DOI: 10.1590/1980-549720190045

NOTES AND INFORMATION / *NOTAS E INFORMAÇÕES*

Interobserver reliability in the classification of pairs of records formed by probabilistic linkage of SISMAMA databases

Confiabilidade interobservadores na classificação de pares formados no relacionamento probabilístico entre bases de dados do SISMAMA

Vania Reis Girianelli[I], Jeane Glaucia Tomazelli[II], Mário Círio Nogueira[III], Camila Soares Lima Corrêa[IV], Eduardo Oliveira de Souza[V], Ligia Gabrielli[VI], Estela M. L. Aquino[VII], Maximiliano Ribeiro Guerra[IV,VIII], Bianca Lucia De Stavola[IX], Isabel dos-Santos-Silva[X], Gulnar Azevedo e Silva[XI]

ABSTRACT: ***Introduction:*** The study assessed interobserver reliability in the classification of record pairs formed during probabilistic linkage of health-related databases, a key step in the methodology validation to be used in a larger on-going study on inequalities in the access to breast and cervical cancer control activities in Brazil (DAAC-SIS). ***Methodology:*** The RecLink software was used to link two databases of the Breast Cancer Control Information System (SISMAMA) in the state of Minas Gerais, Brazil: a reference database, which included 301 screening mammograms with probable benign diagnosis (BI-RADS 3 category) recorded in October 2010, and a database comprising 158,517 mammograms registered in 2011. Subsequently, the 215 pairs of records that were not assigned the maximum RecLink score were independently classified as being true or false by ten independent evaluators from four participating centers. ***Results:*** The Kappa coefficient ranged from 0.87 to 1.00. Six evaluators were in perfect agreement with one or more evaluators from the other centers. The global Kappa was 0.96 (95% confidence interval — 95%CI 0.94 – 0.99). ***Discussion:*** Assessment of interobserver reliability is key to ensuring the quality of the record linkage, and it should be routine practice in studies of this nature. The disclosure of such results contributes to transparency in the conduct of such studies and in the reporting of their findings. ***Conclusion:*** Interobserver reliability in this study was excellent, indicating satisfactory team consistency in the classification of record pairs.

Keywords: Medical Record Linkage. Observer Variation. Health Information Systems. Systems Integration. Breast Neoplasms. Disease prevention. Control.

[I]Departament of Human Rights, Health and Cultural Diversity, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz – Campo Grande (MS), Brazil.
[II]Division of Early Detection and Support to Network Organization, General Coordination of Prevention and Surveillance, Instituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro (RJ), Brazil.
[III]Department of Collective Health, Faculdade de Medicina, Universidade Federal de Juiz de Fora – Juiz de Fora (MG), Brazil.
[IV]Postgraduate Program in Collective Health, Universidade Federal de Juiz de Fora – Juiz de Fora (MG), Brazil.
[V]Fundação Oncocentro de São Paulo – São Paulo (SP), Brazil.
[VI]Institute of Collective Health, Universidade Federal da Bahia – Salvador (BA), Brazil.
[VII]Integrated Program in Gender and Health, Institute of Collective Health, Universidade Federal da Bahia – Salvador (BA), Brazil.
[VIII]Centre de Recherche U900, Epidémiologie des Cancer, Institut Curie – Paris, França.
[IX]Population, Policy and Practice Programme, UCL Great Ormond Street Institute of Child Health – London, England.
[X]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine – London, England.
[XI]Institute of Social Medicine, Universidade do Estado do Rio de Janeiro – Rio de Janeiro (RJ), Brazil.
Corresponding author: Vania Reis Girianelli. Department of Human Rights, Health and Cultural Diversity, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz. Avenida Brasil, 4.036, sala 905, Manguinhos, Zip Code: 21040-361, Rio de Janeiro, RJ, Brazil. Email: vania.girianelli@ensp.fiocruz.br
Conflict of interests: none – **Financial support:** Fundo Newton, partnership between Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (170.011/2015) and Medical Research Council of Great Britain (MR/M026280/1).

RESUMO: ***Introdução:*** O estudo avaliou a confiabilidade interobservadores na classificação de pares de registros formados durante o processo de relacionamento probabilístico, sendo uma das etapas de validação da metodologia a ser utilizada em pesquisa sobre desigualdades de acesso às ações de controle dos cânceres de mama e do colo do útero no Brasil (DAAC-SIS). ***Metodologia:*** O programa RecLink foi usado para relacionar as bases de dados do Sistema de Informação do Controle do Câncer de Mama (SISMAMA) do estado de Minas Gerais, tendo como referência 301 mamografias de rastreamento com resultado provavelmente benigno (categoria BI-RADS 3), registradas em outubro de 2010 e, como comparação, 158.517 mamografias registradas em 2011. Posteriormente, 215 pares de registros, que não obtiveram o escore máximo atribuído pelo RecLink, foram classificados independentemente por dez avaliadores, de quatro centros participantes da pesquisa, como pares verdadeiros ou falsos. ***Resultados:*** O coeficiente Kappa variou de 0,87 a 1,00. Seis avaliadores obtiveram concordância perfeita com um ou mais avaliadores de outros centros. O Kappa global foi 0,96 (intervalo de confiança de 95% — IC95% 0,94 – 0,99). ***Discussão:*** A avaliação interobservadores foi fundamental para garantir a qualidade do processo de relacionamento, e a sua prática deve ser rotina em estudos dessa natureza. A divulgação desses resultados contribui para a transparência na condução e no relato do estudo em curso. ***Conclusão:*** A confiabilidade interobservadores foi excelente, sinalizando homogeneidade satisfatória da equipe na classificação dos pares de registros.

Palavras-chave: Registro médico coordenado. Variações dependentes do observador. Sistemas de Informação em Saúde. Integração de Sistemas. Neoplasias da mama. Prevenção. Controle.

INTRODUCTION

Several Health Information Systems (SIS, acronym in Portuguese) have been developed in the last decades in Brazil to record mortality, morbidity and health care data; however, records belonging to the same individual , cannot easily be identified across these databases because recording of the National Health Card (CNS) number, a unique identification number given to each Brazilian individual, is not yet mandatory in all SIS.

Computer algorithms, based on probabilistic linkage methods, have been developed to help to identify information belonging to the same individual across different SIS. This method uses statistical models to match pairs of records and to score them according to their likelihood of being true pairs. In Brazil, the RecLink software is the most used program[1]. It generates a score, which summarizes the degree of global agreement based on the agreement and disagreement of a set of matched identifier fields[2]. Manual classification of matched pairs that did not obtain a maximum score is, however, still necessary and may vary among different evaluators, because it involves subjective judgement.

Reliability studies of probabilistic linkage are scarce and generally restrict the evaluation to the agreement between the fields of the analyzed databases[3,4]. Assessment of inter-observer reliability is, however, key to monitor and assess the degree of consistency in the classification of pairs of records between the various evaluators contributing to the study.

The present study aimed to evaluate interobserver reliability in the classification of pairs of records formed during probabilistic linkage of data from the Breast Cancer Control Information System (SISMAMA, acronym in Portuguese). This assessment is part of a larger on-going research project to investigate inequalities in the access to breast and cervical cancer control activities in Brazil (DAAC-SIS), and constitutes one of the validation steps of the methodology to be used.

METHODOLOGY

An interobserver reliability study was performed regarding the classification of pairs formed by the RecLink software (version 3.1.6.3160), in the probabilistic linkage between the SISMAMA databases – mammography module, from the State of Minas Gerais, Brazil. The databases analysed comprised only records with a valid CNS number; however, data on this variable was not made available to the evaluators. The reference database included 301 records of women who underwent mammography in October 2010 with a likely benign outcome (BI-RADS 3 category), to whom repeat mammography within six months is recommended[5]. This reference database was linked to a database consisting of 158,517 mammograms registered in 2011, after exclusion of two duplicate records.

For pair formation, the *soundex* code of the woman's first name was used. For score formation, the woman's "full name" and "date of birth" fields and "mother's full name" were used, with the suggested parameters[2]. Only pairs with scores > 0.5 were considered. The pairs with maximum score (17.2) were excluded, and the others were independently analyzed by 10 evaluators, four of whom were from Minas Gerais, two from Bahia, two from Rio de Janeiro, and two from São Paulo).

For each pair of evaluators, Cohen's Kappa coefficient was calculated with the respective 95% confidence intervals (95%CI), and the results were classified as proposed by Byrt[6]. In addition, exact[7] and non-exact[8] global Kappa coefficients were calculated, with their 95%CI estimated using a bootstrap technique with generation of 1,000 random samples based on the original sample (Kappa of each pair of evaluators). These samples were used to generate the sample distribution of the estimate (global Kappa). The lower and upper limits of the global Kappa correspond, respectively, to the estimates of the 2.5 and 97.5 percentiles of the sample distribution. The analyses were performed using the statistical software R[9]. Subsequently, the disagreements among the evaluators were reviewed by the entire team.

The Ethics and Research Committees of the Institute of Social Medicine from *Universidade do Estado do Rio de Janeiro* (Certificate of Presentation for Ethical Assessment – CAAE: 42928415.2.0000.5260) and *Universidade Federal de Juiz de Fora*, Minas Gerais (CAAE: 46844115.7.0000.5147) approved the study.

RESULTS

RecLink formed 281 pairs of records, 66 (23.5%) of which had a maximum score (17.2). The other 215 pairs, with scores ranging from 17.0 to 0.54, were independently classified by 10 evaluators. Only nine pairs (4.2%) obtained discordant classifications. A subsequent review by the entire team revealed that only one pair of records was wrongly classified as true by one evaluator.

The Kappa coefficient for each of the 45 pairs of evaluators ranged from 0.87 to 1.00 (Figure 1), with 80% (36/45) of the pairs having excellent agreement (> 0.92). The remaining pairs involved evaluator 3, with very good agreement (0.87 to 0.90). The agreement was perfect (Kappa = 1.00) for 14 pairs, which involved six of the 10 evaluators (60%), corresponding to at least one evaluator from each of the four centers participating in the study. The exact and non-exact global Kappa was 0.96 ($p < 0.001$, 95%CI 0.94 – 0.99).

DISCUSSION

This study involved the classification of 215 pairs of records by 10 independent evaluators, with dual comparison between them, corresponding to 45 pairs of evaluators. Each center participating in the study had at least one evaluator who obtained perfect agreement with one or more evaluators from other centers.

Interim evaluations, however, should be implemented throughout the DAAC-SIS study to monitor linkage quality. Such care is fundamental to minimize possible losses or inclusion of false pairs, which could introduce biases in the analyses to be performed[10].

The dissemination of these results highlights the efforts that should be made to ensure quality control when conducting record linkage studies across different SIS from the Brazilian Unified Health System (SUS, acronym in Portuguese). In addition, it helps to disseminate the need to incorporate routinely such assessment into similar studies.

CONCLUSION

The study revealed excellent interobserver reliability and demonstrated the team consistency in the classification of record pairs. Assessment of interobserver reliability is a key tool to establish the quality of the record linkage process, and it should be regarded as routine practice in studies of this nature.

Evaluators	T	F	D	Kappa
1 and 2	35	180	0	1.00
1 and 3	29	180	6	0.89
1 and 4	35	180	0	1,00
1 and 5	35	180	0	1.00
1 and 6	34	180	1	0.98
1 and 7	32	180	3	0.95
1 and 8	35	179	1	0.98
1 and 9	35	180	0	1.00
1 and 10	35	180	0	1.00
2 and 3	29	180	6	0.89
2 and 4	35	180	0	1.00
2 and 5	35	180	0	1.00
2 and 6	34	180	1	0.98
2 and 7	32	180	3	0.95
2 and 8	35	179	1	0.98
2 and 9	35	180	0	1.00
2 and 10	35	180	0	1.00
3 and 4	29	180	6	0.89
3 and 5	29	180	6	0.89
3 and 6	28	180	7	0.87
3 and 7	28	182	5	0.90
3 and 8	29	179	7	0.87
3 and 9	29	180	6	0.89
3 and 10	29	180	6	0.89
4 and 5	35	180	0	1.00
4 and 6	34	180	1	0.98
4 and 7	32	180	3	0.95
4 and 8	35	179	1	0.98
4 and 9	35	180	0	1.00
4 and 10	35	180	0	1.00
5 and 6	34	180	1	0.98
5 and 7	32	180	3	0.95
5 and 8	35	179	1	0.98
5 and 9	35	180	0	1.00
5 and 10	35	180	0	1.00
6 and 7	31	180	4	0.93
6 and 8	34	179	2	0.97
6 and 9	34	180	1	0.98
6 and 10	34	180	1	0.98
7 and 8	32	179	4	0.93
7 and 9	32	180	3	0.95
7 and 10	32	180	3	0.95
8 and 9	35	179	1	0.98
8 and 10	35	179	1	0.98
9 and 10	35	180	0	1.00
Global	**27**	**179**	**9**	**0.96**

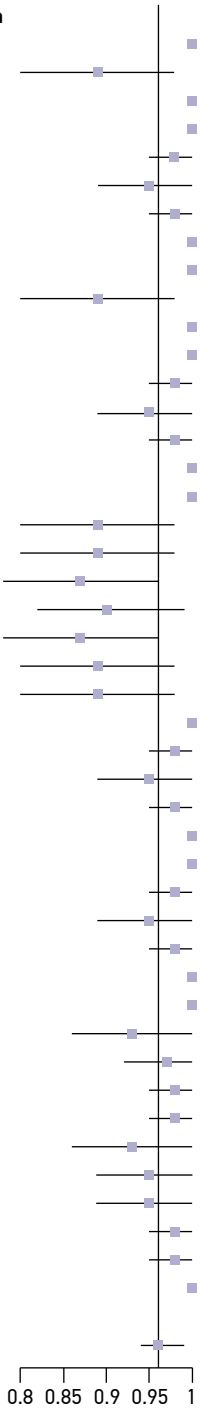

Figure 1. Number of pairs of records classified as true (T), false (F), discordant (D), Kappa coefficient for each of the 45 pairs of evaluators, global Kappa coefficient and respective 95% confidence intervals.

REFERENCES

1. Silva JP, Travassos C, Vasconcellos MMV, Campos LM. Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil. Cad Saúde Colet 2006; 14(2): 197-224.
2. Camargo Jr. KR, Coeli CM. Reclink: aplicativo para o relacionamento de banco de dados implementando o método probabilistic record linkage. Cad Saúde Pública 2000; 16(2): 439-47. http://dx.doi.org/10.1590/S0102-311X2000000200014
3. Oliveira PPV, Silva GA, Curado MP, Malta DC, Moura L. Confiabilidade da causa básica de óbito por câncer entre Sistema de Informações sobre Mortalidade do Brasil e Registro de Câncer de Base Populacional de Goiânia, Goiás, Brasil. Cad Saúde Pública 2014; 30(2): 296-304. http://dx.doi.org/10.1590/0102-311X00024813
4. Guimarães PV, Coeli CM, Cardoso RCA, Medronho RA, Fonseca SC, Pinheiro RS. Confiabilidade dos dados de uma população de muito baixo peso ao nascer no Sistema de Informações sobre Nascidos Vivos 2005-2006. Rev Bras Epidemiol 2012; 15(4): 694-704. http://dx.doi.org/10.1590/S1415-790X2012000400002
5. Brasil. Ministério da Saúde. Instituto Nacional de Câncer. Controle do câncer de mama: documento de consenso. Rio de Janeiro: Instituto Nacional de Câncer; 2004.
6. Byrt T. How good is that agreement? Epidemiology 1996; 7(5): 561.
7. Conger AJ. Integration and generalization of Kappas for multiple raters. Psychol Bull 1980; 88(2): 322-8. https://psycnet.apa.org/doi/10.1037/0033-2909.88.2.322
8. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971; 76(5): 378-82. https://psycnet.apa.org/doi/10.1037/h0031619
9. R Core Team. R: A language and environment for statistical computing [Internet]. Viena, Áustria: R Foundation for Statistical Computing; 2013 [acessado em ago. 2016]. Disponível em: http://www.R-project.org/.
10 Coeli CM. A qualidade do linkage de dados precisa de mais atenção. Cad Saúde Pública 2015; 31(7): 1349-50. http://dx.doi.org/10.1590/0102-311XED010715

Received on: 08/04/2017
Final version presented on: 06/15/2018
Approved on: 07/12/2018

Contribution from authors: Vania Reis Girianelli: development, planning, analysis, interpretation of results and writing of the work. Jeane Glaucia Tomazelli: development, planning, analysis, interpretation of results and writing of the work. Mário Círio Nogueira: analysis, interpretation of results and review of the work. Camila Soares Lima Corrêa: analysis, interpretation of results and review of the work. Eduardo Oliveira Souza: analysis, interpretation of results and review of the work. Ligia Gabrielli: analysis, interpretation of the results and review of the work. Estela M. L. Aquino: analysis, interpretation of the results and review of the work. Maximiliano Ribeiro Guerra: analysis, interpretation of results and review of the work. Bianca Lucia De Stavola: analysis, interpretation of results and review of the work. Isabel dos-Santos-Silva: analysis, interpretation of results and review of the work. Gulnar Azevedo e Silva: conception, planning, analysis, interpretation of results and review of the work.