

Estratégias usadas no relacionamento entre Sistemas de Informações em Saúde para seguimento das mulheres com mamografias suspeitas no Sistema Único de Saúde

Strategies used to link Health Information Systems for the follow-up of women with abnormal mammograms in the Brazilian Public Health System

Jeane Glauca Tomazelliⁱ, Vania Reis Girianelliⁱⁱ, Gulnar Azevedo e Silvaⁱⁱⁱ

RESUMO: *Introdução:* Sistemas de Informação em Saúde (SIS) são instrumentos potenciais para análise da situação de saúde, mas a não obrigatoriedade de preenchimento de um campo comum único dificulta sua integração. O objetivo deste estudo foi descrever as estratégias utilizadas para relacionar bases de dados do Sistema Único de Saúde (SUS) que contenham registros para o controle do câncer de mama e avaliar a adequação da estratégia empregada. *Metodologia:* Foram relacionados probabilisticamente o Sistema de Informação do Controle do Câncer de Mama (SISMAMA), o Sistema de Informação Ambulatorial (SIA, por meio do Boletim de Produção Ambulatorial Individualizado — BPA-I — e da Autorização de Procedimentos Ambulatoriais de Alta Complexidade — APAC), o Sistema de Informação Hospitalar (SIH) e o Sistema de Informação sobre Mortalidade (SIM). A base de referência foram registros de mamografia suspeita e altamente suspeita de malignidade do segundo semestre de 2010. A estratégia de relacionamento incluiu 15 passos. Os registros com Cartão Nacional de Saúde (CNS) ou Cadastro de Pessoa Física (CPF) foram utilizados para estimar a sensibilidade da estratégia, tendo como padrão-ouro os pares de registros identificados nos passos iniciais, que usaram esses campos como chave de blocagem. *Resultados:* A utilização do CNS e do CPF como estratégia de relacionamento permitiu identificar elevada proporção de pares verdadeiros nas bases nas quais existiam essas variáveis: 47,3% nas mamografias de seguimento, 41,4% no SIH e 45,5% na APAC. A sensibilidade da estratégia utilizada foi de 100%. *Conclusão:* O estudo mostrou que as estratégias utilizadas foram satisfatórias e que a utilização do CNS e do CPF permitiu a identificação de muitos pares, mesmo com a ausência de crítica destes e a possibilidade de realizar o relacionamento entre bancos com poucos campos de identificação.

Palavras-chave: Sistemas de Informação em Saúde. Integração de sistemas. Neoplasias de mama.

ⁱInstituto Nacional de Câncer José Alencar Gomes da Silva – Rio de Janeiro (RJ), Brasil.

ⁱⁱEscola Nacional de Saúde Pública Sérgio Arouca, Fundação Oswaldo Cruz – Rio de Janeiro (RJ), Brasil.

ⁱⁱⁱInstituto de Medicina Social, Universidade do Estado do Rio de Janeiro – Rio de Janeiro (RJ), Brasil.

Autor correspondente: Jeane Glauca Tomazelli. Divisão de Detecção Precoce e Apoio à Organização de Rede – Instituto Nacional de Câncer. Rua Marquês de Pombal, 125, Centro, CEP: 22230-240, Rio de Janeiro, RJ, Brasil. E-mail: jtomazelli@inca.gov.br

Conflito de interesses: nada a declarar – **Fonte de financiamento:** nenhuma.

ABSTRACT: Introduction: Health Information Systems are potential instruments to analyze health situation; however, the non-compulsory filling of a single common field makes it difficult to link systems' data. This study aimed to describe and evaluate the adequacy of the strategies used to perform data linkage between databases from the Brazilian Public Health System (SUS) as to records for breast cancer control. **Methods:** The Breast Cancer Control Information Systems (SISMAMA), the Outpatient Information System (SIA, through Individualized Outpatient Service Production — BPA-I — and High-Complexity Outpatient Procedures Authorization Forms — APAC), the Hospital Information System (SIH), and the Mortality Information System (SIM) were linked probabilistically. The baseline was constructed by records with “suspected” and “highly suspected malignancy” from the second half of 2010. The linkage strategy included 15 steps. Registries with the national health service user card (CNS) or social security number (SSN) were used to estimate the sensitivity of the strategy, considering matches between records identified in the initial steps as gold standard, when these fields were used as key for blocking. **Results:** Using CNS and the SSN as a linkage strategy allowed to identify the high proportion of true matches across databases in which these variables were inputted: 47.3% in follow-up mammography records, 41.4% in SIH, and 45.5% in APAC. The sensitivity of the linkage strategy was 100%. **Conclusion:** The study showed that the strategies were satisfactory and the use of CNS and SSN allowed many matches, even without critical proceedings and with the possibility of linkage between databases based on information from only a few identification fields.

Keywords: Health Information Systems. System integration. Breast neoplasms.

INTRODUÇÃO

Os Sistemas de Informação em Saúde (SIS) são instrumentos potenciais para análise da situação de saúde, planejamento, programação e avaliação, pois registram informações epidemiológicas, assistenciais e estatísticas vitais. Implantados em diferentes momentos no Brasil desde a década de 1970, os SIS foram desenvolvidos para cumprir finalidades distintas. Possuem características diversas que vão de sua abrangência, que pode ser universal, como o Sistema de Informação sobre Mortalidade (SIM), até os serviços pertencentes ao Sistema Único de Saúde (SUS), como o Sistema de Informação Ambulatorial (SIA) e o Sistema de Informação Hospitalar (SIH). Entre os SIS não há uniformidade no preenchimento de alguns campos, nem presença de informações que possibilitem a caracterização socioeconômica do indivíduo¹⁻⁴.

Com a integração desses SIS, é possível obter acompanhamento longitudinal dos indivíduos na rede de cuidado assistencial do SUS. Essa integração não pode ser feita diretamente pela não obrigatoriedade de preenchimento de um campo comum único nos SIS que identifique o usuário. Ainda que essa identificação única no SUS seja gradativamente discutida e implementada, por meio do Cartão Nacional de Saúde (CNS)⁵⁻⁷, na prática a não adoção desse identificador em todos os SIS traz dificuldades para a avaliação da atenção em saúde. Dentre os SIS nacionalmente implantados, o CNS é obrigatório apenas para Autorização de Procedimentos de Alta Complexidade (APAC). No ano de 2011, estimou-se que os registros duplicados, ou seja, apresentar mais de um CNS para o mesmo indivíduo, eram superiores

a 30%⁸, o que introduziu mais uma dificuldade ao processo de avaliação. Outros obstáculos foram o processo de cadastramento descentralizado do CNS e problemas de sincronização com a base nacional do cartão, tornando-a não representativa de todos os indivíduos registrados no sistema⁴.

A utilização de metodologias que possibilitem identificar o usuário em diferentes momentos do cuidado é um recurso que vem sendo usado para superar essa dificuldade. O desenvolvimento de rotinas computacionais que relacionem dados dos SIS, de forma a acompanhar a trajetória dos usuários do SUS e avaliar a assistência, mostra-se fundamental para subsidiar o planejamento⁹⁻¹². Gomes Júnior et al.¹³ propuseram rotinas que permitem essa integração usando o módulo de oncologia da APAC. Adaptações dessa rotina¹⁴ e propostas de automatização do algoritmo¹⁵ são desenvolvidas atualmente.

No Brasil, diversos estudos têm aplicado a técnica de relacionamento probabilístico utilizando o *software* RecLink^{®16-23}, e alguns mostram que tal técnica tem mostrado sensibilidade e especificidade satisfatórias¹⁷⁻¹⁹. É importante, contudo, considerar que a acurácia do relacionamento está ligada ao número de identificadores a serem comparados, bem como à qualidade do seu preenchimento¹⁷. De maneira geral, os pares formados que não são verdadeiros (falso-positivos) tendem a ocorrer quando existem poucos campos de comparação, elevada incompletude dos campos usados para comparação e homônimos. Por outro lado, os pares não identificados (falso-negativos) estão mais associados aos erros de digitação dos campos e à informação incorretamente preenchida²⁴.

Existem dois sistemas de informação oficiais no SUS com registros dos exames das ações de detecção precoce para o câncer: o Sistema de Informação do Controle do Câncer do Colo do Útero (SISCOLO) e o Sistema de Informação do Controle do Câncer de Mama (SISMAMA)²⁵. A maior parte dos estudos identificados que empregaram técnica de relacionamento utilizou os módulos “citologia” e “anatomopatológico” das bases de dados do SISCOLO^{20,23,26,27}. Alguns estudos específicos que utilizaram dados do SISCOLO e do SISMAMA vêm sendo conduzidos empregando tanto o *software* RecLink^{®20} quanto os recursos da área de programação ou de outros *softwares* estatísticos^{26,27}. Em apenas um estudo foi feito relacionamento entre as bases do SISCOLO com o SIM²³. Não estão disponíveis na literatura estudos brasileiros que façam o relacionamento com as bases de dados de diferentes SIS com o intuito de seguir as mulheres rastreadas para os cânceres do colo do útero e de mama, e, assim, avaliar as ações de controle dessas doenças.

Este estudo teve por objetivo descrever as estratégias utilizadas no relacionamento probabilístico entre bases de dados dos SIS com registros sobre controle do câncer de mama e avaliar a adequação da estratégia empregada.

MÉTODOS

Realizou-se estudo descritivo sobre a metodologia utilizada para relacionar as bases de dados dos SIS disponíveis para o rastreamento do câncer de mama no município do Rio de

Janeiro. O período de estudo foi de julho de 2010 a dezembro de 2012, iniciando um ano após a implantação do SISMAMA²⁸ e terminando antes de sua substituição pelo Sistema de Informação do Câncer (SISCAN)²⁹.

Foram utilizados os SIS relacionados ao seguimento de mulheres com mamografia suspeita de malignidade. As bases do SISMAMA (exames de mamografia e anatomopatológico de mama), do SIA (procedimentos de investigação diagnóstica e de tratamento) e do SIH (tratamentos cirúrgicos e internações para radioterapia externa) foram cedidas pela Secretaria Municipal de Saúde. Os dados do SISMAMA foram complementados com os da base nacional cedida pelo Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), após análise preliminar identificar ausência de dados para os meses iniciais do estudo. A base do SIM foi cedida pela Assessoria de Dados Vitais da Secretaria Estadual de Saúde (SES-RJ/SVS/CGVS/ADVITAIS) e refere-se a todos os óbitos ocorridos no Estado nesse período.

Da base do SISMAMA municipal foram gerados arquivos mensais, separadamente, para os exames de mamografia e para os exames anatomopatológicos referentes ao período de interesse disponível (maio de 2011 a dezembro de 2012). Da base nacional do SISMAMA foram utilizados apenas os exames de residentes no município do Rio de Janeiro, sendo gerados arquivos mensais das mamografias realizadas de julho de 2010 a dezembro de 2012 e anuais dos exames anatomopatológicos. A disponibilidade do exame anatomopatológico desde janeiro de 2010 permitiu a exclusão dos casos prevalentes³⁰. Foi realizada correspondência entre os campos das duas fontes do SISMAMA (municipal e nacional). Posteriormente, as bases foram unidas por ano e removidas as duplicidades.

A base de referência foi extraída do módulo “mamografia” do SISMAMA referente ao segundo semestre de 2010. Foram elegíveis para o estudo os registros das mulheres residentes no município do Rio de Janeiro com mamografias de rastreamento suspeitas (BI-RADS[®] 4) ou altamente suspeitas (BI-RADS[®] 5) de malignidade³¹ em uma das mamas. Foram excluídos aqueles sem informação de nome da mãe e endereço. Como o tamanho dos campos difere entre as bases utilizadas, foram criados os campos “nome da mulher”, com 30 e 50 caracteres, e “nome da mãe”, com 30 e 45 caracteres. Após a seleção dos registros para a base de referência, os demais compuseram um arquivo denominado “mamografias de seguimento de 2010”, o qual foi utilizado como base de dados de comparação no relacionamento.

O Boletim de Produção Ambulatorial Individualizado (BPA-I) contém procedimentos diagnósticos para câncer de mama (exames anatomopatológicos por biópsia ou peça cirúrgica). Do campo “procedimento realizado” foram extraídos os exames “anatomopatológicos de mama” e “anatomopatológicos de congelamento/parafina exceto colo uterino e mama”, ambos por peça cirúrgica ou por biópsia. O procedimento de congelamento/parafina foi incluído porque o sistema não critica sua Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde (CID). A APAC registra os procedimentos autorizados de quimioterapia, hormonioterapia e radioterapia. A partir do campo “procedimento principal”, foram extraídos os registros que admitem CID para câncer de mama e selecionados os registros que não eram de continuidade.

No SIH foram selecionados, a partir do campo “procedimento realizado”, os registros cirúrgicos: mastectomia — simples e radical com linfadenectomia axilar, segmentectomia/quadrantectomia/setorectomia, setorectomia/quadrantectomia com esvaziamento ganglionar, oncológicos ou não; e internação para radioterapia externa. A competência “agosto de 2011” não estava disponível e não foi possível recuperá-la.

Para o SIA e o SIH, foram selecionados os registros com CID-10 de câncer de mama (C50), neoplasia de mama *in situ* (D05), doença benigna de mama (D24), doença de comportamento incerto (D48.6) e doença do aparelho geniturinário (N60-N64), bem como os campos sem informação. Dos arquivos anuais do SIM foram selecionados os óbitos femininos e não fetais, excluídos aqueles sem dados de nome da mulher e endereço.

Em cada base foi inserido um campo de registro com uma sequência de letras seguidas do ano e da ordem numérica do registro na base, identificando o exame. Em todas as bases foram retirados os acentos e o cedilha. Foram selecionados os registros de sexo feminino e realizou-se revisão manual dos registros de sexo masculino, de modo a identificar possíveis erros de codificação. Realizaram-se alterações para os registros nos quais o nome não era dúbio quanto ao gênero.

As bases anuais de cada SIS, exceto a do módulo de mamografia do SISMAMA 2010, foram unificadas, e as duplicidades residuais, removidas. Os campos em branco e com preenchimento inadequado (números repetidos, caracteres inferiores/superiores ao estabelecido) das variáveis Cadastro de Pessoa Física (CPF) e CNS foram preenchidos com um código criado para referenciar a base, evitando, no relacionamento, formar pares com campos em branco e otimizando o seu processamento de formação.

As análises foram realizadas utilizando o programa R versão 3.1.1³¹.

RELACIONAMENTO ENTRE AS BASES DE DADOS

Utilizou-se, para relacionar, o programa RecLink[®] versão 3.1³², seguindo as etapas padronização, blocagem e pareamento. No pareamento, usaram-se os parâmetros propostos no manual do RecLink^{®32} e os campos “nome”, “nome da mãe” e “data de nascimento” (DT). Os campos “CNS” e “CPF” foram utilizados como chave de blocagem para os passos 1 e 2, respectivamente. O passo 3 incluiu os campos: DT, códigos *soundex* do primeiro e último nome da mulher (PN e UN) e da mãe (PM e UM). Nos demais 12 passos, as estratégias foram progressivamente menos restritas. As bases do SIA e do SIM não possuem o campo “CPF”.

A base de referência foi relacionada inicialmente com as mamografias de 2010, para identificar mulheres que repetiram a mamografia dentro do semestre. O primeiro exame realizado permaneceu na base de referência e os demais foram incluídos em um arquivo denominado “mamografia de seguimento de 2010”. Posteriormente, a base de referência foi relacionada às demais bases, sendo avaliados os pares com escore acima de zero. A cada passo foram mantidos apenas os registros não classificados como pares verdadeiros nas bases de comparação.

A classificação dos pares como verdadeiros, realizada por revisão manual, seguiu os critérios: nome, DT e nome da mãe. Não foram considerados para avaliação de pares os registros em que estavam ausentes dois dos três campos acima, exceto quando o nome da mulher e o endereço (rua, número e bairro) eram idênticos. Foram considerados como pares: situações em que, apesar de o nome da mãe estar totalmente diferente/ausente, os campos “nome da mulher”, “DT” e “endereço” foram iguais/semelhantes; situações em que nome da mãe e/ou endereço ausente(s)/diferente(s), mas o nome foi avaliado como raro (ex.: nomes estrangeiros) e as DT eram iguais. Definiram-se como semelhantes nome e sobrenome raros ou abreviação/ausência do nome do meio ou do sobrenome.

Como os registros anatomopatológicos de mama documentados no SISMAMA devem corresponder aos informados no BPA-I²⁵, realizou-se relacionamento entre essas bases para identificar exames que não estavam no SISMAMA. Como o BPA-I possui poucos campos, inclusive sem o domínio “nome da mãe”, adotou-se a estratégia de usar na chave de blocagem o “Cadastro Nacional de Estabelecimento de Saúde (CNES)”, “data de atendimento/resultado” e “CID-10 do exame” associados aos demais campos disponíveis nessa base (CNS, PN, UN e DT). A CID-10 foi utilizada no pareamento com o nome e a DT. Após esse relacionamento, a base de referência foi associada com os registros do BPA-I que não formaram pares com os anatomopatológicos do SISMAMA, usando-se como critérios de avaliação dos pares a raridade do nome e a regra de que o exame anatomopatológico deveria ter data posterior à da mamografia.

Calculou-se a proporção de incompletude dos campos “CNS” e “CPF” em cada SIS estudado, bem como a respectiva variação percentual (VP) entre 2010 e 2012. Foi apresentado o número de pares identificados entre os exames anatomopatológicos registrados na base do SISMAMA e no BPA-I, e calculou-se o percentual de tipos de procedimentos que não formaram par. Foi apresentado o número de pares formados entre a base de referência e as demais bases, e respectivos número e percentual dos classificados como verdadeiros em cada passo. Calculou-se a sensibilidade da estratégia de relacionamento probabilístico utilizando apenas os registros com CNS e/ou CPF das bases estudadas, tendo como padrão-ouro os pares de registros identificados nos passos 1 e 2.

O estudo foi aprovado pelos Comitês de Ética em Pesquisa do Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro (Parecer nº 1.105.945), do município do Rio de Janeiro (Parecer nº 1.162.544) e do INCA (Parecer nº 1.139.738).

RESULTADOS

A base do SISMAMA do município do Rio de Janeiro possuía 5.565 mamografias e 146 exames anatomopatológicos a mais que a base nacional desse sistema. A revisão do campo “sexo” permitiu identificar 6 registros com codificação errada na mamografia (SISMAMA), 5 no anatomopatológico (SISMAMA), 59 no BPA-I, 40 no SIH, 576 na APAC e 4 no SIM.

A base de referência ficou com 233 mulheres (Figura 1). As mamografias de seguimento, após o processo de limpeza, totalizaram 11.938, em 2010, e 136.382, no período

de 2011 a 2012. A base com menor quantidade de registros foi a de anatomopatológicos do SISMAMA (n = 2.598), seguida do SIH (n = 5.685), dos anatomopatológicos do BPA-I (n = 12.098), da APAC (n = 70.893) e do SIM (n = 147.853).

O campo “CNS” apresentou redução na proporção de incompletude em todos os sistemas, sendo mais expressiva no SIH, passando de 30,6%, no segundo semestre de 2010, para 3,6%, em 2012 (VP = 88,2%). No entanto, em 2012, esse percentual ainda era elevado no SISMAMA (mamografia: 70,6%; anatomopatológico: 88,5%), no SIM (98,5%) e no BPA-I (69,7%). Na APAC tal preenchimento é obrigatório. O campo “CPF” apresentou redução na proporção de incompletude no SIH e nos registros anatomopatológicos do SISMAMA (VP = 34,6 e 6,6%, respectivamente), mas ainda foi elevado no último ano (34,6 e 93,4%, respectivamente). Houve, entretanto, aumento na incompletude desse campo nas mamografias (VP = 17,3%). Os demais sistemas não possuem tal campo (Tabela 1).

No relacionamento entre as bases do BPA-I e do SISMAMA, 99,8% (2.276/2.280) dos pares formados foram classificados como verdadeiros. Dentre os registros não identificados no BPA-I, 546 eram procedimentos específicos para mama (biópsia e peça cirúrgica), correspondendo a uma perda de 19,3% no SISMAMA, já que, nesse sistema, não se espera que o procedimento por “peça cirúrgica ou biópsia exceto colo e mama” seja registrado (Tabela 2).

A Tabela 3 apresenta os pares classificados como verdadeiros no processo de relacionamento. Nas mamografias do segundo semestre de 2010 foram identificadas 9 mamografias de seguimento e 220 para os anos posteriores. Para os exames anatomopatológicos, foram encontrados 30 no SISMAMA e 7 no BPA-I. Em relação ao tratamento, foram identificadas 70 cirurgias no SIH e 455 na APAC. No SIM foram localizados 20 registros.

Tabela 1. Percentual de Cartão Nacional de Saúde e Cadastro de Pessoa Física não preenchido, por base^a e ano, e variação percentual entre 2010 e 2012, município do Rio de Janeiro.

Bases	2010			2011			2012			VP 2010–2012	
	n	% sem inform.		n	% sem inform.		n	% sem inform.			
		CNS	CPF		CNS	CPF		CNS	CPF	CNS	CPF
Mamografia ^b	12.171	82,6	80,7	50.300	80,6	88,7	86.094	70,6	94,7	14,5	-17,3
Anatomopatológico ^b	380	99,7	100,0	1.037	100,0	92,4	1.193	88,5	93,4	11,2	6,6
BPA-I (SIA)	454	100,0	-	6.389	85,5	-	5.255	69,7	-	30,3	-
SIH/SUS	1.029	30,6	52,9	1.874	34,9	49,8	2.818	3,6	34,6	88,2	34,6
APAC (SIA)	13.145	0,0	-	28.515	0,0	-	29.245	0,0	-	0,0	-
SIM	29.539	99,5	-	59.155	99,3	-	59.159	98,5	-	1,0	-

^aApós retirada de registros duplicado no ano; ^bexames da base do Sistema de Informação do Controle do Câncer de Mama (SISMAMA); BPA-I: Boletim de Produção Ambulatorial Individualizado; SIA: Sistema de Informação Ambulatorial; SIH: Sistema de Informação Hospitalar; SUS: Sistema Único de Saúde; APAC: Autorização de Procedimentos Ambulatoriais de Alta Complexidade; SIM: Sistema de Informação sobre Mortalidade.

Na base de referência, 33,0% dos registros tinham o campo “CNS” preenchido, e 44,6%, o CPF. A utilização desses campos como chave de blocagem permitiu identificar, dentre os pares classificados como verdadeiros: 47,3% dos pares das mamografias de seguimento (2011-2012), 41,4% dos pares do SIH e 45,5% dos pares da APAC. Não foi encontrado, com essas chaves, nenhum par verdadeiro com a base das mamografias de seguimento do segundo semestre de 2010 e com os exames anatomopatológicos do SISMAMA. Os passos 3 e 4 também apresentaram elevada proporção de identificação de pares verdadeiros: O relacionamento com as mamografias de seguimento do segundo semestre de 2010 e com os exames anatomopatológicos do SISMAMA, nos quais não houve formação de pares nos passos anteriores, identificou 88,9 e 86,7%, respectivamente. A proporção de pares classificados como verdadeiros com a utilização dos passos 1, 2 e 3 foi de 91,4% para as mamografias de seguimento (2011–2012) e de 92,3% com a base da APAC.

Todos os pares de registros identificados com CNS e CPF (passos 1 e 2, respectivamente) foram detectados na análise complementar, que empregou as estratégias de relacionamento probabilístico utilizadas (sensibilidade = 100%). A Tabela 4 apresenta o número e o percentual de pares detectados em cada passo nesta análise. O passo 3, que apresentou estratégia de blocagem mais restritiva, detectou mais de 90% dos pares verdadeiros em todas as bases avaliadas.

DISCUSSÃO

A proporção de mamografias alteradas (1,9%, 233 / 12.721) encontrada na base de referência foi considerada elevada, já que, para o Brasil, no mesmo ano, esse percentual foi de 1,4%, próximo da avaliação sobre implementação do SISMAMA realizada por Passman et al.³³.

A diferença a mais na base municipal do SISMAMA, apesar de ter dados só a partir de maio de 2011, decorre de esta ter todos os exames realizados no município do Rio de Janeiro,

Tabela 2. Registros anatomopatológico da base do Boletim de Produção Ambulatorial Individualizado antes e depois do relacionamento probabilístico, de julho de 2010 a dezembro de 2012, município do Rio de Janeiro.

Procedimento anatomopatológico	Antes do relacionamento		Pós relacionamento			
			Identificados		Não identificados	
	n	%	n	%	n	%
Peça cirúrgica (mama)	873	7,2	723	82,8	150	17,2
Biópsia (mama)	1.949	16,1	1.553	79,7	396	20,3
Peça cirúrgica ou biópsia (exceto mama ou colo)	9.276	76,7	0	0,0	9.276	100,0
Total	12.098	100,0	2.276	18,8	9.822	81,2

Tabela 3. Pares formados (n1), classificados como verdadeiros (n2) e percentual por passos do relacionamento e sistema de informação.

Passos e chaves de blocagens	Mamografia 2010			Mamografia 2011-12			Anatomop. 2010-12			BPA-I 2010-12			SIH 2010-12			APAC 2010-12			SIM 2010-12		
	n1	n2	%	n1	n2	%		n2	%	n1	n2	%	n1	n2	%	n1	n2	%	n1	n2	%
CNS				83	83	37,7							24	24	34,3	208	208	45,7			
CPF				21	21	9,5							5	5	7,1						
PN + UN + DT + PM + UM	8	8	88,9	97	97	44,1	22	22	73,3							215	215	47,3			
PN + UN + DT				9	8	3,6	4	4	13,3	5	5	71,4	36	36	51,4	20	20	4,4	19	19	95,0
PN + DT + PM + UM	1	1	11,1	2	2	0,9	1	1	3,3							8	8	1,8			
PN + DT + UM																					
PN + DT + PM																					
PN + DT				1						11	0					1	1	0,2	1	1	5
PN + UN + PM + UM				7	6	2,7	2	2	6,7				2	2	2,9	2	2	0,4			
PN + UN + PM				5									1	0	0	3	1	0,2	1		
PN + UN				11	2	0,9				37	1	14,3	3	1	1,4	2			15		
PN + PM + UM				5															2		
PN + UM				6															10		
UN + DT				2	1	0,5	1	1	3,3	2			2	2	2,9				1		
DT				5						101	1	14,3							9		
Total	9	9	100,0	254	220	100,0	30	30	100,0	156	7	100,0	73	70	100,0	459	455	100,0	58	20	100,0

BPA-I: Boletim de Produção Ambulatorial Individualizado; SIH: Sistema de Informação Hospitalar; APAC: Autorização de Procedimentos Ambulatoriais de Alta Complexidade; SIM: Sistema de Informação sobre Mortalidade; CNS: Cartão Nacional de Saúde; CPF: Cadastro Nacional de Pessoa Física; PN: código *soundex* do primeiro nome da mulher; UN: código *soundex* do último nome da mulher; DT: data de nascimento; PM: código *soundex* do primeiro nome da mãe; UM: código *soundex* do último nome da mãe.

enquanto a base nacional possui apenas os de residentes no município. Adicionalmente, também deve-se considerar a possibilidade de perda no fluxo de encaminhamento das bases entre os níveis municipal, estadual e nacional²⁵.

A ausência da base do SIH da competência de agosto de 2011 e a falta de registros no SISMAMA também constituíram uma limitação deste estudo, já que algumas investigações histopatológicas possivelmente não foram incluídas.

A identificação de registros femininos codificados como masculinos indica que a restrição do campo “sexo” pode excluir a formação de pares verdadeiros no processo de relacionamento. Uma alternativa à revisão manual seria a ampliação do número de passos no relacionamento, com inclusão desse campo em alguns passos junto com outros campos na blocagem.

Tabela 4. Número e percentual de pares identificados no relacionamento probabilístico a partir das bases com registro de Cartão Nacional de Saúde e/ou Cadastro de Pessoa Física, por passo e sistema de informação.

Passos	Campos de blocagem	Sistemas de Informação					
		SISMAMA 2010-2012		SIH 2010-2012		APAC 2010-2012	
		n	%	n	%	n	%
3	PN + UN + DT + PM + UM	98	94,2	28	96,6	198	95,2
4	PN + UN + DT	1	1,0	0	0,0	0	0,0
5	PN + DT + PM + UM	1	1,0	0	0,0	5	2,4
6	PN + DT + UM	0	0,0	0	0,0	0	0,0
7	PN + DT + PM	0	0,0	0	0,0	0	0,0
8	PN + DT	0	0,0	0	0,0	0	0,0
9	PN + UN + PM + UM	4	3,8	1	3,4	5	2,4
10	PN + UN + PM	0	0,0	0	0,0	0	0,0
11	PN + UN	0	0,0	0	0,0	0	0,0
12	PN + PM + UM	0	0,0	0	0,0	0	0,0
13	PN + UM	0	0,0	0	0,0	0	0,0
14	UN + DT	0	0,0	0	0,0	0	0,0
15	DT	0	0,0	0	0,0	0	0,0
Total		104	100,0	29	100,0	208	100,0

SISMAMA: Sistema de Informação do Controle do Câncer de Mama; SIH: Sistema de Informação Hospitalar; APAC: Autorização de Procedimentos Ambulatoriais de Alta Complexidade; PN: código *soundex* do primeiro nome da mulher; UN: código *soundex* do último nome da mulher; DT: data de nascimento; PM: código *soundex* do primeiro nome da mãe; UM: código *soundex* do último nome da mãe.

Foram identificados preenchimentos incorretos no CPF (quantidade de dígitos maior ou menor que 11) e repetições do mesmo dígito, dada a ausência de rotina de crítica nos sistemas. Problema com o preenchimento do CPF foi identificado em outro estudo de relacionamento utilizando os registros nacionais da APAC e do SIH¹².

Os quatro passos iniciais permitiram identificar a maioria dos pares verdadeiros. Nenhum exame anatomopatológico, no entanto, foi identificado nos dois passos iniciais, devido à incompletude desses campos na base de dados histopatológicos do SISMAMA e do BPA-I.

A perda de exames anatomopatológicos registrados no SISMAMA foi considerada baixa (19,3%), quando comparada com a base do BPA-I; e só foram identificados pares de registros com procedimentos em peça cirúrgica ou biópsia não específicos da mama, sinalizando adequação da estratégia de relacionamento utilizada.

A estratégia de relacionamento adotada foi efetiva, já que a sensibilidade foi de 100% na análise restrita aos registros com CNS e CPF preenchidos. Adicionalmente, observou-se que essa análise pode contribuir para seleção e/ou priorização das estratégias de blocagem a serem utilizadas. A chave de blocagem com o PN, UM, PM e UM (passo 9), por exemplo, identificou pares, na maioria das bases relacionadas, em que os passos anteriores (passos 6 a 8) não foram capazes de identificar.

Este estudo foi capaz de mostrar a importância dos SIS para avaliar as ações de controle do câncer de mama. Cada um dos sistemas de informação do SUS registra uma etapa da atenção, e a unificação de suas bases de dados permite reconstruir o seguimento das mulheres com mamografia alterada enquanto não se dispõe de um identificador único em todos os SIS do país.

CONCLUSÃO

O estudo mostrou que a utilização do CNS e do CPF permitiu identificar muitos pares, mesmo com ausência de crítica, e realizar o relacionamento entre bancos com poucos campos de identificação e variadas chaves.

A qualificação contínua e progressiva dos SIS é de fundamental importância para avaliar as ações e os programas em saúde. Além de um identificador único obrigatório, a padronização de campos mínimos para qualificação dos registros do indivíduo, como a inclusão de “nome da mãe” no BPA-I, permitirá aumentar a confiabilidade dos resultados. A utilização do BPA-I como fonte de informação, apesar dos poucos campos de identificação, permitiu encontrar alguns registros. A inclusão de um sistema de crítica em campos que fogem ao regramento, como CPF e CNS, não permitindo o registro de dígitos diferentes do estabelecido, otimizaria a utilização das ferramentas de relacionamento. Essa estratégia é um passo muito importante para a vigilância epidemiológica realizada pelos núcleos de informação das Secretarias Municipais e Estaduais de Saúde, responsáveis pelas bases de dados. Sua aplicação poderia contribuir para a melhoria continuada da qualidade desses sistemas e aprimorar a avaliação dos vários programas de atenção em saúde.

REFERÊNCIAS

1. Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Regulação, Avaliação e Controle. Coordenação Geral de Sistemas de Informação. Manual técnico operacional SIA/SUS. Sistema de Informações Ambulatoriais. Brasília: Ministério da Saúde; 2009. 69 p.
2. Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Regulação, Avaliação e Controle. Coordenação Geral de Sistemas de Informação. Manual técnico operacional do Sistema de Informações Hospitalares. Brasília: Ministério da Saúde; 2010. 119 p.
3. Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Regulação, Avaliação e Controle. Coordenação Geral de Sistemas de Informação. APAC. Autorização de Procedimento Ambulatorial. Manual de operação do sistema. Versão 1.1. Brasília: Ministério da Saúde; 2013.
4. Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Regulação, Avaliação e Controle. Sistemas de Informação da Atenção à Saúde. Brasília: Ministério da Saúde; 2015. 166 p.
5. Brasil. Ministério da Saúde. Portaria GM nº 940, de 28 de abril de 2011. Regulamenta o Sistema Cartão Nacional de Saúde (Sistema Cartão). Brasília: Ministério da Saúde; 2011.
6. Brasil. Ministério da Saúde. Secretaria de Atenção à Saúde. Portaria Conjunta nº 2, de 15 de março de 2012. Dispõe acerca do preenchimento do número do Cartão Nacional de Saúde do usuário no registro dos procedimentos ambulatoriais e hospitalares. Brasília: Ministério da Saúde; 2012.
7. Brasil. Ministério da Saúde. Portaria nº 763, de 20 de julho de 2011. Dispõe acerca do preenchimento do número do Cartão Nacional de Saúde do usuário no registro dos procedimentos ambulatoriais e hospitalares. Brasília: Ministério da Saúde; 2011.
8. Brasil. Ministério da Saúde. Conselho Nacional de Secretários da Saúde. Nota técnica 22/2011. Proposta de consolidação do Cartão Nacional de Saúde – Cartão SUS. Atualização das notas técnicas 29/2010 e 32/2010 de 06/08 e 13/09/2010. Brasília: CONASS/progestores; 2011.
9. Magalhães VCL, Costa MCE, Pinheiro RS. Perfil do atendimento no SUS às mulheres com câncer de mama atendidas na cidade do Rio de Janeiro: relacionando os sistemas de informações SIH e APAC-SIA. *Cad Saúde Coletiva*. 2006; 14(2): 375-98.
10. Machado JP, Silveira DP, Santos IS, Piovesan MF, Albuquerque C. Aplicação da metodologia de relacionamento probabilístico de base de dados para a identificação de óbitos em estudos epidemiológicos. *Rev Bras Epidemiol*. 2008; 11: 43-54. <http://dx.doi.org/10.1590/S1415-790X2008000100004>
11. Silveira DP, Artmann E. Acurácia em métodos de relacionamento probabilístico de bases de dados em saúde: revisão sistemática. *Rev Saude Pública*. 2009; 43(5): 875-82. <http://dx.doi.org/10.1590/S0034-89102009005000060>
12. Queiroz OV, Guerra Júnior AA, Machado CJ, Andrade EIG, Meira Júnior W, Acurcio FA, et al. Relacionamento de registros de grandes bases de dados: estimativa de parâmetros e validação dos resultados, aplicados ao relacionamento dos registros de autorizações de procedimentos ambulatoriais de alta complexidade com os registros de sistemas de informações hospitalares. *Cad Saúde Coletiva*. 2010; 18(2): 298-308.
13. Gomes Jr SC dos S, Martino R, Almeida RT. Rotinas de integração das tabelas do sistema de autorização de procedimentos de alta complexidade em oncologia do Sistema Único de Saúde. *Cad Saúde Coletiva*. 2003; 11(2): 231-54.
14. Souza RC, Freire SM. Integração de dados ambulatoriais de quimioterapia e radioterapia registrados nas bases de dados do SUS. In: Congresso Brasileiro de Engenharia Biomédica, 24., 2014. Anais; 2014.
15. Souza RC, Freire SM, Almeida RT. Sistema de informação para integrar os dados da assistência oncológica ambulatorial do Sistema Único de Saúde. *Cad Saúde Pública*. 2010 Jun; 26(6): 1131-40. <http://dx.doi.org/10.1590/S0102-311X2010000600007>
16. Teixeira CLS, Bloch KV, Klein CH, Coeli CM. Método de relacionamento de bancos de dados do Sistema de Informações sobre Mortalidade (SIM) e das autorizações de internação hospitalar (BDAIH) no Sistema Único de Saúde (SUS), na investigação de óbitos de causa mal-definida no Estado do Rio de Janeiro, Brasil, 1998. *Epidemiol Serviços Saúde*. 2006; 15(1): 47-57. <http://dx.doi.org/10.5123/S1679-49742006000100004>
17. Fonseca MGP, Coeli CM, Lucena FFA, Veloso VG, Carvalho MS. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. *Cad Saúde Pública*. 2010 Jul; 26(7): 1431-8. <http://dx.doi.org/10.1590/S0102-311X2010000700022>
18. Migowski A, Chaves RBM, Coeli CM, Ribeiro ALP, Tura BR, Kuschnir MCC, et al. Acurácia do relacionamento probabilístico na avaliação da alta complexidade em cardiologia. *Rev Saúde Pública*. 2011; 45(2): 269-75. <http://dx.doi.org/10.1590/S0034-89102011005000012>

19. Suzuki KMF. O uso de método de relacionamento de dados (record linkage) para integração de informação em sistemas heterogêneos de saúde: estudo de aplicabilidade entre níveis primário e terciário [tese]. Ribeirão Preto: Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo; 2012.
20. Girianelli VR, Thuler LCS, Silva GA. Qualidade do Sistema de Informação do Câncer do Colo do Útero no estado do Rio de Janeiro. *Rev Saúde Pública*. 2009; 43(4): 580-8. <http://dx.doi.org/10.1590/S0034-89102009005000043>
21. 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. <http://dx.doi.org/10.1590/S1519-38292012000300004>
22. Fávero JL, Cerqueira ACB, Fregona G, Prado TN, Werner RCD, Maciel ELN. Prevalência de tuberculose em profissionais da área de enfermagem obtida por método de relacionamento de banco de dados, 2000 a 2008, Espírito santo, Brasil. *Rev Bras Pesq Saúde*. 2012; 14(2): 31-7.
23. Girianelli VR. Fatores Associados ao Risco de Progressão para Câncer do Colo do Útero ou suas Lesões Precursoras em Mulheres com Exame de Papanicolaou Negativo: Um Estudo de Três Anos de Seguimento. Rio de Janeiro: Instituto Nacional de Câncer José Alencar Gomes da Silva; 2008.
24. Coeli CM. A qualidade do linkage de dados precisa de mais atenção. *Cad Saúde Pública*. 2015 Jul; 31(7): 1349-50. <http://dx.doi.org/10.1590/0102-311XED010715>
25. Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação Geral de Ações Estratégicas. Sistemas de Informação do Controle do Câncer de Mama e do Colo do Útero – Manual Gerencial. Rio de Janeiro: CEDC; 2011. 116 p.
26. Freire SM, Almeida RT, Cabral MDB, Bastos EA, Souza RC, Silva MGP. A record linkage process of a cervical cancer screening database. *Comput Methods Programs Biomed*. 2012; 108: 90-101. <https://doi.org/10.1016/j.cmpb.2012.01.007>
27. Cabral MDB. Proposta de relacionamento probabilístico dos registros da base de dados do programa de rastreamento do câncer do colo do útero [tese]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2010.
28. Brasil. Ministério da Saúde. Definir como sistema de informação oficial do Ministério da Saúde, a ser utilizado para o fornecimento dos dados informatizados dos procedimentos relacionados ao rastreamento e a confirmação diagnóstica do câncer de mama, o Sistema de Informação do Controle do Câncer de Mama (SISMAMA). Portaria SAS n° 779 de 31 de dezembro de 2008. Brasília: Ministério da Saúde; 2008.
29. Brasil. Ministério da Saúde. Portaria GM n° 3.394 de 30 de dezembro de 2013. Institui o Sistema de Informação de Câncer (SISCAN) no âmbito do Sistema Único de Saúde (SUS). Brasília: Ministério da Saúde; 2013.
30. Tomazelli J. Avaliação das ações de detecção precoce do câncer de mama no Brasil: uma análise com base nos sistemas de informação em saúde [tese]. Rio de Janeiro: Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro; 2016.
31. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2013 [citado em 31 jul. 2015]. Disponível em: <http://www.R-project.org/>
32. Camargo Jr KR, Coeli CM. ReLink. Rio de Janeiro; 2007.
33. Passman LJ, Farias AM, Tomazelli JG, Abreu DM, Dias MB, Assis M, et al. SISMAMA: implementation of an information system for breast cancer early detection programs in Brazil. *Breast*. 2011; 20 (Suppl 2): S35-9. <https://doi.org/10.1016/j.breast.2011.02.001>

Recebido em: 14/09/2016

Versão final apresentada em: 25/08/2017

Aprovado em: 01/09/2017

