

Qualidade dos sistemas de informações sobre nascidos vivos e sobre mortalidade no Rio Grande do Sul, Brasil, 2000 a 2014

Quality of information systems on live births and mortality in Rio Grande do Sul, 2000 to 2014

Marilyn Agranonik (<https://orcid.org/0000-0003-2699-9628>)¹

Renata Oliveira Jung (<https://orcid.org/0000-0001-8424-6224>)¹

Abstract *This study assesses the quality of the SIM and SINASC information systems in coverage, incompleteness and consistency aspects, as well as the contribution of the linkage for data retrieval. It includes all live births and infant deaths in Rio Grande do Sul from 2000 to 2014. The records were paired by deterministic linkage through the DNV number and, in its absence, by probabilistic linkage. SINASC's coverage rose from 72.2% in 2000 to 98.9%, namely a 37% increase in the number of matched records. All variables in SINASC presented excellent incompleteness throughout the period, except for the number of dead children and maternal occupation. SIM presented poor or very poor incompleteness for most of the variables until 2003. Although it improved, in 2014, six variables still presented regular or poor incompleteness. The linkage procedure greatly reduced the incompleteness for most variables. There was a great variability in terms of consistency: while for gender this percentage was over 97% throughout the period, for another five variables it was still less than 75% in 2014. SINASC presented high coverage level and excellent incompleteness. Problems related to consistency persist. This study shows the linkage technic efficiency to retrieve missing information.*

Key words *Health information systems, Vital statistics, Data accuracy, Birth certificates, Death certificates*

Resumo *O estudo avalia a qualidade do SIM e do SINASC nos aspectos cobertura, incompletude e consistência, bem como a contribuição do “linkage” para a recuperação de dados. Foram analisados nascimentos e óbitos de menores de um ano ocorridos no Rio Grande do Sul entre 2000 e 2014. Os registros foram pareados por “linkage” determinístico através do número da DNV e, na ausência deste, por “linkage” probabilístico. A cobertura do SINASC aumentou 37%, passando de 72,2% em 2000 para 98,9%. O grau de incompletude do SINASC foi excelente para todas as variáveis, exceto quantidade de filhos mortos e ocupação materna. No SIM, até 2003 a maioria das variáveis apresentou preenchimento ruim ou muito ruim. Apesar da melhoria, seis delas ainda possuíam preenchimento regular ou ruim em 2014. Após o “linkage”, a incompletude reduziu-se para grande parte das variáveis. Houve alta variabilidade quanto à consistência: sexo apresentou percentual superior a 97% em todo período, enquanto para outras cinco variáveis essa fração segue inferior a 75% em 2014. Destaca-se a alta cobertura e a excelente incompletude do SINASC. Persistem problemas relacionados à consistência de informações. Evidencia-se a relevância do “linkage” como método para recuperar informações.*

Palavras-chave *Sistemas de informação em saúde, Estatísticas vitais, Confiabilidade dos dados, Declaração de nascimento, Atestado de óbito*

¹ Núcleo de Indicadores Sociais, Fundação de Economia e Estatística. R. Duque de Caxias 1.691, Centro Histórico. 90010-283 Porto Alegre RS Brasil. marilyn.agranonik@gmail.com

Introdução

Os Sistemas de Informação em Saúde (SIS) possibilitam a obtenção de evidências acerca das condições de saúde de uma população. Esses sistemas têm por objetivo armazenar e disponibilizar dados que podem ser utilizados para planejar, organizar e avaliar os serviços de saúde, bem como monitorar e examinar as características epidemiológicas de uma população¹.

O desenvolvimento e a operacionalidade dos SIS trazem consigo a necessidade de assegurar a qualidade de suas informações. A importância de registros completos e fidedignos e da qualidade geral dos dados surge a partir da concepção de que, sem a precisa identificação dos problemas, deficiências e carências de uma comunidade e na ausência de avaliação dos impactos das intervenções, as políticas públicas tornam-se potencialmente inócuas^{2,3}. Sendo assim, dados confiáveis permitem acompanhar a evolução de desfechos de interesse. Além disso, proporcionam subsídios para delinear políticas em saúde e sustentar a tomada de decisões por parte de seus gestores^{1,4}.

O Brasil dispõe de diversos Sistemas de Informação em Saúde, dentre eles o Sistema de Informações sobre Mortalidade (SIM) e o Sistema de Informações sobre Nascidos Vivos (SINASC). Essas bases de dados, gerenciadas pela Secretaria de Vigilância em Saúde do Ministério da Saúde, reúnem os registros dos eventos vitais ocorridos em todo território nacional. Os documentos básicos para a coleta de dados do SIM e do SINASC são, respectivamente, a Declaração de Óbito (DO)⁵ e a Declaração de Nascido Vivo (DNV)⁶, ambos padronizados nacionalmente e distribuídos pelo Ministério da Saúde. Na DNV são registradas informações da mãe, da gestação e da criança, além de dados referentes ao estabelecimento de ocorrência do parto e responsável pelo preenchimento da declaração⁶. A DO inclui dados de identificação do falecido e das condições e causas da ocorrência do óbito. No caso de óbitos de menores de um ano, há uma seção específica na DO incluindo algumas das informações contidas na DNV, como características maternas, do indivíduo e da gestação, além do tipo de óbito (fetal ou não fetal). Desde 1996 as causas de óbito se baseiam na 10ª Revisão da Classificação Internacional de Doenças - CID-10⁵. Os dados de ambas as bases são disponibilizados no site do DATASUS com no máximo dois anos de defasagem.

O monitoramento da qualidade dos dados dos SIS no Brasil não segue um plano regular de avaliações normatizado pelo Ministério da Saú-

de, mas resulta de iniciativas não sistemáticas e isoladas⁷. Especificamente em relação às bases de dados de nascimentos e óbitos, as primeiras ponderações acerca do SINASC ocorreram já na década de 1990, com maior foco na cobertura, consistência e incompletude dos registros a nível municipal⁸⁻¹⁰. Tais estudos, assim como investigações posteriores¹¹⁻¹⁵, em geral descrevem este sistema de forma positiva em relação a estes aspectos. Os resultados, contudo, variam de acordo com as regiões e os períodos de análise.

A avaliação do SIM, considerando apenas os estudos que se restringem a óbitos de menores de um ano, na maioria dos casos é realizada em conjunto com o SINASC para fins de verificação da consistência e recuperação de informações¹⁶⁻²¹. Nestes casos, o uso da técnica de encadeamento de arquivos (*linkage*) permitiu a recuperação de dados do SIM a partir do SINASC, indicando que o preenchimento do segundo tende a ser melhor do que o do primeiro^{19,21}.

Grande parte dos estudos sobre qualidade do SIM e/ou do SINASC trazem análises de regiões e/ou períodos restritos^{9,10,12,13,15-21}. Devido à constatação de grandes disparidades regionais quanto ao preenchimento e à qualidade tanto do SIM como do SINASC^{22,23}, faz-se necessário realizar essa investigação em nível local, porém considerando também as transformações ao longo do tempo. Neste sentido, os objetivos deste estudo, considerando óbitos e nascimentos ocorridos no Rio Grande do Sul (RS), no período de 2000 a 2014, são: (1) Avaliar o grau de *cobertura* do SINASC; (2) Avaliar o grau de *incompletude* de variáveis selecionadas da DNV e da DO; (3) Descrever o *ganho de informações* ao encadear dados do SIM e do SINASC e (4) Descrever o grau de *consistência* entre as informações das duas bases de dados.

Métodos

Bases de dados

Foi realizado um estudo descritivo de série temporal, utilizando informações de todos os nascimentos e óbitos de menores de um ano de vida ocorridos no Rio Grande do Sul entre 2000 e 2014. Este período foi selecionado por dispor de um número mais amplo de variáveis. As informações foram obtidas através do Sistema de Informações de Nascidos Vivos (SINASC) e do Sistema de Informações de Mortalidade (SIM), disponibilizados pelo Núcleo de Informações

em Saúde da Secretaria Estadual da Saúde do Rio Grande do Sul (NIS/DGTI/SES/RS).

Foram avaliadas as seguintes variáveis: (a) características sociodemográficas maternas, como *idade, escolaridade, ocupação, quantidade de filhos vivos e quantidade de filhos mortos*; (b) características da gestação e do parto, como *duração da gestação, tipo de parto, tipo de gravidez e número de consultas pré-natal* (esta última somente na DNV); (c) características do recém-nascido, como *sexo, peso ao nascer, índice de Apgar no 1º e no 5º minuto* (somente para DNV) e *presença de malformações congênitas* (somente para DNV); e (d) características geográficas, como *local e estabelecimento de ocorrência* (do nascimento ou do óbito, dependendo da base de dados). Foi observada também a presença de número da DNV e da DO, bem como valores duplicados para cada um desses registros. Para posterior utilização no encadeamento probabilístico, foram examinadas ainda as variáveis *data de nascimento e nome da mãe*.

Limpeza e encadeamento de arquivos

Inicialmente foram removidos registros duplicados de cada uma das bases de dados, bem como corrigidos erros ortográficos claros na variável *nome da mãe* (remoção de vírgulas, pontos e números e padronização de registros contendo as palavras “ignorado”, “não informado” e “não identificado”).

Para avaliar a cobertura e a consistência das informações, os registros do SIM e do SINASC foram encadeados. Para tanto, foram empregadas duas abordagens de encadeamento de arquivos: a determinística e a probabilística. A primeira foi utilizada para relacionar os registros de cada banco pelo *número da DNV* (variável chave presente em ambas as bases), quando esta informação estava disponível. O segundo método relacionou cada óbito do SIM com o correspondente nascimento no SINASC na ausência do *número da DNV*, quando seu valor estava duplicado, ou quando o *número da DNV* do SIM não possuía correspondente no SINASC. Foram selecionadas três variáveis (*nome da mãe, data de nascimento e sexo*) para o procedimento de bloqueio e comparação dos dados. Posteriormente foi verificado se o peso ao nascer era o mesmo para evitar troca de informações entre irmãos gêmeos.

O método probabilístico implica o cálculo de escores que indicam o grau de concordância entre os registros de um par. Neste estudo, os pares foram classificados como falsos para escores infe-

riores ou iguais a 8; aqueles com escore superior a 8 foram revisados manualmente e reclassificados como verdadeiros ou falsos.

Avaliação do grau de cobertura

O grau de cobertura foi medido através da proporção de óbitos relacionados ao respectivo nascimento, após o encadeamento do SINASC e do SIM. A cobertura superior a 90% foi classificada como excelente, sendo este o valor mínimo preconizado pelo Ministério da Saúde para o uso de dados no cálculo direto de indicadores²⁴.

Avaliação do grau de incompletitude

O grau de incompletitude foi definido como o percentual de informações não preenchidas, ou seja, os ignorados (código “9” nos manuais do SIM e SINASC) ou os registros com campo em branco. No caso das variáveis *número da DO e da DNV, data de nascimento, ocupação materna e estabelecimento de ocorrência do óbito ou nascimento* não há código indicando se a informação é ignorada, portanto, todas são deixadas em branco quando ausentes.

O grau de incompletitude foi classificado, para cada variável, conforme os pontos de corte propostos por Romero e Cunha¹¹: excelente (menor de 5%), bom (5 a 10%), regular (10 a 20%), ruim (20 a 50%) e muito ruim (50% ou mais).

Ganho de informações após o linkage

Posteriormente à junção das duas bases de dados foi mensurada a recuperação de informações ausentes/ignoradas. Tomou-se como base a variável do SINASC dada sua conhecida superioridade em termos de preenchimento em relação ao SIM. Quando disponíveis, valores preenchidos no SIM substituíram os valores ausentes no SINASC.

O ganho de informação, para cada variável, foi definido como a razão entre o número de registros preenchidos depois do *linkage* e antes desse procedimento.

Grau de consistência

O grau de consistência revela a concordância entre as informações do SINASC e do SIM, sendo medido através do percentual de casos com ausência de contradições (códigos diferentes) para cada variável analisada. Foram considera-

dos como consistentes os registros de DNV e DO com preenchimento igual, exceto nos casos onde a informação era ignorada (código “9”) ou não preenchida. Cabe destacar que foram considerados consistentes os registros nos quais a *idade materna* do SIM era até um ano maior do que a do SINASC, uma vez que o valor poderia ter mudado entre os eventos.

A consistência foi avaliada somente para dez variáveis comuns à DNV e à DO: (1) *sexo*, (2) *peso ao nascer*, (3) *duração da gestação*, (4) *tipo de parto*, (5) *tipo de gravidez*, (6) *idade materna*, (7) *escolaridade materna*, (8) *ocupação materna*, (9) *quantidade de filhos vivos* e de (10) *filhos mortos*.

Análise de dados

A tendência de cada variável foi avaliada por meio de regressão polinomial. Para cada variável foram comparados três modelos: linear ($Y=b_0+b_1X$), quadrático ($Y=b_0+b_1X+b_2X^2$) e cúbico ($Y=b_0+b_1X+b_2X^2+b_3X^3$). A variável dependente (Y) representa as proporções calculadas e a independente (X) foi definida como os anos do estudo. Para evitar autocorrelação entre observações, o ano foi centralizado, ou seja, de cada ano foi subtraído o ano médio da série. Para selecionar o modelo de regressão mais adequado observou-se o valor do coeficiente de determinação ajustado (R^2_{ajust}), o valor p e a análise de resíduos²⁵. O nível de significância adotado foi de 5%.

As análises estatísticas e o encadeamento determinístico foram realizados no software STATA, versão 14.0 (StataCorp LP, College Station, Estados Unidos). O encadeamento probabilístico foi obtido através do programa *Link plus* versão 2.0 (Centers for Disease Control and Prevention (CDC), Atlanta, EUA).

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital Ernesto Dornelles.

Resultados

Encadeamento de arquivos

Originalmente estavam registrados como ocorridos no Rio Grande do Sul 2.180.783 nascimentos no SINASC e 28.868 óbitos no SIM (Figura 1). Após a exclusão de quatro registros duplicados no SIM e 261 no SINASC, 28.864 óbitos foram relacionados com 2.180.522 nascimentos.

Todos os registros do SINASC possuíam número de DNV, entretanto 23,8% dos registros de óbito não continham essa informação no período

de estudo. Através do método determinístico, 73,5% dos óbitos foram pareados com os registros do SINASC. Destes, verificou-se que 210 pares eram falsos, pois, apesar do número de DNV no SIM e no SINASC ser igual, o *nome da mãe* e/ou o *ano de nascimento* eram diferentes. Estes registros foram, portanto, avaliados através do relacionamento probabilístico, assim como os registros que não possuíam a variável chave. Constatou-se ainda a presença de duplicação do número da DNV tanto no SINASC (248 casos) como no SIM (30 casos); o pareamento probabilístico foi adotado também nestes casos. Por meio deste segundo tipo de encadeamento de arquivos, foram pareados 18,7% dos óbitos. Ao longo do período de estudo as variáveis *nome da mãe*, *data de nascimento* e *sexo*, utilizadas no encadeamento probabilístico, apresentaram menos de 1,5% de incompletude. No total, 91,5% dos registros foram pareados, não sendo possível reunir 2.464 (8,5%) óbitos de menores de um ano.

Avaliação do grau de cobertura

Verificou-se um aumento de 37,0% na cobertura do SINASC ao longo do tempo, passando de 72,2% em 2000 para 98,9% em 2014 ($p < 0,001$). Isso se deve especialmente ao fato de que, no final do período estudado, um maior número de DOs continha o número da DNV, possibilitando, ainda na fase de encadeamento determinístico, um maior emparelhamento de registros. De fato, no que se refere ao total de pares reunidos, o percentual de registros pareados pelo método determinístico cresceu 88,8%, passando de 52,5% no ano 2000, para 99,1% em 2014. Desde 2004 a cobertura é superior a 90%, ou seja, excelente.

Avaliação do grau de incompletude

Em relação à DNV, o grau de incompletude foi excelente para todas as variáveis selecionadas ao longo de todo período, exceto para *quantidade de filhos mortos* e *ocupação materna* (Tabela 1). Estas exibiram os maiores percentuais de campos ignorados ou em branco em todos os anos: *quantidade de filhos mortos* atingiu 10,1% em 2000, mas caiu para 2,8% em 2014; *ocupação materna*, por sua vez, possui o menor percentual de ignorados registrado em 2004 (4,9%) e o maior em 2011 (11,8%), em uma tendência ascendente ($p = 0,021$).

Destacam-se ainda as variáveis *quantidade de filhos vivos*, *índice de Apgar* e *presença de malformações congênitas*, por apresentarem percentual de ignorados ou em branco de até 5% no período

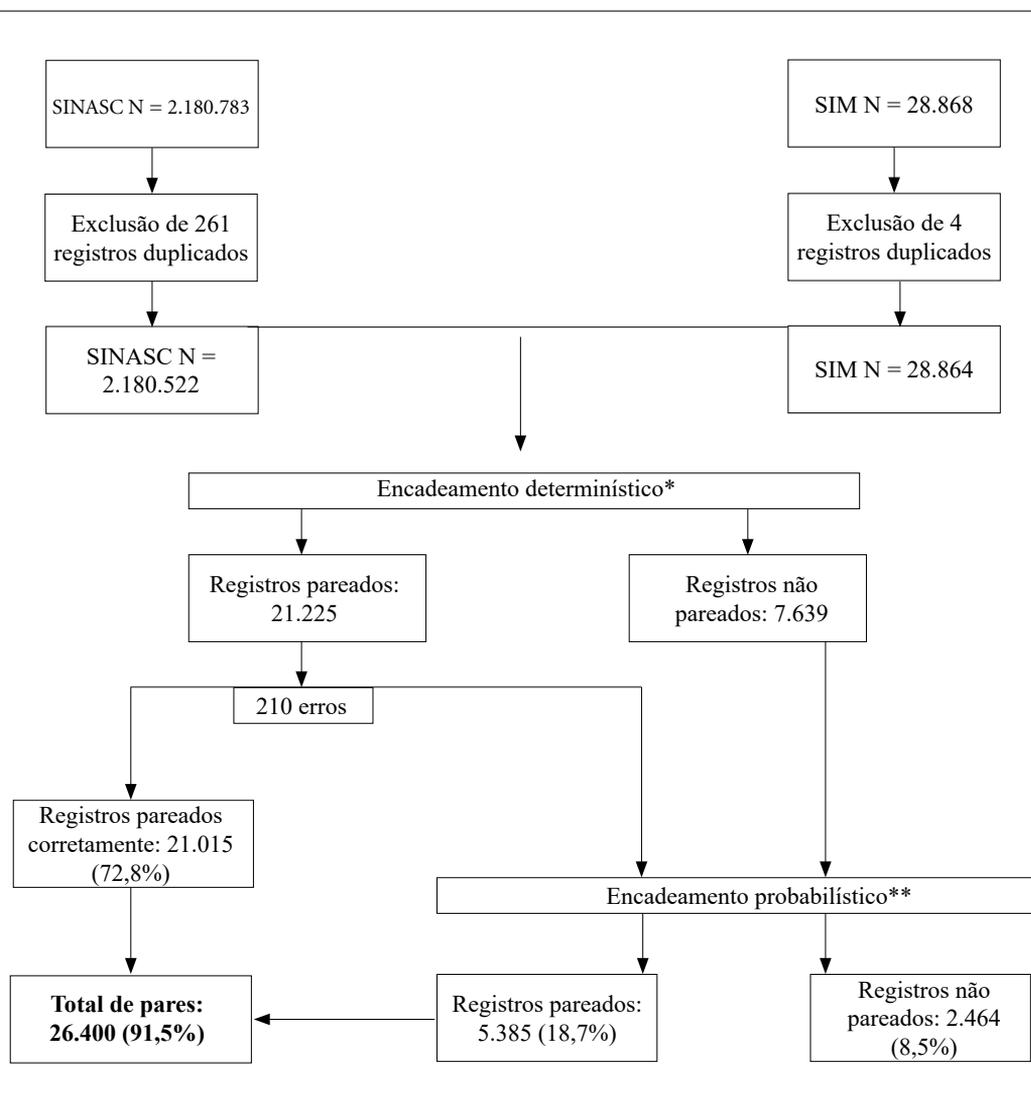


Figura 1. Encadeamento das bases de dados do SINASC e do SIM, através de relacionamento determinístico ou probabilístico de arquivos para óbitos ocorridos no Rio Grande do Sul, 2000-2014.

*Encadeamento determinístico utilizando número da DNV; **Encadeamento probabilístico utilizando as variáveis: nome da mãe, sexo do RN e data de nascimento do RN.

do. Por outro lado, para as variáveis *sexo*, *peso ao nascer*, *data de nascimento*, *tipo de parto*, *tipo de gravidez*, *nome e idade materna* e *local e estabelecimento de ocorrência* foram encontrados menos de 1% de registros incompletos durante os 15 anos avaliados.

Para a maioria das variáveis o grau de incompletitude diminuiu: em 2014 o maior percentual foi de 9,5% para *ocupação materna*. Em relação à tendência da incompletitude, verificou-se um comportamento diferenciado entre as variáveis.

Houve aumento da proporção de incompletitude para *malformações congênitas* ($p = 0,034$), *duração da gestação* ($p = 0,022$) e *ocupação materna* ($p = 0,021$). Para *tipo de gravidez* ($p = 0,893$), *de parto* ($p = 0,810$) e *local de ocorrência* ($p = 0,463$) esse percentual permaneceu constante. A tendência para as demais variáveis foi decrescente na maior parte do período, conforme mostram os modelos de regressão da Tabela 1. Em 2014, 14 das 18 variáveis analisadas atingiram menos de 1% de ignorados ou em branco.

Tabela 1. Percentual de incompletitude de variáveis selecionadas do SINASC para nascimentos ocorridos no Rio Grande do Sul, 2000-2014.

Variáveis / anos	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Δ%	Modelo	R ² _{ajustado}	p*
Nº de nascimentos	175720	160455	155182	149102	152996	147295	141589	133557	135352	133838	133445	137887	139107	141523	143474				
Sexo	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,01	0,02	100,0	Y = 0,0063 - 0,0007X + 0,0001X ² + 0,00003X ³	0,563	0,007
Peso ao nascer	0,26	0,31	0,21	0,17	0,10	0,07	0,05	0,03	0,02	0,02	0,02	0,02	0,01	0,01	0,01	-96,2	Y = 0,04 - 0,02X + 0,002X ²	0,941	< 0,001
Apgar no 1º min	3,35	3,17	2,92	2,60	1,90	2,38	2,33	2,51	2,73	1,87	1,54	1,34	1,02	1,00	0,72	-78,5	Y = 2,09 - 0,17X	0,836	< 0,001
Apgar no 5º min	3,26	2,94	2,77	2,49	1,73	2,26	2,20	2,52	2,67	1,83	1,49	1,29	0,98	0,94	0,70	-78,5	Y = 2,00 - 0,16X	0,793	< 0,001
Malformações congênitas	2,30	0,63	0,50	0,49	0,57	0,66	1,27	0,58	0,40	0,52	1,03	2,82	2,91	2,63	1,34	-41,7	Y = 0,76 + 0,10X + 0,03X ²	0,337	0,034
Data de nascimento	0,00	0,00	0,00	0,00	0,00	0,00	-	-	-	-	-	-	-	-	-	-	Y = 0,0001 - 0,0003X	0,459	0,003
Duração da gestação	0,36	0,32	0,29	0,31	0,13	0,13	0,08	0,10	0,08	0,14	0,16	2,11	2,94	1,55	0,93	158,3	Y = 0,64 + 0,11X	0,290	0,022
Tipo de parto	0,03	0,04	0,01	0,03	0,04	0,03	0,02	0,02	0,03	0,05	0,08	0,02	0,01	0,03	0,03	0,0	-	-	0,810
Tipo de gravidez	0,05	0,07	0,03	0,04	0,03	0,03	0,02	0,05	0,03	0,13	0,12	0,02	0,01	0,03	0,03	-40,0	-	-	0,893
Nº de consultas pré-natal	1,55	0,96	0,69	0,54	0,55	0,47	0,41	0,59	0,45	0,70	0,52	0,60	0,76	0,73	0,36	-76,8	Y = 0,48 + 0,06X + 0,010X ² - 0,003X ³	0,864	< 0,001
Nome da mãe	0,14	0,12	0,03	0,03	0,04	0,02	0,02	0,01	0,01	0,01	0,00	0,00	0,00	0,00	0,02	-85,7	Y = 0,005 - 0,001X + 0,001X ² + 0,0002X ³	0,870	< 0,001
Idade materna	0,24	0,15	0,10	0,08	0,06	0,03	0,02	0,00	0,00	0,00	0,01	0,00	0,01	0,00	0,00	-100,0	Y = 0,01 - 0,01X + 0,002X ² - 0,0002X ³	0,983	< 0,001
Escolaridade materna	1,90	0,98	0,57	0,60	0,55	0,68	0,56	0,75	0,70	0,64	0,61	0,75	0,84	0,67	0,26	-86,3	Y = 0,59 + 0,08X + 0,01X ² - 0,004X ³	0,782	< 0,001

continua

Tabela 1. Percentual de incompletitude de variáveis selecionadas do SINASC para nascimentos ocorridos no Rio Grande do Sul, 2000-2014.

Variáveis / anos	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Δ%	Modelo	R ² ajustado	p*
Nº de nascimentos	175720	160455	155182	149102	152996	147295	141589	133557	135352	133838	133445	137887	139107	141523	143474				
Ocupação materna	7,61	10,06	6,00	7,13	4,92	7,06	5,57	10,30	8,95	8,92	10,04	11,79	10,38	9,85	9,47	24,4	Y = 8,54 + 0,27X	0,294	0,021
Nº de filhos vivos	4,69	4,21	4,04	3,50	2,87	3,43	2,95	2,73	2,55	2,21	2,45	3,56	4,00	3,02	1,82	-61,2	Y = 3,20 - 0,11X	0,334	0,014
Nº de filhos mortos	10,07	8,97	8,65	7,46	6,08	6,87	6,13	5,43	5,20	4,55	4,66	5,73	5,71	4,01	2,77	-72,5	Y = 5,71 - 0,19X + 0,02X ² - 0,01X ³	0,888	< 0,001
Local de ocorrência	-	-	0,00	0,00	0,00	0,00	0,00	0,03	0,00	0,08	0,05	0,00	-	0,00	0,00	-	-	-	0,463
Estabelecimento de ocorrência	0,76	0,71	0,60	0,52	0,41	0,37	0,34	0,35	0,32	0,35	0,41	0,31	0,26	0,28	0,27	-64,5	Y = 0,35 - 0,01X + 0,004X ² - 0,001X ³	0,945	< 0,001

Fonte dos dados brutos: SINASC / Núcleo de Informações em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul. (-): dado numérico igual a 0 (zero) não resultante de arredondamento. Δ%: variação percentual (2014/2000); R² ajustado: coeficiente de determinação ajustado; *Valor p para o teste F do modelo de regressão.

Ao longo de todo o período, as variáveis *peso ao nascer, tipo de parto, índice de Apgar no 1º e no 5º minuto, quantidade de filhos vivos e mortos, ocupação materna e estabelecimento de ocorrência* apresentaram maior número de registros em branco do que com o código “9” (ignorado).

Em relação ao preenchimento da DO, o grau de incompletitude evoluiu de forma positiva no período estudado, com redução da fração de valores ignorados ou em branco para todas as variáveis, quando comparados os valores de 2000 e 2014 (Tabela 2). Entre 2000 e 2003, a maior parte das variáveis analisadas obteve preenchimento ruim ou muito ruim (*quantidade de filhos mortos, quantidade de filhos vivos, ocupação materna, número da DNV, escolaridade e idade materna, peso ao nascer, duração da gestação, tipo de parto e tipo de gravidez*) ou regular (*estabelecimento da ocorrência óbito*).

Apesar do percentual de incompletitude ter diminuído em relação ao início do período, em 2014 as variáveis *escolaridade materna, quantidade de filhos vivos, quantidade de filhos mortos, duração da gestação e idade materna* ainda possuíam escore regular e a *ocupação materna*, classificação ruim. As variáveis *nome da mãe, local e estabelecimento de ocorrência* exibiram tendência de queda no percentual de incompletitude ao longo de todo período. Por outro lado, para *data de nascimento* a tendência foi ascendente ($p < 0,001$). Entretanto, devido ao pequeno número de registros ausentes para essa variável, o grau de incompletitude se manteve excelente, com menos de 1% de ignorados. Para *sexo* a tendência não foi estatisticamente significativa ($p = 0,309$). As demais variáveis mostraram comportamento quadrático, com redução acentuada da proporção de não preenchimento na primeira metade do período e posterior aumento dessa fração, especialmente entre 2013 e 2014 (Tabela 2).

Destacam-se por apresentar, ao longo de todo período, menos de 1% de incompletitude as variáveis: *sexo, data de nascimento e local de ocorrência do óbito*. Para a variável *nome da mãe*, após 2002, identificou-se menos de 0,7% de ignorados ou em branco. O *número da DO, a data de óbito, a idade do falecido, o tipo e a causa básica de óbito* estiveram presentes para todos os registros.

Ao longo de todo o período foram observados mais registros em branco do que com o código “9” (ignorado), exceto para as variáveis *sexo e local de ocorrência do óbito*.

Tabela 2. Percentual de incompletitude de variáveis selecionadas do SIM para óbitos ocorridos no Rio Grande do Sul, 2000-2014.

Variáveis / anos	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Δ%	Modelo	R ² ajustado	p*
Nº de óbitos	2664	2535	2424	2385	2324	2016	1857	1707	1729	1562	1508	1598	1509	1507	1539				
Sexo	0,38	0,16	0,08	0,17	0,30	0,25	0,11	0,29	0,12	0,06	0,33	0,13	0,33	0,73	0,19	-50,0	-	-	0,309
Peso ao nascer	33,56	28,44	24,22	23,94	3,10	1,74	0,70	1,46	0,98	1,86	2,72	6,01	0,66	1,39	0,84	-97,5	Y = 2,11 - 2,08X + 0,36X ²	0,841	< 0,001
Número da DNV	55,37	49,63	43,23	41,30	18,63	18,11	4,68	16,93	11,10	23,37	8,55	11,95	1,39	1,73	1,04	-98,1	Y = 14,7 - 3,6X + 0,31X ²	0,833	< 0,001
Data de nascimento	-	0,08	0,08	-	0,13	-	-	0,35	0,52	0,32	0,33	0,50	0,20	0,40	0,65	-	Y = 0,24 + 0,04X	0,603	< 0,001
Duração da gestação	34,08	28,13	23,10	24,44	5,29	3,52	2,10	2,58	1,85	4,61	3,58	8,14	4,84	7,70	14,55	-57,3	Y = 2,80 - 1,50X + 0,45X ²	0,878	< 0,001
Tipo de parto	33,00	27,57	22,94	23,61	4,30	3,17	1,67	2,40	1,85	4,61	3,25	5,63	1,46	4,51	9,94	-69,9	Y = 2,57 - 1,71X + 0,4X ²	0,869	< 0,001
Tipo de gravidez	31,68	26,39	21,74	22,73	4,00	2,08	1,56	1,82	1,27	3,52	2,92	5,44	0,86	4,05	9,55	-69,9	Y = 2,01 - 1,65X + 0,39X ²	0,867	< 0,001
Nome da mãe	0,90	1,07	1,49	0,67	0,52	0,20	0,43	0,12	0,35	0,19	0,27	0,19	0,27	0,27	0,06	-93,3	Y = 0,47 - 0,07X	0,568	0,001
Idade materna	38,89	33,41	27,15	27,80	9,47	8,58	2,42	1,82	4,05	7,68	4,38	7,32	2,25	5,84	12,35	-68,2	Y = 4,72 - 2,05X + 0,44X ²	0,896	< 0,001
Escolaridade materna	47,15	39,92	32,14	33,75	14,11	15,77	5,65	12,13	9,25	14,47	10,28	12,70	9,01	13,54	19,23	-59,2	Y = 10,64 - 2,01X + 0,46X ²	0,885	< 0,001
Ocupação materna	47,18	40,36	33,17	35,01	15,53	17,91	12,76	18,86	13,59	21,32	14,59	20,59	12,19	19,71	23,91	-49,3	Y = 15,64 - 1,59X + 0,4X ²	0,810	< 0,001
Nº de filhos vivos	43,24	40,47	32,71	37,61	15,66	20,88	16,53	12,48	10,01	13,96	10,48	10,76	4,44	7,76	14,94	-65,4	Y = 13,84 - 2,43X + 0,3X ²	0,884	< 0,001
Nº de filhos mortos	54,95	54,32	41,30	44,86	23,75	37,75	28,33	17,98	12,20	19,01	16,11	15,89	7,69	12,01	18,91	-65,6	Y = 21,26 - 3,1X + 0,31X ²	0,873	< 0,001
Local de ocorrência	0,53	0,39	0,21	0,42	0,22	0,25	0,48	0,29	0,29	0,19	-	0,06	0,27	0,13	-	-	Y = 0,25 - 0,03X	0,491	0,002
Estabelecimento da ocorrência	13,78	11,44	12,71	11,82	9,85	9,97	9,26	9,84	7,69	7,94	8,69	7,20	6,23	7,03	5,20	-62,3	Y = 9,24 - 0,52X	0,903	< 0,001

Fonte dos dados brutos: SIM / Núcleo de Informações em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul.

(-): dado numérico igual a zero não resultante de arredondamento. Δ%: variação percentual (2014/2000); R²_{ajustado}: coeficiente de determinação ajustado; *Valor p para o teste F do modelo de regressão.

Ganho de informação após o linkage

Entre 2000 e 2003, foi possível ampliar em mais de 20% o número de valores preenchidos a cada ano para todas as variáveis, exceto *sexo*. A partir de então, houve uma redução do ganho de informações. Entretanto, ao final do período ocorreu um novo aumento, levando todas as variáveis, exceto *peso ao nascer* e *sexo*, a um ganho de informação entre 10% e 25%.

Para todas as variáveis, exceto *sexo*, a proporção de ganho de informação apresentou comportamento quadrático, com redução desta fração entre 2000 e 2008 e aumento a partir de 2009 (Tabela 3). A variável *sexo* já possuía excelente grau de incompletitude (Tabela 2) e não mostrou tendência significativa para a proporção de ganho de informação no período (p = 0,915).

Em geral, a proporção de registros incompletos reduziu-se acentuadamente após o *linkage*. A partir de 2004, todas as variáveis atingiram grau excelente de incompletitude, exceto *quantidade de filhos mortos* e *ocupação materna*, que permaneceram com escore bom. Considerando as dez variáveis presentes na DNV e na DO, constatou-se que, após a recuperação dos registros ausentes, 99,4% dos nascimentos ainda possuíam duas ou menos informações incompletas; em relação aos óbitos, esse percentual era de 99,5%.

Grau de consistência

O grau de consistência exibiu uma trajetória quadrática no intervalo de tempo em análise, com tendência de elevação nos cinco primeiros anos e estabilização ou leve queda a partir de então (Tabela 4). O nível de consistência de todas as variáveis foi abaixo de 80% entre 2000 e 2003, exceto para *sexo* e *tipo de gravidez*. Embora a concordância entre os dados tenha melhorado desde 2000, a partir de 2012 algumas variáveis vêm apresentando tendência de queda no percentual de registros consistentes. Em particular, entre 2012 e 2014, as variáveis *duração da gestação*, *idade materna*, *escolaridade materna*, *ocupação materna* e *quantidade de filhos mortos* tiveram redução de mais de 10% nesta fração.

Verificou-se grande variabilidade entre as variáveis nesta dimensão. A variável *sexo* se destaca por ter um percentual de registros consistentes superior a 97% em todos os anos e as variáveis *peso ao nascer* e *tipo de gravidez* mostram mais de 90% a partir de 2004. Contudo, as variáveis *duração da gestação*, *escolaridade materna*, *ocupação*

Tabela 3. Ganho de informação (percentual) após o encadeamento das bases de dados SIM e SINASC para óbitos ocorridos no Rio Grande do Sul, 2000-2014.

Variáveis / anos	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Δ%	Modelo	R ² ajustado	p*
Sexo	0,04	-	-	-	0,09	-	-	0,06	-	-	-	0,06	0,07	-	-	-	-	-	0,915
Peso ao nascer	27,3	26,7	23,0	23,3	1,0	0,7	0,1	0,1	0,4	0,1	0,8	3,9	0,2	0,3	0,2	-99,3	Y = 1,15 - 1,93X + 0,32X ²	0,809	< 0,001
Duração da gestação	28,1	26,1	21,2	23,6	2,9	2,2	1,2	1,1	0,9	2,8	1,8	5,0	2,7	6,0	15,4	-45,4	Y = 1,36 - 1,35X + 0,43X ²	0,837	< 0,001
Tipo de parto	27,7	25,7	21,9	23,2	2,2	2,0	0,8	1,2	1,1	3,0	1,6	3,8	0,9	3,6	10,5	-61,9	Y = 1,50 - 1,55X + 0,38X ²	0,830	< 0,001
Tipo de gravidez	25,9	24,0	20,1	21,6	1,9	1,1	0,8	0,6	0,5	1,9	1,2	3,6	0,4	3,2	10,1	-61,1	Y = 0,93 - 1,45X + 0,37X ²	0,831	< 0,001
Idade materna	37,8	34,5	27,5	29,3	7,0	7,2	1,4	0,3	3,3	6,1	2,1	5,3	1,8	5,1	13,3	-64,9	Y = 3,26 - 2,1X + 0,48X ²	0,877	< 0,001
Escolaridade materna	55,4	46,2	35,4	38,9	11,4	14,7	4,1	10,9	7,9	12,8	7,7	10,2	7,9	13,4	22,5	-59,3	Y = 8,48 - 2,47X + 0,62X ²	0,887	< 0,001
Ocupação materna	49,0	44,3	33,5	37,2	10,5	12,3	8,4	13,8	9,4	16,3	7,5	11,9	7,2	16,5	24,6	-49,8	Y = 9,98 - 2,04X + 0,55X ²	0,836	< 0,001
Nº de filhos vivos	45,2	46,5	34,5	43,5	13,0	19,1	15,3	9,3	7,5	12,0	6,7	8,4	3,7	6,7	15,8	-65,0	Y = 11,65 - 2,79X + 0,4X ²	0,845	< 0,001
Nº de filhos mortos	68,8	78,8	46,8	56,8	21,2	45,9	30,9	14,0	8,5	17,3	12,5	13,4	7,0	11,1	20,8	-69,9	Y = 20,21 - 4,36X + 0,54X ²	0,827	< 0,001

Fonte dos dados brutos: SINASC e SIM / Núcleo de Informações em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul.

(-): dado numérico igual a 0 (zero) não resultante de arredondamento. Δ%: variação percentual (2014/2000); R² ajustado: coeficiente de determinação ajustado; *Valor p para o teste F do modelo de regressão.

Tabela 4. Percentual de registros consistentes comparando variáveis selecionadas do SIM e SINASC para óbitos ocorridos no Rio Grande do Sul, 2000-2014.

Variáveis / anos	2014														Modelo	R ² ajustado	P		
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013				2014	Δ%
Nº de pares	1923	2125	2118	2115	2190	1921	1801	1636	1671	1448	1440	1518	1493	1479	1522				
Sexo	98,8	98,4	98,4	98,6	98,4	98,5	98,4	99,2	98,7	99,0	98,3	98,0	98,9	98,2	98,1	-0,7	-	-	0,425
Peso ao nascer	64,0	67,9	70,5	71,8	91,5	91,5	94,0	93,5	92,3	91,4	93,1	87,7	93,8	92,6	94,7	47,9	Y = 91,86 + 1,95X - 0,31X ²	0,822	< 0,001
Duração da gestação	57,3	62,6	66,7	67,5	86,1	87,2	90,6	88,9	90,1	85,9	86,7	78,1	81,8	81,1	73,3	27,8	Y = 88,25 + 1,21X - 0,5X ²	0,849	< 0,001
Tipo de parto	73,4	77,0	79,2	78,9	96,7	96,8	98,2	97,8	98,0	95,3	97,4	94,2	97,0	94,9	89,0	21,3	Y = 97,36 + 1,31X - 0,35X ²	0,835	< 0,001
Tipo de gravidez	74,7	78,4	81,3	80,5	97,5	98,8	98,6	98,8	99,1	96,8	98,3	95,9	98,8	96,1	90,5	21,2	Y = 98,55 + 1,31X - 0,34X ²	0,846	< 0,001
Idade materna	63,3	67,4	72,4	72,4	90,5	89,8	96,5	97,1	94,4	90,1	95,3	91,1	94,8	91,8	84,0	32,6	Y = 94,07 + 1,75X - 0,43X ²	0,875	< 0,001
Escolaridade materna	40,0	48,4	55,2	57,0	76,6	75,0	85,0	76,9	78,7	73,3	77,6	68,8	76,0	73,0	66,0	64,7	Y = 78,34 + 1,69X - 0,53X ²	0,837	< 0,001
Ocupação materna	46,9	50,4	59,2	56,6	77,4	74,0	78,5	58,3	72,7	66,6	71,3	63,4	68,6	63,5	56,8	21,0	Y = 71,73 + 0,65X - 0,4X ²	0,541	0,004
Nº de filhos vivos	34,6	40,0	49,2	47,2	54,5	50,2	47,3	49,9	50,7	47,5	52,4	50,3	49,2	59,6	49,4	42,7	Y = 51,3 + 0,81X - 0,12X ²	0,447	0,011
Nº de filhos mortos	37,4	40,9	50,8	48,8	68,1	53,8	63,4	72,1	77,6	77,6	79,2	76,4	84,2	81,0	74,6	99,4	Y = 70,95 + 3,12X - 0,28X ²	0,894	< 0,001

Fonte dos dados brutos: SINASC e SIM / Núcleo de Informações em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul.

Δ%: variação percentual (2014/2000); R² ajustado: coeficiente de determinação ajustado; *Valor p para o teste F do modelo de regressão.

materna, quantidade de filhos vivos e quantidade de filhos mortos ainda possuíam, em 2014, consistência inferior a 75%.

Discussão

A utilização de dados secundários em pesquisas da área da saúde tem como principal vantagem a economia de tempo e de recursos despendidos²⁶. Entretanto, é preciso assegurar a qualidade das informações a fim de obter indicadores confiáveis. O presente trabalho buscou analisar as informações do SINASC e do SIM quanto aos graus de cobertura, incompletitude e consistência, ao longo de uma série temporal de 15 anos no Estado do Rio Grande do Sul.

Neste estudo, foi observado um aumento gradual da cobertura, medida pelo percentual de registros pareados ao longo do tempo. Verificou-se que a partir de 2004 esta proporção passou a ser superior ao valor mínimo preconizado pelo Ministério da Saúde para uso de dados no cálculo direto de indicadores²⁴. Notou-se ainda o aumento dos pares formados via encadeamento determinístico, resultado do melhor preenchimento da variável chave (*número da DNV*). Em todo o período, entretanto, um considerável número de DOs não foi associado às respectivas DNVs. Este fato pode ter ocorrido por dois motivos: (1) houve falha na emissão de DNV para os referidos registros (subnotificação) ou (2) alguns destes óbitos são referentes a crianças nascidas fora do RS e, como foram avaliados registros de nascimentos ocorridos somente neste estado, essas DNVs podem ter sido registradas em outra localidade.

Em relação à incompletitude, o excelente preenchimento de quase todas as variáveis do SINASC vem sendo reportado por diversos estudos^{1,12-15}. O sistema pode ser considerado, no geral, de boa qualidade, atestando sua potencialidade como fonte de informações para a caracterização dos nascimentos ocorridos no Brasil^{14,22}. Especificamente no Rio Grande do Sul, observou-se uma contínua redução na proporção de valores ignorados para a maioria das variáveis, o que permite inferir que o SINASC vem sendo aprimorado ao longo dos últimos anos.

Quanto às Declarações de Óbito, apesar da melhora ao longo do período, poucas variáveis alcançaram preenchimento excelente, corroborando com os resultados de outros estudos semelhantes em outras regiões do País e períodos de análise^{27,28}. Nos primeiros anos da série, a maioria das variáveis apresentava escore muito ruim,

ruim ou regular de preenchimento, dificultando a caracterização dos óbitos infantis e possíveis análises de suas circunstâncias.

De forma semelhante a outros estudos^{1,11,12,29}, percebeu-se que em todo o intervalo algumas variáveis são sistematicamente piores do que outras, entre elas o *número de filhos vivos e mortos*, a *ocupação materna* e a *escolaridade materna*. Este comportamento foi observado tanto no SIM como no SINASC. Em relação à incompletitude das variáveis *número de filhos vivos e mortos*, Romero e Cunha¹¹ citam como possível fonte do problema a falta de clareza no manual de preenchimento, sugerindo que grande parte dos registros em branco na realidade sejam registros de casos com zero filhos. Com efeito, constatou-se que, no SIM, a correlação entre a fração de registros iguais a zero e de registros em branco (exceto ignorados - código "99") para ambas as variáveis é altamente negativa ao longo do período.

O maior grau de registros não preenchidos para a variável *ocupação materna* pode ser explicado pela falta de clareza dos manuais, nos quais não é especificado o período de referência: a instrução é imprecisa ao indicar que deve ser preenchida a ocupação *habitual* da mãe^{11,30}. Além disso, de acordo com as regras de preenchimento tanto da DNV quanto da DO, o código não precisa ser preenchido a menos que seja orientado pela Secretaria da Saúde^{5,6}.

Ainda em relação à clareza metodológica, os manuais especificam que todos os campos devem ser preenchidos, mesmo que com um traço (-) caso a informação não se aplique^{5,6}. Entretanto, observou-se que especificamente nesses casos a informação fica em branco. Verificou-se, por exemplo, para as variáveis *estabelecimento de ocorrência do óbito ou do nascimento* caselas em branco para os eventos que não ocorreram em hospitais ou outros estabelecimentos de saúde, reflexo da ausência de um código identificando neste contexto a informação "não se aplica". Costa e Frias¹² apontam que registros deixados em branco estão associados à falta de cuidado no preenchimento dos documentos, enquanto registros com o código "9" resultam da ausência de informação tanto dos prontuários como de quem fornece os dados (acompanhantes da gestante).

A incompletitude também pode ser atribuída à falta de atenção dada ao preenchimento da DO e da DNV e à classificação de algumas variáveis como menos importantes do que outras^{12,15,29}. Esta distinção tem como resultado a negligência de alguns campos tanto por parte de quem preenche as declarações como dos gestores dos sistemas³¹.

Além disso, o SINASC depara-se com a questão da heterogeneidade de responsáveis pela emissão das declarações^{15,29}, o que exige adequado treinamento e conscientização. Este requisito, contudo, muitas vezes não é atendido. No SIM, mesmo sendo obrigatório que o médico preencha a DO, o percentual de informações incompletas ainda é muito elevado. Alguns autores apontam que deficiências nos cursos médicos em relação à importância do preenchimento dos documentos têm impactos negativos sobre a qualidade dos dados¹⁸.

Ressalta-se ainda que tanto a DNV como a DO foram alteradas em 2011, com a inclusão de novas variáveis e modificações na forma de coleta de outras. No SINASC, o leve aumento no percentual de incompletude de variáveis como *duração de gestação* e *escolaridade materna* neste ano pode estar associado a esta mudança metodológica. No SIM, neste mesmo ano, observou-se uma elevação no percentual de ignorados ou em branco, não apenas nas variáveis cuja metodologia foi alterada, mas também em outras que permaneceram iguais. O uso de um novo formulário pode ter contribuído para esse aumento, na medida em que representa uma mudança operacional, exigindo treinamento e adequação de quem o preenche.

Apesar das falhas no preenchimento das declarações, o uso do método de encadeamento de arquivos permitiu reduzir a incompletude das informações. Entre as dez variáveis avaliadas, oito atingiram excelente grau de preenchimento após 2003. O ganho de informações foi superior a 20% em algumas variáveis na maior parte do período. Entre 2004 e 2009, houve uma redução do ganho de informações, resultante, em grande parte, da melhoria do preenchimento das variáveis antes do encadeamento. De forma semelhante ao apontado por outros autores^{19,21}, um maior percentual de registros ignorados ou em branco no SIM foram preenchidos por informações do SINASC, apesar de, em alguns casos, os dados presentes no SIM terem contribuído para recuperar as informações ausentes no SINASC. Isto se deve claramente ao sistemático melhor preenchimento do SINASC em relação ao SIM, permitindo que os dados daquele completem boa parte das lacunas deste.

É importante notar que, apesar das variáveis utilizadas para o encadeamento das bases de dados possuírem excelente grau de preenchimento ao longo de todo período, foram observados problemas como erros ortográficos e de digitação e incorreções, como, por exemplo, indivíduos com

o mesmo *número de DNV com nome da mãe e datas de nascimento* muito díspares. Esses problemas já haviam sido apontados por outros autores nas bases de dados de outros estados brasileiros³². A partir da metodologia de encadeamento probabilístico utilizada neste trabalho foi possível atenuar em parte estas dificuldades, na medida em que o *software* desconsidera alguns caracteres especiais, bem como possíveis erros ortográficos na avaliação dos pares.

No que se refere à consistência, verificou-se menos de 80% de concordância entre os registros para todas as variáveis, exceto *sexo*, entre 2000 e 2003. Neste mesmo intervalo, foram observados elevados percentuais de registros ignorados ou em branco no SIM para essas variáveis, porém não no SINASC. Portanto, constatou-se que a incompletude, especialmente do SIM, impacta de forma negativa a consistência de informações.

De forma geral, notou-se que as variáveis que apresentam as maiores proporções de registros incompletos e os menores graus de consistência são aquelas que podem ser classificadas como características sociodemográficas maternas (*quantidade de filhos vivos e mortos, ocupação e escolaridade*). Além das nítidas dificuldades metodológicas, segundo Romero e Cunha³⁰ em estudo comparando as UFs quanto à caracterização da condição social do nascimento, o fato de algumas variáveis sociais não serem consideradas como essenciais é uma das principais causas para que seus níveis de preenchimento sejam inferiores.

Este estudo possui algumas limitações. Alguns campos importantes para estudos sobre desfechos perinatais não foram avaliados, entre eles: estado civil materno (SINASC) e *raça/cor* (SIM e SINASC). Os dois foram excluídos das análises por não possibilitarem comparação entre as bases de dados. O primeiro não está presente no SIM. O segundo apresentou mudança apenas no SINASC, passando de *raça/cor* do recém-nascido para *raça/cor* da mãe em 2011, enquanto no SIM permaneceu referente à criança em todo período.

Considerações finais

As falhas no preenchimento da DNV e da DO associam-se a diversos fatores, entre eles a falta de clareza nos manuais, a pluralidade de responsáveis pela emissão da DNV, o maior interesse em algumas variáveis em detrimento de outras e prontuários incompletos. Estas deficiências, em parte, poderiam ser suprimidas mediante conscientização e treinamento tanto dos gestores

dos sistemas como dos responsáveis pelo preenchimento dos documentos, tendo como objetivo atenuar a prática do não preenchimento. Ainda que a incompletude das variáveis, especialmente do SIM, mostre-se como um empecilho para análises mais minuciosas do período avaliado, o encadeamento de arquivos apresentou-se como uma solução eficiente, visto que, após a recuperação de informações o grau de incompletude passou a ser bom ou excelente.

Este estudo evidenciou que o SINASC pode ser considerado uma base de dados confiável. Entretanto, apesar dos claros avanços na cobertura, no preenchimento e na consistência dos registros de ambos os sistemas, o SIM ainda encontra-se aquém dos níveis necessários para uma adequada caracterização dos desfechos em questão. Fica evidente, assim, a necessidade de aperfeiçoamento e de maior atenção à qualidade das informações de saúde, condição necessária para o sucesso de ações de saúde e eficaz alocação de recursos.

Colaboradores

M Agramonik contribuiu na concepção e planejamento do projeto de pesquisa; na obtenção dos dados, análise e interpretação destes; na redação e revisão. RO Jung contribuiu na análise e interpretação dos dados; na redação final e revisão.

Referências

1. Brasil. Ministério da Saúde (MS). *Saúde Brasil 2011: uma análise da situação de saúde e a vigilância da saúde da mulher*. Brasília: MS; 2012.
2. Lima CRA, Schramm JMA, Coeli CM. Gerenciamento da qualidade da informação: uma abordagem para o setor saúde. *Cad. Saúde Coletiva* 2010; 18:19-31.
3. World Health Organization (WHO). *Health Policy and Systems Research: A Methodology Reader*. Geneva: WHO Press; 2012.
4. Brownson RC, Fielding JE, Maylahn CM. Evidence-Based Public Health: A Fundamental Concept for Public Health Practice. *Annu Rev Public Health* 2009; 30:175-201.
5. Brasil. Ministério da Saúde (MS). *Manual de Instruções para o Preenchimento da Declaração de Óbito*. Brasília: MS; 2011.
6. Brasil. Ministério da Saúde (MS). *Manual de Instruções para o preenchimento da Declaração de Nascido Vivo*. Brasília: MS; 2011.
7. 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.
8. Mello-Jorge MHP, Gotlieb SLD, Soboll MLMS, Almeida MF, Latorre MRDO. Avaliação do sistema de saúde sobre nascidos vivos e o uso de seus dados em epidemiologia e estatísticas de saúde. *Rev Saude Publica* 1993; 27(Supl.):1-46.
9. Mello Jorge MHP, Gotlieb SLD, Andrade SM. Análise dos registros de nascimentos vivos em localidade urbana no Sul do Brasil. *Rev Saude Publica* 1997; 31(1):78-89.
10. Mishima FC, Scochi CGS, Ferro MAR, Lima RAG, Costa IAR. Declaração de Nascido Vivo: análise do seu preenchimento no Município de Ribeirão Preto, São Paulo, Brasil. *Cad Saude Publica* 1999; 15(2):387-395.
11. 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.
12. 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):613-24.
13. Silva RS, Oliveira CM, Ferreira DKS, Bonfim CV. Avaliação da completude das variáveis do Sistema de Informações sobre Nascidos Vivos - Sinasc - nos Estados da região Nordeste do Brasil, 2000 e 2009. *Epidemiol Serv Saude* 2013; 22(2):347-352.
14. Oliveira MM, Andrade SSCA, Dimech GS, Oliveira JCG, Malta DC, Neto DLR, Moura L. Avaliação do Sistema de Informações sobre Nascidos Vivos. Brasil, 2006 a 2010. *Epidemiol Serv Saude* 2015; 24(4):629-640.
15. 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.
16. Silva CF, Leite AJM, Almeida NMGS. Linkage entre bancos de dados de nascidos vivos e óbitos infantis em município do Nordeste do Brasil: qualidade dos sistemas de informação. *Cad Saude Publica* 2009; 25:1552-58.
17. Barbuscia DM, Rodrigues-Júnior 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.
18. Silva LP, Moreira CMM, Amorim MHC, Castro DS, Zandonade E. Avaliação da qualidade dos dados do Sistema de Informações sobre Nascidos Vivos e do Sistema de Informações sobre Mortalidade no período neonatal, Espírito Santo, Brasil, de 2007 a 2009. *Cien Saude Colet* 2014; 19(7):2011-2020.
19. Maia LTS, Souza WV, Mendes ACG. A contribuição do linkage entre o SIM e SINASC para a melhoria das informações da mortalidade infantil em cinco cidades brasileiras. *Rev Bras Saude Mater Infant* 2015; 15(1):57-66.
20. Marques LJP, Oliveira CM, Bonfim CV. Avaliação da completude e da concordância das variáveis dos Sistemas de Informações sobre Nascidos Vivos e sobre Mortalidade no Recife-PE, 2010-2012. *Epidemiol Serv Saude* 2016; 25(4):849-854.
21. Mendes ACG, Lima MM, Sá DA, Oliveira LCS, Maia LTS. Uso da metodologia de relacionamento de bases de dados para qualificação da informação sobre mortalidade infantil nos municípios de Pernambuco. *Rev. Bras. Saude Matern Infant* 2012; 12(3):243-249.
22. Szwarcwald CL, Leal MC, Andrade CLT, Souza Júnior PRB. Estimativa da mortalidade infantil no Brasil: o que dizem as informações sobre óbitos e nascimentos do Ministério da Saúde? *Cad Saude Publica* 2002; 18(6):1725-1736.
23. Brasil. Ministério da Saúde (MS). *Saúde Brasil 2005: uma análise da situação de saúde no Brasil*. Brasília: MS; 2005.
24. Rede Interagencial de Informação para a Saúde. *Indicadores básicos para a saúde no Brasil: conceitos e aplicações*. 2ª ed. Brasília: Organização Pan Americana da Saúde (OPAS); 2008.
25. Latorre MR, Cardoso MR. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Rev Bras Epidemiol* 2001; 4(3):145-152.
26. Drummond EF, Machado CJ, Vasconcelos MR, França E. Utilização de dados secundários do SIM, Sinasc e SIH na produção científica brasileira de 1990 a 2006. *Revista Brasileira de Estudos de População* 2009; 26(1):7-19.
27. Ramalho MOA, Frias PG, Vanderlei LCM, Macêdo VC, Lira PIC. Avaliação da incompletude da declaração de óbitos de menores de um ano em Pernambuco, Brasil, 1999-2011. *Cien Saude Colet* 2015; 20(9):289-298.
28. Soares JAS, Horta FMB, Caldeira AP. Avaliação da qualidade das informações em declarações de óbitos infantis. *Rev Bras Saude Matern Infant* 2007; 7(3):289-295.
29. Pedraza DF. Qualidade do Sistema de Informações sobre Nascidos Vivos (Sinasc): análise crítica da literatura. *Cien Saude Colet* 2012; 17(10):2729-2737.
30. Romero DE, Cunha CB. Avaliação do grau de preenchimento e definições das variáveis sociais no Sistema de Informação de Nascidos Vivos. Diferenciais por Unidade Federada. *XIV Encontro Nacional de Estudos Populacionais*; 2004 20-24 de Setembro; Caxambú, MG.

31. Costa JMBS, Frias PG. Avaliação da completitude das variáveis da Declaração de Óbito de menores de um ano residentes em Pernambuco, 1997-2005. *Cien Saude Colet* 2011; 16(Supl. 1):1267-1274.
32. Silva JPL, Travassos C, Vasconcellos MM, Campos LM. Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil. *Cad Saúde Coletiva* 2006; 14(2):197-224.

Artigo apresentado em 18/04/2017
Aprovado em 02/09/2017
Versão final apresentada em 04/09/2017