Vania Reis Girianelli[I,II]

Luiz Claudio Santos Thuler[I,III]

Gulnar Azevedo e Silva[II]

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

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.

[I] Programa de Pós-Graduação em Atenção Oncológica. Instituto Nacional de Câncer. Rio de Janeiro, RJ, Brasil

[II] Departamento de Epidemiologia. Instituto de Medicina Social. Universidade Estadual do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

[III] Departamento de Medicina Geral. Universidade Federal do Estado do Rio de Janeiro. Rio de Janeiro, RJ, Brasil

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

Received: 04/03/2008
Revised: 10/03/2008
Approved: 12/11/2008

INTRODUCTION

Health information system in Brazil consists of several different subsystems with data from sectoral activities. The *Sistema de Informação sobre Mortalidade* (SIM –Mortality Data System) was the first to be created in Brazil in 1976.[8] During the 1990s, several data systems were established to provide information for planning and evaluating health services as part of the *Sistema Único de Saúde* (SUS – Brazilian National Health System).

These systems are a valuable source of information for epidemiological studies reducing research-related costs and time. Yet, the major barrier for their utilization is quality; i.e., inadequate information quality with a great deal of missing and incorrect data.

There have been increased access to and opportunities of evaluating national health data systems in recent years. A recent systematic literature review[11] identified 71 publications of Brazilian database integration related to 40 studies. Of these, 70% were epidemiological studies and, while most had some reference to data quality, only 15% of them actually intended to assess it. Of the articles reviewed, 75% referred to SIM and 57.5% referred to the *Sistema de Informação sobre Nascidos Vivos* (Sinasc –Information System on Live Births). Reclink program was used in 27.5% of studies for probabilistic linkage between databases.

The *Sistema de Informação do Câncer do Colo do Útero* (SISCOLO - Cervical Cancer Data System) is the most recently created information system, developed by *Departamento de Informática do SUS* (DATAUS –SUS Department of Information Technology) together with the *Instituto Nacional do Câncer* (INCA –Brazilian National Institute of Cancer). Established in January 2000, SISCOLO is intended to gather data on identification of women who are SUS users, their demographic and epidemiological information, and information on cytopathology and histopathology tests performed. SISCOLO has continuously developed enabling care programs at the local and state level to follow up women with abnormal test results. Besides being a potential source of information for studies, SISCOLO also helps to subsidize SUS coverage for these tests and allows to evaluating cervical cancer care programs and services.[a] A search of *Biblioteca Virtual em Saúde* (BIREME – Virtual Health Library) databases in November 2007 revealed that the three studies based on SISCOLO had the objective of assessing quality of tests at cytopathology laboratories.[5,6,13]

The objective of the present study was to evaluate data completeness, validity, and sensitivity of SISCOLO database.

METHODS

The study was based on the following sources of data:

- Reference database – consisted of demographic, epidemiological, clinical and laboratory information of a cohort including 2,024 women living in 13 communities served by the *Estratégia Saúde da Família* (ESF – Family Health Strategy) in the cities of Duque de Caxias and Nova Iguaçu, state of Rio de Janeiro, Southeastern Brazil. These women participated in a cross-sectional study[4] conducted between 2001 and 2002 and did not have high-grade intraepithelial lesions of the cervix or cervical cancer (HSIL+) at that time.

- Medical records – information was collected from 13 ESF units participating in the previous study, three reference health units for diagnostic confirmation, and INCA electronic medical records, which is a reference center for cancer treatment in Rio de Janeiro. Data was collected on date and results of cytopathology, colposcopy and histopathology tests performed during the period between the entry date in the study and December 2006, or until diagnostic confirmation of HSIL or cervical cancer (HSIL+). A systematic sample representing 10% of all medical records found at each health unit was reviewed.

- SISCOLO – two SISCOLO databases containing cytopathology and confirmatory tests (colposcopy and histopathology) performed in the state of Rio de Janeiro, available from the Brazilian Ministry of Health DATASUS on May 2007, were studied. For the period between January 2002 and May 2006, data were provided in the version 3.06. For the period between June and December 2006, data were provided in the version 4, which has incorporated the new Brazilian classification for cervical reports.[b] The fields analyzed included only demographic data, test results and women identification. Databases were checked for inconsistencies that were then corrected to avoid potential interferences in the linkage process. Records of the fields "mother's name" and "district of residence" were left blank when they included words indicating missing information or numbers only. In the field "patient name," as the physician's or nurse's name was mistyped in many records, they were manually deleted.

Reclink software program version 3 was used for linkage between databases to identify information of women of the reference cohort. Reclink is an application using a probabilistic record linkage method to estimate the likelihood of a pair of records being from the same individual. Record linkage involves the identification of common fields in both databases and they are scored based upon their probability of matching or differing. This procedure is carried out in three steps.[1,3]

a) Standardization – database fields are prepared for linkage to minimize errors. Fields can be subdivided and adjusted to have the same structure. The program also allows the exclusion of prepositions, punctuation signs, accents, and other symbols.

b) Blockage – logic blocks consisting of one or more fields are created to restrict linkage only to records that have the same content in the related fields selected. Seven strategies of blockage were sequentially applied to minimize loss of true pairs (Figure 1).

---

c) Pairing – construction of scores for different pairs obtained using a certain blockage strategy based on specific criteria for the fields selected as identifiers, because they have greater discrimination power.

The program can calculate scores based on the probability of the matching of two records on one identifier, being a potential true pair, sensitivity (m), or in the event of a potential false pair, false-positive (u). In addition, it can calculate the probability of non-matching on one identifier, being a potential true pair, false-negative ($1 - m$), and in the event of a potential false pair, specificity ($1 - u$). Based on these probabilities, two weighting factors are generated, one for matching and one for non-matching. The weighting factor for matching is calculated using a base 2 logarithm of the likelihood ratio between probabilities m and u and the weighting factor for non-matching is calculated for the remaining probabilities ($1 - m$) and ($1 - u$). The total score of a pair is obtained from the sum of weighting factors attributed after comparing each identifier. The fields selected as identifiers, criteria used and calculated weighting factors are presented in Table 1.

Pairing strategies applied in record linkage and related maximum (full matching on all identifiers) and minimum scores (non-matching on all identifiers) are shown in Figure 1. The greater the number of identifiers, the wider the score range. Date of birth was not included in steps 1 to 4 as it was used as a blockage strategy.

Steps 1 and 2 were more restrictive so few pairs would be created and would be more likely true pairs, thus allowing to be checked. In steps 3 and 4, more pairs were generated during linkage and only those with scores greater than –4.0 were checked. In the following steps, there was a dramatic increase in the number of pairs generated and only those with positive scores were checked.

Thorough manual checking aimed at avoiding misclassifying as true pairs those that did not belong to the same individual. Classification criteria were as follows:

a) same date of birth with identical name and mother's name, or with an abbreviated middle name or one of the middle names missing;

b) same name and mother's name with date of birth with no more than two different digits, or day replaced by month;

c) similar uncommon name or mother's name with same date of birth or address;

d) different name, mother's name or date of birth or missing information in one database, but all remaining fields containing identical or very similar information matching on at least three fields.

Paired records in one step were not included in the following steps, except for records from the reference database when associated to SISCOLO files due to the possibility of a woman undergoing more than one test. Although Reclink version 3 has a feature for identifying duplicity, it was not used.

At each step paired records were saved as files and then put together as a single file using Microsoft Office Access (2003). Each file was associated to the corresponding original SISCOLO database using Reclink, by joining fields related to test results.

To assess SISCOLO data quality, indicators of completeness, validity, and sensitivity were calculated as proposed by the US Centers for Disease Control and Prevention (CDC).[3]

Field completeness was assessed based on the proportion of complete record with no missing information in a given field. Based on criteria described by Mello Jorge et al,[7] this indicator was considered excellent when the proportion of completeness was higher than 90%; good between 70.1% and 90%; and poor when equal to or lower than 70%.

The validity of SISCOLO fields was assessed based on sensitivity where the gold standard was data from medical records (including tests) or from the cross-sectional study (demographic and identification information). In addition, a Bland-Altman plot was constructed to analyze the field "data collection" for it is the most adequate to assess validity of continuous variables as proposed by Szklo & Nieto.[12]

Sensitivity was estimated based on the proportion of tests recorded in the medical records of women in the reference cohort that were identified in SISCOLO. This indicator was interpreted based on the criterion as proposed by Piper et al:[10] high sensitivity when greater than 90%; moderate between 70% and 90%; and low when below 70%.

There were also estimated the related 95% confidence intervals for indicators of field validity and sensitivity.

The study was approved by the Research Ethics Committee of the National Cancer Institute (Protocol No. 074/06).

RESULTS

Completeness of SISCOLO databases containing cytopathology and confirmatory tests (colposcopy and histopathology) was found to be excellent for the fields "mother's name" (98.4% and 98.2%, respectively) and "place of residence" (98.0% and 98.3%, respectively), and good for "district of residence" (84.5% and 89.8%,

Table 1. Parameters used and results of weighting factors by identifier field. State of Rio de Janeiro, Southeastern Brazil, 2002–2006.

Identifier field	Parameters				Weighting factors	
	Algorithm	Sensitivity (%)	Specificity (%)	Minimum matching percentage (%)	Matching	Non-matching
Name	Approximate[a]	92	99	85	6.52	−3.63
Mother's name	Approximate[a]	92	99	85	6.52	−3.63
Date of birth	Character[b]	90	95	65	4.17	−3.25
District of residence	Approximate[a]	92	99	85	6.52	−3.63

[a] Compares sequences of characters based on Levenshtein algorithm, ranging from 1 for full matching to 0 for non-matching, indicated for fields with name information.

[b] Compares sequences of digits, ignoring separators, for pairs of digits in the same position, indicated for fields with complete dates.

respectively). Completeness of the fields "zip code" and "individual taxpayer number" was poor in both databases (Table 2). Since the completion of the field "date of birth" is not required when age is reported, the system assigns a year of birth based on age plus "01/01" for day and month. This set-up was seen in 3.5% of records analyzed. The field "age" is estimated by the system based on the date of birth. Records with age younger than ten years and older than 89 years accounted for 0.2%. Other fields with demographic, and identification information and those related to test data are all required fields and thus there were no records with missing information.

There were checked 19,801 pairs in the record linkage between the reference database and SISCOLO database with cytopathology tests and 556 pairs in the linkage between the reference and the database with confirmatory tests; and 10.6% and 0.9%, respectively, were classified as true pairs. Most true pairs were identified in step 1, accounting for 64.4% for cytopathology tests and 60.0% for confirmatory tests. In step 2, although there can be completion error of the field "city code", it may also indicate migration or misinformation reported

Table 2. Completeness of fields in the Cervical Cancer Data System (SISCOLO) databases. State of Rio de Janeiro, Southeastern Brazil, 2002–2006.

Field	Database with cytopathology tests (%)	Database with confirmatory tests (%)
Mother's name	98.4	98.2
Place of residence	98.0	98.3
District of residence	84.5	89.8
Zip code	30.0	30.8
Individual taxpayer number	0.5	2.0

by women. In this step, the proportion of pairs created was 7.6% and 20.0%, respectively (Table 3).

Steps 1 and 2 were more restrictive, and pairs with very low scores were classified as they showed abbreviations or the exclusion of women's middle name as well as their mother's, or omission of mother's name. In the following less restrictive steps, pairs with similar characteristics, if any, could not be identified as their score was below the cutoff or they did not show other

Table 3. Pairs created, pairs classified as true and score ranges for each step of probabilistic record linkage between Cervical Cancer Data System (SISCOLO) databases and a reference cohort. State of Rio de Janeiro, Southeastern Brazil, 2002–2006.

Step	Database with cytopathology tests					Database with confirmatory tests				
	Number of pairs created	True pairs				Number of pairs created	True pairs			
		Number	%	Cumulated frequency	Score ranges		Number	%	Cumulated frequency	Score ranges
1	1415	1354	64.4	64.4	13.1 to −7.3	3	3	60.0	60.0	2.9 to −7.3
2	751	159	7.6	72.0	13.1 to −7.3	3	1	20.0	80.0	12.5
3	737	117	5.6	77.6	12.8 to −3.6	49	1	20.0	100.0	2.2
4	618	30	1.4	79.0	12.8 to 2.2	49	0	–	100.0	–
5	6118	308	14.6	93.6	16.7 to 0.0	40	0	–	100.0	–
6	4018	102	4.8	98.4	12.8 to 2.3	382	0	–	100.0	–
7	6144	33	1.6	100.0	22.9 to 2.1	30	0	–	100.0	–
Total	19801	2103	100.0	100.0	–	556	5	100.0	100.0	–

Table 4. Sensitivity of fields in the databases with cytopathology and confirmatory tests of the Cervical Cancer Data System (SISCOLO). State of Rio de Janeiro, Southeastern Brazil, 2002–2006.

Field	Database with cytopathology tests		Database with confirmatory tests	
	n	Sensitivity (95% CI)	n	Sensitivity (95% CI)
Woman's name				
Soundex code of the first name	2045	97.2 (96.5;97.9)	5	100.0 (100.0;100.0)
Soundex code of the last name	1911	90.9 (89.6;92.1)	4	80.0 (44.9;100.0)
Mother's name				
Soundex code of the first name	1765	83.9 (82.4;85.4)	5	100.0 (100.0;100.0)
Soundex code of the last name	1573	74.8 (72.9;76.7)	4	80.0 (44.9;100.0)
Date of birth	1705	81.1 (79.4;82.7)	5	100.0 (100.0;100.0)
IBGE code[a] of the city of residence	1871	89.0 (87.6;90.3)	3	60.0 (17.1;100.0)
Soundex code of the district of residence	1420	67.5 (65.5;69.5)	0	-

[a] IBGE: Brazilian Institute of Geography and Statistics

matching fields that would allow classification. There were also found matching addresses and districts with non-matching city code.

The validity of the fields "soundex code of the woman's first name" (97.2%, 95% CI: 96.5;97.9) and "soundex code of the woman's last name" (90.9%, 95% CI: 89.6;92.1) was high, considering that all women included in the reference cohort with records in SISCOLO database with cytopathology tests were identified. Good validity was also seen for the fields "city of residence" (89.0%, 95% CI: 87.6;90.3), "mother's first name" (83.9%, 95% CI: 82.4;85.4), "mother's last name" (74.8%, 95% CI: 72.9;76.7), and "date of birth" (81.1%, 95% CI: 79.4;82.7). The field "soundex code of the district of residence" showed the lowest validity (67.5%, 95% CI: 65.5;69.5) of all fields included in record linkage. In the database with confirmatory tests, only five women were identified, but the woman's first name, mother's first name, and date of birth were correct for all pairs created (Table 4).

A total of 1,147 medical records (56.7%) were retrieved from ESF units. Many records were lost due to inadequate filing and flood, others were discarded as they were inactive due to death, moving or canceling of family enrollment in the program. Of those medical records found, only 636 (55.4%) had records of cytopathology tests. The remaining medical records had only records of visits, especially for diabetes and hypertension management, or were blank. Some tests were recorded only in general record books and could not be identified. From the medical records checked in all health units studied there were identified 1,113 cytopathology tests.

Of 2,103 cytopathology tests identified in SISCOLO, six were duplicated; 31 were performed after diagnostic confirmation of HSIL+, and thus excluded from the analysis; and 831 tests were performed at other health units and not recorded in the medical records of the units

studied. Of the remaining 1,235 tests, 862 (69.8%) were recorded in the medical records, including 197 tests performed at other health units. Of 373 tests identified only in SISCOLO, 157 tests (42.1%) were of women whose medical records were not found at the health units in the study. The other tests were reassessed and their classification was confirmed. Additionally, the medical records showed 251 tests not identified in SISCOLO.

The system sensitivity to identify cytopathology tests was 77.4% (95% CI: 75.0;80.0). Of 2,317 cytopathology tests found in the data sources searched, 89.2% were identified in SISCOLO, 48.0% in the medical records and 37.2% in both sources.

At the three reference units for diagnostic confirmation, no records of colposcopy or histopathology tests of women in the reference cohort were found.

In INCA electronic medical records, 172 patients were identified, of which 80 had records of colposcopy and histopathology tests, corresponding to 173 tests performed during the period studied. Of these 80 patients, only five were identified in SISCOLO, and two underwent tests that were included in the INCA medical records, while all other patients underwent only colposcopy tests at health units not included in the present study. The sensitivity of this database was very low (4.0%, 95% CI: 0.0;21.3).

Of women in the reference cohort, 1,251 (61.8%) had cytopathology, colposcopy or histopathology tests identified in the sources of data studied.

As for validity, sensitivity of the field "date of collection" was 100% in the database with confirmatory tests and 70.3% in the database with cytopathology tests. Sensitivity of the field "test results" was 100% in both databases.

The time interval between date of collection recorded in the database with cytopathology tests and that recorded
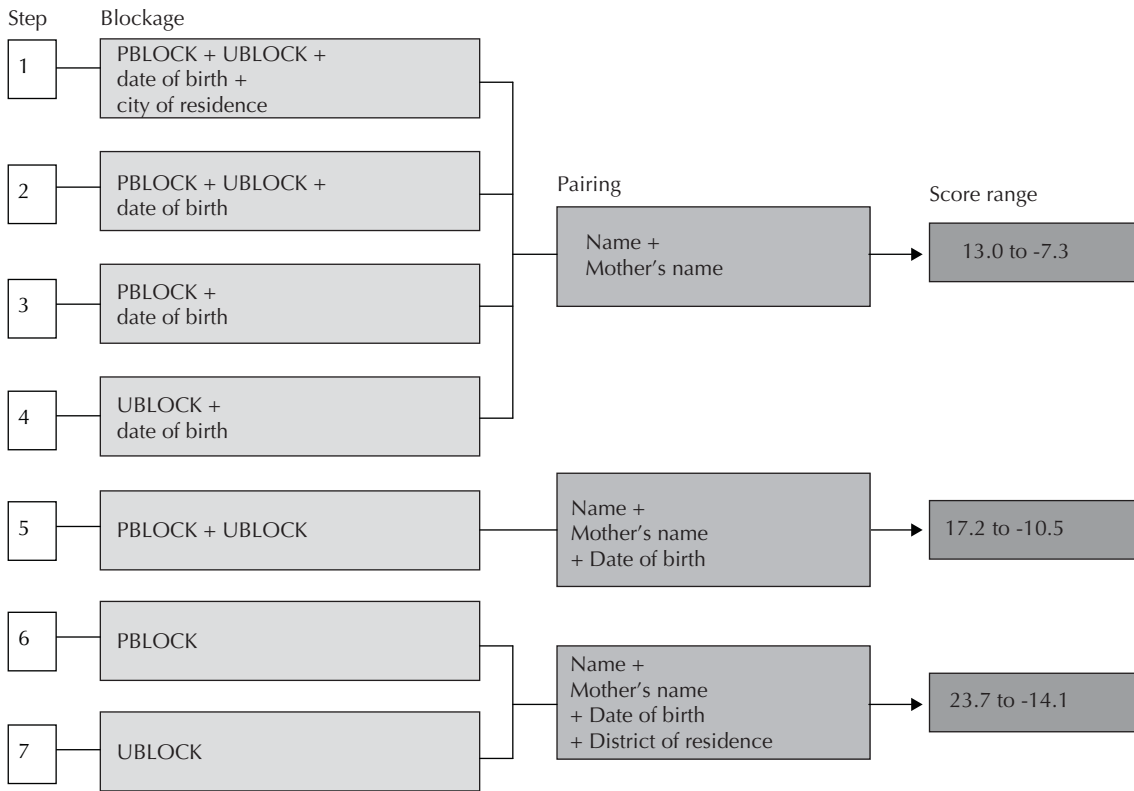
PBLOCK – Soundex code of the first name, UBLOCK – Soundex code of the last name, Soundex code – phonetic 4-digit code consisting of one letter (the word's first letter) and three numbers, and vowels are coded 0 and consonants with similar sounds have the same code (Reclink standardization strategy).

Figure 1. Blockage and pairing strategies applied in record linkage and results of score ranges by consecutive steps. Rio de Janeiro, Southeastern Brazil, 2002–2006.

in medical records was as much as 30 days in 93.5% of cases. A time interval greater than 60 days was seen especially in the second quarter of 2004 and 2006, corresponding to the same period when SISCOLO was updated (Figure 2).

DISCUSSION

In the present study, it was found that SISCOLO contained information of 89.2% of 2,317 cytopathology tests performed in women in the reference cohort.
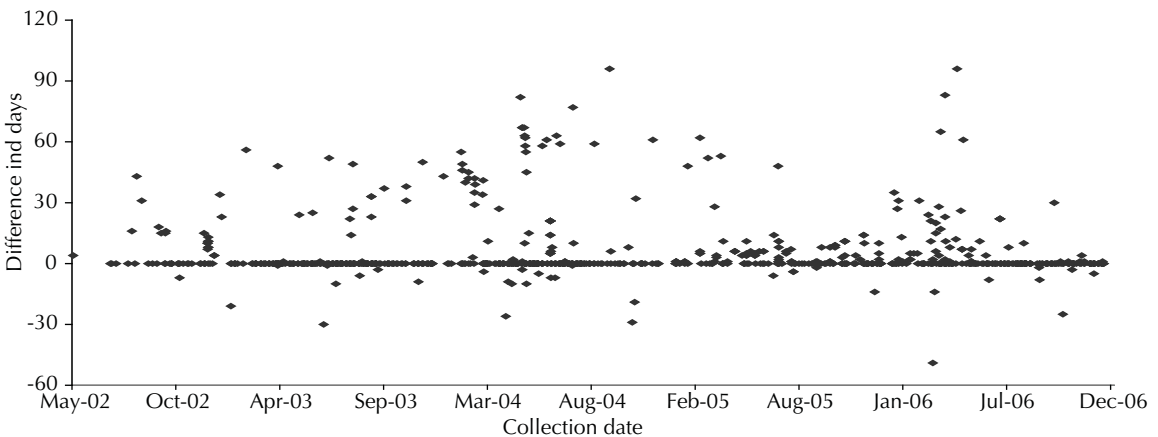
Figure 2. Bland-Altman plot of the time interval time interval between date of collection recorded in the database with cytopathology tests in the Cervical Cancer Data System (SISCOLO) and that in the medical records analyzed. State of Rio de Janeiro, Southeastern Brazil, 2002–2006.

Reclink was essential as no single identifier was available for health information recording. It is expected that linkage between records will be more accurate and uncomplicated with the implementation of a SUS card undergoing in Brazil.

Major difficulties were not encountered for Reclink application; however, manual selection of records was an extremely painstaking and time-consuming task, especially for less restrictive strategies that were required due to missing data and field completion or entry errors. Despite that, more than 70% of pairs were identified in steps 1 and 2, which were the least time-consuming.

As for quality of SISCOLO data, completeness, as well as validity of test results, was found to be excellent for most fields analyzed.

The system sensitivity for the database with cytopathology tests was moderate (77.4%). However, sensitivity is likely to be higher than that found since only information available in medical records was analyzed. Moreover, SISCOLO sensitivity may be different in other population groups or other Brazilian regions as seen for SIM and Sinasc. Despite being mandatory and having longer operation, these databases are still affected by different regional coverage.[8,9]

Sensitivity of the database with confirmatory tests (colposcopy and histopathology) was found to be very low (4.0%) and not yet a valuable tool since most data was not entered into this system. A possible explanation is that these tests are often performed together with other procedures during hospital admissions and covered only through hospital admission authorization. There is a need to make this information mandatory in SISCOLO to allow tracking cases requiring follow-up and treatment. The follow-up module of SISCOLO has undergone improvements and they are expected to enable to generate reports of women with abnormal cytopathology tests requiring follow-up and to allow state and local health managers to provide feedback information to system obtained during active search at the local level.

Information on cervical cancer and precursor lesion screening tests is not readily available at the health units where test specimens are collected, which to some extent make investigations more difficult. In this sense, SISCOLO is a promising tool for epidemiological studies as it could help significantly reducing operational costs and time. It can also be used as an additional instrument to minimize loss to follow-up in cohort studies.

A limitation of SISCOLO is that data available are restricted to SUS users and do not include women undergoing tests in complementary health services. The Brazilian Ministry of Health household survey,[a] conducted in 16 cities during 2002 and 2003, showed that cytopathology tests in SUS ranged from 32.0% in Rio de Janeiro (southeastern) to 54.0% in Aracaju (northeastern). However, the cohort of women in the present study, as they were enrolled in the ESF, they are likely to use more services provided by SUS than the population studied in the survey. Also, as histopathology tests are more costly than cytopathology tests, women will likely turn to SUS services to get them. Unfortunately, the actual number of specialized diagnostic tests performed in Brazil is not known due to deficient information network in reference centers.

In conclusion, the quality of SISCOLO data in the cohort of women studied was good. SISCOLO is an essential instrument for planning and monitoring actions of cervical cancer screening. SUS health services at different level of reference should be encouraged to make better use of SISCOLO and the dissemination of evaluation results can contribute to improve this information system. Further studies are needed to advance SISCOLO development, especially those including a representative sample of laboratories or health units that can help identify sources of errors and missing information.

[a] Brazilian Ministry of Health. Department of Health Care. National Cancer Institute. 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 [cited 2007 Dec 11]. Available from: http://www.inca.gov.br/inquerito

REFERENCES

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/S0102-311X2000000200014

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, Utagawa 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