












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

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

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

**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.

<sup>I</sup>Departamento de Direitos Humanos, Saúde e Diversidade Cultural, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz – Campo Grande (MS), Brasil.

<sup>II</sup>Divisão de Detecção Precoce e Apoio à Organização de Rede, Coordenação Geral de Prevenção e Vigilância, Instituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro (RJ), Brasil.

<sup>III</sup>Departamento de Saúde Coletiva, Faculdade de Medicina, Universidade Federal de Juiz de Fora – Juiz de Fora (MG), Brasil.

<sup>IV</sup>Programa de Pós-Graduação em Saúde Coletiva, Universidade Federal de Juiz de Fora – Juiz de Fora (MG), Brasil.

<sup>V</sup>Fundação Oncocentro de São Paulo – São Paulo (SP), Brasil.

<sup>VI</sup>Instituto de Saúde Coletiva, Universidade Federal da Bahia – Salvador (BA), Brasil.

<sup>VII</sup>Programa Integrado em Gênero e Saúde, Instituto de Saúde Coletiva, Universidade Federal da Bahia – Salvador (BA), Brasil.

<sup>VIII</sup>Centre de Recherche U900, Epidémiologie des Cancer, Institut Curie – Paris, França.

<sup>IX</sup>Population, Policy and Practice Programme, UCL Great Ormond Street Institute of Child Health – Londres, Inglaterra.

<sup>X</sup>Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine – Londres, Inglaterra.

<sup>XI</sup>Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro – Rio de Janeiro (RJ), Brasil.

**Autor correspondente:** Vania Reis Girianelli. Departamento de Direitos Humanos, Saúde e Diversidade Cultural, Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz, Avenida Brasil, 4.036, sala 905, Manguinhos, CEP: 21040-361, Rio de Janeiro, RJ, Brasil. E-mail: vania.girianelli@ensp.fiocruz.br

**Conflito de interesses:** nada a declarar – **Fonte de financiamento:** Fundo Newton, parceria Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (170.011/2015) e Medical Research Council da Grã-Bretanha (MR/M026280/1).

**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 ReLink 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 ReLink 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.

## INTRODUÇÃO

Diversos Sistemas de Informação em Saúde (SIS) foram desenvolvidos nas últimas décadas no Brasil para registrar dados de mortalidade, morbidade e assistência, entretanto eles ainda não são interoperáveis, pois o número do Cartão Nacional de Saúde (CNS), que permitiria a identificação única para cada indivíduo, ainda não se tornou obrigatório em todos os SIS, o que dificulta o relacionamento entre as várias bases de dados.

Programas que utilizam método de relacionamento probabilístico entre bases de dados têm sido empregados para auxiliar a identificação das informações de um indivíduo nos diversos SIS. Esse método usa modelos estatísticos que propiciam a correspondência entre um par de registros. No Brasil, o programa computacional ReLink é o mais empregado<sup>1</sup>. Ele gera um escore, que resume o grau de concordância global com base na concordância e discordância de um conjunto de campos identificadores pareados<sup>2</sup>. A classificação manual dos pares formados que não obtiveram escore máximo, porém, é necessária e, por apresentar componentes subjetivos, pode variar entre diferentes avaliadores.

Estudos de confiabilidade do relacionamento probabilístico são escassos e geralmente restringem a avaliação à concordância entre os campos das bases de dados analisadas<sup>3,4</sup>. A avaliação da confiabilidade interobservadores, no entanto, é fundamental para a busca da homogeneidade na identificação dos pares de registros entre os avaliadores que participam do estudo.

O presente estudo teve como objetivo avaliar a confiabilidade interobservadores na classificação de pares de registros formados durante o processo de relacionamento probabilístico da base do Sistema de Informação do Controle do Câncer de Mama (SISMAMA). Ele insere-se em pesquisa que analisa as desigualdades de acesso às ações de controle dos cânceres de mama e do colo do útero no Brasil (DAAC-SIS), constituindo uma das etapas de validação da metodologia a ser utilizada.

## METODOLOGIA

Foi realizado estudo de confiabilidade interobservadores em relação à classificação dos pares formados, pelo programa computacional RecLink (versão 3.1.6.3160), no relacionamento probabilístico entre as bases de dados do SISMAMA — módulo mamografia, do estado de Minas Gerais. As bases de dados utilizadas foram restritas aos registros que continham o campo CNS preenchido, mas este não foi disponibilizado para os avaliadores. A base de referência incluiu 301 registros de mulheres que realizaram mamografia em outubro de 2010 com resultado provavelmente benigno (categoria BI-RADS 3), para quem se recomenda a repetição de mamografia em seis meses<sup>5</sup>, e a de comparação foi constituída de 158.517 mamografias registradas em 2011, após exclusão de dois registros duplicados.

Para formação dos pares, foi usado o código *soundex* do primeiro nome da mulher. Para a formação do escore, foram utilizados os campos “nome completo” e “data de nascimento” da mulher e “nome completo da mãe”, com os parâmetros sugeridos<sup>2</sup>. Apenas os pares com escores > 0,5 foram considerados. Os pares com escore máximo (17,2) foram excluídos, e os demais, analisados independentemente por 10 avaliadores, sendo quatro de Minas Gerais e dois de cada um dos demais estados (Bahia, Rio de Janeiro e São Paulo).

Para cada dupla de avaliadores, foi calculado o coeficiente Kappa de Cohen com respectivos intervalos de confiança de 95% (IC95%), e os resultados classificados como proposto por Byrt<sup>6</sup>. Adicionalmente, foi calculado o coeficiente Kappa global exato<sup>7</sup> e não exato<sup>8</sup>, e o IC95% foi estimado empregando-se técnica de reamostragem com reposição (*bootstrap*) com geração de 1.000 amostras aleatórias com base na amostra original (Kappa de cada dupla de avaliadores), e essas amostras foram utilizadas para gerar a distribuição amostral da estimativa (Kappa global). Os limites inferiores e superiores do Kappa global correspondem, respectivamente, às estimativas dos percentis 2,5 e 97,5 da distribuição amostral. As análises foram realizadas pelo programa R<sup>9</sup>. Posteriormente, as discordâncias entre os avaliadores foram revisadas por toda a equipe.

O estudo foi aprovado pelos Comitês de Ética e Pesquisa do Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro (Certificado de Apresentação para Apreciação Ética — CAAE: 42928415.2.0000.5260) e da Universidade Federal de Juiz de Fora, Minas Gerais (CAAE: 46844115.7.0000.5147).

## RESULTADOS

O RecLink formou 281 pares de registros, sendo 66 (23,5%) com escore máximo (17,2). Os demais 215 pares, com escore variando de 17,0 a 0,54, foram classificados independentemente por 10 avaliadores. Apenas nove pares (4,2%) obtiveram classificações discordantes. Revisão posterior por toda a equipe revelou que apenas um par foi classificado indevidamente como verdadeiro por um avaliador.

O coeficiente Kappa para cada uma das 45 duplas de avaliadores variou de 0,87 a 1,00 (Figura 1), com 80% (36/45) das duplas tendo concordância excelente ( $> 0,92$ ). As duplas restantes envolveram o avaliador 3, com concordância muito boa (0,87 a 0,90). A concordância foi perfeita (Kappa = 1,00) para 14 duplas, tendo estas envolvido seis dos 10 avaliadores (60%), correspondendo a pelo menos um/a avaliador/a de cada um dos quatro centros participantes do estudo. O Kappa global exato e não exato foi 0,96 ( $p < 0,001$ ; IC95% 0,94 – 0,99).

## DISCUSSÃO

Este estudo envolveu a classificação de 215 pares de registros por 10 avaliadores independentes, com comparação dois a dois entre eles, correspondendo a 45 duplas de avaliadores analisadas. Cada centro participante do estudo contou com pelo menos um/a avaliador/a que obteve concordância perfeita com um ou mais avaliadores de outros centros.

Avaliações intermediárias, contudo, devem ser implementadas ao longo do estudo DAAC-SIS para monitorar a qualidade. Esses cuidados são fundamentais para minimizar possíveis perdas ou inclusão de falsos pares, o que poderia introduzir vieses nas análises a serem realizadas<sup>10</sup>.

A divulgação desses resultados contribui para documentar os esforços envidados para a garantia de controle de qualidade na condução de estudos que utilizam relacionamento entre diferentes SIS do Sistema Único de Saúde (SUS). Adicionalmente, ajuda a disseminar a incorporação da avaliação em estudos semelhantes.

## CONCLUSÃO

O estudo revelou confiabilidade interobservadores excelente, demonstrando a homogeneidade da equipe na classificação dos pares de registros. A avaliação interobservadores foi fundamental para certificar a qualidade do processo de relacionamento, e a sua prática deve ser rotina em estudos dessa natureza.

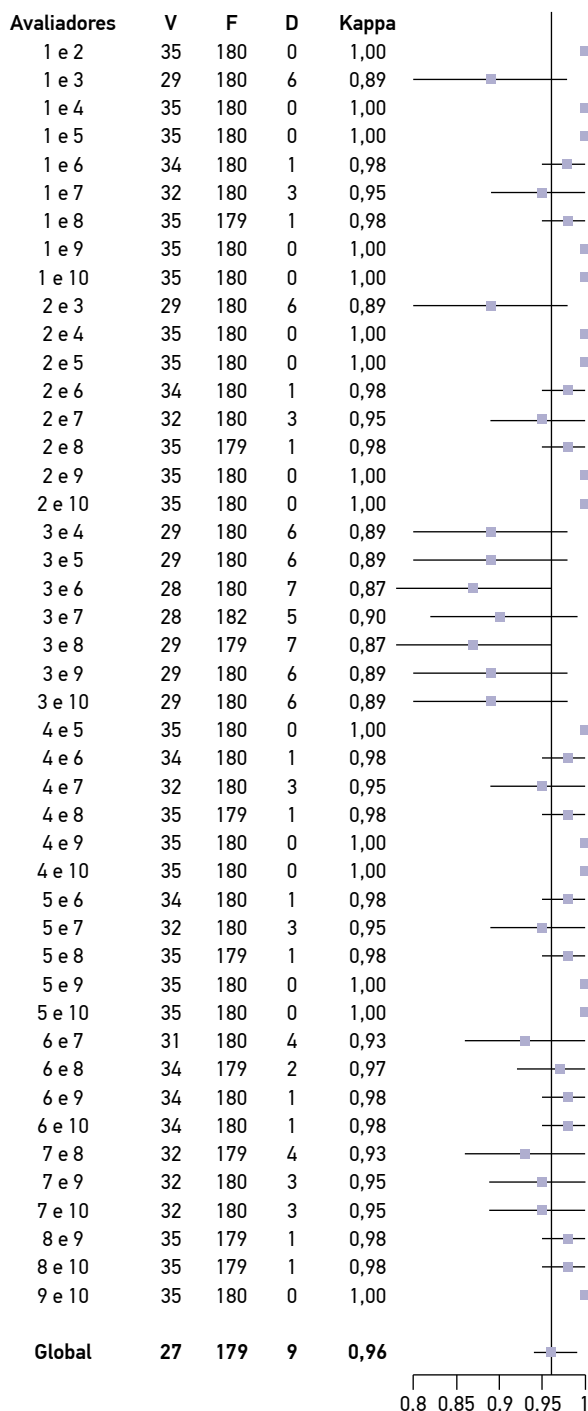


Figura 1. Número de pares classificados como verdadeiros (V), falsos (F), discordantes (D), coeficiente Kappa para cada uma das 45 duplas de avaliadores, coeficiente Kappa global e respectivos intervalos de confiança de 95%.

## REFERÊNCIAS

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/S0102-311X200000200014>
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>

Recebido em: 04/08/2017

Versão final apresentada em: 15/06/2018

Aprovado em: 12/07/2018

**Contribuição dos autores:** Vania Reis Girianelli: concepção, planejamento, análise, interpretação dos resultados e redação do trabalho. Jeane Glucia Tomazelli: concepção, planejamento, análise, interpretação dos resultados e redação do trabalho. Mário Círio Nogueira: análise, interpretação dos resultados e revisão do trabalho. Camila Soares Lima Corrêa: análise, interpretação dos resultados e revisão do trabalho. Eduardo Oliveira Souza: análise, interpretação dos resultados e revisão do trabalho. Lígia Gabrielli: análise, interpretação dos resultados e revisão do trabalho. Estela Maria Motta Lima Leão de Aquino: análise, interpretação dos resultados e revisão do trabalho. Maximiliano Ribeiro Guerra: análise, interpretação dos resultados e revisão do trabalho. Bianca Lucia De Stavola: análise, interpretação dos resultados e revisão do trabalho. Isabel dos-Santos-Silva: análise, interpretação dos resultados e revisão do trabalho. Gulnar Azevedo e Silva: concepção, planejamento, análise, interpretação dos resultados e revisão do trabalho.

