

AD-SISCOLO: a decision-support tool to aid the management of a cervical cancer screening program

Sulafa Yacoub Mohammed Ahmed^{1,2}, Sergio Miranda Freire^{3*}, Tereza Maria Piccinini Feitosa⁴, Lucilia Maria Gama Zardo⁵, Rosmary Terezinha de Almeida²

¹ Department of Electronic Engineering, Faculty of Engineering and Technology, University of Gezira, Wad Madani, Sudan.

² Biomedical Engineering Program, Alberto Luiz Coimbra Institute for Graduate Studies and Research in Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

³ Department of Information Technology and Education in Health, Faculty of Medical Sciences, State University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

⁴ National Cancer Institute José Gomes Alencar da Silva, Planning Division, Rio de Janeiro, RJ, Brazil.

⁵ National Cancer Institute José Gomes Alencar da Silva, Integrated Section on Technology on Citopathology, Rio de Janeiro, RJ, Brazil.

Abstract Introduction: This paper aims to develop a data warehouse (AD-SISCOLO) in order to support the management of the cervical cancer screening program in the municipality of Rio de Janeiro/Brazil. As a part of the management process, the program managers of the municipality perform tedious manual work in order to calculate a series of performance indicators and then take decisions based on them. **Methods:** AD-SISCOLO was implemented using the Pentaho BI Suite Business Intelligence Platform and the MySQL database management system. The indicators to be calculated and visualized in the tool were based on the municipal data of the cytopathology and histopathology tests from January 2012 until December 2014, which was obtained from the Information System of Cervical Cancer (SISCOLO) after a record linkage process. The follow-up indicators were based on a simplified version of the Brazilian guidelines for the cervical cancer screening. **Results:** AD-SISCOLO allows the visualization of a set of test-based and follow-up indicators from different views and dimensions, which enable managers to monitor all the phases of the screening process and to identify the process' failures. **Conclusions:** Compared with the current available environments in Brazil, AD-SISCOLO is unique in its visualization of the follow-up indicators of groups of women, according to their test results and age. Thereby it provides presentation flexibility to suit the program manager's needs.

Keywords Decision-support tools, Data visualization, Data warehouse, Cervical cancer screening, Performance indicators, Healthcare management.

Introduction

Cervical cancer is the fourth most common cancer among women worldwide. There are approximately 288,000 cases of deaths annually. The majority of these cases (80%) occur in low-income countries (World..., 2016). During the past 30 years, cervical cancer mortality

rates have fallen in the majority of the developed countries due to screening and treatment programs. However, these rates have risen or remained unchanged in the majority of the developing countries, where there is limited access to health services, lack of awareness and absence of screening and treatment programs (World..., 2016).

In Brazil, cervical cancer is the third most common cancer among Brazilian women. According to the Brazilian National Cancer Institute, there were an estimated 16,340 new cases of cervical cancer in Brazil in the year 2016 (Brasil, 2015a). The Brazilian National Program Against Cervical Cancer was established in 1998 with the objective of reducing the mortality rates from this cancer (Brasil, 2015b). Despite the existence of the program, there is still a huge variation in the mortality rates among the Brazilian states and even among the municipalities of the same state. For instance, the mortality rates in two municipalities of the state of Rio de Janeiro, Angra dos Reis and Itaguaí, are respectively,



This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ahmed SYM, Freire SM, Feitosa TMP, Zardo LMG, Almeida RT. AD-SISCOLO: a decision-support tool to aid the management of a cervical cancer screening program. Res Biomed Eng. 2018; 34(1):19-30. DOI: 10.1590/2446-4740.08117.

*Corresponding author: Departamento de Tecnologias da Informação e Educação em Saúde, Universidade do Estado do Rio de Janeiro, Av. Prof Manuel de Abreu, 2º andar, CEP 20550-170, Rio de Janeiro, RJ, Brazil. E-mail: sergio@lampada.uerj.br

Received: 11 December 2017/ Accepted: 15 January 2018

1.13 and 10.58 per 100,000 women (Brasil, 2016a). In order to monitor these actions, the Brazilian Ministry of Health created the Cervical Cancer Information System (SISCOLO, in Portuguese), which provides information related to the cytopathology and histopathology tests performed by the Brazilian Public Healthcare System at the three levels of management (federal, state and municipal).

The data from SISCOLO is used to construct a series of performance indicators, which are used in the management and monitoring of the program actions (Brasil, 2013). The indicators may be divided into two groups: those based on the performed tests (test-based indicators) and those based on the follow-up of women (woman-based indicators).

Although the SISCOLO database has data that allows the construction of both types of indicators, the majority of the constructed indicators are test-based (Feitosa and Almeida, 2007; Thuler et al., 2007). Few studies used the woman-based indicators for the evaluation of the Brazilian screening program (Bastos, 2011; Giranelli et al., 2009). This difficulty arises because SISCOLO uses the test rather than the woman as a unit of observation, and there is no unique woman identifier that could be used to link all tests performed on the same woman.

In addition, the Brazilian Ministry of Health provides some tools for the visualization of the performance indicators. The most used is TabNet (Brasil, 2008a), an online tool which permits the user to make a cross tabulation in order to visualize a set of test-based indicators related to the quality of the specimens, Papanicolaou (Pap) and histopathology tests. Other two online tools are the Cervical Cancer Indicators Dashboard (Brasil, 2017a) and Support Room for Strategic Management (SAGE in Portuguese) (Brasil, 2017b), provided by the Brazilian National Cancer Institute and the Brazilian Ministry of Health, respectively. Both tools visualize a limited number of indicators (two in SAGE and five in the Indicators Panel) through a limited number of dimensions. These tools do not enable the monitoring of the indicators in both primary healthcare units and laboratories. In addition, the visualized indicators are not regularly updated. A standalone tool called TabWin (Tabulation for Windows) (Brasil, 2008b) has statistical functionalities, but it requires the downloading of the monthly SISCOLO files and the indicators must be manually configured. The limitations of these tools are some of the obstacles in the management of the screening program, which could reduce its effectiveness (World..., 2013a).

The objective of this work was to develop a user-friendly decision-support tool for the construction and visualization of test and woman-based indicators of the Brazilian Cervical Cancer Screening Program.

Methods

Cervical screening process in Brazil

The type of cervical screening performed in Brazil is an opportunistic one, in which women spontaneously seek health services in a primary healthcare unit especially for maternal and child care (Brasil, 2015b). In the primary healthcare unit, a specimen of the cervical-vaginal material is collected and fixed on a slide. The slide is sent to the laboratory along with a requisition form filled with information about the primary healthcare unit, the woman's personal and socio-demographic information, as well as her clinical status. In the laboratory, an initial assessment is made to decide whether to accept or reject the slide. If the slide is accepted, an assessment is performed in order to determine the adequacy of the specimen, and only satisfactory specimens are subjected to cytopathological analysis. During the cytopathological assessment, all the specimens are analyzed to check the existence of epithelia of a cervix area called the Transformation Zone (TZ), since the majority of the cancer precursor lesions appear in this area (Clark et al., 2012; International..., 2005). The results of the Pap test and the information of the laboratory are filled in the same form that is sent along with the slides. In case of the cytopathological results with high-risk atypical findings, a colposcopy procedure is recommended. During the colposcopy, if necessary, a biopsy is taken in order to perform a confirmatory histopathological test (Brasil, 2011; Sellors and Sankaranarayanan, 2003).

Indicators selection

The first step in the development of the proposed decision-support tool was the selection of the performance indicators to be obtained from the tool.

A health indicator is a single measurement that is reported on regularly and that provides relevant and actionable information about population health and/or health system performance and characteristics. An indicator can provide comparable information, as well as track progress and performance over time (Canadian..., 2017). In this study, we will consider as an indicator any quantity that could be measured in order to help the management of healthcare actions.

A review of the literature was made in order to identify the performance indicators used around the world (Ahmed and Almeida, 2014). The selected set of indicators was classified into two main classes: test-based and woman-based, and then each class was divided into groups. The woman-based indicators will be called follow-up indicators from now on. The majority of the follow-up indicators were identified in the work of Bastos (2011) based on the Brazilian Guidelines for Cervical Cancer Screening for the year 2006 (Brasil, 2006a).

Test-based indicators' class

- *Adequacy of specimens/slides*: Indicators related to the adequacy of the specimens/slides, as well as the presence of material of the transformation zone in the specimens;
- *Pap test*: Indicators related to the results of the Pap test;
- *Histopathology test*: Indicators related to the results of the histopathology test.

Follow-up indicators' class

- *Participation*: Indicators related to the participation of women in the program (within the recommended age of the program);
- *Indicators related to the follow-up of 11 groups of women according to the results of the first Pap test and the recommended approaches of the Brazilian Guidelines for Cervical Cancer Screening for the year 2016* (Brasil, 2016b).

Although the SISCOLO data was collected before the year 2016, the indicators were calculated according to the approach used in the year 2016, considering that the proposed tool should be updated to meet recent and future management needs. Two main steps were performed, which were the analysis and simplification of the recommended clinical approaches, and the formulation of the algorithms for the indicators' calculation.

Analysis and simplification of the recommended clinical approaches

In this step, all the recommended clinical approaches of the Brazilian Guidelines for Cervical Cancer Screening (Brasil, 2016b) were analysed with the aid of a specialist in cytopathology, and another with experience in the Coordination of the Prevention and Early Detection of Cervical Cancer Program. After the analysis, the clinical approaches were simplified according to the available data in SISCOLO, and then the flow diagrams of the simplified approaches were drawn. In order to perform the simplification of the approaches, the following considerations were taken into account:

- Due to the absence of information on women's tests before the year 2012, the first appearance of a specific woman in the SISCOLO data in the observation period was considered as the first test for that woman;
- In the case of a woman with more than one result in the same test, the more severe result was considered (e.g., if a woman had a result of atypical squamous cells of undetermined

significance (ASC-US) and high-grade squamous intraepithelial lesion (HSIL) in the same test, the result of HSIL was considered;

- The results of the colposcopy exam were not considered for the following reasons:
 - Absence of information about the colposcopy exam for 93.9% of the women who underwent a Pap test and their clinical recommendation to undergo a colposcopy exam;
 - According to the guidelines of the year 2016, the results of the colposcopy exam were based on the field information in the updated acquisition form. This information is not available in the SISCOLO data in the observation period (2012-2014);
 - Due to the dismissal of the colposcopy results, if a woman had undergone a histopathology test after the first one, the histopathology test was selected. If the next test was a Pap test, this test was chosen.
- The time interval between tests was calculated by the difference between the release date of the test result and the collection date of the next specimen/biopsy. Each group of women was followed either until the third test or until the recommendation of the clinical approach oriented the woman either to undergo a specific procedure (such as excision of the TZ), or to go to a treatment unit, or to switch to another specific approach. An example of switching to another specific approach would be a woman with initial Pap test results of atypical glandular cells (AGC). If the result of the following histopathological test was cancer, then the woman should follow the specific conduct of women with initial results of cancer;
- The women whose recommended approach was to perform a Pap test within six months were followed until the fourth test;
- In the cases when the guidelines did not make a clear recommendation for a specific condition, the recommendation adopted was related to the most similar condition considering the level of severity of the case;
- The women who were missed or not followed-up were considered to be those women who did not have another test in the SISCOLO during the period of observation;
- The women without a result for the histopathological test were classified into a separate group.

Considering the initial recommended clinical approach for each test result, and the SISCOLO data of the observation period, the year of the first test (entry year) was defined. As an example, the entry year of women with an initial recommended clinical approach of having a Pap test within 3 years was defined to be the year 2012, since the observation period (3 years) does not permit the follow-up of the women who entered in 2013 and 2014. Taking into account the above-mentioned consideration, 11 follow-up groups of women were formed. Table 1 shows each group and its corresponding entry year.

Modeling and implementation of the decision-support tool

The decision-support tool (AD-SISCOLO) is based on a data warehouse (DW), which is commonly used for preparing and displaying the information in a user friendly and easily, accessible way that enables the user to make better and faster decisions (Kimball and Ross, 2011). The AD-SISCOLO was implemented following Kimball's dimensional model, where there are one or more central tables, called fact tables, which contain the process metrics (facts) and are linked to other tables, called dimensional tables, which give a context to the facts (Kimball and Ross, 2011).

AD-SISCOLO was developed in three steps, which are detailed below: Dimensional Modeling; Extract, Transform and Load; and Data Visualization.

Dimensional modeling

The dimensional model was created through the MySQL workbench, which is a visual tool for modeling databases and managing MySQL servers (Oracle..., 2016a). The model consisted of one fact table, which contains the

facts related to the selected groups of indicators. The fact table is surrounded by five dimensional tables, which represent: the test date; the municipality of residence of the screened women; the primary healthcare unit where the specimens were collected; the laboratory in which the specimens were analyzed; and the age group of the women at the time of the specimen's collection. These five dimensions were selected because they enable the managers to continuously monitor and evaluate all the phases of the screening process. In addition, they enable the browsing of the indicators from different views, which could enhance the decision-making process. This dimensional model was used to create the corresponding database in the MySQL Server. Table 2 shows the attributes of each dimension.

Extract, transform and load (ETL) process

The ETL process is a typical process in a data warehouse that extracts data from different sources, processes them and loads the transformed data into the data warehouse tables. In this study, the data were provided by the following sources:

- The database of the National Cervical Cancer Information System (SISCOLO) for the municipality of Rio de Janeiro, from January 2012 to December 2014;
- The auxiliary files containing data for the attributes of the primary healthcare unit, municipality and laboratory dimensions were obtained from the Brazilian Ministry of Health, and the National Register of Health Facilities home pages (Brasil, 2015c; 2015d);

Table 1. The 11 follow-up groups of women according to the initial Pap test result, age, the corresponding initial clinical approach, and entry year in the database.

Groups	Initial clinical approach	Entry year
Without abnormal results		
ASC-US - atypical squamous cells of undetermined significance (< 25 years) or LSIL - low-grade squamous intraepithelial lesion (< 25 years)	Repeat Pap test in 3 years	2012
ASC-US (between 25 and 29 years) and without abnormal results	Repeat Pap test in 12 months	2012
ASC-US (≥ 30 years) and LSIL (≥ 25 years)	Repeat Pap test in 6 months*	2013
ASC-H (high-grade intraepithelial lesions)		
AOI (atypical cells of indeterminate origin)		
AGC (atypical glandular cells)		
HSIL- high-grade squamous intraepithelial lesion (≥ 25 years)	Perform colposcopy exam	2012
HSIL (≤ 24 years)		2013
High-grade intraepithelial lesions, cannot exclude micro-invasion and squamous cell carcinoma		2014**
Adenocarcinoma in situ (AIS) or invasive		

*2014 was not considered as an entry year in this case for two reasons: 1) only women who had the pap test performed in the first six months of 2014 would be eligible to be included; 2) although the recommended approach is to repeat pap test in 6 months, usually the interval is more than that; **For the women who performed the first test in 2014, it is necessary to have at least another cytopathological or histopathological test in the database.

Table 2. Attributes of the dimensions of the DW model.

Dimension	Attributes
Time	Year and month of the test
Municipality	Name, code, name of the health region, code of the health region, and code of the federal unit
Primary healthcare unit	Name, CNES*, name of the municipality, code of the municipality, name of the health region, code of the health region, and code of the federal unit
Laboratory	Name, CNES, CNPJ**, name of the municipality, code of the municipality, name of the health region, code of the health region, and code of the federal unit
Age group	Age group of the woman

* National Register of Health Facilities (CNES, in Portuguese); **National Register of Legal Entities (CNPJ, in Portuguese).

- The data for the time and age group dimensions were obtained from the transformation of the corresponding variables in the SISCOLO database.

Although this study uses the SISCOLO data from the municipality of Rio de Janeiro, the municipality of residence dimension was included in the dimensional model because many women were from other municipalities, and some primary healthcare units from other municipalities sent their specimens to be analyzed in the municipality of Rio de Janeiro.

The fact table contains keys to the dimensional tables as well as the measurable facts that managers would want to examine. The selected indicators are either the facts or are calculated as percentages or ratios of the relevant facts. For example, the indicator “Percentage of exams with squamous cell carcinoma” is calculated through the division of the fact “total number of exams with result squamous cell carcinoma” by the fact “total number of slides with satisfactory specimens, multiplied by 100.

In the ETL process, the monthly archives of SISCOLO data in the Data Base File format (DBF) were imported to the MySQL database manager (Oracle..., 2016b) and consolidated into two tables: one called *citologico* (1,093,904 records) and another called *histologico* (4,303 records), which contain respectively the results of the cytopathology and the histopathology tests of the women. A tool called Pentaho Data Integration (PDI) was used to extract data from the auxiliary files, transform it and feed the data warehouse, by means of a series of scripts that generate the dimensional tables. PDI is one of the tools offered by the Pentaho Business Analytics Platform (PBAP) platform community edition version 6.1.0.1-196 (Pentaho..., 2017). PBAP is a JAVA-based suite of open source Business Intelligence products, which provides data integration, ETL capabilities, online analytical processing services, reporting, dashboarding and data mining.

ETL of the follow-up indicators

The construction of the follow-up indicators required changing the observation unit in SISCOLO from the test to the woman, in order to identify all the tests performed

on each woman who accessed the program. In order to change the observation unit, the literature indicates the use of a record linkage methodology, which aims at bringing together information from two different records that are believed to belong to the same person, family or entity (Fellegi and Sunter, 1969). There are two types of record linkage, deterministic and probabilistic. Deterministic record linkage is used when the entities to be linked are identified by a unique identifier (e.g. National Insurance Number), or when there are several representative identifiers whose quality of data is relatively high. The probabilistic record linkage is used when there are no common identifiers that deterministically classify a link as true or false, which is the case of the SISCOLO database. Both the *citologico* and *histologico* tables were submitted to a probabilistic record linkage process in order to identify all records in both tables belonging to the same woman. In this way, each woman received a unique identifier, and all tests performed on the same woman were associated to her identifier. The record linkage process followed the same methodology as explained in Freire et al. (2012).

After the linkage process, the *citologico* and *histologico* tables were modified through the creation and population of all required variables for the follow-up process. Those variables were created according to the simplified flow diagrams for each group in Table 1. The created algorithms for all the 11 groups of follow-up indicators were translated into a JAVA program which was executed in order to load the fact table.

Data visualization

The most popular way of organizing the facts according to the dimensions is a data cube. The data cube is a pictorial representation for a fact of interest in the n-dimensional space where the dimensions are represented along the axes of the cube and the measures are shown in each point in the cube. The cubes were generated through the Pentaho Schema Workbench, version 3.10.0 (Hitachi..., 2015) and are visualized through the Saiku Analytics plugin, version 3.8.8 (Meteorite.bi..., 2016). It provides the end-user with a simple interface, which only requires basic computer skills. The graphical

user interface (GUI) is equipped with standard security, where the user has to log in and there is an option to define various roles to control the access to the tool. After logging in, the GUI provides interactive features to perform the decision-support queries that managers normally need to address. Such features involve the drill down, roll up and slicing and dicing of data. The GUI permits the visualization of the indicators in terms of a cross-table that displays a distribution of the values of the indicator(s) on one or more dimensions. In addition, it offers a set of about 20 different graphs and tools for basic descriptive statistics. The results of the queries can be saved in different file formats such as PDF, CSV, XLS and others.

Preparation of the pre-defined reports

In order to visualize the follow-up indicators, another tool of PBIS called Pentaho Report Designer (PRD) (Hitachi..., 2017), version 5.4.1.0.130, was used to create the pre-defined reports. The PRD reports were prepared in a way that made it easy for the manager to monitor the main follow-up indicators. The simplified flow diagrams for the groups of Table 1 in the section “*Analysis and simplification of the recommended clinical approaches*” were used to design simple, compact diagrams for the visualization of the indicators.

The project’s home page (Ahmed et al., 2017) contains more information about AD-SISCOLO, such as: the AD-SISCOLO model along its data dictionary; the list of indicators available in the cubes generated through the Pentaho Schema Workbench; a tutorial on how to query the cubes and see the designed reports in the Pentaho User Console; and instructions on how to generate and install a similar AD-SISCOLO after having performed a record linkage of the tables of SISCOLO.

Ethics statement

This study was approved by the Research Ethics Committees of: 1) Hospital Universitário Clementino Fraga Filho of Universidade Federal do Rio de Janeiro/Brazil (CEP/HCFE-UFRJ – CAAE: 39106514.0.0000.5257); 2) Municipal Secretary of Health of Rio de Janeiro/Brazil (CEP/SMS-RJ – CAAE: 39106514.0.3001.5279). Given the size of the dataset used in this study and the fact that the data is retrospective, the ability to give consent was not feasible for all patients. For the purposes of this study, the information about the patients was anonymized and de-identified prior to analysis, through the removal from the database of all detailed identifying personal and demographic data of patients and health professionals. In addition, the computer used for handling the data was not connected to the network and used only by authorized persons.

Results

The AD-SISCOLO dimensional model is shown in Figure 1, with the fact table (fatos) connected to the dimensional tables (tempo - time, municipio_residencia - municipality, u_coleta - primary healthcare unit, laboratorio, and faixa_etaria – age-group) through the foreign keys (id_tempo, municipio_residencia, id_u_coleta, cnes, and faixa_etaria) respectively. In addition to the foreign keys, the fact table contains the measured facts. As an example, the first three facts are norm, asc_us and lsil, which represent the number of cytopathological tests with normal results, atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesion respectively, for each combination of the primary key values of the dimensional tables.

The cubes and reports of AD-SISCOLO are accessed through the Pentaho User Console, a web interface of Pentaho Business Analytics Platform (Universidade..., 2017). Basically, after logging into the Pentaho User Console, users can access the cubes via the Saiku Analytics plugin and then choose which cube they want to analyze, choose the metrics, the dimensions, and the respective levels in the hierarchy. The metrics are then shown in a tabular form in the graphical interface (Figure 2). It is also possible to display the results in a variety of charts, which are available in the Saiku plugin.

Through the entry screen of the Pentaho User Console, the pre-defined reports of the follow-up indicators (Figure 3) may be accessed. By clicking on the “Browse Files” button, a list of reports is presented and, after double clicking one of them, the respective report is shown, where the user may choose the period and municipality of reference.

Figure 4 shows a pre-defined report for the follow-up of women with the Pap test result of adenocarcinoma in situ or invasive adenocarcinoma.

Discussion

The AD-SISCOLO offers a visualization of a set of test-based and follow-up indicators from different views and dimensions. This multidimensional visualization of the indicators enables managers to easily monitor and evaluate all the phases of the screening process, and to identify the process’ failures (World..., 2002). This is corroborated by the International Agency for Research on Cancer that states that the integration of test-based and woman-based indicators is essential for the evaluation of the program’s overall effectiveness (International..., 2005). Several national and international studies and guidelines highlighted the essential role of those indicators in the detection and control of cancer

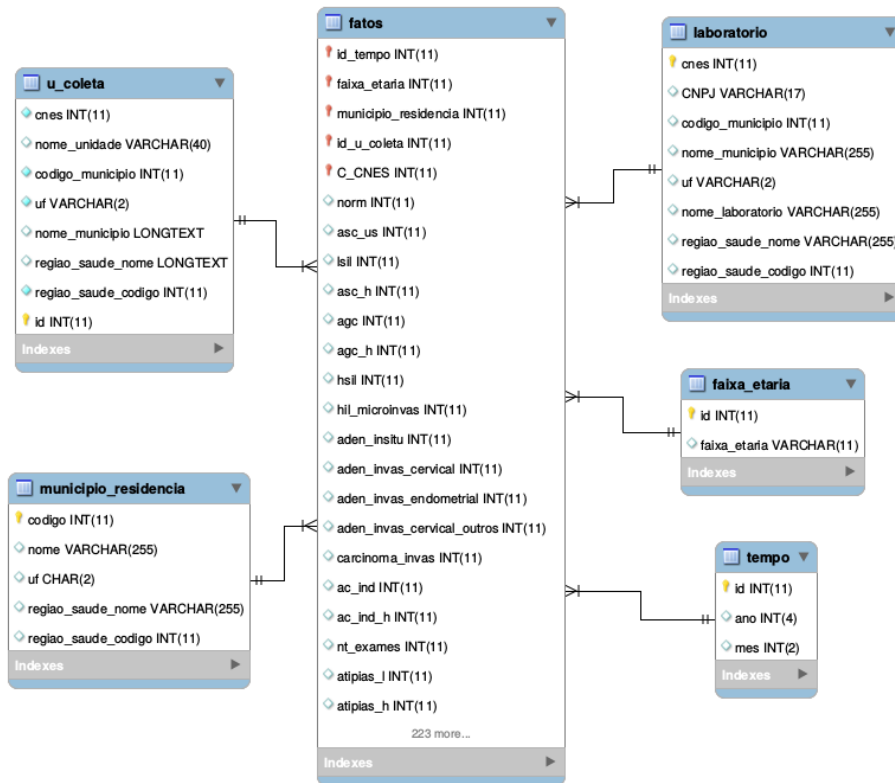


Figure 1. Dimensional model of the AD-SISCOLO.

Saiku Analytics

Cubos: Indicadores Histopatológicos

Medidas: % NIC III

Colunas: tempo

Linhas: UF, Nome da região de saúde, Nome do laboratório

Filtros: (empty)

UF	Nome da região de saúde	Nome do laboratório	% NIC III
RJ	METROPOLITANA I	HOSPITAL MARIO KROEFF	9,1
		IFF FIOCRUZ	35,5
		MS HOSPITAL DE IPANEMA	16,7
		MS HOSPITAL FEDERAL DA LAGOA	,0
		MS INCA SITEC SERV CITOPATOLOGIA	23,0
		SMS HOSPITAL MUNICIPAL DA PIEDADE AP 32	31,7
		UFRJ INSTITUTO DE GINECOLOGIA	29,6
SP	SAO PAULO	LAB AFIP	,0

Figure 2. Saiku Analytics GUI. This example shows the “% NIC III” (from the Histopatologic indicators’ cube) metric for the year 2012 and the laboratory name, municipality and state levels of the dimension Laboratorio.

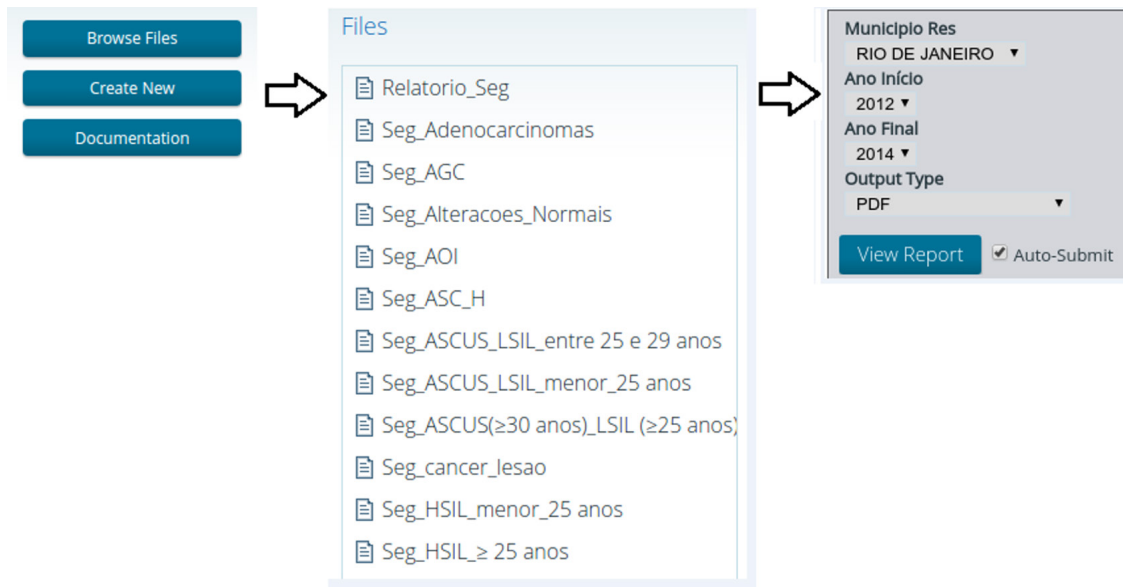


Figure 3. Steps to access and visualize a pre-defined report. A list of pre-defined reports is presented after clicking on the “Browse Files” button. After choosing one of the reports, it is shown on the screen and the user may then change the period and/or the municipality of reference.

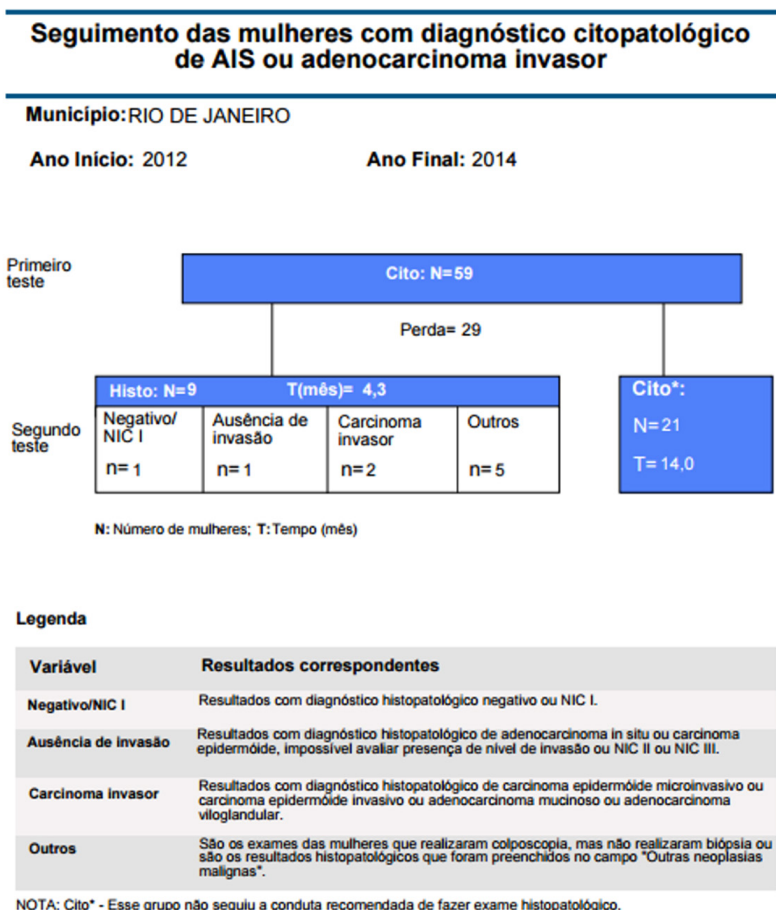


Figure 4. Output screen for the follow-up indicators for women with initial Pap test results of adenocarcinoma in situ or invasive adenocarcinoma, in the municipality of Rio de Janeiro in the period 2012-2014.

and cancer precursor lesions (Bastos, 2011; Feitosa and Almeida, 2007; Public..., 2009; The National..., 2013).

In addition, the methodology to obtain a number of follow-up indicators from the SISCOLO raw data and the Brazilian guidelines for cervical cancer screening and their visualization by groups of women, according to their test results and age, are the major contributions of this DW. These indicators are visualized in a compact diagram that follows the logic used in the follow-up diagrams of the Brazilian Guidelines for Cervical Cancer Screening (Brasil, 2016b). This follow-up should be performed systematically for all the women who enrolled in the program, not only for women with positive test results (World..., 2014). AD-SISCOLO does follow one of WHO recommendations (World..., 2014) that states that, whenever possible, research should be focused on filling gaps in information based on the needs of the country, and should generate data to guide decision-making. To this end, AD-SISCOLO offers functions not available in the current Brazilian tools that are based on SISCOLO data (TabNet, TabWin, SAGE, Cervical Cancer Indicators Dashboard), namely: availability of follow-up indicators, based on the latest Brazilian guidelines, and OLAP tool features, such as drill down, roll up and slicing and dicing of data.

Cross-sectional Brazilian studies highlighted the importance of the indicators of the laboratories internal quality control in identifying any nonconformity from the moment the specimens arrive at the laboratory until the release of the results (Araujo et al., 2015; Tobias et al., 2016). These indicators could be obtained from some tools provided by the Brazilian Ministry of Health (Brasil, 2008a; 2008b; 2017a; 2017b) indirectly, where the managers have to perform manual work in order to calculate them. On the other hand, the AD-SISCOLO enables the user to directly obtain them, which reduces the managers' effort and time as well as enables a systematic evaluation of the performance of the laboratories.

The literature describes tools to support clinical decision regarding the diagnosis or treatment of the premalignant and early stage malignant conditions, but is limited in the description of systems to support the screening actions (Agency..., 2013). A recent review of data and information systems for cervical cancer screening (Drummond et al., 2017) essentially lists recommendations for the development of such systems, such as: use of standards, minimum data sets, system integration, local and international collaboration, and avoidance of vertical systems for a single purpose. These recommendations have long been recognized by the health informatics community.

Although there are systems for cervical cancer screening, such as one developed in Argentina (World..., 2013b), to the best of our knowledge, we do not know

of any detailed description of or public access to the functionalities of such systems.

The exploratory analysis of the SISCOLO data shows that the percentages of completeness of the variables related to the women's clinical information increased by approximately double compared with that observed in the study of Bastos (2011). In addition, comparing the findings of the follow-up indicators with those presented by Bastos (2011) for the State of Rio de Janeiro, it could be observed that 10% of the target population was screened annually, which was similar to that obtained in the State, and was a third of the goal established by the Brazilian Ministry of Health (Brasil, 2006b). This value could be partially explained by the fact that Rio de Janeiro is the municipality with 53.6% of its population covered by private health plans (Agência..., 2015). The percentage of loss to follow-up for women eligible to repeat a Pap test in six months was 43.1% in the municipality of Rio de Janeiro and 63.7% in the State of Rio de Janeiro. The average time to perform the test was similar, around 10 months, for both the municipality and the State. On the other hand, the average time to perform a histopathology test for the women with positive test results in the second Pap test was 6.5 months, which was longer than the 4 months observed for the State. The flow diagrams used in this study were based on the Brazilian Guidelines for Cervical Cancer Screening, updated in 2016. Therefore, the findings in this study should be interpreted with caution since the DW data set covers a period previous to 2016, when the clinical approaches for the problem was slightly different.

It is worth mentioning that the majority of women in the group without atypical findings in the first test had only one test in the period of observation, which is an expected situation, since the majority of these women were not undergoing the test for the first time and the interval for a second test is three years, which spans more than the observed period of time. There was a subgroup of women in this group who presented a second test with atypical findings with high risk for cancer or even cancer in an interval of 15 months. This is the type of finding which should be seen as an alert, since it is an unexpected situation considering the natural history of the disease (Forsmo et al., 1997).

In this work, the authors were allowed to use only the data for the municipality of Rio de Janeiro in the period from 2012 to 2014. In order to expand the use of the tool at the State or National level of management the SISCOLO data of the whole State/Country should be included. Considering this limitation, the analysis of the indicators for other municipalities should be considered with caution, since it reflects only the proportion of tests that are performed in Rio de Janeiro. In the same vein,

there is a need to complement the dataset in order to have a longer period of follow up. If the system were continuously updated, the entry year would not be as important as it is in this study because there would be a period of time that would provide a more representative picture of the cervical cancer screening status.

AD-SISCOLO is not difficult to install and run, but it requires an information technology (IT) technical staff to keep it running. However the update of AD-SISCOLO with new data requires, besides the permission to the data access, that a new record linkage and the population of the database be performed, which are also activities that demand the support of an IT technical staff.

AD-SISCOLO provides flexible visualization features for both test-based and follow-up indicators that overcomes some of the limitations of the current tools provided by the Brazilian Ministry of Health. The DW offers an aggregated view of the follow-up process by groups of clinical recommendations, which makes it a valuable complement for the other tools that focus on the individual follow-up of women with atypia. AD-SISCOLO may be seen as a decision support tool for managers since it increases their capacity to monitor, evaluate, control and plan the program actions by means of new and accessible indicators.

Acknowledgements

The authors thank TWAS-CNPq Postgraduate Fellowship Programme (Grant: 190023/2012-0), as well as the Gezira University, Wad Madni, Sudan for the work leave of the first author to carry out her doctoral course in Brazil. The authors are also grateful to the *Coordenação das Linhas de Cuidado das Doenças Crônicas Não Transmissíveis da Secretária Municipal de Saúde do Rio de Janeiro* for the provision of the data used in this study.

References

- Agency for Healthcare Research and Quality – AHRQ. Decision support tools for screening and treatment decisions in early cancer. Rockville: AHRQ; 2013 [cited 2017 Dec 9]. Available from: <https://effectivehealthcare.ahrq.gov/ehc/products/529/1749/cancer-decision-support-tools-protocol-131112.pdf>
- Agência Nacional de Saúde Suplementar – ANS. Beneficiários de planos privados de saúde, por cobertura assistencial (Brasil - 2003-2011). Rio de Janeiro: ANS; 2015 [cited 2017 Dec 9]. Available from: http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Perfil_setor/Caderno_informacao_saude_suplementar/2015_mes12_caderno_informacao.pdf
- Ahmed SYM, Almeida RT. A review of performance indicators for evaluation of cervical cancer screening programs. In: Anais do XXIV Congresso Brasileiro de Engenharia Biomédica; 2014; Uberlândia, Brasil. Rio de Janeiro: SBEB; 2014 [cited 2017 Dec 9]. Available from: http://www.canal6.com.br/cbeb/2014/artigos/cbeb2014_submission_207.pdf
- Ahmed SYM, Freire SM, Almeida RT. Indicadores de eficácia para a avaliação das ações de um programa de rastreamento do câncer do colo de útero. Rio de Janeiro: UERJ; 2017 [cited 2017 Dec 9]. Available from: <http://www.lampada.uerj.br/adsiscolo>
- Araujo MLC Jr, Santana DA, Almeida LB, Quintana SBS, Silva GRF, Fonseca RCSP. Quality in cytopathology: an analysis of the internal quality monitoring indicators of the Instituto Nacional de Câncer. J Bras Patol E Med Lab. 2015; 51(2):102-7. <http://dx.doi.org/10.5935/1676-2444.20150018>.
- Bastos EA. Estimativa da Efetividade do Programa de Rastreamento do Câncer do Colo do Útero no Estado do Rio de Janeiro [dissertation]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 2011 [cited 2017 Dec 9]. Available from: http://www.peb.ufrj.br/teses/Tese0143_2011_08_12.pdf
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Coordenação de Prevenção e Vigilância. Nomenclatura brasileira para laudos cervicais e condutas preconizadas: recomendações para profissionais de saúde. 2nd ed. Rio de Janeiro: INCA; 2006a [cited 2017 Dec 9]. Available from: http://bvsm.s.saude.gov.br/bvs/publicacoes/Nomenclaturas_2_1705.pdf
- Brasil. Ministério da Saúde. Gabinete do Ministro. Portaria nº 399/GM de 22 de fevereiro de 2006. Divulga o Pacto pela Saúde 2006 – Consolidação do SUS e aprova as Diretrizes Operacionais do Referido Pacto. Diário Oficial da República Federativa do Brasil, Brasília, fev. 2006b [cited 2017 Dec 9]. Available from: http://bvsm.s.saude.gov.br/bvs/saudelegis/gm/2006/prt0399_22_02_2006.html
- Brasil. Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde – DATASUS. Sistema de informação do câncer do colo do útero e mama: informações estatísticas. Brasília; 2008a [cited 2017 Dec 9]. Available from: <http://w3.datasus.gov.br/siscam/index.php?area=0401>
- Brasil. Ministério da Saúde. Departamento de Informática do Sistema Único de Saúde – DATASUS. Tab para Windows – TabWin. Brasília; 2008b [cited 2017 Dec 9]. Available from: <http://www2.datasus.gov.br/DATASUS/index.php?area=060805>
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer – INCA. Sistemas de informação do controle do câncer de mama e do colo do útero: manual gerencial. Rio de Janeiro: INCA; 2011 [cited 2017 Dec 9]. Available from: http://www1.inca.gov.br/inca/Arquivos/manual_gerencial_2011.pdf
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Sistema de informação do câncer: manual preliminar para apoio à implantação. Rio de Janeiro: INCA; 2013 [cited 2017 Dec 9]. Available from: http://bvsm.s.saude.gov.br/bvs/publicacoes/inca/siscan_manual_preliminar.pdf
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Câncer do colo do útero. Rio de Janeiro: INCA; 2015a [cited 2017 Dec 9]. Available from: http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home++/colo_uterio/definicao
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Controle do câncer do colo

- do útero, histórico das ações. Rio de Janeiro: INCA; 2015b [cited 2017 Dec 9]. Available from: http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_nacional_controle_cancer_colo_uterio/historico_acoes
- Brasil. Ministério da Saúde. Departamento de Informática do SUS. Cadastro Nacional de Estabelecimentos de Saúde. Brasília; 2015c [cited 2017 Dec 9]. Available from: <http://cnes.datasus.gov.br/>
- Brasil. Ministério da Saúde. Departamento de Informática do SUS. Tabelas nacionais. Brasília; 2015d [cited 2017 Dec 9]. Available from: <ftp://ftp.datasus.gov.br/territorio/tabelas>
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Atlas de Mortalidade por Câncer. Rio de Janeiro: INCA; 2016a [cited 2017 Dec 9]. Available from: <http://www1.inca.gov.br/vigilancia/mortalidade.asp>
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Coordenação de Prevenção e Vigilância – Conprev. Divisão de Detecção Precoce e Apoio à Organização de Rede. Diretrizes brasileiras para o rastreamento do câncer do colo do útero. 2nd ed. Rio de Janeiro: INCA; 2016b [cited 2017 Dec 9]. Available from: http://www1.inca.gov.br/inca/Arquivos/Diretrizes_para_o_Rastreamento_do_cancer_do_colo_do_uterio_2016_corrigeo.pdf
- Brasil. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva – INCA. Painel de indicadores do câncer do colo de útero. Rio de Janeiro: INCA; 2017a [cited 2017 Dec 9]. Available from: http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_nacional_controle_cancer_colo_uterio/indicadores/
- Brasil. Ministério da Saúde. Sala de Apoio a Gestão Estratégica – SAGE. Prevenção e tratamento de câncer de colo e mama. Brasília; 2017b [cited 2017 Dec 9]. Available from: <http://sage.saude.gov.br/>
- Canadian Institute for Health Information – CIHI. CIHI Health Indicators. Canada; 2017 [cited 2017 Dec 9]. Available from: <https://www.cihi.ca/en/cihi-health-indicators>
- Clark I, Whiting P, Twin J, Docker J, Ferguson J. Pathology handbook. 5th ed. Canberra: Buckner Printing Company; 2012.
- Drummond JL, Were MC, Arrossi S, Wools-Kaloustian K. Cervical cancer data and data systems in limited-resource settings: challenges and opportunities. *Int J Gynaecol Obstet.* 2017; 138(Suppl 1):33-40. PMID:28691330. <http://dx.doi.org/10.1002/ijgo.12192>.
- Feitosa TMP, Almeida RT. Pap smear screening for the control of cervical cancer in Minas Gerais State, Brazil, 2002. *Cad Saude Publica.* 2007; 23(4):907-17. PMID:17435888. <http://dx.doi.org/10.1590/S0102-311X2007000400018>.
- Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc.* 1969; 64(328):1183-210. <http://dx.doi.org/10.1080/01621459.1969.10501049>.
- Forsmo S, Buhaug H, Skjeldestad F, Haugen O. Treatment of pre-invasive conditions during opportunistic screening and its effectiveness on cervical cancer incidence in one Norwegian county. *Int J Cancer.* 1997; 71(1):4-8. PMID:9096657. [http://dx.doi.org/10.1002/\(SICI\)1097-0215\(19970328\)71:1<4::AID-IJC2>3.0.CO;2-Y](http://dx.doi.org/10.1002/(SICI)1097-0215(19970328)71:1<4::AID-IJC2>3.0.CO;2-Y).
- Freire SM, Almeida RT, Cabral MDB, Bastos EA, Souza RC, Silva MGP. A record linkage process of a cervical cancer screening database. *Comput Methods Programs Biomed.* 2012; 108(1):90-101. PMID:22341207. <http://dx.doi.org/10.1016/j.cmpb.2012.01.007>.
- Girianelli VR, Thuler LCS, Silva GA. Qualidade do sistema de informação do câncer do colo de útero no estado do Rio de Janeiro. *Rev Saude Publica.* 2009; 43(4):580-8. PMID:19618027. <http://dx.doi.org/10.1590/S0034-89102009005000043>.
- Hitachi Vantara Community. Pentaho Schema Workbench. 2015 [cited 2017 Dec 9]. Available from: <https://sourceforge.net/projects/mondrian/files/schema%20workbench/>
- Hitachi Vantara Community. Pentaho Report Designer. 2017 [cited 2017 Dec 9]. Available from: <https://sourceforge.net/projects/pentaho/files/Report%20Designer/>
- International Agency for Research on Cancer – IARC. Cervix cancer screening: IARC Handbook of Cancer Prevention. 1st ed. Lyon: IARC Press; 2005 [cited 2017 Dec 9]. Available from: <http://www.iarc.fr/en/publications/pdfs-online/prev/handbook10/>
- Kimball R, Ross M. The data warehouse took kit: the complete guide to dimensional modeling. 2nd ed. New York: John Wiley & Sons; 2011.
- Meteorite.bi Company. Saiku analytics. 2016 [cited 2017 Dec 9]. Available from: <http://www.meteorite.bi/products/saiku>
- Oracle Corporation. MySQL WorkBench. 2016a [cited 2017 Dec 9]. Available from: <https://dev.mysql.com/downloads/workbench/>
- Oracle Corporation. MySQL database. 2016b [cited 2017 Dec 9]. Available from: <http://www.mysql.com>
- Pentaho Corporation. Pentaho Suite Business Intelligence. 2017 [cited 2017 Dec 9]. Available from: <http://www.pentaho.com>
- Public Health Agency of Canada – PHAC. Performance monitoring for cervical cancer screening programs in Canada. Canada; 2009 [cited 2017 Dec 9]. Available from: <https://www.canada.ca/content/dam/phac-aspc/migration/phac-aspc/cd-mc/cancer/pmccspc-srpccuc/pdf/cervical-eng.pdf>
- Sellors JW, Sankaranarayanan R. Colposcopy and treatment of cervical intraepithelial neoplasia: a beginners' manual Lyon: IARC Press; 2003.
- The National Cervical Screening Programme of Republic of Ireland – NCSP. Guidelines for quality assurance in cervical screening. 2nd ed. Limerick; 2013 [cited 2017 Dec 9]. Available from: <https://www.cervicalcheck.ie/fileupload/QualityAssurance/NCSS-PUB-Q-1%20Guidelines%20for%20Quality%20Assurance%20in%20Cervical%20Screening.pdf>
- Thuler LCS, Zardo LMG, Zeferino LC. Perfil dos laboratórios de citopatologia do Sistema Único de Saúde. *J Bras Patol E Med Lab.* 2007; 43(2):103-14. <http://dx.doi.org/10.1590/S1676-24442007000200006>.
- Tobias AHG, Amaral RG, Diniz EM, Carneiro CM. Quality indicators of cervical cytopathology tests in the public service in Minas Gerais, Brazil. *Rev Bras Ginecol Obstet.* 2016; 38(2):65-70. PMID:26883863. <http://dx.doi.org/10.1055/s-0035-1571175>.

Universidade do Estado do Rio de Janeiro – UERJ. AD-SISCOLO: rastreamento do câncer de colo de útero: indicadores de desempenho. Rio de Janeiro; 2017 [cited 2017 Dec 9]. Available from: <http://pentaho.lampada.uerj.br:8080/pentaho/Login>

World Health Organization – WHO. Cervical cancer screening in developing countries. Geneva; 2002 [cited 2017 Dec 9]. Available from: <http://apps.who.int/iris/bitstream/10665/42544/1/9241545720.pdf>

World Health Organization – WHO. Monitoring national cervical cancer prevention and control programmes: quality control and quality assurance for visual inspection with acetic acid (VIA)-based programmes. Geneva; 2013a [cited 2017 Dec 9]. Available from: http://apps.who.int/iris/bitstream/10665/79316/1/9789241505260_eng.pdf

World Health Organization – WHO. Compendium of innovative health technologies for low-resource settings: assistive devices: ehealth solutions: medical devices. Geneva; 2013b [cited 2017 Dec 9]. Available from: http://apps.who.int/iris/bitstream/10665/202537/2/9789241509992_eng.pdf

World Health Organization – WHO. Comprehensive cervical cancer control, a guide to essential practice. 2nd ed. Geneva; 2014 [cited 2017 Dec 9]. Available from: http://apps.who.int/iris/bitstream/10665/144785/1/9789241548953_eng.pdf

World Health Organization – WHO. Screening as well as vaccination is essential in the fight against cervical cancer. Geneva; 2016 [cited 2017 Dec 9]. Available from: <http://www.who.int/reproductivehealth/topics/cancers/fight-cervical-cancer/en/>