

A contribuição do *linkage* entre o SIM e SINASC para a melhoria das informações da mortalidade infantil em cinco cidades brasileiras

The contribution of the linkage between the SIM and SINASC to improving information on infant mortality in five Brazilian cities

Lívia Teixeira de Souza Maia ¹
Wayner Vieira de Souza ²
Antonio da Cruz Gouveia Mendes ³

¹⁻³ Centro de Pesquisas Aggeu Magalhães. Av. Professor Moraes Rego, s/n - Campus da UFPE. Cidade Universitária. Recife, PE, Brasil. CEP: 50.670-420. E-mail: livia_tsouza@yahoo.com.br

Abstract

Objectives: to analyze the contribution of the linkage between the Mortality Information System (SIM) and the Live Births Information System (SINASC) to improving the quality of data on infant mortality (IM) in five cities, one in each Brazilian macro-region.

Methods: an analysis was conducted of the deaths of children aged under one year registered in the SIM in 2005 and live births in 2004 and 2005 included in the SINASC. The data bases were related using linkage. A statistical association was found between the linkage of data and cities and IM components, as well as the degree of completeness of the variables before and after linkage.

Results: Ninety percent of deaths were related to the live birth declaration, with the cities of Belém and Goiânia having the lowest proportions of paired records. Porto Alegre had the highest percentage of paired records, followed by Guarulhos and Recife. Both city of residence of the child and death certificate component were found to have a significant influence on the quality of information. Of the 2,464 incomplete fields for variables common to both systems, 92% were recovered after linkage of the databases.

Conclusions: employing this technique made it possible to improve vital statistics and this demonstrates the potential of this strategy for surveillance of infant mortality at health services.

Key words Infant mortality, Information systems, Epidemiological surveillance

Resumo

Objetivos: analisar a contribuição do *linkage* entre o Sistema de Informações sobre mortalidade (SIM) e o Sistema de informações sobre nascidos vivos (SINASC) para a melhoria da qualidade dos dados de mortalidade infantil (MI) em cinco cidades, sendo uma de cada macrorregião brasileira.

Métodos: analisaram-se os óbitos de menores de um ano registrados no SIM em 2005 e os nascidos vivos em 2004 e 2005 contidos no SINASC. As bases de dados foram relacionadas através do *linkage*. Verificou-se a associação estatística entre o relacionamento dos dados e as cidades e os componentes da MI, bem como o percentual de incompletude das variáveis antes e após o *linkage*.

Resultados: 90% dos óbitos foram relacionados à declaração de nascido vivo, sendo as cidades de Belém e Goiânia onde se obteve a menor proporção de registros pareados. Porto Alegre apresentou maior percentual de pares, seguida de Guarulhos e Recife. Verificou-se uma influência significativa entre a cidade da criança e o componente do óbito na qualidade da informação. Dos 2464 campos incompletos para as variáveis comuns aos sistemas, 92% foram recuperados após o relacionamento das bases.

Conclusões: a utilização dessa técnica possibilitou qualificar das estatísticas vitais e demonstrou o potencial de aplicação dessa estratégia na vigilância do óbito infantil nos serviços de saúde.

Palavras-chave Mortalidade infantil, Sistemas de informação, Vigilância epidemiológica

Introdução

A morte e o nascimento constituem os primeiros eventos vitais a serem registrados de forma sistemática na história. No Brasil, o Ministério da Saúde implantou o Sistema de Informações sobre Mortalidade (SIM) em 1975 e o Sistema de Informações sobre Nascidos Vivos (SINASC) em 1990. Apesar do longo período de funcionamento, estes sistemas ainda apresentam inconsistências que, apesar de não comprometerem a análise global dos eventos, necessitam ser solucionadas de modo a torná-los uma fonte de informação confiável ao planejamento e avaliação do impacto das ações de atenção à saúde.^{1,2}

Apesar da ampliação da cobertura e da melhora na qualidade dos dados dos SIM e SINASC, especialmente nos últimos dez anos, ainda constata-se importantes desigualdades na qualidade das informações de nascimentos e óbitos de menores de um ano no Brasil.³⁻⁵ Tais diferenças são principalmente observadas ao se comparar a qualidade e a cobertura destes sistemas nas regiões brasileiras, na qual se constata que o Norte e Nordeste apresentam qualidade inferior em relação às Regiões Sul, Sudeste e Centro-Oeste. Discrepâncias assim também se observam entre municípios de grande e de pequeno porte, com melhor informação nas cidades grandes.⁶

Com relação ao SIM, problemas como o erro na classificação do tipo de óbito e da causa básica da morte, a subnotificação, o sub-registro e a presença de variáveis ignoradas ou não preenchidas, ainda comprometem a qualidade e a confiabilidade dos dados, e consequentemente, a obtenção de dados reais sobre a mortalidade infantil.^{1,7-9}

A boa cobertura do SINASC em diversas áreas do país tem sido apontada na literatura,^{5,7} entretanto ainda persiste a preocupação com a melhoria tanto da completude das variáveis quanto da confiabilidade das informações.⁹

A análise de completude das variáveis é considerada uma importante dimensão da avaliação da qualidade das estatísticas vitais e reflete a falta de cuidado e de importância dada ao preenchimento pelos profissionais de saúde, ausência de informação nos prontuários médicos e até o desconhecimento de certas informações pelos acompanhantes da mulher/criança.^{10,11}

Com os avanços tecnológicos ocorridos na área da informática a crescente disponibilidade de largas bases de dados informatizadas na saúde, a aplicação da metodologia de relacionamento de banco de dados ou *linkage* tem assumido importância por permitir monitorar eventos e/ou aumentar a quanti-

dade e a qualidade das informações disponibilizadas.^{5,12}

O desenvolvimento de programas computacionais que permitem a integração de diferentes bases de dados tem contribuído sobremaneira para a difusão dessa metodologia. Destacando-se três, de distribuição gratuita: o *Reclink*, software desenvolvido por pesquisadores brasileiros;¹³ o *Febri*, criado pela Universidade Nacional Australiana;¹⁴ e o *Link Plus*, desenvolvido e adotado pelos *Centers for Disease Control and Prevention* (CDC) de Atlanta-GA, Estados Unidos da América.¹⁵

O método do *linkage* permite a identificação de um mesmo indivíduo em diferentes bancos de dados em saúde, com crescente utilização nas pesquisas em saúde para investigar diversos objetos de estudos, tais como pesquisas etiológicas, estudos sobre migrantes e avaliação dos serviços de saúde.¹²

As necessidades dos gestores e pesquisadores da área da Saúde em integrar as informações associadas à ausência de identificador unívoco às bases de dados, pela forma como foram estruturados os sistemas de informações em saúde no Brasil, determinam a necessidade de integrar dados desses sistemas pelo emprego do método probabilístico.^{12,16}

Quanto à mortalidade infantil, a integração do SIM e do SINASC tem como potencialidades: fornecer elementos para avaliar cobertura e qualidade das informações; estabelecer associação entre variáveis da declaração de nascido vivo (DNV) e a mortalidade infantil; e finalmente, permitir a plena utilização de dados oficiais secundários, disponibilizados por estes sistemas, em estudos analíticos longitudinais com baixos custos operacionais.^{5,17,18}

O presente trabalho teve como objetivo verificar a contribuição do *linkage* entre as bases de dados do SIM e do SINASC para a qualidade das informações sobre a mortalidade infantil em cinco cidades brasileiras, com ênfase no aspecto da completude das variáveis comuns aos dois sistemas.

Métodos

Foi realizado um estudo descritivo de corte transversal no qual foram relacionadas as bases de dados do SIM e do SINASC pelo método de *linkage*. Analisaram-se os óbitos de menores de um ano ocorridos em 2005 registrados no SIM e os nascidos vivos em 2004 e 2005 registrados no SINASC. Optou-se pelos dois anos dos dados do SINASC para que houvesse a possibilidade de resgatar todas as DNV referentes aos óbitos infantis do ano de 2005, que incluem os nascimentos no ano anterior.

Foi selecionada uma cidade de cada macroregião brasileira para o estudo e de acordo com os seguintes critérios: porte populacional entre 1.200.000 e 1.600.000 habitantes em 2005, segundo dados do Instituto Brasileiro de Geografia e Estatística (IBGE)¹⁹ e 100% de cobertura dos sistemas SIM e SINASC sendo classificadas, portanto, como “adequadas” de acordo o Sistema de Monitoramento dos Indicadores de Mortalidade Infantil (MONITORIMI).²⁰ Assim, foram selecionadas as cidades de Belém, Recife, Guarulhos, Porto Alegre e Goiânia.

O *linkage* de bases de dados é um instrumento metodológico que possibilita o relacionamento de fontes de informação diferentes em um só registro,²¹ tendo sido utilizado para esse estudo os métodos determinístico e probabilístico.

No relacionamento determinístico a unificação dos registros é realizada por meio de um identificador unívoco presente nas diferentes bases de dados. Na inexistência desse campo único, faz-se necessário a aplicação do método probabilístico que utiliza rotinas automatizadas para sua execução baseando-se em campos comuns (ex: nome, data de nascimento) presentes em ambos os bancos de dados com o objetivo de identificar a probabilidade de um par de registros pertencer à mesma pessoa.^{12,21}

Nesse estudo foi executado inicialmente o método determinístico considerando como variável unificadora o número da DNV, uma vez que se trata de uma variável comum aos dois sistemas e de preenchimento obrigatório para os óbitos infantis no SIM. Nesta etapa foram utilizados os programas: Epi Info, versão 6.04d de 1997, produzido pelo CDC e planilhas eletrônicas do Microsoft Office Excel 2003. No Epi Info aplicou-se o comando “Join” do programa “MERGE file” pelo número da DNV. Com produto dessa etapa obteve-se o primeiro banco dos dados concatenados.

Para *linkage* dos registros não pareados deterministicamente recorreu-se ao método probabilístico por meio do *Reclink III*[®] versão 3.0.4.4005. Esse programa permite associar arquivos com base no relacionamento probabilístico de registros e encontra-se disponível gratuitamente na internet.^{12,13}

Executou-se o método probabilístico através da estratégia de múltiplos passos, associada a uma revisão manual dos pares duvidosos, buscando classificá-los como pares verdadeiros ou não pares. As rotinas automáticas utilizadas foram: padronização, relacionamento e combinação dos arquivos.

A primeira rotina teve por objetivo padronizar os arquivos visando minimizar o efeito de erros ou diferenças de grafia sobre o pareamento das

variáveis selecionadas (nome da mãe, município de residência, data de nascimento e sexo da criança). Dois novos bancos de dados padronizados foram criados a partir das bases originais. Esses bancos foram compostos pelos campos-chaves de cada arquivo, utilizados no processo de blocagem e pareamento além dos campos utilizados nos critérios de decisão sobre verdadeiro *status* de um par na revisão manual.

Os campos-chaves utilizados foram: *soundex* do primeiro nome da mãe, *soundex* do último nome da mãe, nome do município de residência e sexo da criança. As variáveis utilizadas para critério de decisão (inspeção manual dos pares) foram: endereço de residência da mãe, bairro de residência da mãe, complemento endereço de residência, idade da mãe, data do óbito da criança e ano de nascimento da criança.

Na rotina de relacionamento associaram-se os arquivos padronizados a partir dos parâmetros de blocagem e pareamento previamente definidos, tendo como arquivo de comparação os bancos de dados do SINASC por apresentar o maior número de registros e como arquivo de referência o banco de dados do SIM. O processo de relacionamento foi realizado iniciado pela chave de blocagem mais específica passando para as mais sensíveis nos passos subsequentes.

Ainda para essa fase do processamento, foram calculados os escores para estimar a probabilidade de identificação de pares. Foi criado um arquivo de configuração (cln) que continha, inicialmente, os parâmetros sugeridos pelos autores do programa *Reclink III*.^{12,13} Para estimar a probabilidade de identificar um par como verdadeiro quando ele é falso utilizou-se uma fração amostral de 10% e para estimar a probabilidade de identificar um par como verdadeiro quando ele é realmente verdadeiro foi utilizado o programa *CalcParms* que disponibiliza rotinas de estimação que usam o algoritmo *Expectation-Maximisation* (EM).^{12,13} Após a estimação desses parâmetros foram alteradas as configurações dos campos de comparação.

Seguiu-se com a rotina final que consiste na operação de combinação de arquivos, na qual os pares verdadeiros identificados em cada uma das estratégias de relacionamento foram salvos em um arquivo. Consideraram-se pares apenas os que apresentaram todas as variáveis coincidentes, sendo ainda realizada uma revisão manual de todos os identificados.

Por fim, os arquivos das cinco cidades, contendo os registros pareados foram unificados em um único arquivo. Essa base de dados foi analisada segundo o

tipo de *linkage* (determinístico e probabilístico), componente do óbito infantil (neonatal e pós-neonatal) e as cinco cidades estudadas. Os óbitos neonatais são os ocorridos de 0 a 27 dias de vida, e o pós-neonatal os que se deram a partir 28 dias a um ano de vida incompleto.²²

Ao partir da premissa de que quanto melhor a qualidade da informação, maior a chance de sucesso do *linkage*, buscou-se verificar a associação entre o pareamento do SIM e o SINASC e as variáveis independentes: cidades e componente do óbito infantil.

Para verificar a existência de diferenças significativas entre a proporção de pares e não pares resultantes do relacionamento das bases de dados foi aplicado o teste de qui-quadrado (χ^2) com cálculo da respectiva significância estatística ($p < 0,05$), calculado-se o teste χ^2 de tendência para as cidades e o χ^2 Mantel-Haenszel para o componente do óbito. Também se calculou a *Odds Ratio* (OR), estimada em modelo de regressão, a fim de verificar as chances de não relacionamento entre os sistemas e as referidas variáveis.

Para a análise de completude, foram selecionadas variáveis comuns às duas bases de dados: sexo da criança, idade da mãe, raça/cor, escolaridade materna, ocupação da mãe, número de filhos nascidos vivos, número de filhos nascidos mortos, tipo de parto, duração da gestação, peso ao nascer e tipo de gravidez.

Consideraram-se informações incompletas aquelas cujos campos não estavam preenchidos ou com código de ignorado. O percentual de incompletude de cada variável, no SIM e no SINASC, foi calculado antes e depois do *linkage*.

A pesquisa lidou com dados secundários do SIM e SINASC fornecidos pelo Ministério da Saúde mediante termo de cessão e utilização dos bancos de dados e mediante assinatura do termo de responsabilidade e foi realizada dentro dos padrões da ética científica, tendo sido submetida e aprovada pelo Comitê de Ética em Pesquisa (Registro no CAAE – 0067.0.095.000-07).

Resultados

A estruturação dos bancos de dados e das etapas do *linkage* é apresentada na Figura 1. Dos 1685 óbitos de menores de um ano registrados nas cinco cidades estudadas, foi possível relacionar 90,0% Declaração de Óbito (DO) à sua respectiva DNV (Tabela 1).

Porto Alegre foi o município que apresentou maior percentual de registros de óbito pareados (96,7%). Proporção acima de 90% também foi encontrada nas cidades de Guarulhos e Recife. Em

Belém e Goiânia se alcançou menor parcela de pares, 81,6% e 83,2%, respectivamente (Tabela 1).

Em relação ao tipo de *linkage*, 74,5% dos registros de óbito foram pareados pelo método determinístico, enquanto que o relacionamento probabilístico resultou em um percentual de pareamento de 25,5% (Tabela 1).

Em todas as cidades, com exceção de Guarulhos, verificou-se um predomínio do número de registros pareados por meio do *linkage* determinístico. Em Porto Alegre e em Recife o percentual de pares formados de forma determinística ultrapassou 95%, chegando a 99,6% e 97,0%, respectivamente. Em Belém essa proporção foi de 85,2% e de 71,3% em Goiânia. Em Guarulhos apenas 16,8% dos dados foram concatenados deterministicamente. Nessa cidade a maior contribuição foi dada pelo método probabilístico, responsável por 83,2% dos pares (Tabela 1).

Constatou-se uma associação significativa entre a cidade de nascimento da criança e o componente do óbito infantil com a qualidade da informação. Tomando-se Porto Alegre como referência, em função da maior proporção de pares, observa-se que Belém apresentou maior chance de não pareamento entre o SIM e o SINASC (OR=6,67), seguida por Goiânia (OR=5,95). Os valores da *Odds Ratio* demonstram que, para o conjunto das cinco cidades, existe uma chance de 62% (OR=1,62) de não concatenar a DO à respectiva DNV nas mortes do período pós-neonatal (Tabela 2).

Na análise de completude das onze variáveis comuns ao SIM e ao SINASC, observou-se que dos 2464 campos ignorados ou não preenchidos, restaram apenas 193 (8%) após o relacionamento das bases de dados, sendo possível recuperar 2271 registros (92%). Esse incremento na completude das informações dessas variáveis foi mais expressivo para o SIM, que passou de uma média de 12% de incompletude para 1,5% (Tabela 3 e Figura 2).

Entre as cidades, em Goiânia e Guarulhos observou-se o maior incremento na qualidade das informações. Em Goiânia, o conjunto das variáveis comuns aos dois sistemas atingia a uma média de incompletude no SIM de 28,4% antes do *linkage*, passando a um percentual de 5,1% de incompletude após o relacionamento dos registros. Na cidade de Guarulhos, a proporção de incompletude destas variáveis, que era de 21,9%, atingiu um percentual de 0,6% após o *linkage*. Nessas cidades, destacam-se as variáveis paridade materna (número de filhos nascidos vivos e mortos), raça/cor; ocupação da mãe e escolaridade materna, cujo percentual de ignorados/não preenchidos ultrapassavam 30% no SIM,

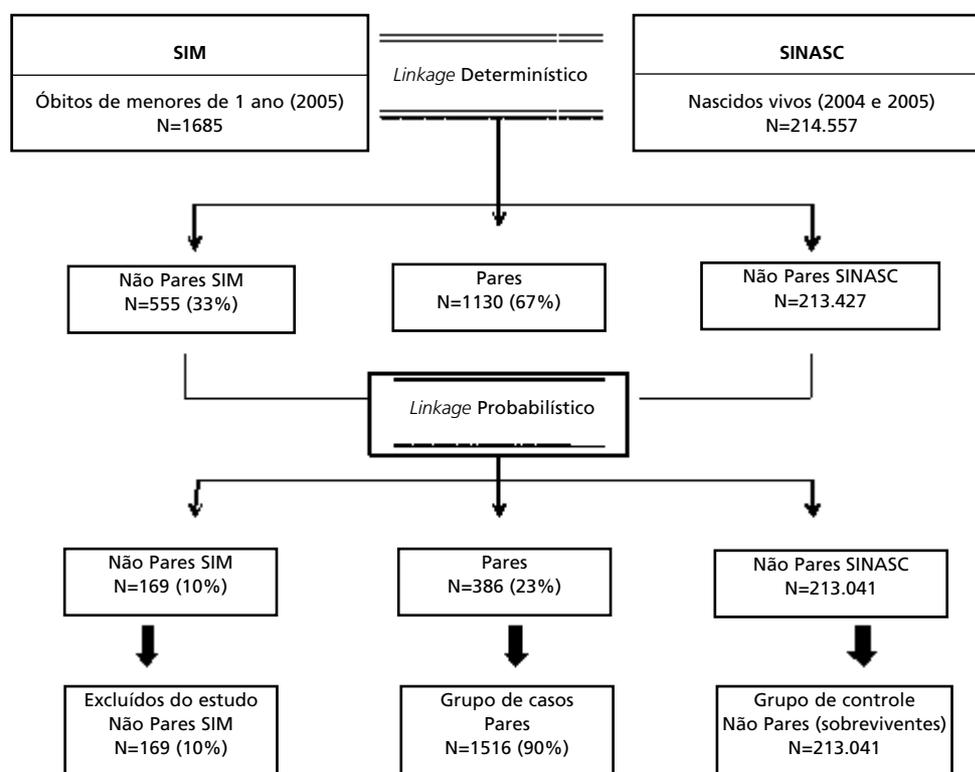
tendo uma redução significativa no percentual após o relacionamento das bases de dados (Tabela 3).

No SINASC, a variável raça/cor foi a que apresentou maior percentual de campos ignorados/não preenchidos (24,6%), atingindo um percentual de 79,6%, na cidade de Guarulhos, e de 40,8%, em Goiânia, tendo sido reduzido a 3,4% na média das cinco cidades, após o linkage. Quanto ao SIM,

observou-se que as variáveis, número de filhos nascidos mortos e ocupação da mãe, apresentaram os mais elevados percentuais de incompletude antes do relacionamento dos dados, (21,9% e 19,3%, respectivamente), os quais diminuíram para 0,2% e 7,3%, respectivamente, após a realização do linkage (Figura 2).

Figura 1

Estruturação dos bancos de dados e das etapas do linkage entre SIM e SINASC – Determinístico e Probabilístico.



SIM = Sistema de Informações sobre mortalidade; SINASC = Sistema de informações sobre nascidos vivos.

Tabela 1

Número absoluto e percentual dos óbitos infantis registrados no SIM pareados com o SINASC segundo tipo de *linkage* e cidade estudada. 2005.

Cidade	Total Óbitos	Óbitos Pareados		Pares – Tipo de <i>linkage</i>			
				Determinístico		Probabilístico	
		n	%	n	%	n	%
Recife	385	368	95,6	357	97,0	11	3,0
Goiânia	268	223	83,2	159	71,3	64	28,7
Belém	472	385	81,6	328	85,2	57	14,8
Guarulhos	316	304	96,2	51	16,8	253	83,2
Porto Alegre	244	236	96,7	235	99,6	1	0,4
Total	1685	1516	90,0	1130	74,5	386	25,5

SIM = Sistema de Informações sobre mortalidade; SINASC = Sistema de informações sobre nascidos vivos.

Tabela 2

Número absoluto, percentual e *Odds Ratio* (OR) dos óbitos infantis registrados segundo pareamento entre o SIM e o SINASC por cidade e componente do óbito. 2005.

Variável	Par		Não Par		Total		<i>Odds Ratio</i> (OR)
	n	%	n	%	n	%	
Cidade*							
Belém	385	81,6	87	18,4	472	28,0	6,67
Recife	368	95,6	17	4,4	385	22,8	1,36
Guarulhos	304	96,2	12	3,8	316	18,8	1,16
Goiânia	223	83,2	45	16,8	268	15,9	5,95
Porto Alegre	236	96,7	8	3,3	244	14,5	1,00
Total	1516	90,0	169	10,0	1685	100,0	
Componente**							
Neonatal	1031	91,5	96	8,5	1127	66,9	1,00
Pós Neonatal	485	86,9	73	13,1	558	33,1	1,62
Total	1516	100,0	169	100,0	1685	100,0	

* χ^2 (tendência)= 17,3 ($p<0,001$); ** χ^2 (Mantel-Haenszel)= 8,61 ($p=0,003$); SIM = Sistema de Informações sobre mortalidade; SINASC = Sistema de informações sobre nascidos vivos.

Tabela 3

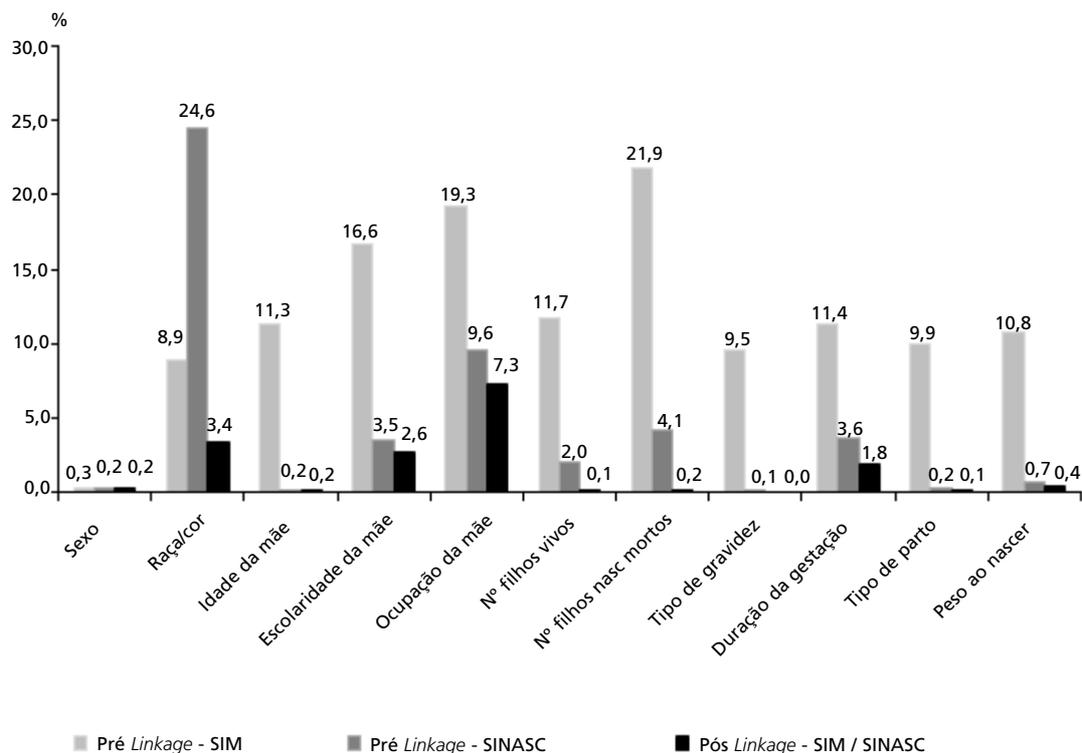
Percentual de incompletude (campos ignorados/sem preenchimento) das variáveis comuns ao SIM e ao SINASC, antes e após o linkage dos bancos de dados segundo cidade. 2005.

Variáveis	Belém		Recife		Guarulhos		Goiânia		Porto Alegre	
	Pré ¹	Pós ²								
Sexo	0,5	0,5	0,8	0,3	0,3	0,3	0,0	0,0	0,0	0,0
Raça/Cor	1,0	0,0	1,9	0,3	0,3	0,3	40,8	16,6	0,4	0,0
Idade da mãe	2,3	0,0	0,5	0,0	34,2	0,0	19,3	0,9	0,4	0,0
Escolaridade da mãe	9,6	0,0	1,9	0,5	37,8	2,0	31,4	10,3	2,5	0,4
Ocupação da mãe	2,1	0,0	0,8	0,0	41,4	1,0	34,5	18,4	17,8	16,9
Nº filhos vivos	0,5	0,0	1,4	0,0	4,6	0,3	50,7	0,0	1,3	0,0
Nº filhos nascidos mortos	41,3	0,0	4,1	0,0	7,6	1,0	55,6	0,0	0,8	0,0
Tipo de gravidez	2,1	0,0	0,3	0,0	27,6	0,0	17,5	0,0	0,0	0,0
Duração da gestação	2,6	0,0	0,8	0,0	28,9	2,0	24,2	7,2	0,4	0,0
Tipo de parto	2,6	0,0	0,5	0,0	28,6	0,0	17,5	0,4	0,4	0,0
Peso ao nascer	2,9	0,0	0,3	0,0	29,6	0,0	21,1	2,2	0,0	0,0

Pré¹ = antes do linkage; Pós² = após realização do linkage; SIM = Sistema de Informações sobre mortalidade; SINASC=Sistema de informações sobre nascidos vivos.

Figura 2

Percentual de incompletude (campos ignorados/sem preenchimento) das variáveis comuns ao SIM e ao SINASC, antes e após o linkage dos bancos de dados nas cinco cidades estudadas. 2005.



SIM = Sistema de Informações sobre mortalidade; SINASC = Sistema de informações sobre nascidos vivos.

Discussão

Os resultados do presente estudo mostraram o elevado número de pares formados entre a DO e respectiva DNV, após a utilização da técnica de *linkage*. Mesmo nas cidades de Belém e Goiânia, onde se verificaram os mais baixos percentuais de pareamento dessa análise, os achados da pesquisa foram expressivos quando comparados a estudos anteriores no âmbito nacional, que obtiveram proporções de pareamento de registros que variaram de cerca de 40% a aproximadamente 70%.^{17,23,24} Outros estudos utilizaram a busca de informações em cartórios de registro civil, além do relacionamento dos bancos de dados, como estratégia.^{25,26}

Esse elevado percentual de sucesso no *linkage* para o conjunto das cinco cidades parece estar associado à boa qualidade das informações do SIM e do SINASC nesses municípios.

Outro aspecto, provavelmente relacionado à qualidade das informações é a proporção de pares obtidos por meio do *linkage* determinístico, uma vez que esse resultado depende do preenchimento da variável unívoca (número da DNV) no SIM e no SINASC, cujo preenchimento é obrigatório para os óbitos de menores de um ano. Nesse sentido se destacam Porto Alegre e Recife, com mais de 95% de pares formados de forma determinística.

De maneira oposta, salientam-se os achados de Guarulhos, onde se verificou uma elevada quantidade de DO sem o preenchimento do número da DNV, tendo como consequência o baixíssimo percentual de pareamento pelo método determinístico. Nessa cidade mais de 80% dos registros só puderam ser pareados por meio do *linkage* probabilístico.

O Estado de São Paulo tem como peculiaridade o fato de que as estatísticas vitais são gerenciadas pela Fundação Sistema Estadual de Análise de Dados (SEAD) vinculada à Secretaria Estadual de Planejamento e Desenvolvimento Regional do Estado. O sistema realiza a integração das informações sobre nascidos vivos e óbitos do Registro Civil e das respectivas DNV e DO dos 645 municípios paulistas.²⁷ Entretanto o fato de apresentar diferenças e especificidades no processamento dos dados pode estar relacionado à baixa qualidade no preenchimento de algumas variáveis, a exemplo do número da DNV.

O município de nascimento da criança esteve associado à qualidade dos registros. Belém e Goiânia apresentaram maior chance estatisticamente significativa de não pareamento entre o SIM e o SINASC. Embora Recife e Porto Alegre, cidades do Nordeste

e do Sul do país, respectivamente, tenham apresentado semelhanças nos resultados do *linkage*, os achados deste estudo atestam uma diferença estatisticamente significativa, para o conjunto das cinco cidades, entre as chances de pareamento dos registros e o município.

Apesar da progressiva melhora nas estatísticas vitais, ainda constata-se no Brasil um gradiente regional importante, com contrastes relevantes entre as Regiões Norte e Nordeste e o Centro-Sul.^{6,26} Segundo Andrade e Szwarcwald⁶ esses resultados enfatizam as desigualdades em saúde da população brasileira, que se refletem não apenas na variação dos níveis da mortalidade infantil, mas também na desigualdade de adequação das informações para o seu cálculo. Ainda segundo esses autores, os problemas de estimação são mais presentes, justamente, nas áreas com as piores situações de saúde, detendo também maior precariedade das informações.⁶

No tocante aos componentes do óbito infantil, constatou-se que os óbitos neonatais apresentaram melhor qualidade dos registros vitais, tanto quanto ao percentual de pares, como quanto ao tipo do *linkage*, que foi confirmada pela maior proporção de registros relacionados pelo método determinístico. Esses resultados reforçam o pressuposto de que quanto mais próximos os eventos nascimento e morte, mais oportuna é a investigação e consequentemente, mais consistente a informação. Ademais, um fator que também pode estar relacionado à qualidade da informação é o fato de que a maior parte dos óbitos neonatais ocorre em ambiente hospitalar.²⁸

Esses achados corroboram estudo realizado em Pernambuco no qual, nos municípios população maior que 500 mil habitantes, mais de 80% dos óbitos neonatais foram pareados de forma determinística.²⁸

Uma das mais importantes contribuições do *linkage* foi a melhoria da completude das informações, possibilitando o resgate das variáveis comuns ao SIM e ao SINASC que constavam em um dos bancos de dados como ignoradas ou não preenchidas. Esse aporte ocorreu mais expressivamente do SINASC para o SIM, de maneira semelhante ao que foi identificado por Mendes *et al.*²⁸

Outros estudos também apontam que as informações de nascimento (SINASC) ainda apresentam qualidade superior aos dados de óbito registrados no SIM, o que pode estar relacionado, entre outras razões, ao acesso aos serviços de saúde e também aos próprios objetivos dos sistemas.^{3,28-30}

Os resultados desta pesquisa confirmam o potencial de aplicação desse método para a melhoria da

qualidade das estatísticas vitais, para o aprimoramento dos sistemas de informação em saúde e para a estratégia na vigilância do óbito infantil. Sendo uma ferramenta cuja incorporação na rotina dos serviços

de saúde é absolutamente factível, quer seja pelo baixo custo operacional, quer seja pela facilidade metodológica face aos avanços da tecnologia.

Referências

1. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde. In: Saúde Brasil 2011: uma análise da situação de saúde e a vigilância da saúde da mulher. Monitoramento de indicadores de qualidade dos Sistemas de Informações sobre Mortalidade (SIM) e Nascidos Vivos (Sinasc), nos anos 2000, 2005 e 2010. Brasília, DF; 2012. p. 401-18.
2. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Análise de Situação em Saúde. Saúde Brasil 2010: Uma análise da situação de saúde e de evidências selecionadas de impacto de ações de vigilância em saúde. Brasília, DF; 2011.
3. Almeida MF, Alencar GP, Schoeps D. Sistema de Informações sobre Nascidos Vivos – Sinasc: Uma Avaliação de sua Trajetória. In: A experiência brasileira em sistemas de informação em saúde.. Brasília, DF: Ministério da Saúde, Organização Pan-Americana da Saúde, Fundação Oswaldo Cruz. 2009.
4. Mello Jorge MHP, Laurenti R, Gotlieb SLD. O sistema de Informações sobre Mortalidade – SIM, Concepção, Implantação e Avaliação. In: A experiência brasileira em sistemas de informação em saúde. Brasília, DF: Ministério da Saúde, Organização Pan-Americana da Saúde, Fundação Oswaldo Cruz; 2009.
5. Melo Jorge MHP, Laurenti R, Gotlieb SLD. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC. Ciênc Saúde Coletiva. 2007; 12 (3): 643-54.
6. Andrade CLT, Szwarcwald CL. Desigualdades socioespaciais da adequação das informações de nascimentos e óbitos do Ministério da Saúde, Brasil, 2000-2002. Cad Saúde Pública. 2007; 23: 1207-16.
7. Romero DE, Cunha CB. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema de Informação sobre Nascidos Vivos, 2002. Cad Saúde Pública. 2007; 23: 701-14.
8. Theme Filha 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 Saúde Pública. 2004; 20 (Supl. 11): 83-91.
9. Pedraza, F D. Qualidade do Sistema de Informações sobre Nascidos Vivos (Sinasc): análise crítica da literatura. Ciênc Saúde Coletiva. 2012; 17 (10): 2729-37.
10. 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 Saúde. 2013, 22 (2): 347-52.
11. Costa JMBS, Frias PG. Avaliação da completude das variáveis da Declaração de Nascido Vivo de residentes em Pernambuco, Brasil, 1996 a 2005. Cad Saúde Pública. 2009; 25 (3): 613-24.
12. Camargo Jr KR, Coeli CM. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilistic Record linkage. Cad Saúde Pública. 2000; 16 (2): 439-47.
13. Reclink. Relacionamento probabilístico de registros [dados na Internet]. Brasil. [Acesso em 25 ago. 2014]. Disponível em: <http://www.iesc.ufjf.br/reclink/>
14. Febrl. A Freely Available Record Linkage System with a Graphical User Interface.[homepage on the Internet]. Austrália. [access 25 ago 2014]. Available from: <http://data-mining.anu.edu.au/linkage.html>
15. LinlPlusCenters for Disease Control and Prevention. National Program of Cancer Registries. Link Plus [homepage on the Internet]. Atlanta: CDC [access 2008 jul 14]. Available from: <http://www.cdc.gov/cancer/npccr/tools/registryplus/lp.htm>
16. Queiroz OV, Guerra Junior AA, Machado CJ, Andrade ELG, Meira Júnior W, Acúrcio FA, Santos Filho W, Cherchiglia ML. A construção da Base Nacional de Dados em Terapia Renal Substitutiva (TRS) centrada no indivíduo: relacionamento dos registros de óbitos pelo subsistema de Autorização de Procedimentos de Alta Complexidade (Apac/SIA/SUS) e pelo Sistema de Informações sobre Mortalidade (SIM) - Brasil, 2000-2004. Epidemiol Serv Saúde. 2009; 18 (2): 107-20.
17. Silva CF, Leite AJM, Almeida NMGS, Gondim RC. Fatores de risco para mortalidade infantil em município do Nordeste do Brasil: linkage entre bancos de dados de nascidos vivos e óbitos infantis - 2000 a 2002. Rev Bras Epidemiol. 2006; 9 (1): 69-80.
18. Maia LTS, Souza WV; Mendes ACG. Diferenciais nos fatores de risco para a mortalidade infantil em cinco cidades brasileiras: um estudo de caso-controle com base no SIM e no SINASC. Cad Saúde Pública. 2012; 28 (11): 2163-76.
19. IBGE (Instituto Brasileiro de Geografia e Estatística). Estimativas de População 2005. [acesso em 5 nov 2008]. Disponível em: < <http://www.ibge.gov.br/home/>>
20. Fundação Oswaldo Cruz. Sistema de Monitoramento dos Indicadores de Mortalidade Infantil (MonitorIMI). [acesso em 10 nov 2008]. Disponível em: <<http://www.monitorimi.cict.fiocruz.br/fontes.htm#>>
21. Machado C J. A literature review of record linkage procedures focusing on infant health outcomes. Cad Saúde Pública. 2004; 20 (2): 362-71.
22. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Secretaria de Atenção à Saúde. Manual de vigilância do óbito infantil e fetal e do Comitê de Prevenção do Óbito Infantil e Fetal. Ministério da Saúde, Secretaria de Vigilância em Saúde, Secretaria de Atenção à Saúde. 2 ed. Brasília, DF; 2009. p. 96.
23. Martins EF, Velásquez-Meléndez G. Determinantes da mortalidade neonatal a partir de uma coorte de nascidos

- vivos, Montes Claros, Minas Gerais, 1997-1999. *Rev Bras Saúde Matern Infant.* 2004; 4 (4): 405-12.
24. Nascimento EMR, Costa MCN, Mota ELA, Paim JS. Estudo de fatores de risco para óbitos de menores de um ano mediante compartilhamento de bancos de dados. *Cad Saúde Pública.* 2008; 24 (11): 2593-602.
25. Moraes Neto OL, Barros MBA, Martelli CMT, Silva AS, Cavenaghi SM, Siqueira Jr JB. Diferenças no padrão de ocorrência da mortalidade neonatal e pós-neonatal no Município de Goiânia, Brasil, 1992-1996: análise espacial para identificação das áreas de risco. *Cad Saúde Pública.* 2001; 17 (5): 1241-50.
26. Moraes Neto OL, Barros MBA. Fatores de risco para mortalidade neonatal e pós-neonatal na Região Centro-Oeste do Brasil: linkage entre bancos de dados de nascidos vivos e óbitos infantis. *Cad Saúde Pública.* 2000; 16 (2): 477-85.
27. SEADE. Fundação Sistema Estadual de Análise de Dado. [Internet]. Brasil. [acesso em 10 set 2014]. Disponível em: <http://www.seade.gov.br/>
28. 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 Saúde Matern Infant.* 2012; 12 (3): 243-9.
29. Soares JAS, Horta FMB, Caldeira AP. Avaliação da qualidade das informações em declarações de óbitos infantis. *Rev Bras Saúde Matern Infant.* 2007; 7: 289-95.
30. Frias PG, Pereira PMH, Andrade CLT, Lira PIC, Szwarcwald CL. Avaliação da adequação das informações de mortalidade e nascidos vivos no Estado de Pernambuco, Brasil. *Cad Saúde Pública.* 2010; 26: 671-81.

Recebido em 6 de junho de 2014

Versão final apresentada em 12 de dezembro de 2014

Aprovado em 6 de janeiro de 2015