

## Métodos para avaliar a completitude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática

Methods for assessing the completeness of data in health information systems in Brazil: a systematic review

Lourani Oliveira dos Santos Correia <sup>1</sup>  
Bruna Merten Padilha <sup>2</sup>  
Sandra Mary Lima Vasconcelos <sup>2</sup>

**Abstract** *The scope of this study was to review the literature on the methods used to assess the completeness of the data in health information systems. Searches were conducted in the Medline, SciELO and Lilacs databases using the key words “quality,” “data quality,” “quality of information” and “information systems.” It included articles, dissertations and theses that assessed the fulfillment of variables. In the 19 studies reviewed, the epidemiological rationality systems were the most evaluated (90.5%). Studies on completeness are still scarce (one per year on average). Methods vary from isolated analysis to the evaluation of more than one dimension of quality. The term completeness is the word most used to characterize the fulfillment of variables. The Romero and Cunha score is the most frequent. The SIS databases and Datasus site were the sources and means of access to the most common data. Studies with data from the Southeast of Brazil were predominant. The indicators that measure the completeness vary from simple relative frequency to indices and synthesis-indicators. The heterogeneity of methods lacks standardization and dissemination of information to enable comparisons between the studies.*

**Key words** *Health evaluation, Data quality, Information quality, Information systems, Systematic review*

**Resumo** *Este estudo teve como objetivo revisar a literatura acerca dos métodos aplicados para avaliar a completitude dos dados dos sistemas de informação em saúde. Foram pesquisadas as bases de dados Medline, SciELO e Lilacs usando os descritores “qualidade”, “qualidade dos dados”, “qualidade da informação” e “sistemas de informação”. Incluíram-se artigos, dissertações e teses que avaliaram o preenchimento das variáveis. Nos 19 estudos revisados os sistemas de racionalidade epidemiológica, foram os mais avaliados (90,5%). Estudos sobre completitude ainda são escassos (média de um/ano). Os métodos variam desde a sua análise isolada à avaliação de mais de uma dimensão de qualidade. O termo completitude é o mais utilizado para caracterizar o preenchimento das variáveis. O escore de Romero e Cunha é o mais frequente. As fontes e o meio de acesso aos dados mais comuns foram os bancos de dados dos SIS e o sítio do Datasus. Predominaram estudos com dados da região Sudeste e Brasil. Os indicadores que mensuram a completitude variam de uma simples frequência relativa a índices e indicadores-síntese. A heterogeneidade de métodos carece de padronização e divulgação de informações que permitam comparações entre os estudos.*

**Palavras-chave** *Avaliação em Saúde, Qualidade dos dados, Qualidade da informação, Sistemas de Informação, Revisão sistemática*

<sup>1</sup> Programa de Pós-Graduação em Nutrição, Faculdade de Nutrição, Universidade Federal de Alagoas (UFAL). Av. Lourival Melo Mota, Cidade Universitária. 57072-900 Maceió AL Brasil. lourani@hotmail.com  
<sup>2</sup> Laboratório de Nutrição em Cardiologia, Faculdade de Nutrição, UFAL.

## Introdução

Diversas áreas da ciência que necessitam de informação para tomada de decisão<sup>1</sup> têm demonstrado interesse crescente em avaliar a qualidade da informação<sup>2</sup>. Contudo, embora não haja ainda consenso na literatura sobre definições teóricas e operacionais, há uma tendência a estudá-la sob três grandes vertentes: uma que trata da qualidade ou valor transcendente (ou valor filosófico, ou metafísico) da informação, outra que se baseia nos aspectos intrínsecos e outra nos atributos contingenciais<sup>3</sup>. Em cada uma dessas vertentes existem atributos específicos que podem ser monitorados e avaliados.

Paim *et al.*<sup>3</sup> declaram que, nos aspectos intrínsecos, os atributos de qualidade são inerentes ao produto-informação (dado, documento ou texto) que ao serem tratados em termos precisos e identificáveis, podem ser mensurados e quantificados. A completeza é um desses atributos e resulta da inclusão de todos os dados necessários para responder a uma questão de determinado problema. No âmbito dos Sistemas de Informação em Saúde (SIS), essa dimensão é também denominada de completitude e pode ser entendida como “o grau em que os registros de um sistema de informação possuem valores não nulos”<sup>4</sup>.

Nas últimas quatro décadas, foram implantados e/ou implementados no Brasil vários SIS, o que ampliou o uso da informação para a gestão do setor saúde e consolidou uma rede de informações composta por sistemas de racionalidade epidemiológica, de assistência à saúde (produção de serviços), monitoramento de programas de saúde, gerenciamento de serviços, entre outros<sup>5</sup>. Os dados gerados por estes sistemas compõem uma grande base nacional que tem como finalidade primordial a produção de indicadores de saúde que retratem as condições de saúde da população no que diz respeito ao processo saúde-doença e aos aspectos administrativos dos serviços de saúde<sup>6</sup>.

Com o avanço da microinformática e a efetivação do processo de descentralização das ações de saúde, o processamento desses dados passou a ser de responsabilidade das Secretarias Municipais de Saúde (SMS)<sup>7</sup>. Atualmente, a alimentação das bases de dados nacional com os dados produzidos nas SMS ocorre via internet, com periodicidade regulamentada em portarias ministeriais. Cabe ao Ministério da Saúde (MS) a consolidação e a disponibilização desses dados no sítio do Departamento de Informática do Sistema Único de Saúde (Datasus) (<http://www.datasus.gov.br>),

proporcionando amplo acesso aos gestores, pesquisadores, profissionais de saúde e a sociedade em geral.

Embora as informações dos SIS constituam relevantes fontes de dados secundários que tem como uma das principais vantagens a ampla cobertura populacional, é comum à maioria dos autores a recomendação para que se observe a qualidade dos dados registrados<sup>8-11</sup>, especialmente no que diz respeito à cobertura dos eventos, à confiabilidade das informações e à completitude dos dados<sup>9,11,12</sup>. Essa premissa deve-se ao fato de que, em todas as etapas do ciclo de produção da informação podem ocorrer problemas na qualidade desta envolvendo produtores, gestores e usuários<sup>13</sup>. Se estas questões não forem levadas em consideração, o conhecimento gerado com base nesses dados pode não representar adequadamente a realidade estudada<sup>9</sup>.

O gerenciamento da qualidade da informação dos SIS do Brasil ainda não é realizado de forma organizada e sistemática, avaliando-se apenas partes do ciclo de produção da informação. As iniciativas são pontuais e mais frequentes nas regiões Sul e Sudeste do país. Os métodos utilizados são diversos e em 90% das análises são priorizadas as dimensões confiabilidade, validade, cobertura e completitude. A completitude é a segunda dimensão mais avaliada nas análises referente ao SIM, ao Sinasc e ao Sinan<sup>1</sup>.

Ainda que se verifique um aumento do interesse em se avaliar a completitude, ele ainda é insuficiente para dar conta do grande volume e da diversidade de dados que são produzidos pelos atuais SIS. O monitoramento e avaliação dessa dimensão é uma importante ferramenta para averiguar se o preenchimento dos instrumentos de coleta e/ou banco de dados é realizado de forma adequada, contribuindo para identificar fragilidades e potencialidades dos dados produzidos pelo SIS e com isso ampliar o seu uso e/ou recomendar estratégias visando a melhoria da qualidade da informação.

É nesse contexto que este trabalho teve como objetivo revisar a literatura acerca dos métodos aplicados para avaliar a completitude dos dados dos SIS do Brasil. Espera-se fornecer elementos que reforcem a discussão sobre a instituição de uma política nacional de avaliação que englobe todos os aspectos que influenciam na qualidade da informação, bem como estimular a adoção desses métodos por pesquisadores, gestores e profissionais de saúde de modo a ampliar o número de estudos dessa dimensão de qualidade da informação.

## Métodos

Para identificação de artigos publicados em revistas científicas foram consultadas as bases de dados bibliográficas *Medical Literature Library of Medicine On-Line* (Medline) via PUBMED, *Scientific Electronic Library Online* (SciELO) e Literatura Latino-Americana e do Caribe (Lilacs) via Biblioteca Virtual em Saúde (BVS). Adicionalmente, realizou-se busca manual nas referências dos artigos selecionados.

A busca eletrônica foi realizada utilizando-se os seguintes descritores: “sistemas de informação”, “qualidade”, “qualidade de dados” e “qualidade da informação”. Limitou-se a busca aos limites de assunto “humanos”, aos idiomas português e inglês e às publicações realizadas nos últimos 10 anos.

Incluíram-se artigos, dissertações e teses que avaliaram o preenchimento das variáveis dos sistemas de informações de abrangência nacional. Foram excluídos editoriais; artigos de revisão; análises epidemiológicas descritivas que utilizaram indicadores do próprio sistema ou complementadas com outras fontes de dados que não foi precedida de avaliação da completitude; estudos em duplicidade nas bases de dados; estudos de avaliação de outras dimensões de qualidade sem incluir a completitude e análises abordando aspectos operacionais dos SIS.

Para descrever os estudos quanto à execução e à divulgação dos resultados foram utilizadas as seguintes variáveis: afiliação institucional do primeiro autor; categoria do estudo; procedência; ano e periódico de publicação.

O detalhamento dos métodos empregados foi descrito a partir da síntese das seguintes informações: sistema de informação, fontes de dados, meios de acesso, local e período de realização, dimensão de qualidade avaliada, unidade de análise selecionada, indicadores empregados, escore para classificação, variáveis utilizadas e análise estatística. A dimensão de qualidade e os métodos aplicados pelos estudos para realizar as avaliações foram categorizados segundo as definições conceituais propostas por Lima et al.<sup>1</sup>. Quando os estudos analisaram mais de uma dimensão de qualidade foram considerados apenas os dados referentes à avaliação da completitude.

## Resultados

Foram identificados 972 estudos, dos quais 941 obtidos da busca eletrônica e 31 do exame na lista de referências.

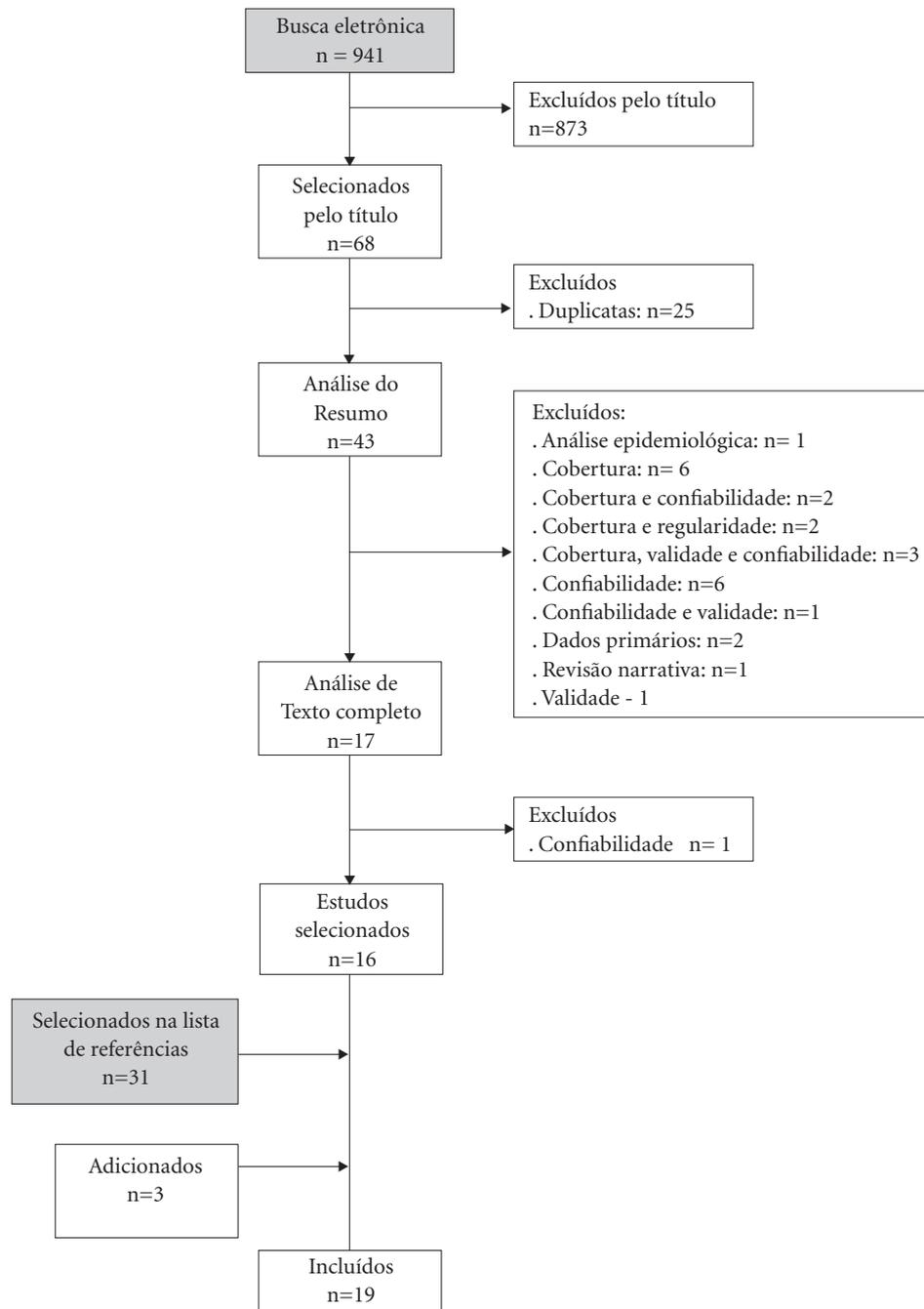
Dos estudos identificados nas bases de dados 873 foram excluídos após análise do título, uma vez que não descreviam aspectos relacionados à avaliação da qualidade dos dados dos SIS, permanecendo 68 artigos para análise dos resumos. Desses, excluiu-se 25 por estarem em duplicata. Após averiguação dos 43 resumos restantes, selecionou-se 17 para leitura de texto completo, visto que 23 tratavam de avaliação de outros atributos de qualidade que não a completitude, um realizou apenas análise epidemiológica, um de revisão narrativa e dois utilizaram dados primários para análise qualitativa da qualidade dos dados. Procedeu-se a exclusão de mais um artigo subsequentemente à leitura por avaliar apenas a confiabilidade dos SIS, o que contabilizou 16 estudos da busca *on line*.

Sequencialmente, realizou-se busca manual nas referências dos 16 estudos selecionados e três foram adicionados à lista de estudos desta revisão, que ficou composta por 18 artigos e uma dissertação. O processo de seleção dos estudos pode ser visualizado na Figura 1.

Dos dezenove estudos revisados, seis estudaram o SIM; dois o SIM/Sinasc; quatro o Sinasc; cinco o Sinan; um o Sishiperdia e um o Siscolo. As publicações só surgiram a partir do ano de 2005, com um pequeno incremento observado em 2009, entretanto ainda em número escasso, um estudo/ano, em média. Foram encontrados predominantemente em periódicos da área de saúde pública ou saúde coletiva, representando 61% (Quadro 1).

Quanto à afiliação institucional do primeiro autor, 84,2% estão vinculados a instituições de ensino superior, predominando universidades federais. Apenas 15,8% dos autores são oriundos dos serviços de assistência à saúde, das esferas federal<sup>14</sup>, estadual<sup>15</sup> e municipal<sup>16</sup>. Um número expressivo (36,8%) destes estudos foi produto de programas de pós-graduação, incluindo Doutorado<sup>17</sup>, Mestrado<sup>14,18-20</sup> e Especialização<sup>15,21</sup>.

O objetivo de verificar exclusivamente a completitude foi comum em aproximadamente 80% dos estudos (n = 15). Além de avaliar a qualidade dos dados, Glatt<sup>14</sup>, Malhão et al.<sup>22</sup>, criaram e aplicaram indicadores e/ou índices de avaliação da completitude; Vanderlei et al.<sup>23</sup> avaliaram os preditores do óbito neonatal; Costa e Frias<sup>24,25</sup> averiguaram a evolução do adequado preenchi-



**Figura 1.** Fluxograma do processo de seleção dos estudos.

Fonte: Autora, 2013

mento das variáveis do SIM e SINASC entre os municípios; e, Almeida et al.<sup>12</sup> examinaram se as informações disponíveis nos SIS são suficientes para avaliar o perfil e a magnitude da mortalidade perinatal, comparando a disponibilidade de

informações sobre nascidos vivos, óbitos fetais e neonatais precoces no Registro Civil, SIM e SINASC.

A avaliação da qualidade dos dados referente aos sistemas de informação de racionalidade epi-

**Quadro 1.** Características dos estudos incluídos na revisão.

Autor e ano	Periódico	SIS	Fonte de dados	Forma de acesso	Unidade de análise	Abrangência	Período avaliado	Escore adotado
Glatt, 2005 <sup>14</sup>	-	Sinan	BD de AIDS	Programa Nacional de DST/AIDS	Aids Criança e Adulto	Brasil	1980 a 2004	Sinan, 2006; Glatt, 2005
Romero; Cunha, 2006 <sup>26</sup>	Cad Saude Publica	SIM	BD do SIM	Datasus	Óbitos < 1 ano	Brasil	1996 a 2001	Romero; Cunha, 2006
Almeida et al., 2006 <sup>12</sup>	Rev bras epidemiol	SIM e Sinasc	BD do SIM e do Sinasc	Datasus	Óbitos NP e F, Nascidos vivos	Brasil	2002	Piper et al., 1993
Romero; Cunha, 2007 <sup>27</sup>	Cad Saude Publica	Sinasc	BD do Sinasc	Datasus	Nascidos vivos	Brasil	2002	Romero; Cunha, 2006
Pedrosa et al., 2007 <sup>28</sup>	Rev Assoc Med Bras	SIM	BD do SIM, cópias DO e pront. Hosp.	SMS e Hospitais	Óbitos NP hospit.	Cidade de Maceió, AL	2001 a 2002	NR
Moreira; Maciel, 2008 <sup>19</sup>	J bras pneumol	Sinan	BD de Tuberculose	Datasus, SINANWEB	Casos not. de Tuberculose	Estado do Espírito Santo	Jan./2001 a Set./2005	Sinan, 2006
Cequeira et al., 2008 <sup>18</sup>	UFES Rev Odont	Sinan	BD do Sinan	Datasus, SINANWEB	Casos not. Gest. HIV+ e crianças expostas	Estado do Espírito Santo	2001 a 2006	Glatt, 2005
Oliveira et al., 2009 <sup>15</sup>	Epidemiol Serv Saude	Sinan	BD e rel. do Sinan, Fichas de Not/Inv e Prontuários	Datasus, SES e Unidades de Saúde	Casos invest. de febre tifóide	Estado da Bahia	2003 a 2006	Sinan, 2006
Costa; Frias, 2009 <sup>24</sup>	Cad Saude Publica	Sinasc	BD do Sinasc	Datasus	Nascidos vivos	Estado de Pernambuco	1996 a 2005	Romero; Cunha, 2007; Costa; Frias, 2009
Girianelli et al., 2009 <sup>17</sup>	Rev Saúde Públ	Siscolo	Estudo transversal, pront. e BD do Siscolo	SMS e Datasus	Exames citológicos e histopatológicos	Cidades de Duque de Caxias e Nova Iguaçu, RJ	2002 a 2006	Mello Jorge et al., 2006
Malhão et al., 2010 <sup>22</sup>	Epidemiol Serv Saude	Sinan	BD do Sinan	Programa Nacional de Tuberculose	Casos notificados de Tuberculose	Brasil	2001 a 2006	Romero; Cunha, 2006; Malhão, 2010

continua

demiológica (SIM, Sinasc e Sinan) foi realizada em 90,5% dos estudos. Apenas dois sistemas de informação de monitoramento de programas de assistência à saúde tiveram a completude dos seus dados mensurados: o Sishiperdia e o Siscolo.

Dentre as várias unidades de análise disponíveis nos SIS de racionalidade epidemiológica, foram priorizadas os nascidos vivos; os óbitos (menores de um ano, fetais, mulheres com câncer de mama e suicídio); e, os casos notificados de AIDS em

Quadro 1. continuação

Autor e ano	Periódico	SIS	Fonte de dados	Forma de acesso	Unidade de análise	Abrangência	Período avaliado	Escore adotado
Macente; Zandonade, 2010 <sup>29</sup>	J Bras Psiquiatr	SIM	BD do SIM	Datasus	Óbitos por suicídio	Estado do Espírito Santo	1996 a 2007	Romero; Cunha, 2007
Vanderlei et al., 2010 <sup>23</sup>	Rev Bras Saude Mat Infant	Sinasc	BD do Sinasc	SES	Nascidos vivos	Cidade de Recife, PE	1995 a 2006	NR
Zillmer et al., 2010 <sup>21</sup>	Rev Gaucha Enferm	Sishiperdia	Ficha de cadastro	Unidade Básica de Saúde	Cadastro de hipertensos e diabéticos	Cidade de Pelotas, RS	Dez./2002 a Nov./2008	NR
Costa; Frias, 2011 <sup>25</sup>	Cien Saude Colet	SIM	BD do SIM	Datasus	Óbitos < 1 ano	Estado de Pernambuco	1997 a 2005	Romero; Cunha, 2006; Costa; Frias, 2011
Barbuscia; Rodrigues, 2011 <sup>16</sup>	Cad Saude Publica	SIM e SINASC	BD do SIM e SINASC	Fundação SEADE e SES	Óbitos perinatais e NP, Nascidos vivos	Cidade de Ribeirão Preto, SP	2000 a 2007	Almeida et al., 2006; Piper et al., 1993
Almeida et al., 2011 <sup>30</sup>	Rev Saúde Públ	SIM	BD do SIM	Fundação SEADE e SES	Óbitos fetais	Cidade de São Paulo, SP	1º sem. de 2008	Costa; Frias, 2011
Silva et al., 2011 <sup>20</sup>	Rev Esc Enferm USP	SINASC	BD do SINASC	SES	Nascidos vivos	Estado do Paraná	2000 a 2005	Silva et al., 2011
Felix et al., 2012 <sup>31</sup>	Cien Saude Colet	SIM	BD do SIM	Datasus	Óbitos por CA de mama	Cidade do Espírito Santo	1998 a 2007	Romero; Cunha, 2006, 2007

Notas: \* Dissertação localizada na base de dados LILACS com arquivo completo adquirido após contato com a autora. BD = Banco ou base de dados. SIM: Sistema de Informação Mortalidade. Sinasc: Sistema de Informação sobre Nascidos Vivos. Sinan: Sistema de Informação de Agravos de Notificação. Siscolo: Sistema de Informação do Câncer de Colo de Útero. Sishiperdia: Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos. NR = Não referido. DO = Declaração de óbito. NP = Neonatal Precoce. F = Fetais. SES – Secretaria Estadual de Saúde. SMS – Secretaria Municipal de Saúde.

Fonte: Autora, 2013.

crianças e adultos, febre tifoide, gestantes HIV positivo, crianças expostas ao HIV e tuberculose (Quadro 1).

Em 94,7% dos estudos os autores utilizaram os bancos de dados dos SIS, dos quais 61% acessaram-no via *internet*<sup>12,15,17-19,24,25,27,29,31</sup>. Todavia, Zillmer et al.<sup>21</sup> transcreveram os dados das fichas de cadastro do Sishiperdia arquivadas na unidade básica de saúde. Os demais estudos tiveram acesso às informações nas secretarias estadual<sup>16,20,23,30</sup> e municipal de saúde<sup>28</sup> e na Coordenação Na-

cional do Programa em estudo, (DST/AIDS<sup>14</sup> e Tuberculose<sup>22</sup>). E ainda, três estudos coletaram dados em mais de uma fonte de informação<sup>15,17,28</sup> (Quadro 1).

A maioria dos estudos foi realizada com dados procedentes da região sudeste (36,8%). O Brasil e o Nordeste foram contemplados com 26,3% das análises, ambos ocupando a 2ª posição. No Sul, apenas dois estudos acerca da completude dos dados do Sinasc e do Sishiperdia foram publicados no período analisado. Predominou a análise

se de dados organizados em série histórica com períodos iniciados a partir da década de 2000. O estudo sobre a completitude dos casos notificados de AIDS foi o que apresentou o período mais longo analisado: 23 anos. Testes estatísticos aplicados aos resultados foram encontrados em 42,1% dos estudos. O mais comum foi a regressão linear simples, em decorrência de que grande parte dos estudos analisou a tendência da completitude por meio de série temporal (Quadro 1).

Mais da metade dos autores (n = 12) optaram por avaliar exclusivamente a qualidade do preenchimento das variáveis. Diversos termos foram utilizados para designar esse preenchimento, sendo completitude (n = 9) e completude (n = 5) os mais frequentes. Dentre os estudos que avaliaram mais de uma dimensão de qualidade destacaram-se as seguintes: consistência<sup>14,26,27</sup>, duplicidade de registros<sup>14</sup>, acessibilidade<sup>26</sup>, oportunidade<sup>26</sup>, clareza metodológica<sup>26,27</sup>, validade<sup>17,28</sup>, confiabilidade<sup>28,30</sup> e cobertura<sup>12</sup>.

Como os instrumentos de coleta de dados são específicos para cada SIS, entre os cinco sistemas avaliados, identificou-se 290 tipos de variáveis selecionadas. De um modo geral, foi mais estudada a completitude das variáveis de identificação individual (escolaridade, idade, sexo, raça/cor, estado civil); as que são definidas nos manuais e instrucionais dos SIS como variáveis essenciais ou de preenchimento obrigatório, ou ainda aquelas de interesse do próprio pesquisador para atender aos objetivos do estudo.

Para mensurar a magnitude da completitude, o cálculo do percentual de campos incompletos foi utilizado em 63% dos estudos. Consideraram-se “campo completo” aquele cujo preenchimento é feito com “categoria distinta daquelas indicativas de ausência de dados”<sup>14</sup>, e “incompleto”, o campo preenchido com o código de informação “ignorada” definido nos instrucionais ou manuais de preenchimento dos sistemas e, aqueles que são deixados “em branco” pelo profissional responsável pelo preenchimento ou pelo digitador quando do processamento das informações. Quando os autores analisaram diretamente o documento original de coleta de dados, as questões que apresentaram caligrafia ilegível<sup>21,28</sup> ou opções de repostas preenchidas simultaneamente<sup>28</sup> também foram consideradas incompletas.

Os escores utilizados para classificar o grau de completitude eram distintos e apresentaram intervalos de classes bem variados, tanto quando se avalia a qualidade dos dados pelo aspecto da completeza ou pela ausência de informação (Quadro 2).

O escore elaborado por Romero e Cunha<sup>26,27</sup> foi utilizado em 36,8% (7/19) dos estudos. Em segundo lugar com 21,1% (4/19) das citações, o recomendado pela Coordenação Nacional do Sinan (14) e em terceiro, com 15,8 % (3/19) o proposto por Costa e Frias<sup>9</sup>. Também foram encontrados escores elaborados pelo próprio autor<sup>9,14,20</sup>. Os estudos de Pedrosa et al.<sup>28</sup>, Vanderlei et al.<sup>23</sup> e Zillmer et al.<sup>21</sup> descreveram a frequência relativa do preenchimento das variáveis sem, contudo, classificar os resultados obtidos (Quadro 1). Adaptações dos escores propostos são realizadas livremente pelos autores. Para os casos em que os SIS já atingiram uma qualidade excelente dos dados produzidos, novos escores estão sendo definidos para se adequar ao atual padrão de qualidade do sistema<sup>20</sup>.

Quanto às técnicas aplicadas, verificou-se que foi empregada, em todos os estudos, a análise descritiva de indicadores do próprio banco de dados. Destes, 26,3% usaram-na exclusivamente; 36,8% adicionaram à análise de série temporal; 15,8% verificaram a concordância entre avaliadores e a busca ativa de registros, respectivamente; 10,5%, à análise descritiva utilizando indicadores de distintos bancos de dados; e, 5,3%, realizaram respectivamente, a técnica de comparação com critérios, e, o relacionamento de registros usando a *linkage* inter e intrabanco de dados. Convém ressaltar que alguns autores<sup>17,26-28</sup> empregaram até três técnicas diferentes para avaliar a qualidade dos dados dos SIS.

## Discussão

A despeito da compreensão de que a avaliação da qualidade envolve aspectos multidimensionais, esta revisão se limitou a investigar os métodos utilizados para avaliar apenas uma dessas dimensões. A qualidade e a pertinência das informações produzidas, com intuito de conhecer as condições de saúde da população, podem estar comprometidas quando se verifica variáveis com preenchimento inadequado. Dados incompletos inviabilizam a avaliação de outras dimensões de qualidade<sup>27</sup> e o uso de técnicas como *linkage*<sup>17</sup>, que contribuem para melhorar a qualidade do preenchimento e expandir o escopo do uso dessas informações em estudos epidemiológicos e na tomada de decisões.

Embora haja crescente interesse na avaliação da completitude, ainda é muito pequeno o número de estudos publicados. Outrossim, estudos sobre qualidade dos dados são mais frequentes

**Quadro 2.** Escores utilizados para classificação do grau de completitude.

Classificação do campo	Escores	Autor/Ano	Referência
Completo	Excelente: $\geq 90\%$ ; Bom: entre 70,1 e 90%; Péssimo: $\leq 70\%$	Mello-Jorge et al, 1996 <sup>32</sup>	17
	Excelente - acima de 90%; Regular - 70 a 89%; Ruim - Abaixo de 70%	Sinan, 2006 <sup>33</sup>	14, 15, 18, 19
	Excelente - acima de 95%; Bom - 90 a 95%; Regular - 70 a 90%; Ruim - 50 a 70% e Muito ruim - abaixo de 50%	Costa; Frias, 2009 <sup>24</sup>	23, 24, 30
	Boa - menor que 10%; Regular - de 10 a 29%; Precária - $\geq 30\%$	Piper et al., 1993 <sup>34</sup>	12, 16
Incompleto ou sem informação	Excelente - $\geq 5\%$ ; Bom - 5 a 10%; Regular - 10 a 20%; Ruim - 20 a 50%; Muito ruim - 50 e mais	Romero; Cunha, 2006 <sup>26</sup>	22-28
	Excelente - menor que 1%; Boa - de 1 a 2,99%; Regular - de 3 a 6,99%; Ruim - $\geq 7\%$	Silva et al., 2011 <sup>20</sup>	20

Fonte: Autora, 2013.

nas regiões sul e sudeste do país<sup>1</sup>, que são mais desenvolvidas e onde os SIS já superaram esses problemas requerendo avaliação de outros aspectos de qualidade. Apenas dois estudos foram financiados pelo MS, o que traduz o baixo investimento das instâncias gestoras que priorizam iniciativas relativas à cobertura do SIM e ao preenchimento da causa básica de óbito<sup>25</sup>. Outro aspecto que pode ter limitado a localização dos estudos é a priorização do uso de palavras-chave em detrimento aos Descritores em Ciências da Saúde (DeCS) para descrever o conteúdo dos estudos<sup>35,36</sup>.

A menor frequência de estudos com abrangência municipal e a pequena participação de autores vinculados aos serviços de saúde traz à tona a questão já levantada por Branco desde 1996<sup>6</sup> sobre a necessidade de se dotar os municípios de capacidade técnica, administrativa e institucional, para garantir “maior envolvimento destes

com a produção, processamento e análise das informações, o que tenderia em muito melhorar a qualidade dos dados”. O menor volume de informações produzidas no nível municipal facilitaria a identificação de falhas no preenchimento e a correção se daria de forma mais ágil, devido à maior proximidade com a fonte de captação de eventos. Também possibilitaria a realização de estudos mais detalhados, melhor monitoramento da coleta de dados e o desenvolvimento de novos mecanismos para se adquirir mais consistência nas informações registradas<sup>7,23</sup>.

O uso de diferentes métodos e técnicas para avaliar a completitude inviabilizam sínteses e comparações dos resultados dos estudos entre mesmos SIS. Esse cenário é idêntico ao descrito por Lima et al.<sup>1</sup>, ao revisar amplamente as iniciativas brasileiras para avaliar a qualidade dos dados dos SIS, e por Pedraza<sup>37</sup>, ao analisar criticamente a literatura sobre a qualidade dos dados

do SINASC. Esta situação remete à necessidade de mais estudos nessa área que contribuam com a sistematização e a divulgação destes métodos, observando as especificidades de cada SIS e identificando as contribuições que cada um deles pode trazer ao conhecimento e à melhoria da qualidade da informação.

Independente de a completitude ter sido avaliada de forma isolada ou em conjunto com outras dimensões de qualidade, a magnitude dos campos preenchidos ou incompletos (“em branco” e/ou “ignorados”) foi mensurada, inicialmente, utilizando-se a frequência relativa. Embora Malhão et al.<sup>22</sup> e Glatt<sup>14</sup> tenham criado e aplicado índices nas suas avaliações, ressaltaram que estas medidas-resumo são limitadas, pois não refletem o grau de preenchimento entre as variáveis. Outrossim, não permitem avaliar pequenos avanços ou retrocessos na qualidade dos mesmos. Portanto, é necessário aprimorar a construção destas medidas, estabelecer parâmetros e validá-los<sup>14</sup> para que possam ser amplamente utilizados.

As múltiplas definições para campo “incompleto” refletem a baixa clareza metodológica dos manuais de preenchimento dos SIS, já identificadas no SIM, no Sinasc<sup>26,27,32</sup> e no SISHIPERDIA<sup>38</sup>. Recomendações atuais referentes a esses problemas apontam para a necessidade de padronização e divulgação das definições<sup>1</sup>, inclusive adotando conceitos internacionais e adequação na elaboração de itens nos formulários de coleta<sup>32</sup> permitindo comparações mais fidedignas. Consta-se então que a clareza metodológica é outra dimensão de qualidade que precisa ser avaliada para todos os SIS. Entretanto, dispor apenas de instruções claras e padronizadas não é garantia de informação de qualidade<sup>25</sup>. É preciso que os profissionais de saúde envolvidos no processo sejam devidamente capacitados a respeito de como coletar esses dados<sup>32</sup>, bem como sejam conscientizados sobre a importância da qualidade dos dados.

A maioria dos autores considerou “incompleto” aqueles campos que apresentaram a categoria “ignorada” ou “em branco”. Todavia, a variável definida como “ignorada” traz muitos questionamentos pela falta de clareza com que é abordada nos instrucionais de preenchimento<sup>27</sup>. Há vários tipos de interpretações que vão desde o dado ignorado por parte do informante, à resposta sem informação ou à resposta que, apesar de preenchida, não era aceita pelo sistema e passava a ser identificada como “ignorada”<sup>39</sup>. Já as variáveis “em branco” são interpretadas como um reflexo da falta de cuidado e da pouca importância dada pelo profissional no preenchimento do instru-

mento de coleta<sup>24</sup>. Glatt<sup>14</sup> acrescentou à definição de campo “incompleto”, o preenchido com numeral zero, data ignorada ou termo que indique ausência de dados (não consta, não informado, etc.). Por isso, maior atenção deve ser dada quando se analisa banco de dados que apresenta campos com perguntas abertas. Quando se analisou o instrumento original de coleta de dados também foram considerados “incompletos” campos com respostas simultâneas<sup>28</sup> ou aqueles cuja letra do profissional que o preencheu estava ilegível gerando dúvidas<sup>21</sup>.

Uma vantagem que poucos pesquisadores de países da América do Sul dispõem<sup>27</sup> é a coleta de dados para o estudo a partir de *download* dos bancos de dados e/ou a tabulação desses dados diretamente do sítio do Datasus. Embora exista esta facilidade de acesso, as análises de completitude a partir dessas informações refletem a qualidade do preenchimento após o envio de dados para a base nacional. Esses dados estão sujeitos a erros inerentes ao próprio sistema decorrentes de falha na execução da transferência<sup>15</sup> e aos procedimentos para limpeza do banco de dados realizadas pela Coordenação Nacional dos Programas e/ou sistemas de informação antes da sua disponibilização<sup>14</sup>.

A restrição de variáveis para acesso público no sítio do Datasus é uma limitação para estudos com alguns sistemas de informação. Enquanto no SIM e Sinasc pode-se efetuar *download* dos bancos de dados de todas as variáveis, exceto as de identificação do indivíduo, no Sinan, Sishiperdia e Siscolo só é possível realizar tabulação das variáveis selecionadas pelo Datasus, Secretaria de Assistência à Saúde (SAS) ou Secretaria de Vigilância em Saúde (SVS), ocasionando redução no escopo dessas avaliações.

Em síntese, observou-se que a completitude ainda é uma das dimensões de qualidade pouco explorada nos SIS do Brasil. Os escassos estudos realizados evidenciam que os problemas observados são comuns a quase todos os sistemas de informação, ocasionando, em algumas situações, limitações no uso dessas informações para a tomada de decisões. Estudos semelhantes aos identificados, com diversos níveis de complexidade, são factíveis de serem realizados pelas equipes responsáveis pela coleta, processamento e retroalimentação das informações nos três níveis de gestão do SUS e universidades, auxiliando na divulgação das informações coletadas e na melhor completitude dos dados.

Estudos dessa natureza contribuem para identificar inadequações estruturais nos instru-

mentos de coleta de dados; carência de treinamento dos operadores em decorrência de atualizações do sistema de informação; revisão de documentos técnicos referentes à classificação de variáveis, entre outras. Podem ainda, se constituir em estratégias para melhoria da qualidade dos dados e informações, ao estimular o uso integrado de bancos de dados, possibilitando o resgate de informações incompletas ou inconsistentes.

### **Colaboradores**

LOS Correia trabalhou na concepção, pesquisa nas bases de dados e redação do artigo; BM Padilha na pesquisa nas bases de dados e revisão final do artigo; SML Vasconcelos na concepção, redação e revisão final do artigo.

## Referências

- Lima CRA, Schramm JMA, Coeli CM, Silva MEM. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. *Cad Saude Publica* 2009; 25(10):2095-2109.
- Calazans ATS. Qualidade da informação: conceitos e aplicações. *TransInformação* 2008; 20(1):29-48.
- Paim I, Nehmy RMQ, Guimarães CG. Problemática do conceito "Qualidade" da Informação. *Perspect Ci Inf* 1996; 1(1):111-119.
- English L. *Improving data ware house and business information quality*. New York: John Wiley & Sons; 1999.
- Coeli CM, Carmargo KRCJ, Sanches KRB, Cascão AM. Sistemas de informação em saúde. In: Medronho RA, Carvalho DM, Bloch KV, Luiz RR, Werneck GL. *Epidemiologia*. São Paulo: Atheneu; 2009. p. 525-530.
- Branco MAF. Sistema de informação em saúde no nível local. *Cad Saude Publica* 1996; 12(2):267-270.
- Almeida MF. Descentralização de sistemas de informação e o uso das informações a nível municipal. *Inf Epidemiol SUS* 1998; 7(3):28-33.
- Barros SG, Chaves SCL. A utilização do sistema de informações ambulatoriais (SIA-SUS) como instrumento para caracterização das ações de saúde bucal. *Epidemiol Serv Saude* 2003; 12(1):41-51.
- Costa JMB, Frias PG. Avaliação da completude das variáveis da Declaração de Nascido Vivo de residentes em Pernambuco, Brasil, 1996 a 2005. *Cad Saude Publica* 2009; 25(3):613-624.
- Luquetti LB, Laguardia J. Confiabilidade dos dados de atendimento odontológico do Sistema de Gerenciamento de Unidade Ambulatorial (SIGAB) em Unidade de Básica de Saúde do Município do Rio de Janeiro. *Epidemiol Serv Saude* 2009; 3(18):255-264.
- Theme MM, Gama SGN, Cunha CB, Leal MC. Confiabilidade do Sistema de Informações sobre Nascidos Vivos Hospitalares no Município do Rio de Janeiro, 1999-2001. *Cad Saude Publica* 2004; 20(Supl. 1):S83-S91.
- Almeida MF, Alencar GP, Novaes HD, Ortiz LP. Sistema de informação e mortalidade perinatal: conceito e condições de uso em estudos epidemiológicos. *Rev Bras Epidemiol* 2006; 9(1):56-68.
- Strong DM, Lee YW, Wang RY. 10 potholes in the road to information quality. *Computer IEEE* 1997; 30(8):38-46.
- Glatt R. *Análise da qualidade da base de dados de AIDS do sistema de informação de agravos de notificação (SINAN) [dissertação]*. Rio de Janeiro: Fundação Oswaldo Cruz; 2005.
- Oliveira MEP, Soares MRAL, Costa MCN, Mota ELA. Avaliação da completude dos registros de febre tifóide notificados no Sinan pela Bahia. *Epidemiol Serv Saude* 2009; 18(3):219-226.
- Barbuscia DM, Rodrigues AL. Completude da informação nas Declarações de Nascido Vivo e nas Declarações de Óbito, neonatal precoce e fetal, da região de Ribeirão Preto, São Paulo, Brasil, 2000-2007. *Cad Saude Publica* 2011; 27(6):1192-1200.
- Girianelli VR, Thuler LCS, Silva GA. Qualidade do sistema de informação do câncer do colo do útero no estado do Rio de Janeiro. *Rev Saude Publica* 2009; 43(4):580-588.
- Cerqueira ACB, Sales CMM, Lima R, Silva MZ, Vieira RCA, Brioschi AP, Fagundes Moreira-Silva S, Lima LHM, Espinosa Miranda A, Nória Maciel EL. Completude do sistema de informação de agravos de notificação compulsória de gestante HIV positivo entre 2001 e 2006, no Espírito Santo, Brasil. *UFES Rev Odont* 2008; 10(1):33-37.
- Moreira CMM, Maciel ELN. Completude dos dados do Programa de Controle da Tuberculose no Sistema de Informação de Agravos de Notificação no Estado do Espírito Santo, Brasil: uma análise do período de 2001 a 2005. *J Bras Pneumol* 2008; 34(4):225-229.
- Silva GF, Aidar T, Mathias TAF. Qualidade do Sistema de Informações de Nascidos Vivos no estado do Paraná, 2000 a 2005. *Rev Esc Enferm USP* 2011; 45(1):79-86.
- Zillmer JGV, Schwartz E, Muniz RM, Lima LM. Avaliação da completude das informações do hiperdia em uma Unidade Básica do Sul do Brasil. *Rev Gaucha Enferm* 2010; 31(2):240-246.
- Malhão TA, Oliveira GP, Codenotti SB, Moherdau F. Avaliação da completude do Sistema de Informação de Agravos de Notificação da Tuberculose, Brasil, 2001-2006. *Epidemiol Serv Saude* 2010; 19(3):245-256.
- Vanderlei LCM, Simões FTPA, Vidal SA, Frias PG. Avaliação de preditores do óbito neonatal em uma série histórica de nascidos vivos no Nordeste brasileiro. *Rev Bras Saude Mater Infant* 2010; 10(4):449-458.
- Costa JMBS, Frias PG. Avaliação da completude das variáveis da Declaração de Nascido Vivo de residentes em Pernambuco, Brasil, 1996 a 2005. *Cad Saude Publica* 2009; 25(3):613-624.
- Costa JMBS, Frias PG. Avaliação da completude das variáveis da declaração de óbitos de menores de um ano residentes em Pernambuco, 1997-2005. *Cienc Saude Colet* 2011; 16(Supl. 1):1267-1274.
- Romero DE, Cunha CB. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). *Cad Saude Publica* 2006; 22(3):673-681.
- Romero DE, Cunha CB. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informações sobre Nascidos Vivos, 2002. *Cad Saude Publica* 2007; 23(3):701-714.
- Pedrosa LDC, Sarinho SW, Ximenes RA, Ordonha MR. Qualidade dos dados sobre óbitos neonatais precoces. *Rev Assoc Med Bras* 2007; 53(5):389-394.
- Macente LB, Zandonade E. Avaliação da completude do sistema de informação sobre mortalidade por suicídio na região Sudeste, Brasil, no período de 1996 a 2007. *J Bras Psiquiatr* 2010; 59(3):173-181.
- Almeida MF, Alencar GP, Schoeps D, Minuci EG, Silva ZP, Ortiz LP, Novaes HMD, Alencar AP, Raspatini PR, Santos PC. Qualidade das informações registradas nas declarações de óbito fetal em São Paulo, SP. *Rev Saude Publica* 2011; 45(5):845-853.

31. Félix JD, Zandonade E, Amorim MHC, Castro DS. Avaliação da completude das variáveis epidemiológicas do Sistema de Informação sobre Mortalidade em mulheres com óbitos por câncer de mama da Região Sudeste - Brasil (1998 a 2007). *Cienc Saude Colet* 2012; 17(4):945-953.
32. Mello-Jorge MHP, Laurenti R, Gotlieb SLD. Avaliação dos sistemas de informação em saúde do Brasil. *Cad Saude Colet* 2010; 18(1):7-18.
33. Sistema de Informação de Agravos de Notificação (SINAN). Relatórios gerenciais. [homepage on the Internet]. [acessado 2006 ago 3]. Disponível em: <http://dtr2004.saude.gov.br/sinanweb/index.php?name=-completitude#>
34. Piper JM, Mitchell Junior EF, Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee Birth certificates using maternal and newborn hospital records. *Am J Epidemiol* 1993; 137(7):758-768.
35. Oliveira EFB, Oliveira HB, Azevedo JLM, Fagundes DJ. Avaliação de descritores na angiologia e cirurgia vascular em artigos publicados em dois periódicos nacionais. *Acta Cir Bras* 2003; 18(1):1-14.
36. Pereira TA, Monteiro EFS. Terminologia DeCS e as novas regras ortográficas da língua portuguesa: orientações para uma atualização. *Acta Cir Bras* 2012; 27(7):509-514.
37. Pedraza DF. Qualidade do Sistema de Informações sobre Nascidos Vivos (SINASC): análise crítica da literatura. *Cien Saude Colet* 2012; 17(10):2729-2737.
38. Jardim AI, Leal AMO. Qualidade da informação sobre diabéticos e hipertensos registrada no Sistema HIPERDIA em São Carlos-SP. *Physis* 2009; 19(2):405-417.
39. Cruz MM, Toledo LM, Santos EM. O sistema de informação de AIDS do município do Rio de Janeiro: suas limitações e potencialidades enquanto instrumento da vigilância epidemiológica. *Cad Saude Publica* 2003; 19(1):81-89.

---

Artigo apresentado em 15/03/2013

Aprovado em 31/03/2013

Versão final apresentada em 01/04/2013