Research note

Assessing the completeness and agreement of variables of the Information Systems on Live Births and on Mortality in Recife-PE, Brazil, 2010-2012

doi: 10.5123/S1679-49742016000400019

Lays Janaina Prazeres Marques[1]
Conceição Maria de Oliveira[2]
Cristine Vieira do Bonfim[3]

[1]Universidade Federal de Pernambuco, Programa de Pós-Graduação Integrado em Saúde Coletiva, Recife-PE, Brasil
[2]Centro Universitário Maurício de Nassau, Departamento de Saúde, Recife-PE, Brasil
[3]Fundação Joaquim Nabuco, Diretoria de Pesquisas Sociais, Recife-PE, Brasil

Abstract

Objective: to assess the information completeness and agreement on infant deaths. **Methods**: this was an evaluation study with descriptive design using data of the Information System on Live Births (Sinasc) and Mortality Information System (SIM) of residents in Recife-PE, Brazil, in 2010-2012; the deterministic records linkage was used to combine the data on infant deaths and live births. **Results**: of the 837 infant deaths registered on SIM, 811 (96.9%) were linked; the completeness obtained was above 95% on SIM and 98% on Sinasc; the agreement varied from 0.762 (substantial) to 0.997 (excellent) for the intraclass correlation coefficient, and it was excellent for Kappa index (>0.80). **Conclusion**: Sinasc and SIM presented excellent completeness and agreement for most of the variables analyzed. The relationship between the databases is a tool that can be used by the health services of the municipalities to improve the vital statistics information systems.

Key words: Infant Mortality; Vital Statistics; Information Systems; Epidemiology, Descriptive.

*Research funded by the National Council for Scientific and Technological Development (CNPq)/Ministry of Science, Technology and Innovation (MCTI): Process No. 144065/2013-4

Correspondence:
Cristine Vieira do Bonfim – Fundação Joaquim Nabuco, Diretoria de Pesquisas Sociais, Rua Dois Irmãos, No. 92, Ed. Renato Carneiro Campos, Apipucos, Recife-PE, Brasil. CEP: 52071-440
E-mail: cristine.bonfim@uol.com.br

Introduction

The infant mortality rate is a health indicator that reflects the life conditions of the population and the quality of maternal and infant health care.[1,2] Although being widely used, most of low and middle income countries do not have enough information to calculate mortality rates among infants under one year old.[3]

The Information Systems on Live Births (*Sinasc*) and on Mortality (*SIM*) were released by the Brazilian Ministry of Health in 1976 and 1990, respectively, and are important data sources to the monitoring of infant mortality. The Certificate of Live Birth (CLB) and the Death Certificate (DC) are the documents that feed those information systems, and provide data to calculate indicators on health, epidemiology and demography.[1-3]

> *The access to reliable data allows a more valid analysis of the situation of births, deaths and their determinants.*

Database linkage is an strategy which aims to improve the quality of information: the use of integrated databases favors the recovery of incomplete or incoherent records.[4-6] The linkage usage depends on factors such as data coverage and quality, so the identification of the same individual in different databases can be linked to a unique record, enabling the information gain to the systems compared.[7,8]

The use of linkage techniques in researches related to the characteristics of deaths and births in Brazil improves the completeness and reliability of the information provided by *Sinasc* and *SIM*.[9,10] Some studies used linkage to identify risk factors associated to infant and neonatal mortality,[8,11,12] to verify the information quality on live births and infant death,[5,7] and to assess the infant mortality rate.[13]

The access to reliable data allows a more valid analysis of the situation of births, deaths and their determinants. The availability of quality information favors the analysis of health situation and the planning of actions to reduce infant mortality. This study aimed to assess the information completeness and agreement on infant deaths.

Methods

This is an evaluation study with descriptive design, using data of all infant deaths and live births of residents in Recife-PE, occurred from 2009 (live births) to 2012, recorded on *Sinasc* and *SIM* databases.

To combine the data on infant deaths and live births, we used the deterministic record linkage, with the program Epi Info version 6.04d. The linkage is performed from a variable which is common to the different data sources to unify the records in one database, completing the blank spaces and correcting the incorrect data.[5,14]

The following identification fields were adopted to combine *Sinasc* and *SIM*: number of the CLB, mother's name and date of birth. To avoid misclassification of false positives and/or false negatives, the pairs formed were manually verified, using the variables 'address', 'sex' and 'birth weight'.[6]

For each variable analyzed on *Sinasc* and *SIM*, a pre-linkage and post-linkage filling was conducted, according to the scores proposed by Romero and Cunha,[15] who consider the incompleteness as the proportion of ignored/blank spaces, adopting the following criteria: excellent (<5%); good (5 to 9.9%); regular (10 to 19.9%); poor (20 to 49.9%) and very bad (≥50%).

Pearson chi-square test (χ^2) was used to verify the existence of significant differences between the completeness proportions of the variables in common between *SIM* and *Sinasc* and the database that resulted from the linkage. The agreements of qualitative and discrete quantitative variables were analyzed using Kappa index and the intraclass correlation coefficient (ICC), respectively. The parameters used as reference points to classify the Kappa index and the ICC were: excellent agreement (0.80 to 1.00), substantial (0.60 to 0.79), moderate (0.40 to 0.59), reasonable (0.20 to 0.39), weak (0 to 0.19) and no agreement (<0).[16] The significance level adopted was 5%. The analyses were performed with the program R for Windows® version 3.2.2.

The study project was approved by the Research Ethics Committee of the Joaquim Nabuco Foundation (CAAE: 27491014.6.0000.5619) in March 10th 2014, and by the Municipal Health Department of Recife.

Results

From January 1st 2009 to 31st December 2012, 88,988 live births were registered on *Sinasc*; 837 infant death were registered on *SIM* from 1st January

2010 to 31 December 2012. It was possible to link 811 (96.9%) Death Certificates to their respective Certificates of Live Birth. Out of the 26 (3.1%) non-linked records, 15 (1.8%) presented problems in one of the identification variables – mother's name –, with different spellings between the databases; seven variables (0.8%) did not present the CLB number on *SIM* and four (0.4%) presented differences between both databases regarding this field (Figure 1).

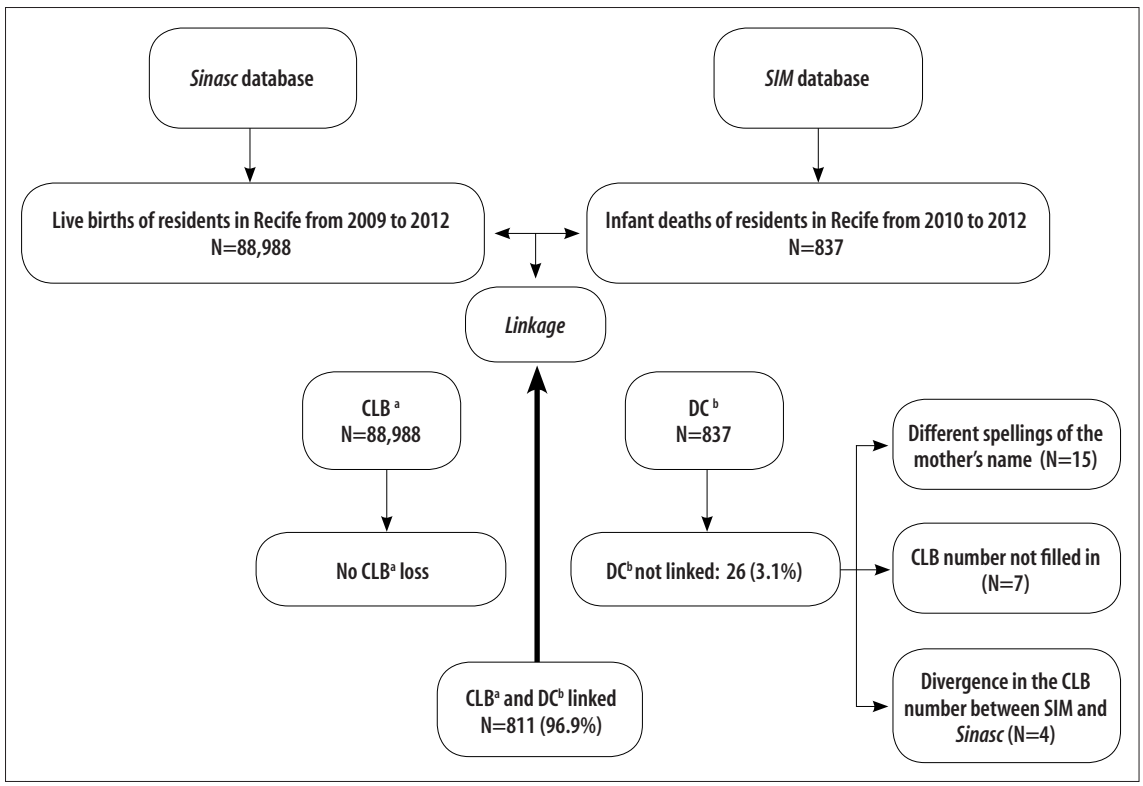
In the pre-linkage phase, a low percentage of incompleteness on *Sinasc* was observed; it was lower than the one found on *SIM*. All the variables presented completeness above 95% on *SIM* and 98% on *Sinasc*, and were classified as excellent. In the post-linkage phase, it was possible to recover the incomplete fields and complete all the variables, achieving 99 to 100% of completeness, remaining as excellent (Table 1).

When comparing the completeness proportion between the variables on *Sinasc* and post-linkage databases, only the variable 'number of children dead' presented statistically significant difference ($p<0.05$). In the analyses between *SIM* and post-linkage databases, statistically significant differences were observed for all the variables, except for 'sex' (Table 1).

According to Kappa index, the agreement was excellent for all variables (Kappa index >0.80). The highest agreement assessed by ICC, classified as excellent, was identified for the variable 'birth weight' (ICC=0.997), whilst the smallest agreement, classified as substantial, was found for 'number of live children' (ICC=0.762) (Table 2).

Discussion

A high linking proportion was observed between both information systems, higher than 95%. A recent survey (2015) on linkage between *SIM* and *Sinasc* to improve information on infant mortality also identified a percentage higher than 95% in Recife-PE.[5]



a) CLB: Certificate of Live Birth
b) DC: Death Certificate

Figure 1 – Flowchart of the linkage between the Information Systems on Live Births (Sinasc) and on Mortality (*SIM*) in the municipality of Recife, Pernambuco, 2010-2012

Table1 – Completeness of the Information Systems on Live Births (Sinasc) and on Mortality (SIM), before and after the linkage, in the municipality of Recife, Pernambuco, 2010-2012

Variables	Pre-*linkage*						Post-*linkage* SIM-Sinasc	
	SIM			*Sinasc*				
	N	%	p [a]	N	%	p [a]	N	%
Related to the mother and to the pregnancy								
Mother's age	805	99.3	0.04	–	100.0	NA [b]	–	100.0
Mother's education level	800	98.7	<0.01	809	99.7	1.00	810	99.9
Number of live children	801	98.8	<0.01	–	100.0	NA [b]	–	100.0
Number of dead children	776	95.7	<0.01	800	99.9	<0.01	–	100.0
Length of pregnancy	783	96.6	<0.01	807	99.5	0.13	–	100.0
Type of pregnancy	799	98.5	<0.01	–	100.0	NA [b]	–	100.0
Related to the newborn and delivery								
Sex	801	98.8	0.81	801	98.8	0.81	803	99.0
Birth weight	805	99.3	0.04	–	100.0	NA [b]	–	100.0
Type of delivery	799	98.5	<0.01	–	100.0	NA [b]	–	100.0

a) Pearson chi-square test. The p-value refers to the comparison between Sinasc and SIM with the database that resulted from the linkage.

b) NA: not applicable, because the analyzed proportions are the same.

Table 2 – Analysis of the agreement between the variables that are common to the Information Systems on Live Births (Sinasc) and on Mortality (SIM) in the municipality of Recife, Pernambuco, 2010-2012

Qualitative variables	n	Kappa index	95%CI [a]
Mother's education level	725	0.844	0.811;0.876
Length of pregnancy	671	0.808	0.776;0.840
Type of pregnancy	793	0.917	0.872;0.961
Type of delivery	783	0.942	0.918;0.965
Sex	800	0.973	0.957;0.988
Quantitative variables	**n**	**ICC [b]**	**95%CI [a]**
Mother's age	746	0.993	0.992;0.994
Birth weight	763	0.997	0.996;0.997
Number of live children	663	0.762	0.732;0.790
Number of dead children	731	0.857	0.837;0.875

a) 95%CI: 95% Confidence Interval

b) ICC: intraclass correlation coefficient. The p-value was <0.001 for all the variables. The agreement was obtained from the 811 records linked, after the linkage between *Sinasc* and *SIM*.

Among the factors that contributed to the adequacy of the related information, we can mention: advances on *SIM* and *Sinasc* coverage and regularity,[17,18] improvements on DC and CLB completness,[19] besides the consolidation of infant and fetal mortality surveillance in Recife-PE.[2]

One of the identification variables (mother's name) presented different spellings in some records, between the databases. These cases make the linkage between DC and CLB more difficult.[6] Another important group of non-linked records concerns the CLB number, which was absent or divergent. It revealed a deficiency in collecting this piece of information, and pointed to the need of improving this data completeness.[4]

For all the variables that *Sinasc* and *SIM* have in common, the completeness was excellent. An evaluation on *Sinasc*, conducted nationwide, showed that this system presents high completeness and low percentage of ignored/blank spaces.[19] Almost all the deaths recorded occurred within hospitals, making the search for medical records of the mother and the newborn easier.[17] This

fact probably allowed data recovery of *SIM* from *Sinasc* records, with an increment of completeness percentage after the linkage.

The fields of all variables analyzed in both CLB and DC presented agreement from substantial to excellent. This finding confirms the improvement in the adequacy level of vital information,[20] shows the acceptable quality of data related to vital events and reassures the use of these information systems as assessment tools of health situation.[10]

It is necessary to incorporate the continuous completeness analysis of DC and CLB so the assessment of the information adequacy may contribute to vital statistics improvement. We suggest the use of linkage in the routine of municipality health services, considering its low operational cost, ease of execution and potential improvement in the quality of vital statistics information systems.

Authors' contributions

All the authors equally contributed to the study conception and design, data analysis, drafting and approval of the final version of the manuscript, and declared to be responsible for all aspects of the work, ensuring its accuracy and integrity.

References

1. Mello Jorge MHP, Laurenti R, Gotlieb SLB. Avaliação dos sistemas de informação em saúde no Brasil. Cad Saude Colet. 2010;18(1):7-18.
2. Frias PG, Szwarcwald CL, Souza Júnior PRB, Almeida WS, Lira PIC. Correção de informações vitais: estimação da mortalidade infantil, Brasil, 2000-2009. Rev Saude Publica. 2013 dez;47(6):1048-58.
3. Szwarcwald CL, Frias PG, Souza Júnior PRB, Almeida WS, Morais Neto OL. Correction of vital statistics based on a proactive search of deaths and live births: evidence from a study of the North and Northeast regions of Brazil. Popul Health Metr. 2014;12:16.
4. Barreto JOM, Nery IS. Óbitos infantis em um estado do Nordeste brasileiro: características e evitabilidade. Tempus Actas Saude Colet. 2015 set;9(3):9-19.
5. 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 Matern Infant. 2015 jan-mar;15(1):57-66.
6. Coeli CM. A qualidade do *linkage* de dados precisa de mais atenção. Cad Saude Publica. 2015 jul;31(7):1349-50.
7. Correia LOS, Padilha BM, Vasconcelos SML. Métodos para avaliar a completitude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática. Cienc Saude Coletiva. 2014 nov;19(11):4467-78.
8. Santos SLD, Silva ARV, Campelo V, Rodrigues FT, Ribeiro JF. Utilização do método *linkage* na identificação dos fatores de risco associados à mortalidade infantil: revisão integrativa da literatura. Cienc Saude Coletiva. 2014 jul;19(7):2095-104.
9. Schoeps D, Almeida MF, Raspantini PR, Novaes HMD, Silva ZP, Lefevre F. SIM e SINASC: representação social de enfermeiros e profissionais de setores administrativos que atuam em hospitais no município de São Paulo. Cienc Saude Coletiva. 2013 maio;18(5):1483-92.
10. Frias PG, Szwarcwald CL, Lira PIC. Avaliação dos sistemas de informações sobre nascidos vivos e óbitos no Brasil na década de 2000. Cad Saude Publica. 2014 out; 30(10):2068-80.
11. Lansky S, Friche AAL, Silva AAM, Campos D, Bittencourt SDA, Carvalho ML, et al. Pesquisa Nascer no Brasil: perfil da mortalidade neonatal e avaliação da assistência à gestante e ao recém-nascido. Cad Saude Publica. 2014;30 supl 1:S192-207.
12. Gaiva MAM, Fujimori E, Sato APS. Mortalidade neonatal em crianças com baixo peso ao nascer. Rev Esc Enferm. USP. 2014 out;48(5):778-86.
13. Morais CAM, Takano OA, Souza JSF. Mortalidade infantil em Cuiabá, Mato Grosso, Brasil, 2005: comparação entre o cálculo direto e após o *linkage* entre bancos de dados de nascidos vivos e óbitos infantis. Cad Saude Publica. 2011 fev; 27(2):287-94.
14. Coeli CM, Pinheiro RS, Camargo Júnior KR. Conquistas e desafios para o emprego das técnicas de *record linkage* na pesquisa e avaliação em saúde no Brasil. Epidemiol Serv Saude. 2015 out-dez;24(4):795-802.
15. Romero DE, Cunha CB. Avaliação da qualidade das variáveis epidemiológicas e demográficas do Sistema

de Informações sobre Nascidos Vivos, 2002. Cad Saude Publica. 2007 mar;23(3):701-14.

16. Landis JR, Koch GC. A medida da concordância entre observadores para dados categóricos. Biometrics. 1977 mar;33(1):159-74.

17. Ramalho MOA, Frias PG, Vanderlei LCM, Macedo VC, Lira PIC. Avaliação da incompletitude de óbitos de menores de um ano em Pernambuco, Brasil, 1999-2011. Cienc Saude Coletiva. 2015 set;20(9):2891-8.

18. Rodrigues M, Bonfim C, Portugal JL, Frias PG, Gurgel IGD, Costa TR, et al. Análise espacial da mortalidade infantil e adequação das informações vitais: uma proposta para definição de áreas prioritárias. Cienc Saude Coletiva. 2014 jul;19(7):2047-54.

19. Oliveira MM, Andrade SSCA, Dimech GS, Oliveira JCG, Malta DC, Rabello Neto DL, et al. Avaliação do Sistema de Informações sobre nascidos vivos. Brasil, 2006 a 2010. Epidemiol Serv Saude. 2015 out-dez;24(4):629-40.

20. Frias PG, Szwarcwald CL, Lira PIC. Estimação da mortalidade infantil no contexto de descentralização do Sistema Único de Saúde (SUS). Rev Bras Saude Matern Infant. 2011 out-dez;11(4):463-70.

Received on 06/03/2016
Approved on 20/05/2016