


Análise da aplicação de uma rotina determinística para a identificação de gestações múltiplas no Sistema de Informações sobre Nascidos Vivos*

doi: 10.5123/S1679-49742020000200001


Analysis of the application of a deterministic routine for identifying multiple pregnancies on SINASC, Brazil


Análisis de la aplicación de una rutina determinista para la identificación de gestaciones múltiples en el Sinasc, Brasil


Fernanda Pinheiro Aguiar¹ -  orcid.org/0000-0003-0197-1354

Patrícia Viana Guimarães Flores² -  orcid.org/0000-0001-5074-5113

Luís Carlos Torres Guillen¹ -  orcid.org/0000-0001-5246-733X

Helena Pereira da Silva Santos¹ -  orcid.org/0000-0002-5712-590X

Luís Guilherme Santos Buteri Alves¹ -  orcid.org/0000-0002-5810-8474

Kenneth Rochel de Camargo Jr.³ -  orcid.org/0000-0003-3606-5853

Rejane Sobrino Pinheiro¹ -  orcid.org/0000-0002-3361-3626

Cláudia Medina Coeli¹ - orcid.org/0000-0003-1757-3940

¹Universidade Federal do Rio de Janeiro, Programa de Pós-Graduação em Saúde Coletiva, Rio de Janeiro, RJ, Brasil

²Ministério da Saúde, Hospital Federal de Bonsucesso, Rio de Janeiro, RJ, Brasil

³Universidade do Estado do Rio de Janeiro, Instituto de Medicina Social, Rio de Janeiro, RJ, Brasil

Resumo

Objetivo: analisar a aplicação de rotina determinística na identificação de gemelares no Sistema de Informações sobre Nascidos Vivos (Sinasc) do Brasil. **Métodos:** de duplicação do Sinasc e *linkage* com a base de mortalidade (óbitos fetais) do estado do Rio de Janeiro, 2007-2008; empregou-se rotina determinística mediante chave baseada em informações maternas e do nascimento (Sinasc), complementada por revisão manual. **Resultados:** dos 433.874 registros do Sinasc, 9.036 (2,1%) estavam classificados como nascidos vivos de gravidez múltipla; após implementação da rotina proposta, foram reclassificados como gemelares 385 registros e 286 como não gemelares; a acurácia da informação de gestação múltipla no Sinasc foi boa (sensibilidade=95,8%; especificidade=99,9%); a aplicação da rotina, sem a etapa de revisão manual, aumentou a sensibilidade em 4,2%, sem mudança importante da especificidade. **Conclusão:** a despeito de a acurácia da informação sobre gestação múltipla no Sinasc ser boa, sugere-se o emprego dessa rotina como opção para melhorar a classificação dos gemelares.

Palavras-chave: Sistemas de Informação em Saúde; Integração de Sistemas; Avaliação em Saúde; Gravidez Múltipla.

*Artigo originado da tese de Fernanda Pinheiro Aguiar, intitulada 'Idade e escolaridade maternas ao nascimento e mortalidade na infância', defendida junto ao Programa de Pós-Graduação em Saúde Coletiva/Instituto de Estudos em Saúde Coletiva da Universidade Federal do Rio de Janeiro (IESC/UFRJ), em 22 de maio de 2018. Os autores recebem bolsa de pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico/Ministério da Ciência, Tecnologia, Inovações e Comunicações (CNPq/MCTIC) (Cláudia Medina Coeli, Processos nº 447199/2014-5 e nº 305545/2015-9; Rejane Sobrino Pinheiro, Processo nº 309728/2012-6; e Kenneth Rochel de Camargo Jr., Processo nº 300686/2013-7) e da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (Cláudia Medina Coeli, Processo nº E-26/203.195/2015; e Kenneth Rochel de Camargo Jr., Processo nº E-26/102.900/2012), bolsa de Iniciação Científica da FAPERJ (Helena Pereira da Silva Santos, Processo nº E-26/202.575/2016; e Luís Guilherme Santos Buteri Alves, Processo nº E-26/219.275/2016) e bolsa de Doutorado Nota DEZ da FAPERJ (Fernanda Pinheiro Aguiar, Processo nº E-26/400.418/2016).

Endereço para correspondência:

Fernanda Pinheiro Aguiar – Rua Pajuçara, nº 600, apto. 302, Cocotá, Ilha do Governador, Rio de Janeiro, RJ, Brasil.

CEP: 21910-300

E-mail: aguiarfernandap@gmail.com



Introdução

O Sistema de Informações sobre Nascidos Vivos (Sinasc) foi implantado com o propósito de reunir informações referentes aos nascimentos ocorridos em todo o território nacional. Desde 1990, o Sinasc tem-se mostrado relevante na caracterização e situação dos partos e dos nascimentos, bem como na identificação de grupos de risco/vulnerabilidade entre mães e crianças.^{1,2}

A gravidez múltipla é fator de risco para desfechos negativos ao nascimento.³⁻⁶ O risco de morte de bebês múltiplos chega a ser 12 vezes maior, comparado ao mesmo risco para aqueles nascidos de gestação única. A principal explicação para tal diferença está no aumento da proporção de prematuridade e do crescimento intrauterino restrito em gemelares.^{7,8}

Crianças de gestação múltipla também apresen-

O Sinasc possui uma variável indicativa da quantidade de filhos na mesma gestação; porém, erros de preenchimento levam a um equívoco na classificação da informação sobre gemelaridadeil.

tam maior risco de desenvolvimento de condições indesejáveis no longo prazo, como paralisia cerebral, comprometimento cognitivo, atraso da linguagem, dificuldades de aprendizagem, problemas psiquiátricos e sociocomportamentais.^{7,8} Assim, é importante identificar nascimentos múltiplos em estudos que empregam bases de dados de estatísticas vitais.

Cada gemelar é incluído na base de nascimentos em registro separado, com número de identificação próprio. Adicionalmente, o Sinasc possui uma variável indicativa da quantidade de filhos na mesma gestação; porém, erros de preenchimento levam a um equívoco na classificação da informação sobre gemelaridade.^{1,9,10} O *linkage* de bancos de dados vem sendo utilizado para melhoria da qualidade da informação, mediante sua recuperação e confirmação em uma mesma base (processo de identificação de duplicidades) ou entre bases diferentes.¹¹⁻¹⁴

O objetivo deste trabalho foi avaliar a aplicação de uma rotina determinística visando identificar gestações múltiplas na base do Sinasc do estado do Rio de Janeiro, nos anos de 2007 e 2008.

Métodos

Foi realizado estudo descritivo para avaliação da melhora da informação sobre gestações múltiplas na base do Sinasc, com a aplicação de rotina determinística (*linkage* interno).

Foram utilizados dados do Sinasc (N=433.882) do estado do Rio de Janeiro referentes aos anos de 2007 e 2008. Os registros de óbitos fetais (N=372) da base de dados do Sistema de Informações sobre Mortalidade (SIM) também foram consultados, na busca manual de gemelares, quando havia indicação de gestação múltipla no Sinasc mas apenas um registro de nascido vivo.

A rotina determinística baseou-se em quatro processos: (i) comparação dos registros (*linkage* interno à base de dados), empregando-se chave determinística composta pelas informações maternas (*soundex* do primeiro nome da mãe, *soundex* do segundo nome da mãe; *soundex* do último nome da mãe) e do nascimento (data de nascimento completa; código do estabelecimento de nascimento); (ii) comparação automática do endereço de residência, empregando-se uma rotina baseada na distância de edição de Levenshtein; (iii) busca manual de gemelares no SIM; e (iv) revisão manual.

Inicialmente, foi realizado o pré-processamento da base do Sinasc, visando a eliminação de registros com número da Declaração de Nascido Vivo duplicado.

Registros que apresentaram a mesma chave determinística foram avaliados segundo a informação sobre gestação (única; múltipla) no Sinasc. Os registros com classificação concordante, ou seja, que coincidiram no valor da chave e apresentavam indicação de gravidez múltipla (chave+/Sinasc+), e aqueles com classificação discordante, ou seja, que coincidiram no valor da chave mas apresentavam indicação de gravidez única (chave+/Sinasc-), tiveram os endereços comparados automaticamente. Quando os endereços coincidiam por completo, os registros foram classificados como gravidez múltipla. Quando os endereços eram discordantes, foi realizada etapa de revisão manual para classificação final. Nesta etapa, informações sobre nome completo da mãe, idade materna, local de nascimento, tipo de parto e de gravidez foram utilizadas pelo pesquisador para definir sua classificação como gestação múltipla ou não.

Quando a chave não identificou gemelaridade e a informação constante no Sinasc era de gravidez única (chave-/Sinasc-), os registros foram classificados como não gemelares. Para os registros com indicação de gra-

videz múltipla, em que a chave não identificou o registro como gemelar (chave-/Sinasc+), foi realizada busca manual na base de óbitos fetais do SIM para confirmação da situação de gemelar, uma vez que bebês da mesma gestação poderiam ser encontrados em sistemas de informações diferentes. Os que não foram encontrados na base de óbitos fetais foram submetidos à revisão manual.

Foram avaliados os registros que tiveram mudança de classificação (gestação múltipla ou única) após a aplicação da rotina completa (chave determinística, comparação de endereço, busca na base de óbitos fetais e revisão manual de pares). A classificação após a aplicação da rotina completa foi considerada padrão ouro para as análises de acurácia, tanto das informações sobre gemelares registradas no Sinasc como da classificação obtida por meio da aplicação de uma rotina reduzida, baseada tão somente nas informações contidas no Sinasc e na concordância da chave determinística, sem realização dos demais procedimentos (comparação de endereço, busca na base de óbitos fetais e revisão manual de pares). Neste caso, foram classificados como gestação múltipla os registros do Sinasc que apresentaram indicação de gestação múltipla ou que apresentaram registros concordantes na chave determinística. Os registros que não apresentaram informação de gestação múltipla no Sinasc e, simultaneamente, sem pares identificados pela chave determinística, foram considerados não gemelares. Calcularam-se a sensibilidade, especificidade e valor preditivo positivo, e respectivos intervalos de confiança de 95% (IC_{95%}).

Os programas PostgreSQL 9.2 e Stata 12 foram empregados, respectivamente, para a implementação da rotina de *linkage* determinístico e para as análises.

O projeto do estudo, fundamentado em dados secundários disponibilizados pela Secretaria de Estado de Saúde e Defesa Civil do Rio de Janeiro e desenvolvido de acordo com os princípios da ética na pesquisa, foi submetido ao Comitê de Ética em Pesquisa do IESC/UFRJ como emenda ao projeto 'Registro Integrado de Saúde: avaliando longitudinalmente a morbimortalidade de uma coorte de nascidos vivos e de suas mães - Fase 1' e aprovado em 3 de outubro de 2012 – Certificado de Apresentação para Apreciação Ética (CAAE) nº 07534512.9.0000.5286.

Resultados

Dos 433.882 registros de nascidos vivos do estado do Rio de Janeiro nos anos de 2007 e 2008, oito foram excluídos por duplicidades e 9.036 (2,1%) se encontravam

classificados como gravidez múltipla no Sinasc; destes últimos, 8.136 apresentaram concordância na chave determinística (chave+/Sinasc+). Após implementação da rotina e conferência automática dos endereços, 6.508 registros, que apresentaram mesmo endereço, foram classificados automaticamente como gemelares; e 1.628, que apresentaram endereços diferentes, classificados como gemelares após revisão manual (Figura 1).

Todos os 385 registros que apresentaram chave+/Sinasc- foram classificados como gemelares: 260 com mesmo endereço foram classificados automaticamente, e 125 após revisão manual (Figura 1).

Foram identificados 900 registros nos quais a rotina não apontou gemelaridade e a informação no Sinasc era de gravidez múltipla (chave-/Sinasc+). Destes, 78 registros encontrados após busca na base de óbitos fetais. Para os demais 738, a revisão manual identificou 452 gemelares e 370 não gemelares.

Houve 424.537 registros na categoria chave-/Sinasc(-); 9.051 foram classificados como gemelares e 424.823 como não gemelares, com mudança do *status* inicial em 671.

A acurácia da informação de gestação múltipla no Sinasc, quando comparada à classificação derivada da aplicação da rotina completa, foi de sensibilidade=95,8% (IC_{95%} 95,3;96,2%), especificidade=99,9% (IC_{95%} 99,9;99,9%) e valor preditivo positivo=95,9% (IC_{95%} 95,5;96,3%) (Tabela 1).

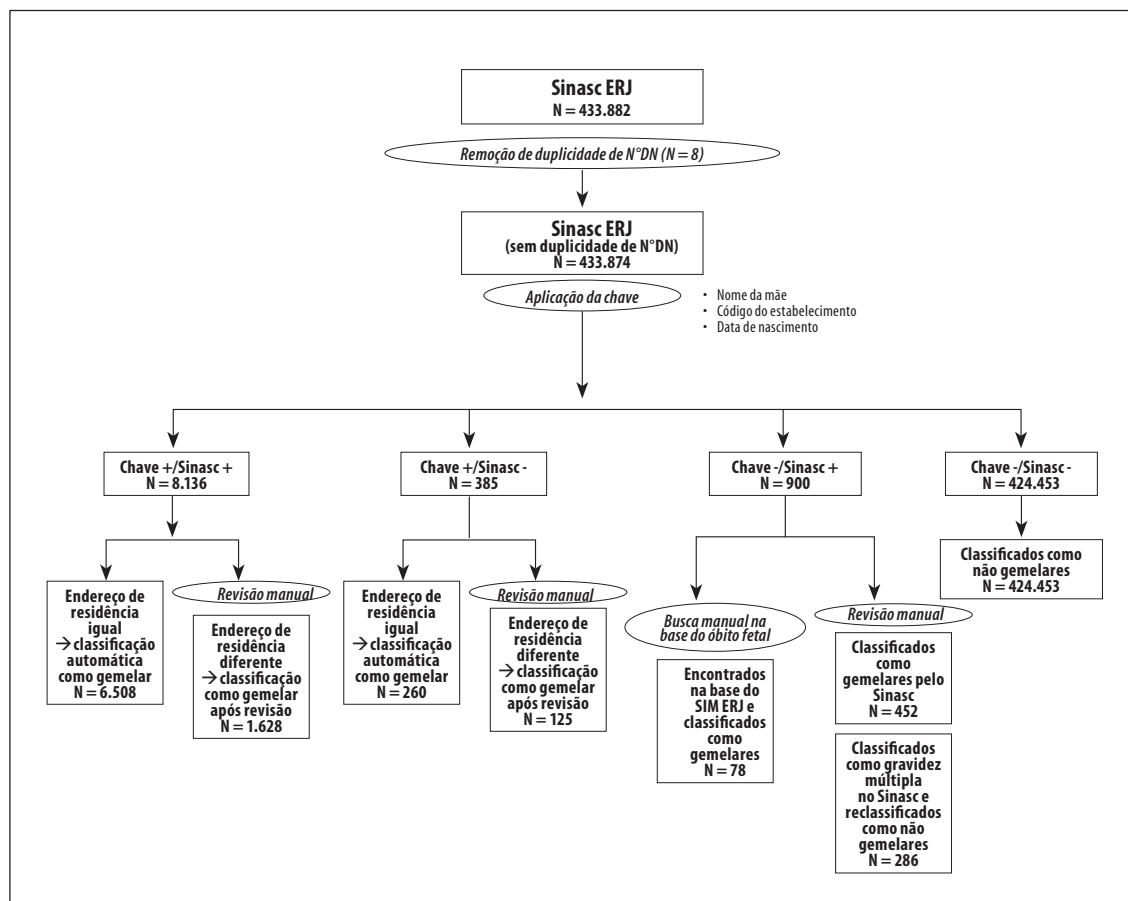
Já na aplicação da rotina sem revisão manual, a acurácia foi de sensibilidade=100,0%, especificidade=99,9% (IC_{95%} 99,9;99,9%) e valor preditivo positivo=96,1% (IC_{95%} 95,7;96,4%) (Tabela 2).

Discussão

O estudo utilizou-se de rotina determinística, a qual permitiu a melhor classificação da informação sobre gestações múltiplas no Sinasc, evitando tanto erros falso-positivos como falso-negativos. A classificação equivocada de uma gestação múltipla, como duplicidade em processos de relacionamento de dados, representa um desafio no desenvolvimento de algoritmos para registros eletrônicos de saúde.¹⁵⁻¹⁷

A cobertura e a qualidade dos dados do Sinasc são fundamentais para sua confiabilidade como fonte de informações substanciais na avaliação e pesquisa em saúde.^{18,19}

Ainda que se tenha observado boa qualidade da informação sobre tipo de gravidez no Sinasc do es-



ERJ: Estado do Rio de Janeiro.

Figura 1 – Fluxograma da aplicação da rotina determinística para a identificação de gestações múltiplas no banco de dados do Sistema de Informações sobre Nascidos Vivos do estado do Rio de Janeiro, 2007-2008

Tabela 1 – Acurácia da informação sobre gemelares registrada no banco de dados do Sistema de Informações sobre Nascidos Vivos do estado do Rio de Janeiro, 2007-2008

Sinasc ^a	Rotina determinística (padrão ouro)		Total
	Gemelar	Não gemelar	
Gemelar	8.666	370	9.036
Não gemelar	385	424.453	424.838
Total	9.051	424.823	433.874

Sensibilidade = 95,8% (IC_{95%} 95,3;96,2%)
 Especificidade = 99,9% (IC_{95%} 99,9;99,9%)
 Valor preditivo positivo = 95,9% (IC_{95%} 95,5;96,3%)

a) Sinasc: Sistema de Informações sobre Nascidos Vivos.

Tabela 2 – Acurácia da informação sobre gemelares após aplicação de uma rotina determinística automática, sem revisão manual, no banco de dados do Sistema de Informações sobre Nascidos Vivos do estado do Rio de Janeiro, 2007-2008

Rotina sem revisão manual no Sinasc ^a	Rotina determinística (padrão-ouro)		Total
	Gemelar	Não gemelar	
Gemelar	9.051	370	9.337
Não gemelar	–	424.537	424.537
Total	9.051	424.823	433.874

Sensibilidade = 100,0%

Especificidade = 99,9% (IC_{95%} 99,9;99,9%)

Valor preditivo positivo = 95,9% (IC_{95%} 95,7;96,4%)

a)Sinasc: Sistema de Informações sobre Nascidos Vivos.

tado do Rio de Janeiro, resultado concordante com a literatura,^{1,20,21} a aplicação da rotina é útil e de fácil execução. Porém, dadas as características particulares dos gemelares, estudos sobre desfechos neonatais, em geral, excluem os registros desse grupo, que devem ser analisados separadamente.^{22,23} A baixa frequência de gemelares em relação ao total de nascimentos faz com que mudanças no número de casos, mesmo quando pequenas em termos absolutos, tenham importância relativa. Ademais, a classificação de Robson passou a ser divulgada no Sinasc, sendo a informação acurada sobre gemelaridade necessária para a adequada categorização das mulheres.²⁴

Técnicas de *linkage* de bases de dados, determinísticas ou probabilísticas, vêm sendo utilizadas para melhorar a qualidade da informação.^{12,25} Rotinas determinísticas apresentam ótimo desempenho quando a qualidade dos dados é boa.^{26,27} seu processamento é rápido e podem ser empregadas sem revisão manual dos *links* formados.

A rotina desenvolvida nesta pesquisa incluiu etapa de revisão manual, viável apenas para bases de pequeno ou médio volume em situações de discordância das informações. Nas situações de bases de maior volume de registros, a aplicação tão somente da chave, sem a realização da etapa de revisão manual, aumenta a sensibilidade para a identificação de gemelares sem alteração importante da especificidade e do valor pre-

ditivo positivo. Uma alternativa intermediária seria o processamento manual apenas dos registros não identificados pela chave, embora, para estes registros, exista informação de gravidez múltipla no Sinasc.

Como limitação do presente estudo, cumpre destacar a não revisão manual de todos os casos. Entretanto, é muito baixa a probabilidade de equívoco na classificação do tipo de gravidez quando há concordância entre a chave e o Sinasc.

Ainda que o incremento com a recuperação de gemelares pareça pequeno, o custo de sua realização é baixo diante da possibilidade de melhora da informação. Sugere-se que a rotina proposta passe a ser empregada usualmente, em especial nos estudos sobre desfechos neonatais em gemelares.

Contribuição das autoras

Aguiar FP e Coeli CM foram responsáveis pela concepção, estruturação e análise dos dados. Aguiar FP, Coeli CM, Flores PVG, Guillen LCT, Santos HPS, Alves LGSB, Camargo Jr KR e Pinheiro RS contribuíram com a análise e interpretação dos dados, redação das versões preliminares e revisão crítica do manuscrito. Todos os autores aprovaram a versão final do manuscrito e declaram-se responsáveis por todos os aspectos do trabalho, garantindo sua precisão e integridade.

Referências

1. 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* [Internet]. 2004 [citado 2019 out 21];20(Supl.1):S83-91. Disponível em: <http://www.scielo.br/pdf/csp/v20s1/09.pdf> . doi:10.1590/S0102-311X2004000700009
2. 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* [Internet]. 2009 mar [citado 2019 out 21];25(3):613-24. Disponível em: <http://www.scielo.br/pdf/csp/v25n3/16.pdf> . doi: 10.1590/S0102-311X2009000300016
3. Morais Neto OL, Barros MBA. Risk factors for neonatal and post neonatal mortality in the Central-West region of Brazil: linked use of life-birth and infant death records. *Cad Saúde Pública* [Internet]. 2000 abr-jun [citado 2019 out 21];16(2):477-85. Disponível em: <http://www.scielo.br/pdf/csp/v16n2/2097.pdf> . doi: 10.1590/S0102-311X2000000200018
4. Silva CF, Leite AJM, Almeida NMGS, Gondim RC. Fatores de risco para a 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* [Internet]. 2006 mar [citado 2019 out 21];9(1):69-80. Disponível em <http://www.scielo.br/pdf/rbepid/v9n1/04.pdf> . doi: 10.1590/S1415-790X2006000100009
5. Ramos HÂDC, Cuman RKN. Risk factors for prematurity: document search. *Escola Anna Nery* [Internet]. 2009 abr-jun [citado 2019 out 21];13(2):297-304. Disponível em: <http://www.scielo.br/pdf/ean/v13n2/v13n2a09.pdf> . doi: 10.1590/S1414-81452009000200009
6. Silva VFG. Complicações na gestação de gemelar. Fertilização in vitro versus espontânea. Instituto de Ciências Biomédicas Abel Salazar [Internet]. Porto Universidade do Porto; 2013. Disponível em: <https://repositorio-aberto.up.pt/handle/10216/72427>
7. Shinwell ES, Haklai T, Eventov-Friedman S. Outcomes of mMultiplets. *Neonatology* [Internet]. 2009 [cited 2019 Oct 21];95(1):6-14. Available from: <https://www.karger.com/Article/Abstract/151750> . doi: 10.1159/000151750
8. Cooke RWI. Does neonatal and infant neurodevelopmental morbidity of multiples and singletons differ? *Seminars in Fetal & Neonatal Medicine* [Internet]. 2010 Dec [cited 2019 Oct 21];15(6):362-6. Available from: [https://www.sfnjournal.com/article/S1744-165X\(10\)00047-8/fulltext](https://www.sfnjournal.com/article/S1744-165X(10)00047-8/fulltext) . doi:10.1016/j.siny.2010.06.003
9. Barbuscia DM, Rodrigues-Júnior AL. Completeness of data on live birth certificates and death certificates for early neonatal and fetal deaths in the Ribeirão Preto Region, São Paulo State, Brazil, 2000-2007. *Cad Saúde Pública* [Internet]. 2011 Jun [cited 2019 Oct 21]; 27(6):1192-200. Available from: <http://www.scielo.br/pdf/csp/v27n6/16.pdf> . doi: 10.1590/S0102-311X2011000600016
10. Oliveira MM, Andrade SSCA, Dimech GS, Oliveira JCG, Malta DC, Rabelo Neto, DL, et al. Avaliação do Sistema de Informações sobre nascidos vivos. Brasil, 2006 a 2010. *Epidemiol Serv Saúde* [Internet]. 2015 out-dez [citado 2019 out 21];244(4):629-40. Disponível em: <http://www.scielo.br/pdf/ress/v24n4/2237-9622-ress-24-04-00629.pdf> . doi: 10.5123/S1679-49742015000400005
11. Méray N, Reitsma JB, Ravelli AC, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *Journal of Clinical Epidemiol* [Internet] ogy. 2007 Sep [cited 2019 Oct 21];60(9):883-91. Available from: [https://www.jclinepi.com/article/S0895-4356\(06\)00500-2/fulltext](https://www.jclinepi.com/article/S0895-4356(06)00500-2/fulltext) .doi: 10.1016/j.jclinepi.2006.11.021
12. 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. *Ciênc Saúde Coletiva* [Internet].

- 2014 jul [citado 2019 out 21];19(7):2011-20. Disponível em: <http://www.scielo.br/pdf/csc/v19n7/1413-8123-csc-19-07-02011.pdf> . doi: 10.1590/1413-81232014197.0892201313.
13. Bartholomay P, Oliveira G.P, Pinheiro RS, Vasconcelos AMN. Melhoria da qualidade das informações sobre tuberculose a partir do relacionamento entre bases de dados. *Cad Saúde Pública* [Internet]. 2014 nov [citado 2019 out 21];30(11):2459-70. Disponível em: <http://www.scielo.br/pdf/csp/v30n11/0102-311X-csp-30-11-2459.pdf> doi: 10.1590/0102-311X00116313
 14. Souza MaiaMaia LTS, Souza WV, MendesCruz ACG.M. A contribuição do linkage entre SIM e SINASC para a melhoria das informações da mortalidade infantil em cinco cidades brasileiras. *Revista Brasileira de Saúde Mater rno-Infant* [Internet] il. 2015 mar [citado 2019 out 21];15(1):57-66. Disponível em: <http://www.scielo.br/pdf/rbsmi/v15n1/1519-3829-rbsmi-15-01-0057.pdf> . doi: 10.1590/S1519-38292015000100005
 15. Baldwin E, Johnson K, Berthoud H, Dublin S. Linking mothers and infants within electronic health records: a comparison of deterministic and probabilistic algorithms. *Pharmaco epidemiology and Drug Saf* [Internet]ety. 2015 Jan [cited 2019 Oct 21];24(1):45-51. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pds.3728> . doi: 10.1002/pds.3728
 16. Harron KL, Doidge JC, Knight HE, Gilbert RE, Goldstein H, Cromwell DA, Meulen JHet al. A guide to evaluating linkage quality for the analysis of linked data. *International Journal of Epidemiol* [Internet]ogy. 2017 Oct [cited 2019 Oct 21];, 46(5):1699-710. Available from: <https://academic.oup.com/ije/article/46/5/1699/4107250> . doi: 10.1093/ije/dyx177
 17. Harper G. Linkage of maternity hospital episode statistics data to birth registrationand notification records for birth sin England 2005-2014: qQuality assurance of linkage of routine data for singleton and multiple births. *BMJ Open* [Internet]. 2018 Mar [cited 2019 Oct 21];8(3):e017898. Available from: <https://bmjopen.bmj.com/content/8/3/e017898>. long . doi: 10.1136/bmjopen n-2017- 017898
 18. Silva RS, Oliveira CM, Ferreira DKS, Bonfim CV. Avaliação da completude das variáveis do Sistema de Informações sobre Nascidos Vivos — SINASC - nos12 Estados da região Nordeste do Brasil, 2000 e 2009. *Epidemiol Serv Saúde*[Internet]. 2013 abr-jun [citado 2019 out 21];22(2):347-52. Disponível em:<http://scielo.iec.gov.br/pdf/ess/v22n2/v22n2a16.pdf> . doi: 10.5123/S1679-49742013000200016
 19. Mello 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* [Internet]. 2007 maio-jun [citado 2019 out 21];12(3):643-54. Disponível em: <http://www.scielo.br/pdf/csc/v12n3/14.pdf> . doi: 10.1590/S1413- 81232007000300014
 20. Gabriel GP, Chiquetto L, Morcillo AM, Ferreira MC, Bazan IG, Daolio LD, et al. Evaluation of data on live birth certificates from the Information System on Live Births (Sinasc) in Campinas, São Paulo, 2009. *Revista Paulista de Pediatr* [Internet] ia. 2014 Sep [cited 2019 Oct 21];32(3):183-8. Available from: http://www.scielo.br/pdf/rpp/v32n3/en_0103-0582-rpp-32-03-0183.pdf . doi: 10.1590/0103-0582201432306
 21. Bonilha EA, Vico ESR, Freitas M, Barbuscia DM, Galleguillos TGB, Okamura MN, et al. Cobertura, completude e confiabilidade das informações do Sistema de Informações sobre Nascidos Vivos de maternidades da rede pública no município de São Paulo, 2011. *Epidemiol Serv Saúde* [Internet]. 2018 [citado 2019 out 21];27(1):e201712811. Disponível em: <http://www.scielo.br/pdf/ress/v27n1/2237-9622-ress-27-01-e201712811.pdf> . doi: 10.5123/s1679-49742018000100011
 22. Gardner MO, Goldenberg RL, Cliver SP, Tucker JM, Nelson KG, Copper RL. The origin and outcome of preterm twin pregnancies. *Obstetrics and Gynecology*. 1995 Apr;85(4):553-7.
 23. Garite TJ, Clark RH, Elliott JP, Thorp JA. Twins and triplets: the effect of plurality and growth on neonatal outcome compared with singleton infants. *American Journal of Obstetrics and Gynecol* [Internet]ogy. 2004 Sep [cited 2019 Oct 21]1;191(3):700-7. Available from: [https://www.ajog.org/article/S0002-9378\(04\)00286-8/fulltext](https://www.ajog.org/article/S0002-9378(04)00286-8/fulltext) . doi: 10.1016/j.

- ajog.2004.03.040Garite TJ, Clark RH, Elliott JP, Thorp JA. Twins and triplets: the effect of plurality and growth on neonatal outcome compared with singleton infants. *American Journal of Obstetrics and Gynecology*. 2001;191(3):700-7.
24. Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. *Saúde Brasil 2017: uma análise da situação de saúde e os desafios para o alcance dos objetivos de desenvolvimento sustentável*. Brasília, 2018.
25. Pedraza DF. Quality of the Information System on Live Births/Sinasc: a critical analysis of published studies. *Ciênc Saúde Coletiva*. 2012;17(10):2729-37.
26. Coeli CM, Pinheiro RS, Camargo Jr KR. Conquistas e desafios para o emprego das técnicas de record linkage na pesquisa e avaliação em saúde no Brasil. *Epidemiol Serv Saúde*. 2015;24:795-802.
27. Oliveira GP, Bierrenbach AL, Camargo Jr. KR, Coeli CM, Pinheiro RS. Acurácia das técnicas de relacionamento probabilístico e determinístico: o caso da tuberculose. *Rev Saúde Pública*. 2016;50(49)

Abstract

Objective: to evaluate the application of a deterministic routine for identifying multiple pregnancies on the Brazilian Live Birth Information System (SINASC).

Methods: SINASC data deduplication and linkage with the mortality database (fetal deaths) for Rio de Janeiro state for the period 2007-2008; we used a deterministic routine, using a key based on SINASC maternal and birth information, complemented by manual review. **Results:** of the 433,874 SINASC records, 9,036 (2.1%) were classified as multiple pregnancy newborns; after implementing the routine, we reclassified 385 records as twins, and 286 as singletons; accuracy of multiple pregnancy information on the SINASC database was high (sensitivity=95.8%; specificity=99.9%); applying the routine without the manual review process increased sensitivity by 4.2%, with no significant change of specificity. **Conclusion:** despite the accuracy of information regarding multiple pregnancy held on SINASC, we suggest the use of this routine as an option for improving classification of twins.

Keywords: Health Information Systems; Systems Integration; Health Evaluation; Pregnancy, Multiple.

Resumen

Objetivo: evaluación de la aplicación de rutina determinista en la identificación de gestación gemelar en el Sistema de Información sobre Nacidos Vivos (Sinasc) de Brasil. **Métodos:** de duplicación del Sinasc y conexión con la base de mortalidad (muertes fetales), en el estado del Rio de Janeiro, 2007-2008; se empleó una rutina determinista por medio de clave con informaciones maternas y del nacimiento (Sinasc), complementada por revisión manual. **Resultados:** de 433.874 registros del Sinasc, 9.036 (2,1%) estaban clasificados como recién nacidos de gestación múltiple; después de la implementación de la rutina sugerida, 385 registros fueron reclasificados como gemelares y 286 como no gemelares; la exactitud de la información de gestación múltiple en el Sinasc fue buena (sensibilidad=95,8%; especificidad=99,9%); la aplicación de la rutina sin la etapa de revisión manual, aumentó la sensibilidad en 4,2%, sin cambio importante de la especificidad. **Conclusión:** a pesar de la exactitud de la información sobre gestación múltiple en el Sinasc haber sido buena, sugerimos el uso de esta rutina como una opción para mejorar la clasificación de los gemelares.

Palabras clave: Sistemas de Información en Salud; Integración de Sistemas; Evaluación en Salud; Embarazo Múltiple.

Recebido em 20/12/2018

Aprovado em 24/09/2019

Editora associada: Doroteia Aparecida Höfelmann -  orcid.org/0000-0003-1046-3319