Access to colposcopy in the state of São Paulo, Brazil: probabilistic linkage study of administrative data

Acesso à colposcopia no Estado de São Paulo, Brasil: um estudo de relacionamento probabilístico de dados administrativos

Acceso a la colposcopia en el estado de São Paulo, Brasil: estudio de vinculación probabilística con datos administrativos

Abstract

Cervical cancer screening is a multistage process, therefore access to both the primary test and subsequent diagnostic procedures is essential. Considering women undergoing screening on the public health system in the state of São Paulo, Brazil, we aimed to estimate the proportion of women accessing colposcopy within six months of an abnormal smear result. We retrieved records from two administrative databases, the Information System on Uterine Cervical Cancer (SIISCOLO) that contains smear results and the Outpatient Information System of the Brazilian Unified National Health System (SIA/SUS) that records colposcopies. A reference cohort consisted of women, aged 25 years or older, with an abnormal smear result between May 1, 2014, and June 30, 2014. We excluded prevalent cases. We linked the reference cohort and records in the SIA/SUS extending to December 31, 2014. After excluding prevalent cases, 1,761 women with abnormal cytology results were left. A total of 700 (39.8%) women were linked to a colposcopy record within the follow-up period; this dropped to 671 (38.1%) women when follow-up was censored at six months. We could notice a slightly higher attendance in women living in the metropolitan region of São Paulo compared with residents of the rest of the state. We found no association between colposcopy attendance and age or cytology class. These results emphasize that access to colposcopy in the public health system in São Paulo is limited. This compromises the quality of screening, and the issue needs to be prioritized in service planning.

Uterine Cervical Neoplasms; Triage; Colposcopy

Correspondence

L. F. Buss
Faculdade de Medicina, Universidade de São Paulo.
Av. Dr. Enéas Carvalho de Aguiar 470, São Paulo, SP 05403-000, Brasil.
lewisbuss@usp.br

1 Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brasil.
2 Fundação Oncocentro de São Paulo, São Paulo, Brasil.
3 Instituto Nacional de Câncer José Alencar Gomes da Silva, Rio de Janeiro, Brasil.
4 Instituto de Medicina Social, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil.

doi: 10.1590/0102-311X00304820
Introduction

The burden of cervical cancer is largely influenced by differential access to screening. In Brazil, the Brazilian Unified National Health System (SUS) offers free screening with the Papanicolaou (Pap) smear. The access to this test expanded substantially, and in some state capitals over 80% of the target population is covered. However, cervical cancer incidence in Brazil remains high compared with some countries with equivalent coverage.

One explanation for this finding may be the limited access to diagnosis and treatment after an abnormal smear result. Women with cytological results suggestive of a high-grade lesion should be referred directly for colposcopy. If an obstacle exists at the transition from the primary screening test to colposcopy, this would represent an important priority for service planning.

We addressed this question by following a cohort of women with abnormal smear results collected during a routine screening in the state of São Paulo, Brazil. We used a probabilistic technique to link the cytology results, recorded in the Information System on Uterine Cervical Cancer (SISCOLO; a national cervical cancer screening database), to colposcopy records in the Outpatient Information System of the SUS (SIA/SUS; a billing database for outpatient procedures); both available at http://datasus.saude.gov.br/. This study is part of a wider project that investigates inequalities in access to screening services in Brazil.

Methods

Data sources

The SISCOLO is used to monitor cervical cancer screening activities. The results of smears performed on the SUS and the date of reporting are recorded. Patient identifiers in the SISCOLO are the woman’s name, mother’s name, date of birth, address, and the National Health Card (CNS) number. The CNS is a unique patient identifier. Unfortunately, its utility in this regard is limited, since some patients have more than one CNS. Furthermore, it is not an obligatory field in the SISCOLO, with only a roughly 50% completion rate.

The SIA/SUS is an administrative billing database for outpatient procedures. Since March 2014, colposcopy services were required to record individual patient information on this system (Ordinance n. 189, from January 31, 2014). Each colposcopy record contains the date and location of the procedure, patient’s name, date of birth, address, and CNS (obligatory in the SIA/SUS). This is used as a record of the occurrence of a procedure but without clinical specifics.

Selection of the reference cohort

A cohort of women with abnormal cytology results was identified, requiring – according to the Brazilian guidelines – a colposcopy referral. All records of women aged 25 years and older, resident in the state of São Paulo, and with a cytology class more severe than LSIL were retrieved from the SISCOLO. Only women whose results were released between May 1, 2014, and June 30, 2014, were included. This group of women with abnormal cytology will be referred to as the reference cohort.

Timing considerations and prevalent cases

Not all abnormal results in the SISCOLO represent screening smears. A repeat smear is indicated after a colposcopy when the findings were discordant with the original smear, or when the endocervical canal is not visualized. Furthermore, six-monthly smears are recommended for follow-up after treating a lesion. We sought to exclude these non-screening smears from the reference cohort since they represent women within the colposcopy services.

Thus, the SISCOLO lacks the indication for smears, two procedures were applied to best exclude the prevalent cases. First, we aimed to remove all women in the reference cohort who received at least one other abnormal result in the preceding 16 months. To achieve this, records in the reference
cohort were linked with all abnormal smear results in the SISCOLO between January 1, 2013, and April 30, 2014.

Following, we reasoned that women undergoing follow-up (concurrent colposcopy and smear) after treatment in the preceding year may not be excluded by this procedure. The SISCOLO contains the date that the smear result was released, whereas the SIA/SUS records the date the colposcopy was performed. Therefore, these cases would appear as linked records where the colposcopy date preceded the smear date. The reference cohort was linked with the two preceding months in the SIA/SUS (March and April 2014); all records linked to a colposcopy before the smear release were excluded (Figure 1).

Finally, to determine the rate of colposcopy attendance, records in the SIA/SUS between May 1 and December 31, 2014, were included. This allowed a period of six to eight months for recording the colposcopy.

**Data manipulation procedures**

- **Cleaning and de-duplication**

Data pre-processing and cleaning routines were conducted to minimize differences in typing and to standardize entries between the SISCOLO and SIA/SUS. For the women’s names, accents, double spaces, and punctuation were removed. Letters were converted to uppercase and known abbreviations replaced with full names (e.g., Ap → Aparecida). All prepositions in last names (e.g., “de”, “dos” etc.) were removed and full names were split into first, middle (when there was more than one, only the first was retained), and last names. The name fields were searched for strings consistent with the missing values (e.g., “ignorado” and synonyms) and other (more esoteric) entries that would not be consistent with names. The date of birth was split into separate fields for day, month, and year.

**Figure 1**

Linkage procedures and formation of the reference cohort.

*SIA/SUS: Outpatient Information System of the Brazilian Unified National Health System; SISCOLO: Information System on Uterine Cervical Cancer.*
The women’s names were completely recorded in the SISCOLO and missing in five records in the SIA/SUS. Mother’s name, although not used in the linkage procedure, was filled far less reliably. Date of birth was recorded in 100% of cases in both databases. The CNS was 98% and 58% completed in the SIA/SUS and SISCOLO, respectively. When present, the CNS was used as a key for de-duplication. Regarding the high level of missing CNS in the SISCOLO and issues with its performance as a unique identifier, the remaining records were de-duplicated based on the exact agreement of name, date of birth, municipality of residence (Brazilian Institute of Geography and Statistics – IBGE – code), and mother’s name (in the SISCOLO). Only the first record (date of cytology or date of colposcopy) was retained.

• Linkage

First, a gold standard dataset of certain (or at least highly probable) matches between the SISCOLO and SIA/SUS were made, performing a deterministic linkage between records with a completed CNS. The links were manually reviewed. A total of 97.4% agreed on the name, date of birth and address, or name and either date of birth or address. We used this gold standard dataset to calculate m-probabilities: the probability that a particular field (e.g., first name) agrees concerning the two records being truly matched, \( m = P(\text{agreement} \mid \text{match}) \). The value reflects the data entry error rate; if data entry were error-free and completely standardized the m-probability would be 1. Next, the u-probabilities were calculated – the probability that a particular field agrees given that the records are truly unmatched, that is a chance agreement, \( u = P(\text{agreement} \mid \text{non-match}) \) – from the total set of pairwise comparisons, excluding the known matches. The u-probability reflects the uniqueness, or discriminatory power, of the identifier in question.

Complex comparison patterns were generated for agreement on first, middle, and last names with the Jaro-Winkler string comparator. Jaro-Winkler scores range from 0 (perfect dissimilarity) to 1 (perfect similarity) and were divided into five categories: (0.00, 0.75], (0.75, 0.90], (0.90, 0.95], (0.95, 0.99], and (0.99, 1.00). The m- and u-probabilities were frequency adjusted for each category of Jaro-Winkler score.

Match weights were calculated according to the Fellegi-Sunter method. That is, for a given pair of records, the match weight was calculated as the sum of the log-likelihood ratios determined from the probabilities m- and u-, so that:

\[
\log\text{-LR}^+ = \log_2\left(\frac{m}{u}\right) \\
\log\text{-LR}^- = \log_2\left(\frac{(1-m)}{(1-u)}\right)
\]

Where the Log-LR+ is the positive likelihood ratio used for complete or partial field agreement, and Log-LR- is the negative likelihood ratio used for complete field disagreement.

The calculated match weights had a range of -30 to 24. Following a visual inspection of the distribution of match weights for true and false matches, a cut-off of 15 was chosen, above which matches were included without further review, and below which pairs were rejected. This threshold produced an excellent discriminatory capacity (see sensitivity and specificity further). Serial blocks were used to reduce the total number of comparisons. These were based on combinations of the SoundexBR (http://CRAN.R-project.org/package=SoundexBR) phonetic code of first, middle, and last names; date of birth; and the municipality of residence.

The gold standard set was used to estimate the performance of the linkage strategy. The sensitivity for true matches was 96%. The positive predictive value was found to be 96% and the false-positive rate, 4%.

Statistical analyses

Age was categorized into four groups, as well as the cytology class. The municipality of residence was classified as those located in the metropolitan region of São Paulo and those living outside this area. The proportion of women with a linked colposcopy record was calculated and categorical variables were compared between these groups using the chi-squared test.
Women whose cytology results were released at the beginning of the reference period had two months more follow-up compared with those with cytology results released at the end of the reference period. To account for this, the proportion of women that linked with a colposcopy record was calculated within six-month of the cytology result being released.

The linkage algorithm was written and implemented in the R language for statistical computing. Specifically functions from the RecordLinakge and SoundexBR packages were adapted for this specific application. The statistical analysis was performed in R version 3.6.3 (http://www.r-project.org) and Stata 14.1 (https://www.stata.com).

**Ethics**

The project was approved by the University of São Paulo (USP) Ethics Committee (CAAE 49655215.0.0000.0065).

**Results**

We retrieved 2,018 abnormal cytology results reported between May 1, 2014, and June 30, 2014. Of these, 191 were linked with an abnormal result in the preceding 16 months and were excluded. Following linkage with the SIA/SUS, another 66 cases were removed where the colposcopy date preceded the cytology date (Figure 1). This resulted in 1,761 abnormal cytology records. Of these, 700 (39.8%) linked with a subsequent colposcopy record in the SIA/SUS. We observed that 671 (38.1%) records linked with a colposcopy within six months after the release of the cytology result.

Table 1 shows the age, area of residence, and cytology classes of women in the reference cohort according to linkage status. Cytology with atypical cells-unable to exclude a high-grade lesion was over-represented among women with a linked colposcopy. Women resident in the metropolitan area of São Paulo were more likely to have a linked colposcopy record than those living outside of the capital. Figure 2 shows the time from the cytology result to the colposcopy as a cumulative probability plot. Among women undergoing colposcopy, most exams occurred within four months of the release of the abnormal cytology result, with few being performed between four and eight months.

**Discussion**

Among women using the public health system in the state of São Paulo, we found that only 38% of those with abnormal smear results accessed a colposcopy within six months. We observed a higher colposcopy rate for women living in the metropolitan area of São Paulo compared with the rest of the state. We found no clear association between the rate of colposcopy attendance and women's age or cytology class. These results are based on administrative data (SISCOLO and SIA/SUS). Thus, they can be taken to reflect the routine functioning of cervical cancer screening services in the state of São Paulo during the period studied.

A feasibility study of HPV primary testing in São Paulo previously estimated colposcopy attendance. In this study, we offered a single HPV test to women attending routine screening services. We invited those with a positive test for colposcopy and 80% ultimately attended. This markedly greater value (80% vs. 38% in the present study) has several likely explanations. First, we provided additional administrative support — over and above routine conditions — with reminder phone calls and letters sent to women that did not attend initially. Then, the participating centers were primary care units linked with the USP, offering a primary care residency program, and a hospital specialized in women’s health. Thus, likely to perform above the average. Finally, the period of follow-up was substantially longer than in this study. Therefore, the present estimate is likely more representative of the real situation.
Table 1

Characteristics of women, resident in the state of São Paulo, Brazil, with abnormal cytology results recorded on the Information System on Uterine Cervical Cancer (SISCOLO) between May 1, 2014 and June 30, 2014, according to linkage status with colposcopy in the Outpatient Information System of the Brazilian Unified National Health System (SIA/SUS).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Linked colposcopy records (n = 700)</th>
<th>Without linked colposcopy records (n = 1,061)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>208 (29.7)</td>
<td>318 (30.0)</td>
<td>0.442</td>
</tr>
<tr>
<td>35-44</td>
<td>187 (26.7)</td>
<td>254 (23.9)</td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>151 (21.6)</td>
<td>258 (24.3)</td>
<td></td>
</tr>
<tr>
<td>55+</td>
<td>154 (22.0)</td>
<td>231 (21.8)</td>
<td></td>
</tr>
<tr>
<td>Cytology class</td>
<td></td>
<td></td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Atypical cells of undetermined significance (glandular, unknown origin)</td>
<td>147 (21.0)</td>
<td>296 (27.9)</td>
<td></td>
</tr>
<tr>
<td>Atypical cells – possible high-grade lesion (ASC-H, AGUS-H, unknown origin)</td>
<td>356 (50.9)</td>
<td>426 (40.2)</td>
<td></td>
</tr>
<tr>
<td>High-grade squamous lesion (HSIL)</td>
<td>187 (26.7)</td>
<td>311 (29.3)</td>
<td></td>
</tr>
<tr>
<td>Invasive cancer (glandular, squamous, unknown origin)</td>
<td>10 (1.4)</td>
<td>28 (2.6)</td>
<td></td>
</tr>
<tr>
<td>Area of residence</td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td>Greater São Paulo</td>
<td>524 (74.9)</td>
<td>719 (67.8)</td>
<td></td>
</tr>
<tr>
<td>Interior of the state</td>
<td>176 (25.1)</td>
<td>342 (32.2)</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1

Cumulative probability of linkage with a colposcopy record after the release of an abnormal cytology result – women accessing public screening services in the state of São Paulo, Brazil, 2014.
In general, the literature on access to cervical cancer screening in Brazil focused on coverage with the smear test. This line of inquiry used national household surveys and aggregate data in the SISCOLO. Population coverage follows a consistent socioeconomic gradient over the last 30 years. In other countries, colposcopy non-attendance was also associated with greater deprivation. Our study could not explore this relationship due to the very limited sociodemographic information recorded in the SISCOLO and SIA/SUS.

However, we observed a clear divide between attendance in the metropolis (42%) and the rest of the state (34%). We also observed a similar rural-urban divide concerning smear coverage. These findings are particularly relevant regarding the substantially higher mortality from cervical cancer among women living outside the state capitals. The greater density of health services in these cities may help explain this. We considered the proximity to screening services an important factor influencing access to colposcopy.

Previous studies conducted in Brazil emphasized the difficulties in the longitudinal care of women undergoing cervical cancer screening. This is due to a lack of adherence to guidelines on the part of health care professionals and problems in the organization of the health system. Our results are similar to those from a study conducted in the state of Goiás, in which only 35% of women with abnormal cytology results (ASC-H/HSIL) underwent colposcopy. Note that, in this same study, among women with ASC-US/LSIL – for whom the recommended approach is to repeat the smear – 15% were unnecessarily referred for colposcopy.

Studies using aggregate data to estimate the number of colposcopies required at a national level produced divergent results. Using data from 2015, one study showed that twice the number of colposcopies required were performed in Brazil. This calculation was based on the number of Pap smears performed that year and their known positivity rate. However, using different – and potentially more robust – parameters to estimate the necessity for colposcopies in 2017, a second study found that the number of colposcopies performed nationally was 7% less than the required.

Two inter-related problems are likely co-existing: unnecessary colposcopies in women who have no indication and insufficient access to the procedure among those who had. This is a doubly problematic situation. The harm in not investigating women with abnormal smears is self-evident. However, colposcopy itself is not a benign procedure, carrying the inherent risks of bleeding and infection, and the chance of identifying transient lesions. The U.S. Preventive Screening Task Force considers colposcopy use to be a proxy for the harms of cervical cancer screening. Therefore, it should be reserved for and targeted towards those that need it.

In Brazil, cervical cancer screening is conducted opportunistically, putting the onus on individual women and health care providers. However, we observed widespread support for a transition to an organized program. Ideally, this would allow resources that are currently used for over screening to be focused on testing the right women – including at the point of colposcopy referral.

**Strengths and limitations**

The main strength of this study was the use of administrative databases, thus the results reflect the performance of routine screening services and not simply a single unit or trial. Using a gold standard dataset, we could validate our linkage procedure. Furthermore, we could identify only the women with an indication based on an abnormal smear result in the SISCOLO, thereby avoiding our results being distorted by inappropriate use of colposcopy.

We found some limitations. Our reference cohort may have been contaminated with prevalent cases – i.e., those already within secondary care services for cervical cancer. We could not definitively exclude all these cases, largely due to limitations of the available data. However, by removing all women with an abnormal cytology in the preceding 16 months, most were likely excluded. This is because the primary mode of identification of cases, and therefore entry into secondary screening services, is through an abnormal cytology.

The SIA/SUS follow-up period ended on December 31, 2014. This resulted in six to eight months of follow-up depending on the point of entry in the reference cohort. This period was constrained by the original ethical approval for the project. Thus, instabilities and seasonal variation in local service provision are possible, and this could have biased our estimate of access to colposcopy downwards.
if colposcopy provision in the state of São Paulo during the study (March 1, 2014, to December 31, 2014) were subject to additional pressures. However, to the best of our knowledge, the period studied is representative of the typical functioning of colposcopy services in the state of São Paulo. Furthermore, many women in our reference cohort may have eventually accessed colposcopy after the study follow-up period was completed even if, in other contexts, screen-detected cancers were defined as those diagnosed within four months of the primary screening test. Our period of follow-up substantially exceeded four months, especially considering that entry into the cohort was from the date of result release and not of smear collection.

One assumption of our study design was that women accessing primary screening on the public health system did not undergo a colposcopy in a private service. This assumption is supported by results from a recent feasibility study of HPV-based screening in São Paulo, in which less than 1% of women underwent colposcopy in the private sector.

**Conclusion**

In the state of São Paulo, only 38% of women with abnormal smears accessed colposcopy services within six months. Meanwhile, aggregate data suggest that many colposcopies are performed on women without indication. This is a waste of resources and harmful to both under- and over-screened women.
Contributors

L. F. Buss contributed to the study conception, data analysis, interpretation of results, and writing. L. Cury contributed to the study, data acquisition, interpretation of results, and critical review of the manuscript. C. M. Ribeiro contributed to the study conception, interpretation of results, and writing. G. Azevedo e Silva and J. Eluf Neto contributed to the study conception, interpretation of results, and critical review of the manuscript. All authors approved the final version of the manuscript.

Additional informations

ORCID: Lewis Fletcher Buss (0000-0002-9009-9301); Lise Cury (0000-0002-7466-9773); Caroline Madalena Ribeiro (0000-0003-2690-5791); Gulnar Azevedo e Silva (0000-0001-8734-2799); José Eluf Neto (0000-0001-7504-2115).

Acknowledgment

To the Brazilian Graduate Studies Coordinating Board (CAPES).

References


Resumo

A triagem do câncer de colo uterino é um processo que envolve múltiplas etapas. É essencial o acesso ao teste primário e aos procedimentos diagnósticos subsequentes. Com foco nas mulheres que fazem triagem no sistema de saúde pública no Estado de São Paulo, Brasil, buscamos estimar a proporção daquelas que acessam a colposcopia dentro de seis meses após um resultado anormal no teste de Papanicolau. Recuperamos os registros de duas bases de dados administrativos: o Sistema de Informação do Controle do Câncer do Colo do Útero (SISCOLO), que contém os resultados de Papanicolau, e o Sistema de Informação Ambulatorial do Sistema Único de Saúde (SIA/SUS), com os registros das colposcopias. Uma coorte de referência foi constituída de mulheres com idade de 25 anos ou mais com Papanicolau anormal entre 1º de maio de 2014 e 30 de junho de 2014. Excluímos casos prevalentes. Relacionamos a coorte de referência aos registros no SIA/SUS, estendendo até 31 de dezembro de 2014. Após a exclusão dos casos prevalentes, restaram 1.761 mulheres com citologia anormal. Setecentas delas (39,8%) foram relacionadas a um registro de colposcopia dentro do período de seguimento; esse número diminuiu para 671 (38,1%) quando o seguimento foi censurado aos seis meses. Foi observada uma cobertura ligeiramente maior entre mulheres residentes na Grande São Paulo, em comparação com mulheres residentes do interior do estado. Não houve associação entre realização de colposcopia e idade ou classe citológica. Os resultados destacam o acesso restrito à colposcopia no sistema público de saúde no Estado de São Paulo. O cenário compromete a qualidade da triagem, e a questão deve ser priorizada no planejamento dos serviços.

Neoplasias do Colo do Útero; Triagem; Colposcopia

Resumen

Las pruebas de detección del cáncer cervical forman parte de un proceso multietapa. El acceso a tanto el test primario, como a los subsiguientes procedimientos de diagnóstico, es esencial. Considerando a las mujeres que se realizan las pruebas de detección en el sistema público de la salud en el estado de São Paulo, Brasil, el objetivo del estudio fue estimar la proporción de mujeres que acceden a una colposcopia dentro de los seis meses, tras el resultado anormal en una citología. Recuperamos los registros de dos bases de datos administrativas: el Sistema de Información del Cáncer de Cuello de Útero (SISCOLO), que contiene resultados citológicos, y el Sistema de Información Ambulatorial del Sistema Único de Salud (SIA/SUS), que registra colposcopias. La cohorte de referencia consistía en mujeres, con 25 años de edad o mayores, quienes recibieron un resultado anormal en su citología entre el 1º mayo de 2014 y el 30 de junio 2014. Excluimos los casos prevalentes. Vinculamos la cohorte de referencia y los registros en la SIA/SUS, extendiéndolos hasta el 31 de diciembre 2014. Tras excluir los casos prevalentes, quedaron 1.761 mujeres con resultados citológicos anormales; 700 (39.8%), vinculados a un registro de colposcopia dentro del periodo de seguimiento. Esta cifra cayó a 671 (38.1%) cuando el seguimiento fue censurado a los seis meses. Se observó una participación ligeramente superior en mujeres que viven en la región metropolitana de la gran São Paulo, comparada con los residentes del resto del estado. No hubo asociación entre la participación en la colposcopia y edad o tipo de citología. Estos resultados resaltan que el acceso a la colposcopia en el sistema público de São Paulo es limitado. Esto compromete la calidad de las pruebas de detección por lo que se necesita darles prioridad en la planificación de los servicios sanitarios.

Neoplasias del Cuello Uterino; Triaje; Colposcopia