

Vania Reis Girianelli<sup>I,II</sup>

Luiz Claudio Santos Thuler<sup>I,III</sup>

Gulnar Azevedo e Silva<sup>I</sup>

# Qualidade do Sistema de Informação do Câncer do Colo do Útero no estado do Rio de Janeiro

## Quality of Cervical Cancer Data System in the State of Rio de Janeiro, Southeastern Brazil

---

### RESUMO

**OBJETIVO:** Avaliar a qualidade do Sistema de Informação do Câncer do Colo do Útero (Siscolo).

**MÉTODOS:** Estudo descritivo sobre a completitude, validade e sensibilidade dos dados no Siscolo no estado do Rio de Janeiro, com base no seguimento de uma coorte de 2.024 mulheres entre 2002 e 2006. As participantes eram residentes em comunidades assistidas pela Estratégia Saúde da Família nos municípios de Duque de Caxias e Nova Iguaçu (RJ). As duas bases de dados do Siscolo, referentes aos exames citopatológicos e aos exames confirmatórios (colposcopia e histopatologia), foram comparadas a dados obtidos em uma base de referência de pesquisa e prontuários médicos. O gráfico de Bland-Altman foi utilizado para analisar as variáveis contínuas. Para o relacionamento entre os bancos de dados foi utilizado o programa computacional Reclink.

**RESULTADOS:** A completitude do sistema foi excelente para os campos “nome da mãe” e “logradouro de residência”, boa para “bairro de residência” e péssima para “CEP” e “CPF”. Quanto à validade, a sensibilidade do campo “data da coleta” foi de 100% para os exames confirmatórios e de 70,3% para os exames citopatológicos. Já para o campo “resultados dos exames”, a sensibilidade foi de 100% em ambos os exames. A sensibilidade do sistema em identificar os exames citopatológicos foi de 77,4% (IC 95%: 75,0;80,0), enquanto para os exames confirmatórios (colposcopia e histopatologia) foi de 4,0% (IC 95%: 0,0;21,3).

**CONCLUSÕES:** Os dados do Siscolo foram considerados de boa qualidade, em particular para os campos relacionados aos exames citopatológicos. O uso dos dados de colposcopia e histopatologia não foi satisfatório devido ao seu escasso registro no sistema.

**DESCRITORES:** Neoplasias do Colo do Útero, prevenção & controle. Programas de Rastreamento. Sistemas de Informação. Informática em Saúde Pública. Epidemiologia Descritiva.

<sup>I</sup> Programa de Pós-Graduação em Atenção Oncológica. Instituto Nacional de Câncer. Rio de Janeiro, RJ, Brasil

<sup>II</sup> Departamento de Epidemiologia. Instituto de Medicina Social. Universidade Estadual do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

<sup>III</sup> Departamento de Medicina Geral. Universidade Federal do Estado do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

**Correspondência | Correspondence:**

Vania Reis Girianelli  
Instituto Nacional de Câncer  
Coordenação de Prevenção e Vigilância  
R. dos Inválidos, 212 - 3º andar – Centro  
20231-020 Rio de Janeiro, RJ, Brasil  
E-mail: vaniagirianelli@yahoo.com.br

---

## ABSTRACT

**OBJECTIVE:** To evaluate quality of a cervical cancer data system.

**METHODS:** Descriptive study on the completeness, validity, and sensitivity of data of the Cervical Cancer Data System (SISCOLO) in the State of Rio de Janeiro, Southeastern Brazil, based on the follow-up of a cohort of women, carried out between 2002 and 2006. The cohort consisted of 2,024 women living in communities served by the Family Health Program in the cities of Duque de Caxias and Nova Iguaçu. Two databases from the Siscolo, including cytopathology and confirmatory testing (colposcopy and histopathology) were compared to data from a reference database and medical records. The Bland-Altman plot was used to analyze continuous variables. The linkage between databases was analyzed using the RecLink software program.

**RESULTS:** The completeness of the data system was considered excellent with respect to the fields "mother's name" and "street address;" good for "district of residence" and poor for "zip code" and "individual taxpayer number". In regard to validity, sensitivity of the field "date of collection" was 100% and 70.3% for confirmatory and cytopathology tests, respectively, while sensitivity of the field "test results" was 100% for both tests. The sensitivity of the system to identifying cytopathology tests was 77.4% (95% CI: 75.0;80.0) whereas for confirmatory tests was 4.0% (95% CI: 0.0;21.3).

**CONCLUSIONS:** Data quality of SISCOLO was considered good, particularly for the fields related to cytopathology testing. The use of colposcopy and histopathology data was inadequate due to small number of cases registered in the System.

**DESCRIPTORS:** Uterine Cervical Neoplasms, prevention & control. Mass Screening. Information Systems. Public Health Informatics. Epidemiology, Descriptive.

---

## INTRODUÇÃO

O sistema de informação em saúde no Brasil é composto por diferentes subsistemas que compilam dados referentes às atividades setoriais. O Sistema de Informação sobre Mortalidade (SIM) foi o primeiro a ser desenvolvido no País, em 1976.<sup>8</sup> Na década de 1990, diversos sistemas de informação foram desenvolvidos para subsidiar o planejamento e avaliação dos serviços de saúde do Sistema Único de Saúde (SUS).

Os dados acumulados nesses sistemas representam uma importante fonte de informação para estudos epidemiológicos, possibilitando a redução de custo e tempo despendido nas pesquisas. Entretanto, um dos obstáculos para seu uso é a possibilidade de sua qualidade ser deficiente, com muitos dados faltantes e incorretos.

O acesso e as possibilidades de análise dos sistemas nacionais de informação em saúde têm avançado nos últimos anos. Em recente revisão sistemática da literatura,<sup>11</sup> foram identificadas 71 publicações de integração de bases de dados brasileiras referentes a 40 estudos. Destes, 70% foram classificados como estudos

epidemiológicos e, embora a maioria tenha referido algum aspecto relacionado à qualidade da informação, apenas 15% objetivaram avaliar a qualidade dos dados. Dentre os arquivos pesquisados, 75% eram referentes ao SIM e 57,5% ao Sistema de Informação sobre Nascidos Vivos (Sinasc). O programa computacional ReLink foi utilizado em 27,5% dos estudos para o relacionamento probabilístico entre as bases de dados.

Um dos sistemas de informação mais recentes, desenvolvido pelo Departamento de Informática do SUS (DATASUS) em parceria com o Instituto Nacional de Câncer (INCA), foi o Sistema de Informação do Câncer do Colo do Útero (Siscolo). Implantado em janeiro de 2000, este Sistema destina-se ao armazenamento de dados sobre identificação da mulher, informações demográficas, epidemiológicas e dos exames citopatológicos e histopatológicos realizados no SUS. O Siscolo tem sido aperfeiçoado continuamente, possibilitando que as coordenações dos programas municipais e estaduais acompanhem as mulheres com resultados dos exames alterados. Esse sistema foi elaborado

para subsidiar o pagamento desses exames pelo SUS e propiciar a avaliação dos serviços e programas de controle e assistência do câncer do colo do útero,<sup>a</sup> sendo também uma fonte potencial de informação em pesquisas. Em levantamento realizado em novembro de 2007 nas bases de dados da Biblioteca Virtual em Saúde, da BIREME, os três estudos que utilizaram esse sistema de informação tiveram como objetivo avaliar a qualidade dos exames citopatológicos processados nos laboratórios de citopatologia.<sup>5,6,13</sup>

O objetivo do presente estudo foi avaliar completude, validade e sensibilidade das informações da base de dados do Siscolo.

## MÉTODOS

As fontes de dados utilizadas foram as seguintes:

- Base de dados de referência – Constituiu-se de informações demográficas, epidemiológicas, clínicas e laboratoriais relativas a uma coorte de 2.024 mulheres, residentes em 13 comunidades assistidas pela Estratégia Saúde da Família (ESF) dos municípios de Duque de Caxias e Nova Iguaçu (RJ). Essas mulheres participaram de um estudo transversal<sup>4</sup> conduzido entre 2001 e 2002 e não apresentaram lesão de alto grau ou câncer do colo do útero (LA+) naquela ocasião.
- Prontuários médicos – Foram consultadas 13 unidades de saúde da ESF, nas quais haviam sido avaliadas inicialmente; três unidades de saúde de referência para confirmação diagnóstica; e o prontuário eletrônico do Instituto Nacional do Câncer (INCA), referência para tratamento de câncer no estado. Foram compilados dados sobre data e resultados dos exames citopatológicos, colposcópicos e histopatológicos realizados entre a data de entrada no estudo e dezembro de 2006, ou até a confirmação diagnóstica de lesão de alto grau ou câncer do colo do útero (LA+). Uma amostra sistemática correspondendo a 10% dos prontuários localizados em cada unidade de saúde foi reavaliada.
- Siscolo – Foram utilizadas as duas bases de dados desse sistema, referentes aos exames citopatológicos e aos exames confirmatórios (colposcopia e histopatologia) realizados no estado do Rio de Janeiro, disponibilizadas pelo Datasus, Ministério da Saúde, em maio de 2007. Para o período de janeiro de 2002 a maio de 2006, os dados foram fornecidos na versão 3.06. Para o período entre junho e dezembro de 2006 os dados foram

fornecidos na versão 4, que incorpora a nova nomenclatura brasileira para laudos cervicais.<sup>b</sup> Os campos analisados restringiram-se aos dados demográficos, resultados dos exames e identificação das mulheres. As bases de dados foram inspecionadas para identificar inconformidades e corrigi-las para evitar possíveis interferências no processo de relacionamento. Os registros dos campos “nome da mãe” e “bairro de residência” foram deixados em branco quando apresentavam palavras para designar sem informação ou apenas algarismos. No campo “nome da paciente” o nome do médico ou do enfermeiro estava indevidamente digitado após o nome da mulher em vários registros, sendo necessário apagar manualmente.

O programa Reclink versão 3 foi utilizado para relacionamento entre as bases de dados para identificar informações das mulheres da coorte de referência. Este programa baseia-se na técnica de relacionamento probabilístico, estimando o quanto é provável que um par de registros refira-se ao mesmo indivíduo. O relacionamento de registros é feito por meio da identificação dos campos em comum existentes nas bases de dados, sendo gerado um score a partir da probabilidade de concordância e discordância entre eles. Esse processo é realizado em três etapas.<sup>1,2</sup>

a) Padronização – estruturação dos campos das bases de dados, para prepará-las para o relacionamento e minimizar a ocorrência de erros. Os campos podem ser subdivididos e convertidos para terem a mesma estrutura. O programa também possibilita a exclusão de preposições, sinais de pontuação, acentos e outros símbolos.

b) Bloqueio – criação de blocos lógicos, constituídos por um ou mais campos, de forma a restringir o relacionamento apenas aos registros que apresentam o mesmo conteúdo nos respectivos campos selecionados. Sete estratégias de bloqueio foram sequencialmente empregadas para minimizar a perda de pares verdadeiros (Figura 1).

c) Pareamento – construção de escores para os diferentes pares obtidos numa determinada estratégia de bloqueio, a partir dos parâmetros especificados para os campos selecionados como identificadores, pois têm maior poder de discriminação.

O programa calcula os escores a partir da probabilidade de o identificador concordar entre dois registros, dado que se trata de potencial par verdadeiro, sensibilidade ( $m$ ), e em caso de potencial par falso, falso-positivo ( $u$ ). Ainda, é calculada a probabilidade de o identificador

<sup>a</sup> Ministério da Saúde. Secretaria de Assistência à Saúde. Instituto Nacional de Câncer. Viva mulher. Câncer do colo do útero: informações técnico-gerenciais e ações desenvolvidas. Rio de Janeiro; 2002.

<sup>b</sup> Ministério da Saúde. Secretaria de Atenção à Saúde. Instituto Nacional de Câncer. Coordenação de Prevenção e Vigilância. Nomenclatura brasileira para laudos cervicais e condutas preconizadas: recomendações para profissionais de saúde. Rio de Janeiro; 2006.

discordar, dado que se trata de potencial par verdadeiro, falso negativo ( $1 - m$ ), e em caso de potencial par falso, especificidade ( $1 - u$ ). A partir destas probabilidades, dois fatores de ponderação são gerados, um para concordância e outro para discordância. O fator de ponderação de concordância é calculado por meio do logaritmo de base 2 da razão de verossimilhança entre as probabilidades  $m$  e  $u$  e o de discordância para as demais ( $1 - m$ ) e ( $1 - u$ ). O escore total de determinado par é obtido a partir da soma dos fatores de ponderação atribuídos após a comparação de cada identificador. Os campos selecionados como identificadores, os parâmetros utilizados e os valores calculados para os fatores de ponderação são apresentados na Tabela 1.

As estratégias de pareamento utilizadas no processo de relacionamento e os respectivos escores máximo (concordância total para todos os identificadores) e mínimo (discordância em todos os identificadores) são apresentados na Figura 1. Quanto maior o número de identificadores, maior a variação dos escores. Nos passos de 1 a 4, a data de nascimento não foi incluída porque foi utilizada como estratégia de blocagem.

Os dois primeiros passos foram mais restritivos para que fossem formados poucos pares, com maior possibilidade de serem verdadeiros e, conseqüentemente, propiciassem sua inspeção. Nos passos 3 e 4, com o aumento quantitativo de pares formados no processo de relacionamento, foram inspecionados apenas aqueles que apresentaram escore acima de  $-4,0$ . Nos passos subseqüentes, com o aumento quantitativo expressivo de pares, foram inspecionados apenas aqueles que apresentaram escore positivo.

O processo de inspeção manual foi rigoroso, com a intenção de não classificar como verdadeiros pares que não representavam um mesmo indivíduo, e obedeceu aos seguintes critérios de classificação:

1. data de nascimento idêntica, com nome e nome da mãe idênticos, ou com o nome do meio abreviado ou um dos nomes do meio ausente;
2. nome e nome da mãe idênticos, com data de nascimento diferindo em até dois dígitos, ou troca do dia pelo mês;
3. nome ou nome da mãe incomum semelhante, com data de nascimento ou endereço idêntico;
4. nome, nome da mãe, ou data de nascimento diferindo ou sem informação numa das bases de dados, mas os demais campos contendo informações idênticas ou muito próximas com concordância em pelo menos três campos.

Os registros pareados num passo não foram incluídos nos passos subseqüentes, exceto para os registros da base de dados de referência, quando relacionados aos arquivos do Siscolo, devido à possibilidade de a mulher

realizar mais de um exame. Embora a versão 3 do programa Reclink apresente rotina para identificação de duplicidades, não foi utilizada.

A cada passo os registros pareados foram salvos em arquivos e posteriormente unidos para formar um arquivo único, por intermédio do programa Microsoft Office Access (2003). Em seguida, cada arquivo foi associado à base de dados original do Siscolo correspondente, utilizando o programa Reclink, agregando os campos relacionados aos resultados dos exames.

Para avaliar a qualidade dos dados do Siscolo, foram calculados os indicadores de completude, validade e sensibilidade, conforme proposto pelos *Centers for Disease Control and Prevention* (CDC).<sup>2</sup>

A completude dos campos foi mensurada pela proporção de registros preenchidos, sem informação ignorada, em um determinado campo. Este indicador foi interpretado como excelente quando o percentual de completude foi superior a 90%; bom, entre 70,1% e 90%; e péssimo, quando igual ou inferior a 70%, com base nos critérios utilizados por Mello Jorge et al.<sup>7</sup>

Para análise da validade dos campos do Siscolo foi calculada a sensibilidade, tendo como padrão-ouro os dados oriundos dos prontuários consultados (relativos aos exames) ou do estudo transversal (dados demográficos e de identificação). Adicionalmente, o gráfico de Bland-Altman foi utilizado para analisar o campo “data da coleta” por ser o mais adequado para avaliar a validade de variáveis contínuas conforme referido por Szklo & Nieto.<sup>12</sup>

A sensibilidade do sistema foi estimada considerando a proporção de exames registrados nos prontuários médicos das mulheres da coorte de referência, que foram identificados no Siscolo. Este indicador foi interpretado conforme critério proposto por Piper et al.<sup>10</sup> elevado quando superior a 90%; moderado, entre 70% e 90%; e baixo quando inferior a 70%.

Para os indicadores de validade dos campos e de sensibilidade do sistema também foram calculados os respectivos intervalos com 95% de confiança.

A pesquisa foi aprovada pela Comissão de Ética em Pesquisa do Instituto Nacional de Câncer (registro nº 074/06).

## RESULTADOS

A completude das bases de dados dos exames citopatológicos e confirmatórios (colposcopia e histopatologia) do Siscolo foi excelente para os campos “nome da mãe” (98,4% e 98,2% respectivamente) e “logradouro de residência” (98,0% e 98,3% respectivamente), e boa para o “bairro de residência” (84,5% e 89,8%, respectivamente). Já a completude dos campos

**Tabela 1.** Parâmetros utilizados e resultados dos fatores de ponderação por campo identificador. Estado do Rio de Janeiro, 2002-2006.

Campo identificador	Algoritmo	Parâmetro			Fator de ponderação	
		Sensibilidade (%)	Especificidade (%)	Proporção mínima de concordância (%)	Concordância	Discordância
Nome	Aproximado <sup>a</sup>	92	99	85	6,52	- 3,63
Nome da mãe	Aproximado <sup>a</sup>	92	99	85	6,52	- 3,63
Data de nascimento	Caractere <sup>b</sup>	90	95	65	4,17	- 3,25
Bairro de residência	Aproximado <sup>a</sup>	92	99	85	6,52	- 3,63

<sup>a</sup> Compara seqüências de caracteres com base no algoritmo distância de Levenshtein, variando de 1 para correspondência total a 0 para discordância total, indicado para campos com informação sobre nome.

<sup>b</sup> Compara seqüências de dígitos, ignorando separadores, para pares de dígitos na mesma posição, indicado para campos que trazem data completa.

“Código de endereçamento postal (CEP)” e “Cadastro de Pessoa Física (CPF)” foi péssima em ambas as bases de dados (Tabela 2). Uma vez que o campo “data de nascimento” não é obrigatório se a idade for informada, o sistema atribui uma data a partir da idade e compõe com “01/01” para dia e mês. Em 3,5% dos registros foi observada esta estrutura nas bases de dados. O campo “idade” é calculado pelo sistema de acordo com a data de nascimento. Registros com idade inferior a dez anos e superior a 89 anos corresponderam a 0,2% nas bases de dados. Os demais campos com dados demográficos, de identificação e relacionados aos dados dos exames são de preenchimento obrigatório e não apresentaram registros sem informação.

Foram inspecionados 19.801 pares no relacionamento entre a base de dados de referência e a base de dados de exames citopatológicos do Siscolo, e 556 pares com a base de dados dos exames confirmatórios, tendo sido classificados como pares verdadeiros 10,6% e 0,9%, respectivamente. A maioria dos pares classificados foi identificado no passo 1, correspondendo a 64,4% para a base dos exames citopatológicos e 60,0% para a base dos exames confirmatórios. O passo 2, embora possa expressar erro de preenchimento do campo código do município, também pode significar migração, ou declaração errônea por parte da mulher. Neste passo o percentual de pares formados foi 7,6% e 20,0%, respectivamente (Tabela 3).

Os passos iniciais (1 e 2) são mais restritivos, sendo classificados pares com escore muito baixo por apresentarem abreviaturas ou exclusão dos nomes do meio da mulher e/ou da mãe, ou omissão no nome da mãe. Nos passos menos restritivos, se houve pares com estas características, não puderam ser identificados por apresentarem escore abaixo do limiar estabelecido, ou não disporem de outros campos concordantes que possibilitassem a classificação. Adicionalmente, foram observados endereços e bairros concordantes com código do município discordante.

A validade do campo “código soundex do primeiro nome da mulher” (97,2%, IC 95%: 96,5; 97,9) e do “código soundex do último nome da mulher” (90,9%, IC 95%: 89,6;92,1) foi alta, considerando que todas as mulheres pertencentes à coorte de referência com registros na base dos exames citopatológicos do Siscolo tenham sido identificadas. Também apresentaram boa validade os campos relacionados ao município de residência (89,0%, IC 95%: 87,6;90,3), primeiro nome da mãe (83,9%, IC 95%: 82,4;85,4), último nome da mãe (74,8%, IC 95%: 72,9;76,7) e data de nascimento (81,1%, IC 95%: 79,4;82,7). O campo “código soundex do bairro de residência” foi o que apresentou menor validade (67,5%, IC 95%: 65,5;69,5), dentre os campos utilizados no processo de relacionamento. Na base dos exames confirmatórios, apenas cinco mulheres foram identificadas, mas o primeiro nome da mulher, primeiro nome da mãe e data de nascimento estavam corretos em todos os pares formados (Tabela 4).

No levantamento realizado nas unidades de saúde da ESF, foram localizados 1.147 prontuários (56,7%). Muitos foram perdidos devido a enchentes e ao arquivamento inadequado, outros foram descartados

**Tabela 2.** Completitude dos campos das bases de dados do Sistema de Informação do Câncer do Colo do Útero. Estado do Rio de Janeiro, 2002-2006.

Campos	Base de dados dos exames citopatológicos (%)	Base de dados dos exames confirmatórios (%)
Nome da mãe	98,4	98,2
Endereço de residência	98,0	98,3
Bairro de residência	84,5	89,8
CEP	30,0	30,8
CPF	0,5	2,0

CEP: Código de endereçamento postal

CPF: Cadastro de pessoa física

**Tabela 3.** Pares formados, pares classificados como verdadeiros e variação dos escores para cada passo do relacionamento probabilístico entre as bases de dados do Sistema de Informação do Câncer de Colo do Útero e da coorte de referência. Estado do Rio de Janeiro, 2002-2006.

Passos	Base de dados dos exames citopatológicos					Base de dados dos exames confirmatórios				
	Número de pares formados	Pares verdadeiros			Variação dos escores	Número de pares formados	Pares verdadeiros			Variação dos escores
Número	%	Frequência acumulada		Número		%	Frequência acumulada			
1	1415	1354	64,4	64,4	13,1 a -7,3	3	3	60,0	60,0	2,9 a -7,3
2	751	159	7,6	72,0	13,1 a -7,3	3	1	20,0	80,0	12,5
3	737	117	5,6	77,6	12,8 a -3,6	49	1	20,0	100,0	2,2
4	618	30	1,4	79,0	12,8 a 2,2	49	0	-	100,0	-
5	6118	308	14,6	93,6	16,7 a 0,0	40	0	-	100,0	-
6	4018	102	4,8	98,4	12,8 a 2,3	382	0	-	100,0	-
7	6144	33	1,6	100,0	22,9 a 2,1	30	0	-	100,0	-
Total	19801	2103	100,0	100,0	-	556	5	100,0	100,0	-

**Tabela 4.** Sensibilidade dos campos da base de dados dos exames citopatológicos e confirmatórios do Sistema de Informação do Câncer do Colo do Útero. Estado do Rio de Janeiro, 2002-2006.

Campo	Base de dados dos exames citopatológicos		Base de dados dos exames confirmatórios	
	n	Sensibilidade (IC 95%)	n	Sensibilidade (IC 95%)
Nome da mulher				
Código <i>soundex</i> do primeiro nome	2045	97,2 (96,5;97,9)	5	100,0 (100,0;100,0)
Código <i>soundex</i> do último nome	1911	90,9 (89,6;92,1)	4	80,0 (44,9;100,0)
Nome da mãe				
Código <i>soundex</i> do primeiro nome	1765	83,9 (82,4;85,4)	5	100,0 (100,0;100,0)
Código <i>soundex</i> do último nome	1573	74,8 (72,9;76,7)	4	80,0 (44,9;100,0)
Data de nascimento	1705	81,1 (79,4;82,7)	5	100,0 (100,0;100,0)
Código do IBGE do município de residência	1871	89,0 (87,6;90,3)	3	60,0 (17,1;100,0)
Código <i>soundex</i> do bairro de residência	1420	67,5 (65,5;69,5)	0	-

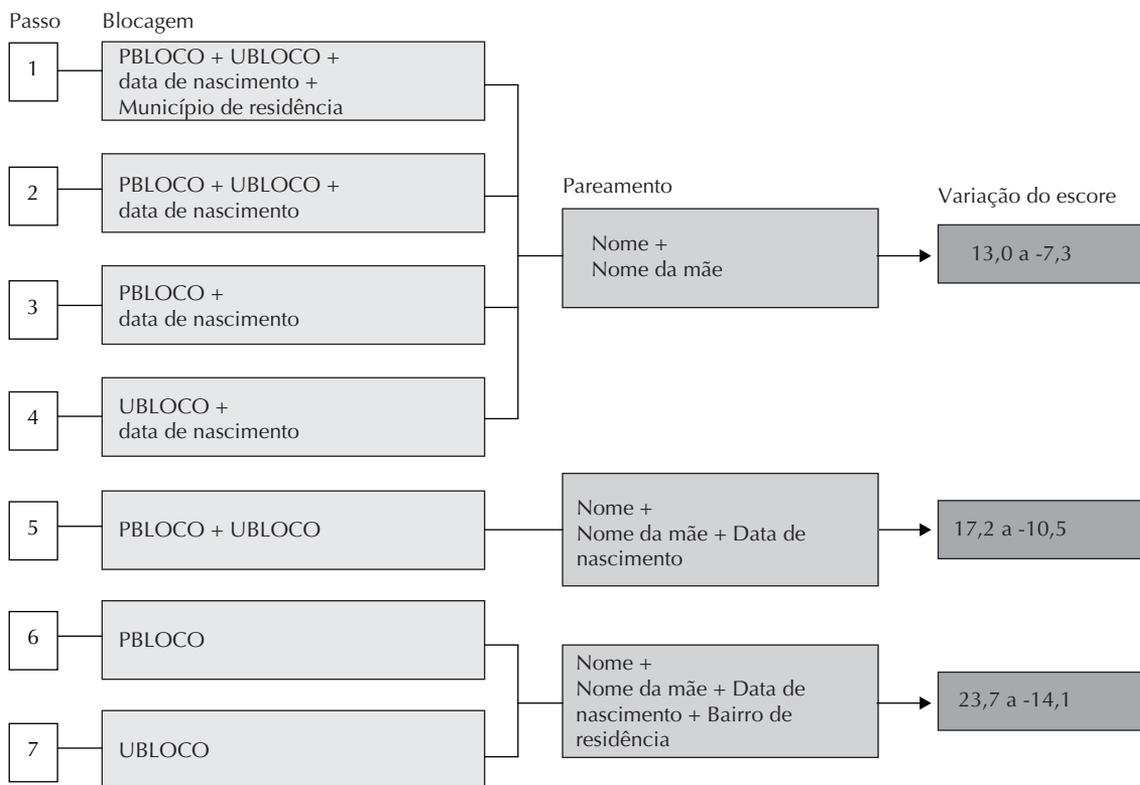
em função de estarem inativos por óbito, mudança de endereço ou cancelamento de cadastro da família. Dentre os prontuários localizados, apenas 636 (55,4%) tinham registros dos exames citopatológicos realizados. Os demais prontuários continham apenas registros de consultas, principalmente para controle de diabetes e hipertensão, ou estavam em branco. Alguns registros dos exames foram feitos apenas em livros de controle, inviabilizando sua identificação. Dentre os prontuários consultados em todas as unidades de saúde avaliadas foram identificados 1.113 exames citopatológicos.

Dos 2.103 exames citopatológicos identificados no Siscolo, seis eram repetidos, 31 foram realizados após confirmação diagnóstica de LA+, e por isto excluídos da análise, e 831 referiam-se a exames realizados em outras unidades de saúde, portanto não foram registrados nos prontuários das unidades avaliadas. Dos 1.235 exames restantes, 862 (69,8%) estavam registrados nos prontuários, incluindo 197 exames que

foram realizados em unidades de saúde não avaliadas. Dentre os 373 exames identificados apenas no Siscolo, 157 (42,1%) correspondiam aos exames das mulheres cujos prontuários não foram localizados nas unidades de saúde avaliadas. Os demais exames foram reavaliados e tiveram a classificação confirmada. Adicionalmente, os prontuários tinham registro de 251 exames não identificados no Siscolo.

A sensibilidade do sistema na identificação de exames citopatológicos foi de 77,4% (IC 95%: 75,0;80,0). Dentre os 2.317 exames citopatológicos localizados nas fontes de dados consultadas, 89,2% foram identificados no Siscolo, 48,0% nos prontuários consultados e 37,2% em ambas as fontes.

Nas três unidades de referência para confirmação diagnóstica, não foram localizados registros de exames colposcópicos ou histopatológicos das mulheres pertencentes à coorte de referência.

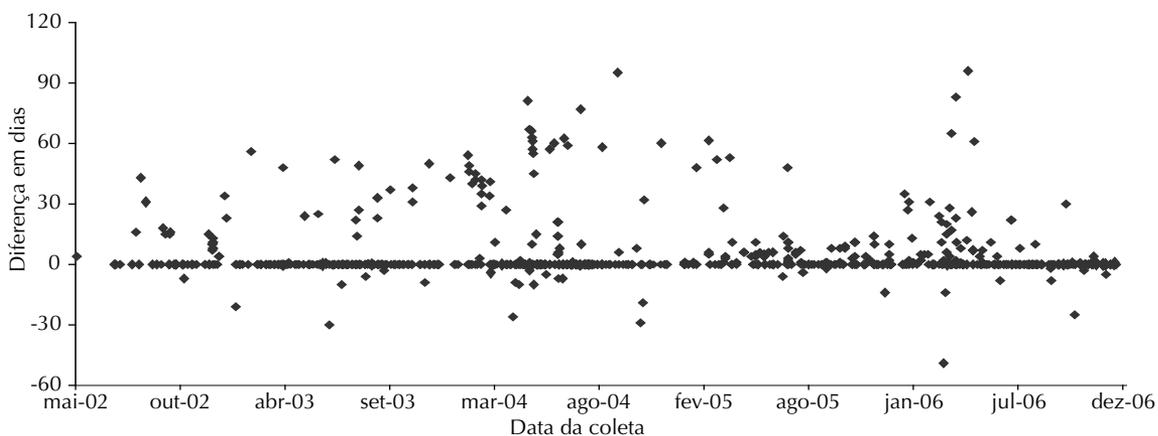


PBLOCO – Código soundex do primeiro nome, UBLOCO – Código soundex do último nome, Código Soundex - código fonético constituído por quatro dígitos, sendo o primeiro correspondente à primeira letra da palavra e os demais por dígitos numéricos, em que as vogais têm código 0 e as consoantes com sons similares o mesmo código (estratégia de padronização disponibilizada pelo Reclink).

**Figura 1.** Estratégias de blocagem e pareamento empregadas no processo de relacionamento entre registros e resultado da variação dos escores por passos sequenciais. Estado do Rio de Janeiro, 2002-2006.

No prontuário eletrônico do hospital do INCA foram identificadas 172 pacientes, das quais 80 tinham registros de exames colposcópicos e histopatológicos, correspondendo a 173 exames realizados no período

avaliado. Dessas 80 pacientes, apenas cinco foram identificadas no Siscolo, das quais duas realizaram exames que constavam em prontuário do hospital do INCA, enquanto as demais realizaram apenas exames



**Figura 2.** Gráfico de Bland-Altman da diferença entre a data de coleta dos exames citopatológicos do Sistema de Informação do Câncer do Colo do Útero e dos prontuários analisados. Estado do Rio de Janeiro, 2002-2006.

colposcópicas em unidades de saúde não avaliadas no estudo. A sensibilidade dessa base de dados foi muito baixa (4,0%, IC 95%: 0,0;21,3).

Dentre as mulheres pertencentes à coorte de referência, 1.251 (61,8%) tiveram exames citopatológicos, colposcópicos ou histopatológicos identificados nas fontes de dados consultadas.

Quanto à validade, a sensibilidade do campo “data da coleta” foi de 100% na base de dados dos exames confirmatórios e 70,3% na base de dados dos exames citopatológicos. Para o campo “resultados dos exames” a sensibilidade foi de 100% em ambas as bases de dados.

A diferença observada entre a data da coleta na base dos exames citopatológicos e no registro em prontuário foi de até 30 dias em 93,5% dos casos. Uma diferença acima de 60 dias ocorreu principalmente no segundo trimestre de 2004 e 2006; este último coincidiu com o período de mudança da versão do Siscolo (Figura 2).

## DISCUSSÃO

No presente estudo, foi possível identificar que o Siscolo continha informações de 89,2% dos 2.317 exames citopatológicos realizados pela coorte de mulheres de referência.

A utilização do programa Reclink foi indispensável frente à inexistência de um identificador único para o registro de informações em saúde. Espera-se que com a implantação do cartão SUS, processo em curso no País, o relacionamento entre registros se torne mais ágil e preciso.

A operacionalização do programa não apresentou maiores dificuldades, porém a seleção manual de registros foi muito trabalhosa e demorada, particularmente para as estratégias menos restritivas, que se tornaram necessárias devido à ausência de dados ou a erros de preenchimento ou de digitação. Apesar dos problemas observados, mais de 70% dos pares foram identificados nos dois primeiros passos, que foram os mais rápidos de serem realizados.

Quanto à qualidade dos dados do Siscolo, a completude verificada foi excelente para maioria dos campos analisados, bem como a validade para os resultados dos exames realizados.

A sensibilidade do sistema para a base de dados dos exames citopatológicos foi moderada (77,4%). No entanto, é provável que a sensibilidade seja maior do que a observada, pois apenas as informações disponíveis

nos prontuários foram consideradas. Além disso, é possível que a sensibilidade do Siscolo seja diferente em outros grupos populacionais ou em outras regiões do País, como ocorre com o Sistema de Informação sobre Mortalidade (SIM) e o Sistema de Informação sobre Nascidos Vivos (Sinasc). Esses sistemas, apesar de serem obrigatórios e terem maior tempo de implantação, ainda apresentam variação regional de cobertura.<sup>8,9</sup>

Já a sensibilidade para a base de dados dos exames confirmatórios (colposcopia e histopatologia) foi muito baixa (4,0%) e, portanto, ainda não é útil, visto que a maioria dos dados não foi digitada no sistema. Uma das possíveis razões para isto é que parte desses exames pode compor um conjunto de procedimentos prestados durante internação, em que os recursos financeiros são repassados por meio da Autorização de Internação Hospitalar. Faz-se necessário, portanto, estabelecer mecanismos que garantam que esta informação seja obrigatória no Siscolo de forma a permitir o seguimento dos casos rastreados que necessitam de acompanhamento e de tratamento. O módulo de seguimento do Siscolo, que tem sido aperfeiçoado, poderá propiciar melhorias na medida em que possibilita a geração de relatórios das mulheres com exames citopatológicos alterados cujo seguimento não foi concluído, permitindo que os gestores municipais e estaduais realmente o sistema com as informações oriundas de busca ativa realizada pelo nível local.

As informações dos exames de rastreamento do câncer do colo do útero e suas lesões precursoras nem sempre são acessíveis nas unidades de saúde onde é realizada a coleta do material, o que, de certa forma, tem dificultado a condução de estudos. Nesse sentido, o uso do Siscolo em estudos epidemiológicos é promissor, proporcionando significativa redução do custo operacional e de tempo. Além disso, este sistema também poderá ser utilizado de forma complementar para minimizar as perdas de seguimento em estudos de coorte.

Uma limitação do Siscolo refere-se aos dados estarem restritos à população usuária do SUS, não englobando as mulheres que realizam os exames em serviços de saúde suplementar. Em inquérito domiciliar<sup>a</sup> conduzido pelo Ministério da Saúde, entre 2002 e 2003, em 16 cidades brasileiras, a realização de exames citopatológicos na rede SUS variou de 32,0% no Rio de Janeiro a 54,0% em Aracaju. No entanto, é provável que a coorte de mulheres analisada no presente estudo, por ser cadastrada na ESF, utilize mais os serviços oferecidos pela rede SUS do que a população daquele inquérito.<sup>a</sup> Ao mesmo tempo, os exames histopatológicos têm custo superior àqueles dos exames citopatológicos, e por esta razão é provável que seja maior a procura pela

<sup>a</sup> Ministério da Saúde. Secretaria de Atenção à Saúde. Instituto Nacional de Câncer. Inquérito Domiciliar sobre Comportamentos de Risco e Morbidade Referida de Doenças e Agravos não Transmissíveis. Brasil, 15 capitais e Distrito Federal 2002–2003. Rio de Janeiro; 2003 [citado 2007 dez 11]. Disponível em: <http://www.inca.gov.br/inquerito>

rede SUS para sua realização. Infelizmente, ainda não é possível mensurar a quantidade real de exames diagnósticos especializados realizados no País por falhas no processo de informação na rede de referência.

Conclui-se que os dados do Siscolo avaliados na coorte de mulheres estudadas foram de boa qualidade. Este sistema é essencial para o planejamento e monitoramento das ações de rastreamento do câncer do colo do

útero. A ampliação do uso do Siscolo pelos serviços de saúde nos diversos níveis de referência do SUS deve ser estimulada e a divulgação dos resultados de avaliações realizadas pode representar um passo importante para o seu aperfeiçoamento. A realização de outros estudos será importante para o aprimoramento do Siscolo, sobretudo considerando uma amostra representativa de laboratórios ou de unidades de saúde e assim auxiliar na identificação de fontes de erros e de perda de informação.

## REFERÊNCIAS

1. Camargo Jr KR, Coeli CM. Reclink: Aplicativo para o relacionamento de base de dados, implementando o método probabilistic record linkage. *Cad Saude Publica*. 2000;16(2):439-47. DOI: 10.1590/S0102-311X200000200014
2. Coeli CM, Camargo Jr KR. Avaliação de diferentes estratégias de blocagem no relacionamento probabilístico de registros. *Rev Bras Epidemiol*. 2002;5(2):185-96. DOI: 10.1590/S1415-790X2002000200006
3. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN. Updated Guidelines for Evaluating Public Health Surveillance Systems Recommendations from the Guidelines Working Group. *MMWR Morb Mortal Wkly Rep*. 2001 [citado 2007 dez 11];50(RR13):1-35. Disponível em: <http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm>
4. Girianelli VR, Thuler LCS, Szklo M, Donato A, Zardo LM, Lozana JA, et al. Comparison of HPV DNA tests and liquid based cytology with conventional cytology for the early detection of cervix uteri cancer. *Eur J Cancer Prev*. 2006;15(6):504-10. DOI: 10.1097/01.cej.0000220630.08352.7a
5. Longatto Filho A, Almeida DCB, Adura PJD, Marzola VO, Cavaliere MJ. Influência da qualidade do esfregaço cérvico-vaginal na detecção de lesões intra-epiteliais. *Folha Med*. 2002;121(2):79-83.
6. Maeda MYS, Loreto C, Barreto E, Cavaliere MJ, Utgawa ML, Sakai YI, et al. Estudo preliminar do SISCOLO-Qualidade na rede de saúde pública de São Paulo. *J Bras Patol Med Lab*. 2004;40(6):425-9. DOI: 10.1590/S1676-24442004000600011
7. Mello Jorge MHP, Gotlieb SLD, Oliveira H. O Sistema de Informações sobre Nascidos Vivos: primeira avaliação dos dados brasileiros. *Inf Epidemiol SUS*. 1996;5:15-48.
8. Mello Jorge MHP, Laurenti R, Gotlieb SLD. Análise da qualidade das estatísticas vitais brasileiras: a experiência de implantação do SIM e do SINASC. *Cienc Saude Coletiva*. 2007;12(3):643-54. DOI: 10.1590/S1413-81232007000300014
9. Paes NA. Avaliação da cobertura dos registros de óbitos dos Estados brasileiros em 2000. *Rev Saude Publica*. 2005;39(6):882-90. DOI: 10.1590/S0034-89102005000600003
10. Piper JM, Mitchell Jr EF, Snowden M, Hall C, Adams M, Taylor P. Validation of 1989 Tennessee Birth Certificates Using Maternal and Newborn Hospital Records. *Am J Epidemiol*. 1993;137(7):758-68.
11. Silva JPL, Travassos C, Vasconcellos MM, Campos LM. Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil. *Cad Saude Coletiva*. 2006;14(2):197-224.
12. Szklo M, Nieto FJ. *Epidemiology beyond the basics*. Gaithersberg: Aspen Publishers; 2000.
13. Thuler LCS, Zardo LM, Zeferino LC. Perfil dos laboratórios de citopatologia do Sistema Único de Saúde. *J Bras Patol Med Lab*. 2007;43(2):103-14. DOI: 10.1590/S1676-24442007000200006

---

Pesquisa financiada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq – Processo nº 476941/2006-7; Edital Universal).

Artigo baseado em tese de doutorado de Girianelli VR apresentada ao Instituto Nacional de Câncer em 2008.