialmente treinados para essa tarefa, procurará identificar dificuldades na aplicação dos instrumentos. Restrições impostas pelas instituições ou deficiências nos processos de trabalho, falhas de preenchimento, limitações do entendimento das perguntas e variáveis, obstáculos de acesso às fontes primárias ou aos participantes das entrevistas etc. são algumas das dificuldades que a supervisão tentará solucionar. A revisão por amostra sistemática de formulários preenchidos poderá reduzir erros e falhas na aplicação dos instrumentos de coleta. Nessa revisão se podem comparar os formulários preenchidos com prontuários clínicos ou outra fonte primária ou mesmo repetir entrevistas quando for o caso, no intuito de verificar a consistência e assegurar a qualidade dos dados.

Embora não exista consenso sobre definições para qualidade e seus parâmetros no campo da informação, os atributos mais comumente arrolados na avaliação da qualidade dos dados incluem: *cobertura*, *completude* ou *completitude* e *confiabilidade*. Outros elementos devem ser considerados: que os dados estejam acessíveis (acessibilidade); que sejam válidos (validade) – relativo ao desempenho operacional dos instrumentos e dos testes diagnósticos, e com grau adequado de precisão das medidas; que sejam obtidos oportunamente, isto é, quando deles se necessitem (oportunidade); que sejam coerentes (coerência) – que não possuam valores contraditórios ou conflitantes, e que não exista duplicidade de dados que exista possibilidade de serem rastreados até as suas fontes originais (rastreadibilidade), e que tenham adequada reprodutibilidade – a aplicação repetida do instrumento de medida nas mesmas condições oferecerá resultado do mesmo valor.

A cobertura dos dados é um atributo de qualidade especialmente importante, porque sempre se deseja que o conjunto de dados coletados represente de maneira significativa a realidade que se quer revelar. Como cobertura entende-se a inclusão de todos os indivíduos, participantes da pesquisa, ou das unidades de análise que se planejou incluir. Tratando-se de dados secundários, obtidos de sistemas de informação em saúde – registros de coleta contínua – este aspecto é essencial. Os resultados da análise de dados secundários poderão ser subestimados ou superestimados no cálculo dos indicadores caso o sistema de informação não tenha sido capaz de captar todos os eventos ocorridos em um período de referência e em uma dada localidade. Por exemplo, para calcular o coeficiente de mortalidade infantil espera-se que todos os óbitos de menores de 1 ano tenham sido registrados e que o número de nascidos vivos na mesma área e período corresponda de fato ao número de nascimentos ocorridos. Não sendo assim, será necessário realizar estimativas ou aproximações ao valor do coeficiente por métodos indiretos, com possível impacto sobre a interpretação dos resultados.

Também parte integrante dos procedimentos para assegurar a qualidade dos dados em pesquisa epidemiológica são os cuidados que se devem tomar na formação de *bancos de dados*, para o que hoje se empregam largamente os recursos de informática. Uma vez codificados os dados dos formulários podem ser digitados ou, como convém em pesquisas de maior complexidade ou com maior número de casos, registrados diretamente em meio eletrônico durante a coleta. Para isto, equipamentos de informática e o desenvolvimento de programas informatizados tornaram-se importantes auxiliares. Porém isso não dispensa que se realize a revisão do banco de dados antes de iniciar a análise: erros de digitação ocorrem e, portanto, se faz necessária uma avaliação do preenchimento de dados das variáveis e da sua consistência.

Segurança de dados é um tema que provoca atualmente intenso debate em todo o mundo. A ampliação de acesso aos bancos de dados, possibilitado sobretudo pela internet, tem lamentavelmente ampliado as práticas de uso indevido dos recursos tecnológicos da informática. Esse aspecto nos remete aos requisitos da ética em pesquisa, em especial do sigilo e da confidencialidade que devem ser rigorosamente atendidos por quem pesquisa, pelas instituições públicas e privadas que realizam censos e inquéritos e por todos que gerenciam os sistemas de informática de registro contínuo de dados. A identificação nominal dos indivíduos participantes e das instituições que fornecem dados deve ser preservada, e os dados devem ser utilizados exclusivamente para os fins a que se destinam, de acordo com sua aprovação em Comitê de Ética em Pesquisa e, no Brasil, segundo a Resolução n.º 196/96 do Conselho Nacional de Saúde.

## Considerações finais

O estudo das características dos dados contribui para a promoção da qualidade das informações epidemiológicas e, portanto, para melhorar o seu potencial de aplicação na formulação de políticas, nas ações de saúde e na avaliação das intervenções. A produção dessas informações cresce em importância no Brasil, e a disponibilidade de dados dos sistemas de informática em saúde tem atuado como importante motivação para pesquisas epidemiológicas.

Para tanto, a disponibilidade de infraestrutura de captação e registro de dados, com o emprego extensivo dos recursos de tecnologia da informática e da informática nos serviços, representa um fator positivo. Além disso, os programas informatizados possibilitam processar enorme volume de dados em velocidade cada vez maior, e os computadores têm hoje capacidade de armazenamento e transmissão de dados bastante acima das necessidades. De um lado, essa tecnologia, na qual a rede mundial internet ocupa lugar de destaque, tem ampliado a capacidade de registro, acesso e análise de dados, mas, por outro lado, suas facilidades não significam necessariamente melhor qualidade de acesso e análise de dados, mas, por outro lado, a dificuldade de acesso e análise de dados, por outro lado, ocupa lugar de destaque, tem ampliado a capacidade de armazenamento e transmissão de dados bastante acima das necessidades. De um lado, essa tecnologia, na qual a rede mundial internet ocupa lugar de destaque, tem ampliado a capacidade de registro, acesso e análise de dados, mas, por outro lado, suas facilidades não significam necessariamente melhor qualidade de dados. Nesse particular, o trabalho do profissional que os obtém e registra primariamente continua sendo o ponto angular de sustentação da cadeia de produção de dados e informações em Epidemiologia.

## Referências bibliográficas

Alzarqui M, Moia E, Spinelli H. Sistemas de Informação em Saúde: de sistemas cerrados a la ciudadanía social. Un desafío en la reducción de desigualdades en la gestión local. *Cad Saude Publica*, Rio de Janeiro, 22(12): 2693-2702, 2006.

É possível avaliar o nível de cobertura dos dados de um sistema de informação, realizando-se:

- A comparação do número de eventos captados com o número de eventos estimados (*i. e.*, obtidos de forma indireta por meio de estimativas baseadas em dados censitários ou outros);
- A comparação do registro obtido com a frequência do mesmo evento captado em outro sistema de informação, como, por exemplo, a comparação do número de nascimentos proveniente dos dados do registro civil e do SINASC;
- A comparação de eventos semelhantes, como, por exemplo, a comparação do número de nascidos vivos do SINASC com o número de partos registrados no SIH/SUS.

Técnicas conhecidas como *linkage* de bancos de dados também são usadas para avaliar a cobertura dos dados de mais sistemas de informação em saúde. Esta técnica consiste em identificar o mesmo indivíduo em duas ou mais bases de dados, considerando-se os "valores" de uma ou mais variáveis identificadoras sejam utilizadas na comparação das bases (p. ex., nome, sexo, data e local de nascimento). Estudo que avaliou a validade dos dados sobre nascidos vivos (Almeida *et al.*, 2006) utilizou essa técnica. Outros métodos de relacionamento de bancos de dados em saúde, como o de captura-recaptura, têm sido aplicados com sucesso, possibilitando realizar estimativas de indicadores epidemiológicos (Maia-Elkhoury *et al.*, 2007).

Estudos sobre a completude e a confiabilidade dos dados de vários sistemas de informação em saúde brasileiros, a exemplo do SINASC e do SIM, têm sido conduzidos, indicando a boa qualidade dos dados para a aplicação de informações epidemiológicas sobre a situação de saúde e em pesquisas epidemiológicas (Silva *et al.*, 2001; Almeida *et al.*, 2006; Laurenti, Mello-Jorge & Gottlieb, 2008). Na avaliação da completude dos dados observa-se em que extensão os formulários possuem valores não nulos para variáveis essenciais. Isto é, observa-se a medida de frequência para a ausência de registro ou da ocorrência de dados ignorados. A elevada proporção de dados ignorados ou faltantes compromete por completo a sua qualidade e consequentemente a análise e interpretação. Considera-se um excelente grau de completude quando a ausência de registro válido situa-se abaixo de 5%; uma proporção de registros ausentes superior a 30% é considerada muito elevada e até o ponto em que não se comprometa a validade dos resultados, os dados ausentes de uma variável em um banco de dados podem passar a ter valores atribuídos ou imputados. Fazê-lo significa adotar critérios e regras uniformes e que não introduzam vieses no cálculo de parâmetros e indicadores; por exemplo, o valor médio ou mediano de uma série numérica dos valores de uma variável pode ser usado para "preencher" os vazios deixados pelo preenchimento incompleto de um formulário ou protocolo.

No estudo de Almeida *et al.* (2006), sobre a validade de dados da Declaração de Nascido Vivo registrados no SINASC, foi realizada avaliação de acordo com informações obtidas das mães em entrevistas domiciliares e nos prontuários das mães e recém-nascidos nos hospitais, analisando-se o grau de concordância dos resultados como medida de confiabilidade e concordando-se uma concordância elevada quando o valor igual ou maior que 90%. Medidas de validade (sensibilidade e especificidade) foram calculadas para a identificação da presença/ausência de cada uma das condições do recém-nascido.