COPPE
UFRJ
Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia

DEVELOPMENT OF A DATA WAREHOUSE TO SUPPORT THE EVALUATION OF A CERVICAL CANCER SCREENING PROGRAM

Sulafa Yacoub Mohammed Ahmed

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Biomédica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Biomédica.

Orientadores: Rosimary Terezinha de Almeida
Sergio Miranda Freire

Rio de Janeiro
Março de 2017

DEVELOPMENT OF A DATA WAREHOUSE TO SUPPORT THE EVALUATION OF A CERVICAL CANCER SCREENING PROGRAM

Sulafa Yacoub Mohammed Ahmed

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA BIOMÉDICA.

Examinada por:

____

Prof^{a}. Rosimary Terezinha de Almeida, PhD.

____

Prof. Renan Moritz Varnier Rodrigues de Almeida, PhD.

____

Prof^{a}. Yara Lucia Mendes Furtado de Melo, DSc.

____

Prof. Fábio Bastos Russomano, DSc.

____

Dr^{a}. Liz Maria de Almeida, DSc.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2017

Ahmed, Sulafa Yacoub Mohammed

Development of a Data Warehouse to Support the Evaluation of a Cervical Cancer Screening Program/Sulafa Yacoub Mohammed Ahmed. – Rio de Janeiro: UFRJ/COPPE, 2017.

XVI, 137 p.: il.; 29,7 cm.

Orientadores: Rosimary Terezinha de Almeida
Sergio Miranda Freire

Tese (doutorado) – UFRJ/ COPPE/ Programa de Engenharia Biomédica, 2017.

Referências Bibliográficas: p. 101-111.

1. Cervical cancer screening. 2. Performance indicators. 3. Data warehouse. I. Almeida, Rosimary Terezinha de *et al*. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Biomédica. III. Título.

DEDICATION

This thesis is lovingly dedicated to my great family, whose support, encouragement, and constant love have sustained me throughout my life.

To “Kaltoum Bit Mahmoud” for the great lessons, love, joy, and wonderful moments we spent together.

To the memories of Prof. Abdallah Said Ahmed

&

To my beloved country

ACKNOWLEDGMENTS

God for the unconditional love.

Sincere gratitude to my supervisors Dr. Rosimary Terezinha de Almeida and Dr. Sergio Miranda Freire who have put their valuable experience and knowledge at my disposal, encourage, help and support me during every moment of these four years.

Teaching and administrative staff of the Biomedical Engineering Department.

Friends of the LESS laboratory especially Alfonso, Karla, Leticia and Rogério for the unforgettable nice moments we spent together.

Dr. Tereza Feitosa and Dr. Lucilia Zardo for the valuable discussions and help during the project.

Gretha, Telma, Aunt Jane e Mrs. Doris for the kindness, warm hugs and lifelong friendship.

Dr. Noha Hassan and Dr. Mohamed Sanad for the great example of a real human being.

Brazil, the amazing country that embraced me and gave me a chance to pass through one of the valuable experiences in my life.

CNPq-TWAS for the DSc fellowship that makes this thesis possible.

Gezira University for the all support since I was BSc student until now.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

DESENVOLVIMENTO DE UM ARMAZÉM DE DADOS PARA A AVALIAÇÃO DE UM PROGRAMA DE RASTREAMENTO DO CÂNCER DO COLO DO ÚTERO

Sulafa Yacoub Mohammed Ahmed

Março/2017

Orientadores: Rosimary Terezinha de Almeida
Sergio Miranda Freire

Programa: Engenharia Biomédica

O objetivo desse trabalho foi desenvolver um armazém de dados (AD) para apoiar a gestão do programa de rastreamento do câncer do colo do útero no município do Rio de Janeiro. O processo de gestão do programa exige o trabalho manual e tedioso para calcular muitos dos indicadores de desempenho. O AD desenvolvido foi implementado usando a plataforma de *business intelligence* Pentaho BI Suite e o gerenciador de banco de dados MySQL. Os indicadores a serem calculados e visualizados no AD foram baseados nos dados dos testes citopatológicos e histopatológicos do município do Rio de Janeiro de janeiro 2012 a dezembro de 2014, obtidos do Sistema de Informação do Câncer de Colo do Útero (SISCOLO). O AD permite a visualização de um conjunto de indicadores de desempenho baseados nos dados dos testes e das mulheres a partir de diferentes visões e dimensões, o que permite monitorizar todas as fases do processo de rastreamento e identificar as falhas. Comparado aos ambientes disponíveis, o AD é único na visualização dos indicadores de acompanhamento das mulheres, de acordo com o resultado do teste e a idade. Desta forma, o AD oferece flexibilidade de apresentação dos indicadores para atender às necessidades dos gerentes do programa.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

DEVELOPMENT OF A DATA WAREHOUSE TO SUPPORT THE EVALUATION OF A CERVICAL CANCER SCREENING PROGRAM

Sulafa Yacoub Mohammed Ahmed

March/2017

Advisors: Rosimary Terezinha de Almeida
Sergio Miranda Freire

Department: Biomedical Engineering

The aim of this work was to develop a data warehouse (DW) 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 test-based and woman-based performance indicators. The developed DW was implemented using the Pentaho BI Suite business intelligence platform and the MySQL database manager. The indicators to be calculated and visualized in the DW were based on the municipal data of cytopathology and histopathology tests from January 2012 until December 2014, which was obtained from the Cervical Cancer Information System (SISCOLO). The developed DW allowed the visualization of a set of test-based and woman-based indicators from different views and dimensions, which enable managers to monitor all the phases of the screening process and to identify the process' failures. Compared with the current available environments, this DW is unique in the visualization of the follow-up indicators of group of women, according to their test results and age. Thereby it provides presentation flexibility to suit the program manager's needs.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES

LIST OF ABBREVIATIONS

AGC	Atypical glandular cells of undetermined significance – possibly non-neoplastic
AGC-H	Atypical glandular cells of undetermined significance – cannot exclude high-grade intraepithelial lesion
AIS	Adenocarcinoma in situ
ASC-H	Atypical squamous cells of undetermined significance – cannot exclude high-grade intraepithelial lesion
ASC-US	Atypical squamous cells of undetermined significance – possibly non-neoplastic
AOI-H	Atypical squamous cells of undetermined significance with unknown origin – cannot exclude high-grade intraepithelial lesion
AOI-non-neoplastic	Atypical squamous cells of undetermined significance with unknown origin – possibly non-neoplastic
BVS	*Biblioteca Virtual em Saúde*
CEP	Postal address code (*Código de Endereçamento Postal*, in Portuguese)
CIN I	Cervical Intraepithelial Neoplasia grade 1
CIN II	Cervical Intraepithelial Neoplasia grade 2
CIN III	Cervical Intraepithelial Neoplasia grade 3
CMV	Cytomegalovirus
CNES	National Register of Health Facilities (*Cadastro Nacional de Estabelecimentos de Saúde*, in Portuguese)
CNPJ	National Register of Legal Entities (*Cadastro Nacional de Pessoas Jurídicas*, in Portuguese)
CPF	Individual Identification Number (*Cadastro de Pessoas Físicas,* in Portuguese)
DA	Data Access
DATASUS	Health Informatics Department of the Brazilian Ministry of Health *(Departamento de Informática do Sistema Único de Saúde*, in Portuguese)
DW	Data warehouse

ER	Entity Relationship
ETL	Extract-transform-load
ETZ	Excision of transformation zone
GUI	Graphical user interface
HCFF-UFRJ	*Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro*
HPV	Human papillomavirus
HSIL	High-grade intraepithelial lesion
IARC	International Agency for Research on Cancer
INCA	Brazilian National Cancer Institute (*Instituto Nacional de Câncer*, in Portuguese)
LSIL	Low-grade squamous intraepithelial lesions
MIQ	Internal Quality Monitoring of the Cytopathological Examination (*Monitoramento Interno da Qualidade do Exame citopatológico*, in Portuguese)
PBIS	Pentaho BI Suite
PDI	Pentaho Data Integration
PRD	Pentaho Report Designer
PSW	Pentaho Schema Workbench
SAGE	Strategic Management Supporting Hall (*Sala de Apoio à Gestão Estratégica*, in Portuguese)
SCJ	Squamocolumnar junction
SIM	Hospital Information System (Sistema de Informações Hospitalares, in Portuguese)
SIM	Information System on Mortality (*Sistema de Informação Sobre Mortalidade*, in Portuguese)
SIA	Outpatient Information System (*Sistema de Informação Ambulatorial*, in Portuguese)
SISCOLO	Cervical Cancer Information System (*Sistema de Informação do Câncer do Colo do Útero*, in Portuguese)
SISCAN	Cancer Information System (*Sistema de Informação do Câncer*, in Portuguese)

SISMAMA	National Program for Control of Breast Cancer (*Sistema de Informação do Câncer de Mama*, in Portuguese)
SMS-RJ	Municipal Secretary of Health of municipality of Rio de Janerio (*Secretaria Municipal de Saúde do Rio de Janeiro,* in Portuguese)
SQL	Structured query language
TabNet	*Tabulador para Internet*
TABWIN	*Tabulador para windows*
TZ	Transformation zone
WHO	World Health Organization

CHAPTER 1 – INTRODUCTION

Cervical cancer is the second 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, where the access to cervical cancer screening is poor or does not even exist (IARC, 2015; WHO, 2016). In some regions of the world, cervical cancer threatens women's lives more than the pregnancy-related causes. This condition not only affects the women’s health and lives, but also their children, families, and communities (WHO, 2004). According to the Brazilian National Cancer Institute (INCA, in Portuguese), there were an estimated 16,340 new cases of cervical cancer in Brazil for the year 2016. This is the third most common type of cancer among women in Brazil and the first and second most common type in the North and Northeast regions, respectively (BRAZIL, 2015a).

Cervical cancer screening is a public health intervention to detect and treat individuals with increased probability of having either the disease itself or a precursor of the disease (WHO, 2014). In order to achieve this goal, an organized screening program should be established to offer diagnostic services, follow-up and treatment (WHO, 2007).

There are several diagnostic tests that can be used for cervical cancer screening. The cytopathologic Pap smear test is the only one that has been used in large populations and has been tremendously successful in reducing the incidence and mortality rates by cervical cancer (WHO, 2012). According to the World Health Organization (WHO), the detection of cervical abnormalities by applying a cytopathologic Pap smears test could reduce the development of cervical cancer by up to 80% (WHO, 2002).

In the 1960s and 1970s, the high-income countries’ incidence rates were similar to those of the developing countries nowadays. The decline of the incidence and mortality rates in the majority of the high-income countries was due to the establishment of effective screening programs (GAKIDOU *et al.*, 2008). On the other hand, some of the high-income countries showed an increasing incidence and mortality rates. This increment was due to the lack of effective screening, the low population coverage and the poor quality of the Pap smear test (LAZCANO-PONCE *et al.*, 1998).

There are many elements for an effective screening program, among them: the ability of the program to ensure a high level of coverage for the target population; the availability of high quality laboratory services and well-trained healthcare professionals (including smear collectors, cytopathologists, colposcopists and program managers);

rapid transportation of specimens to the laboratory; the quality control of cytopathologic reading; a well organized system to deliver the test results to women in an understandable way, as well as follow up and adequate treatment guarantee. In order to manage the screening program, these elements should be monitored and evaluated regularly (WHO, 2002).

The monitoring of screening programs requires a well-organized information system to record all the phases of the screening. These phases should be recorded in a way that allows the reconstruction of the pathway of the women attending the program as well as the provision of means of evaluation of the program in many dimensions. The most frequent dimensions found in the literature are population coverage; adequacy of Pap smear specimens; screening test results; and follow-up of positive tests (IARC, 2005; WHO, 2004; WHO, 2007). Monitoring and evaluation of these dimensions requires a construction of sets of measurable performance indicators for each dimension. In general, these indicators could be classified into test-based and woman-based, where their integration is essential for the evaluation of the program overall effectiveness (IARC, 2005; MILLER, 1992).

The test-based indicators are mainly related to the test production and deal with the test as a unit of observation. The adequacy of Pap smear specimens and test results are considered to be some of the core test-based indicators for detection and control of the cancer precursor lesions. At least one-half to two-thirds of false negatives is the result of inadequate specimens (IARC, 2005). On the other hand, monitoring of the percentages of the test results over time has been used as an indicator by program managers to ensure that the program has achieved its goal of detection of precancerous lesions as well as those actions be taken in cases of the increase of abnormal or advanced lesions (CPAC, 2011).

The woman-based indicators deal with woman as a unit of observation, which permit the follow-up of the women during the screening process as well as the calculation of the population coverage. The follow-up of women with positive test results ensures that they are referred to the specialized services as well as the determination of whether further treatment or rehabilitation is required. In the Latin American countries, the lack of follow-up of women with positive tests is one of the main barriers against achieving the required performance of the screening programs (GAGE *et al.*, 2003; IARC, 2005; ROBLES, 2004; SANTIAGO *et al.*, 2003). The indicator for population coverage refers to the proportion of women in the target population actually screened at least once during the recommended interval by the screening program. The number of performed Pap smear tests rather than the number of women could be used as a proxy of the coverage, but this is not an indication of the real coverage, since this number may include women outside the target age group, and women screened more than once (NCCSPRI, 2014; WHO, 2004; WHO, 2014).

In Brazil, the National Program Against Cervical Cancer was established in 1998 by the Brazilian Ministry of Health. The program actions were initiated to formulate guidelines and structures to achieve the goal of early detection of cancer and its precursor lesions and monitor the quality of care for women in all stages of disease (prevention, early detection, treatment and rehabilitation) as well as to offer proper treatment of the disease and its precursor lesions (BRAZIL, 2017). Despite 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 Itaguai, are respectively, 1.13 and 10.58 per 100,000 women (BRAZIL, 2016a). This variation led the program managers to evaluate and monitor the program actions at the three levels of management (federal, state and municipal). In order to monitor and evaluate these actions, the Cervical Cancer Information System (SISCOLO, in Portuguese) was established in 1999 by INCA in a partnership with the Health Informatics Department of the Brazilian Ministry of Health (DATASUS, in Portuguese). The SISCOLO system provides information related to the cytopathologic and histopathologic tests performed by the Brazilian Public Healthcare System. The data from SISCOLO are used to construct a series of performance indicators, which are used in the management and monitoring of the program actions (BRAZIL, 2013a).

Several authors have been using the SISCOLO database to construct test-based indicators such as THULER et al. (2007), Cabral *et al*. (2008), Feitosa *et al.* (2007), Souza *et al.* (2012). Although the SISCOLO database has data that allows the construction of test-based indicators as well as woman-based indicators, the majority of the constructed indicators are test-based. Due to the difficulty of identifying all the test results performed by the same women, there are few studies that used the woman-based indicators for the evaluation of the Brazilian screening program (BASTOS, 2011; GIRIANELLI *et al.*, 2009). This difficulty arises because SISCOLO uses the test rather than woman as a unit of observation. There is no unique woman identifier that could be used to link all tests performed by the same woman.

The construction of woman-based indicators requires 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 (FELLEGI *et al.*, 1969; NEWCOMBE, 1967). The record linkage methodology is based on bringing together information from two different records that are believed to belong to the same person, family or entity. 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.

Cabral (2010) applied the Fellegi-Sunter probabilistic record linkage methodology to a sample of SISCOLO for the State of Rio de Janeiro. The proposed methodology showed the potential of the SISCOLO database to construct performance indicators to evaluate the effectiveness of the screening. FREIRE *et al.* (2012) linked the records of SISCOLO from 2006 to 2009, with sensitivity above 90% and specificity near 100%. That study showed that it is possible to integrate SISCOLO to produce the required performance indicators for the evaluation of the program actions taking a woman as a unit of observation. This integrated database was used by Bastos (2011) to estimate the effectiveness of the screening program for the State of Rio de Janeiro by means of a set of woman-based indicators.

These cross-section studies revealed the situation of the program in a specific period. However, the continuous evaluation of the program requires the use of an automatic evaluation tool. In order to achieve this goal, the Brazilian Ministry of Health provides some tools with different visualization techniques of the performance indicators (BRAZIL, 2013a; BRAZIL, 2015b; BRAZIL, 2015c; BRAZIL, 2015d). Although these environments are considered useful evaluation tools, they lack the visualization of the woman-based indicators. In addition, they have some limitations that obstruct the evaluation process. In some of these environments, the program managers need to make an effort to put the indicators in a form that could help decision-making. Some of those environments provide the indicators in an absolute form, which does not allow the comparative analysis between primary healthcare units, laboratories and municipalities. Another limitation is that they offer a very limited graphical representation (few charts and graphs). Since none of these environments contains all the required indicators for the evaluation process, managers need to consult more than one environment.

Due to the mentioned limitations of the available online tools, the local managers have to perform tedious manual work in order to construct and visualize the performance indicators in the desired form. They download monthly SISCOLO data files and then manually construct the indicators using a spreadsheet, statistical tools or a stand-alone tool offered by the Ministry of Health (BRAZIL, 2011a; BRAZIL, 2016b). This manual process is time-consuming and requires personnel with some basic knowledge of data management, which is usually not the case. In addition, this process does not provide a friendly user interface for the continuous monitoring of the indicators. These limitations are considered one of the main obstacles in the management of the screening program.

From what was mentioned above, there is a necessity to develop a user-friendly environment for the automatic construction and visualization of both woman-based and test-based indicators. Nowadays, data warehouses (DW) are commonly used for preparing and displaying information in a friendly way. DW could be a very useful tool in the monitoring and evaluation of screening program actions.

The thesis was organized in seven chapters. The second chapter describes the main and the specific objectives of the thesis. Chapter 3 provides a background about cervical cancer, the evaluation of the screening programs, the record linkage process and the data warehouse as well as a historical view of the Brazilian Cervical Cancer Screening Program. Chapter 4 presents the literature review about the performance indicators used for the evaluation of screening programs worldwide and in Brazil, as well as the identified Brazilian tools for the visualization of the performance indicators. Chapter 5 presents a description of the whole DW building process, including the selection of the indicators, the description, cleanness and record linkage of data and the DW building. Chapter 6 presents the visualization of the performance indicators through the graphical interface of the DW. Chapter 7 discusses the DW visualization of the indicators, the limitations and future developments for the proposed DW as well as the conclusions drawn from the study.

CHAPTER 2 – OBJECTIVES

The objective of this study is to develop a data warehouse to support the evaluation of a cervical cancer screening program by enabling the automatic construction of test-based as well as woman-based indicators using the SISCOLO data.

Specific objectives:

- Define a set of performance indicators that could be estimated using the SISCOLO data.
- Create a model to support the development of a data warehouse.
- Develop and implement algorithms to estimate the performance indicators.
- Create a graphical user interface (GUI) for the visualization of the indicators.

CHAPTER 3 – THEORETICAL BACKGROUND

3.1 Cervical cancer

Cancer refers to any one of a large number of diseases characterized by the development of abnormal cells that divide uncontrollably and have the ability to spread and destroy normal body tissue. Cervical cancer is a type of cancer that occurs in the cells of the cervix, the lower part of the uterus that connects to the vagina (MAYOCLINIC, 2015).

Cervix is a cylindrically shaped organ located in the lower third part of the uterus that extends into the vagina. The cervix is divided into ectocervix and endocervix. The ectocervix is the part of the cervix that projects into the vagina and is covered by non-keratinized stratified squamous epithelium similar to that of the vagina (Figure 1). In contrast, the endocervix (endocervical canal) is a luminal cavity within an opening orifice (external os), which communicates with the uterine cavity. The transitional area between the squamous epithelium of the vagina and the columnar epithelium of the endocervix is called the squamo columnar junction (SCJ). The SCJ is the result of a continuous remodeling process resulting from uterine growth, cervical enlargement, and hormonal status (IARC, 2005; SELLORS *et al.*, 2003). During this process, the original SCJ migrates from it is initial position toward the ectocervix. Migration of SCJ occurs in large part by a process termed squamous metaplasia, in which the columnar endocervical epithelium is replaced by a stratified squamous epithelium. The area of the cervix where this transformation takes place is referred to as the transformation zone (TZ). The TZ has, for unknown reasons, a unique susceptibility to be infected by the human papillomavirus (HPV) and the majority of cervical cancer cases (90%) are initiated in this area (CLARK *et al.*, 2012; FORBES *et al.*, 1999; IARC, 2005).

Although HPV is essential to the transformation of cervical epithelial cells, the infection of high-risk HPV is necessary, but may not be sufficient for the development of cervical cancer. A variety of additional factors may contributes to the process of cervical cancer development, among them: sexually transmitted viruses like Cytomegalovirus (CMV), the human herpes virus 6 (HHV-6), and HHV-7; Chlamydia infections; smoking; a weak immune system; or a family history of cervical cancer (ACS, 2015; BURD, 2003).

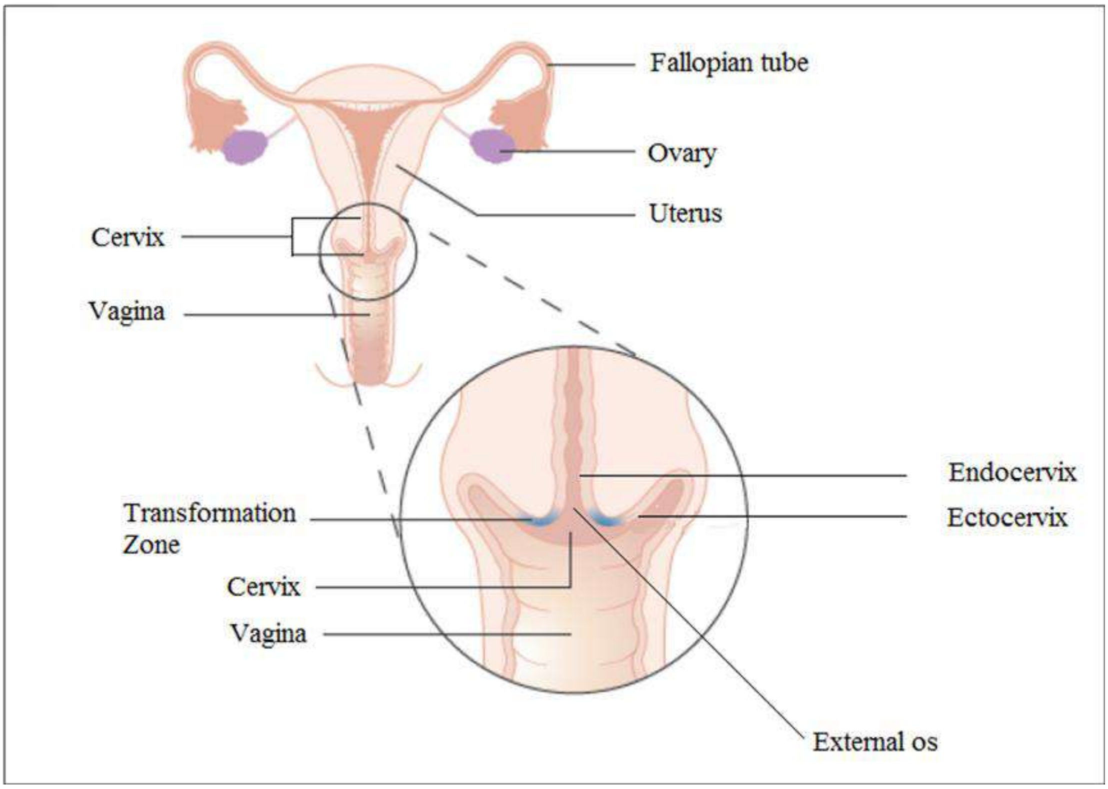


Figure 1: Gross anatomy of the uterine cervix. Source: Adapted with permission from Cancer Research UK (2015).

Early stages of cervical cancer may not cause signs or symptoms, even though it can be detected through a Pap smear test. The Pap smear is a test that looks for changes in cells of cervix (BURD, 2003). In order to report the results of Pap smear test in a unified format, a standard reporting system should be used.

The reporting system for the Pap smear test results has evolved and been refined over time. The most common reporting system is the Bethesda System, which was introduced in 1988. The Bethesda System was developed to establish standards of cytopathologic classification to reduce the diagnostic doubt between benign cellular changes and real atypias as well as to reflect an advanced understanding of cervical neoplasia. Inclusion of statements regarding the adequacy of the specimen and the lack of epithelia of TZ was a significant innovation of the Bethesda System. The Bethesda System was modified in 1991, 2001 and recently in 2014 (BURD, 2003; NAYAR *et al.*, 2015).

3.2 Cervical cancer screening programs

Cervical cancer screening is a systematic application of a test to identify and treat cervical abnormalities (IARC, 2005). There are many types of screening tests, such as liquid-based cytopathology and visual inspection with acetic acid, but the Pap smear test is a commonly used one. In the cases of a positive Pap smear test result, histopathological analysis is performed to confirm the results obtained by the Pap smear test (WHO, 2002; WHO, 2007).

There are two types of screening programs: organized and opportunistic. Organized screening refers to programs in which a target population has been identified and strategies are developed and implemented to provide screening tests for the specific population. In addition, mechanisms are available for the systematic follow-up of women with positive tests as well as for the provision of adequate treatment. Opportunistic screening refers to services provided to women upon request or to women who are already in a health facility while seeking for other services, without any efforts to reach a particular population. Organized screening has many advantages over opportunistic screening: greater impact on cervical cancer control; more cost-effective; less over-screening and over-treatment (WHO, 2004; WHO, 2007).

Regardless of the screening type, three core factors are required to ensure that the program is performed effectively: policy formulation; training; monitoring and evaluation of the program actions (WHO, 2004).

Formulation of policies for screening should involve key stakeholders to formulate clear policies regarding the needs and health priorities of the population. The recommended policies should cover the following issues: target age group, frequency of screening, population coverage; access to health care providers; integrated healthcare services (WHO, 2004).

The goal of training is to ensure that there are sufficient expert personnel to attract women to services, screen eligible women with an appropriate test, and treat women with positive test results. Plans for training should be based on programmatic goals with special attention given to achieving coverage and maintaining the quality of care. In addition, plans for training should specify who, what, how, where, and when training will be approached, as well as how much it will cost (WHO, 2004).

The monitoring and evaluation of a screening program should include the following targets: screening coverage, management of women with positive tests; results of the screening test, false negative rate; missed groups in the target population; and cancers which are detected during the interval between consecutive screening. In order to do this assessment, performance indicators should be constructed and monitored (IARC, 2005; MILLER, 1992).

3.3 Performance indicators

Indicators are a way of seeing the big picture by looking at a small piece of it (SCHIRNDING, 2002). In this context, a performance indicator can be defined as a variable that measures one aspect of a program directly linked to the program's objectives (WHO, 2013).

The types and levels of indicators depend on the types of decisions to be made. Useful indicators should be applicable to both user and data providers. There are several factors that affect the choice of indicators, such as: purpose of their use and target audience. Indicators can be used for problem definition, policy formulation, policy implementation and evaluation. Sometimes the same indicators can serve many purposes, while in other situations separate sets of indicators may be needed.

In general, the criteria of use of performance indicators for local evaluation purposes evaluation are different from those for international purposes. According to Schirnding (2002) the local performance indicators should:

- be relevant both to individual citizens and to local government;
- reflect local circumstances;
- be based on information that can be easily collected;
- show trends over a reasonable period of time;
- be meaningful, clear and easy to understand;
- lead to the setting of targets or thresholds.

The form in which an indicator is presented can have important consequences for decision-making. An indicator can be measured at one time, over several times or continuously, to show changes in a parameter. Indicators can be presented in a variety of statistical forms, for example: as simple frequencies or magnitudes (number of deaths, number of people with health outcomes of interest); as rates (mortality and morbidity), as ratios; as measurements of rate change; or in other more complex forms. When relevant and possible, data should be disaggregated, for example, by age and sex, geographical area, socioeconomic status, urban-rural divide and national and sub-national levels (SCHIRNDING, 2002).

In the case of the evaluation of cervical cancer screening programs, indicators such as the number of smears taken, number and proportion of positive smears, number of women referred for diagnosis and therapy, number of invasive cervical cancers diagnosed and number of pre-cancer and benign lesions detected should be obtained. Such data must be analyzed by age to confirm that women in the target age group are being selected and that subsequent management is appropriate. Indicators like the incidence rate of cervical cancer, number of advanced cases and mortality from the disease are considered some of the simplest forms of evaluation. More detailed evaluation requires the identification of all women who develop invasive cervical cancer or high-grade lesions in the target population and documentation of their screening history (MILLER, 1992; WHO, 2004).

3.4 Brazilian Cervical Cancer Screening Program

Historical background

The Brazilian National Program Against Cervical Cancer was established in 1998 by the Brazilian Ministry of Health, aiming to reduce the incidence and mortality of the disease (BRAZIL, 2017). During August and September of the year 1998, a wide national campaign known as “The first intensification phase” was performed. At this time more than 3 million women were mobilized to undergo cytopathological tests. The coordination of the program was officially transferred to INCA in the year 1999 and in the same year, the SISCOLO was established. In 2002, the second intensification phase was performed, where 3.9 million women were examined, with priority given to the age group between 35 and 49 years (BRAZIL, 2016c).

In 2005, the Ministry of Health launched the national polices for Oncological Care Network (Rede de Atenção Oncológica, in Portuguese) which set the control of cervical and breast cancers as a fundamental component to be provided in the state and municipal health plans (BRAZIL, 2005). In the year 2006, the Brazilian Cervical Screening Guidelines were updated, focusing on the screening of women between 25 and 59 years old (BRAZIL, 2006a). In 2010, through Ordinance No. 310/2010 (BRAZIL, 2010a) the Ministry of Health instituted a working group for evaluating the National Program for the Control of Cervical Cancer. The group discussed the advances and challenges in six areas: management; access and coverage of the screening; quality of cytopathology tests; access and quality of treatment; indicators of the impact of the cervical cancer program; and new control technologies. The conclusions and recommendations were gathered in the plan of actions to reduce the incidence and mortality of cervical cancer (BRAZIL, 2010b).

In 2011, Brazilian Cervical Screening Guidelines were updated by focusing on the screening of women between 25 and 64 years old, but including specific approaches of women at 20 years old or younger (BRAZIL, 2011b). In 2012, to improve the quality and reliability of the cytopathological test, the Quality Management Manual for Cytopathology Laboratory was published by the Brazilian National Institute of Cancer and the Ministry of Health. This manual presents some important indicators for the monitoring of the laboratory test results, which assess the overall and individual performance (BRAZIL, 2012a).

In 2013, Ordinance No. 3.394/2013 instituted the Cancer Information System (SISCAN, in Portuguese), a web-based version that integrates the Information Systems of Cervical Cancer (SISCOLO) and Breast Cancer (SISMAMA, in Portuguese) (BRAZIL, 2013b). In the same year, by means of Ordinance no. 3.388 / 2013, the Ministry of Health redefined the National Qualification in Cytopathology in the prevention of cervical cancer (QualiCito, in Portuguese) (BRAZIL, 2013c). The QualiCito consists of the definition of quality standards and the quality evaluation of the cervical cytopathological examination through the monitoring by SUS managers of the performance of the public and private laboratories that provide services to the SUS (BRAZIL, 2017).

In 2014, the Ministry of Health, through the National Immunization Program, began the campaign of vaccination of girls between 11 and 13 years against the HPV virus. In 2016, the Brazilian Cervical Screening Guidelines were updated, where the approaches for women at the age of 20 years or younger were extended to women of the age of 24 years or younger. In addition, other clinical approaches were introduced or updated according to the age of the woman (BRAZIL, 2016d). In addition, a revised edition of the Quality Management Manual for Cytopathology Laboratory was launched (BRAZIL, 2016c).

The screening process

The type of screening performed in Brazil is opportunistic, in which women spontaneously seek health services in a primary healthcare unit especially for maternal and child care. In the primary healthcare unit, a specimen of the cervicovaginal material is collected and fixed on a slide. The slide is sent to the laboratory along with a requisition form (see Annex 1) filled in with information about the primary healthcare unit, and the woman's personal information, socio-demographic and clinical status.

In the laboratory, an initial assessment is made to decide whether to accept or reject the slide. In a case of rejection, the slide is returned back to the primary healthcare unit with a report explaining the reasons for rejection. The reasons for a slide to be rejected are: identification of the slide and/or specimen does not coincide with the information in the requisition form; absence or misidentification of the slide and/or bottle; a damaged or missed slide; reasons pertaining to the laboratory; and other reasons.

If the slide is accepted, further evaluation is performed to analyze the adequacy of the specimen in the slide for the cytopathologic analysis and the presence of epithelia in the specimen. Regarding the specimen adequacy, if the specimen is classified as unsatisfactory for analysis, then no further cytopathologic analysis will be performed. The reasons for a specimen to be classified as unsatisfactory are: presence of acellular or hypocellular material (< 10% of the specimen); impaired reading (> 75% of specimen) due to the presence of blood, pus cells, artifacts due to desiccation of the specimen, external contaminants, intense cellular overlay, and other reasons.

The cytopathologic analysis is only performed on the satisfactory specimens. In this work we will use the term “performed tests” to refer to the tests performed on slides with satisfactory specimens.

In the analysis of the epithelia, three types of epithelial cells may be encountered in the specimen: squamous cells (ectocervical); glandular cells (endocervical) and metaplastic cells. The presence of glandular and metaplastic cells ensures that the specimens has representative elements of the TZ.

The results of the performed tests are reported according to the Brazilian nomenclature for cervical reports which is based on the Bethesda reporting system for the year 2001 and adapted to the Brazilian scenario (BRAZIL, 2012b).

According to the Brazilian nomenclature for 2016 (BRAZIL, 2016d), the Pap smear tests may show the following descriptive diagnosis: within normal limits in the examined material; benign cellular changes; and atypical cells.

The atypical cells could be:

- Atypical cells of undetermined significance:
    - Squamous:
        - Possibly non-neoplastic (ASC-US);
        - Cannot exclude high-grade intraepithelial lesion (ASC-H).
    - Glandular (AGC):
        - Possibly non-neoplastic (AGC-non-neoplastic);
        - Cannot exclude high-grade intraepithelial lesion (AGC-H).
    - With unknown origin (AOI, in Portuguese):
        - Possibly non-neoplastic (AOI- non-neoplastic);
        - Cannot exclude high-grade intraepithelial lesion (AOI-H).
- In squamous cells:
    - Low-grade squamous intraepithelial lesions (LSIL), including infection by HPV and cervical intraepithelial neoplasia grade I (CIN I);
    - High-grade Intraepithelial lesion (HSIL), including cervical intraepithelial neoplasia grades II and III (CIN II and CIN III);
    - High-grade Intraepithelial lesions, cannot exclude microinvasion;
    - Squamous cell carcinoma.
- In glandular cells:
    - Adenocarcinoma in situ (AIS);
    - Invasive adenocarcinoma: cervical, endometrial and without further specifications;
    - Other neoplastic malignancies.

For each test result, the Brazilian Guidelines for Cervical Cancer Screening (BRAZIL, 2016d) recommend a specific clinical approach. The recommended routine for screening in Brazil is a triennial one (i.e. repeat a Pap smear test every three years after two consecutive normal tests performed with an interval of one year). Box 1 presents the recommended clinical approaches for each cytopathological test result.

The recommended clinical approaches are varied depending on the risk of cancer, or of a pre-invasive lesion that requires treatment. The recommended clinical approaches for women with cytopatholgical test results of low risk atypical findings (ASC-US and LSIL) vary according to the age of woman. The recommendation for young women (<25 years old) is to repeat a Cytopathology test after 3 years, while for the woman at the age of 25 or older, the recommended time to repeat the cytopathology test depends on the initial diagnosis (ASC-US or LSIL) and the age of woman, as presented in Box 1. If the result of the repeated cytopathology test is negative (without atypical findings), a follow up the screening routine is recommended. If the result of the repeated Pap test is positive, it is recommended to make a colposcopy exam, which is done by using a special magnifying device called a colposcope to look at, vagina, and cervix. During a colposcopy exam, it is important to visualize the SCJ in its entire circumference. Sometimes, even if the SCJ is not totally visible, the physician can detect if there are changes in the surface of the cervix (colposcopy with changes). In this case, a biopsy is recommended for further histopathologic investigations (BRAZIL, 2016b; SELLORS *et al.*, 2003).

In the case of the cytopathological results with high-risk atypical findings for cancer or pre-invasive lesions, an immediate colposcopy exam it recommended. During the colposcopy, if necessary, a biopsy is taken in order to perform a confirmatory histopathology test. The biopsy is sent to the laboratory along with a requisition form (see Annex 2) filled with information about the primary healthcare unit, colposcopy diagnosis, and the women's personal information, socio-demographic and clinical status. If high-risk atypical findings are confirmed, or it is not possible to exclude invasive or pre-invasive lesions, an excisional procedure is recommended. In the other situations, a follow-up is recommended for each woman using a further cytopathology test and/or colposcopy exam (BRAZIL, 2016d). An example of the recommended clinical approaches for women with cytopathological test results of ASC-H is presented in Figure 2, where the second test/procedure is defined after the evaluations of the visibility of SCJ and the types of the colposcopic findings. The detailed description of the recommended clinical approaches for all the test results and their flow diagrams can be found in the Brazilian Guidelines for Cervical Cancer Screening for the year 2016 (BRAZIL, 2016d).

Box 1: The initial recommended clinical approaches for the cytopathological results

Cytopathological results	Age group	Initial approach
Within normal limits, benign cellular changes, microbiological findings and endometrial cells	All	Follow the triennial routine
Squamous atypical cells of undetermined significance, possibly non-neoplastic (ASC-US)	< 25 years	Repeat cytopathology test in 3 years
	25≤ age ≤29	Repeat cytopathology test in 12 months
	≥30	Repeat cytopathology test in 6 months
Squamous atypical cells of undetermined significance, cannot exclude High-grade intraepithelial lesion (ASC-H)	All	Perform colposcopy
Atypical cells of undetermined significance with unknown origin (AOI)	All	Perform colposcopy
Atypical glandular cells of undetermined significance (AGC)	All	Perform colposcopy
Low-grade squamous intraepithelial lesions (LSIL)	< 25 years	Repeat cytopathology test in 3 years
	≥25 years	Repeat cytopathology test in 6 months
High-grade intraepithelial lesion (HSIL)	< 25 years	Perform colposcopy
	≥25 years	Perform colposcopy
Squamous cell carcinoma or high-grade intraepithelial lesions, cannot exclude micro-invasion	All	Perform colposcopy
Adenocarcinoma in situ (AIS) or invasive adenocarcinoma	All	Perform colposcopy

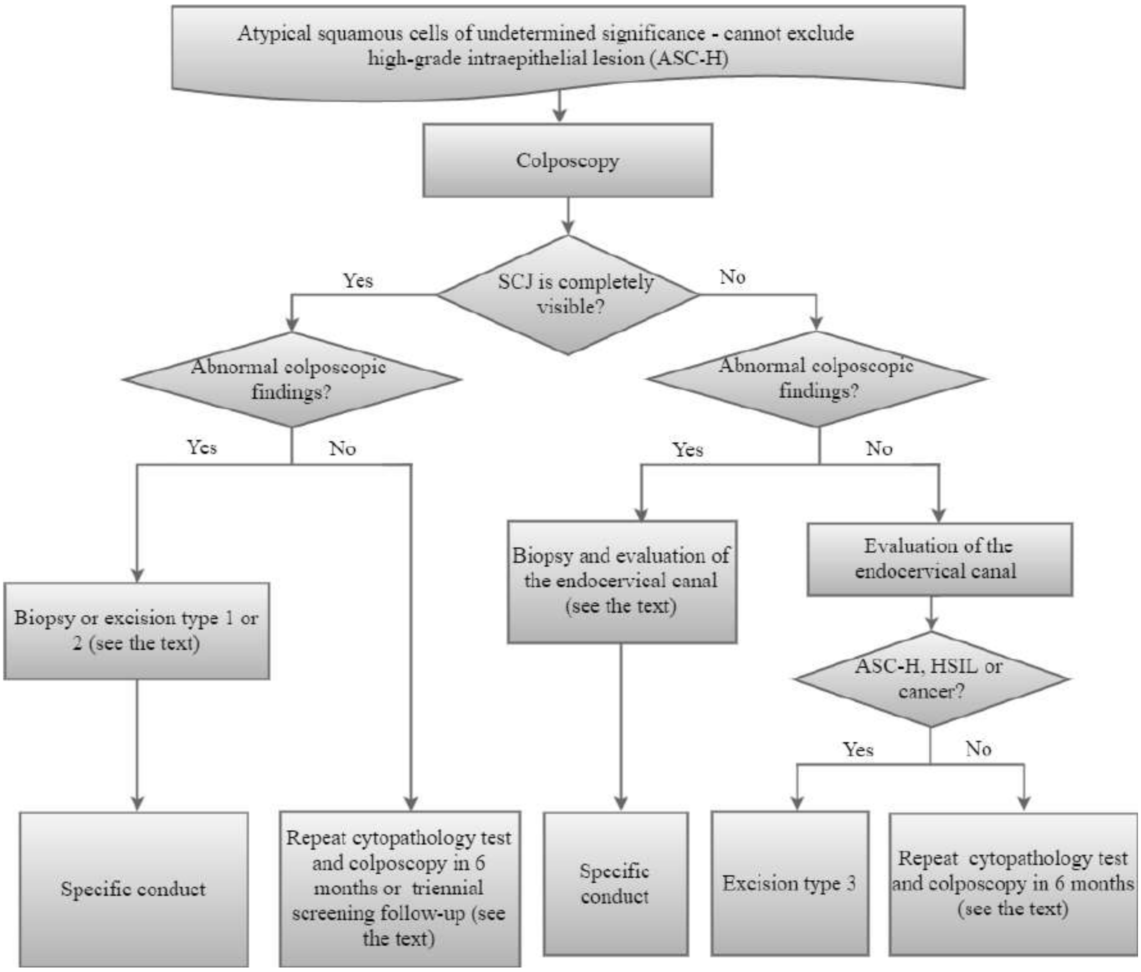

Atypical squamous cells of undetermined significance - cannot exclude high-grade intraepithelial lesion (ASC-H)
Colposcopy
SCJ is completely visible?
Yes
No
Abnormal colposcopic findings?
Yes
No
Abnormal colposcopic findings?
Yes
No
Biopsy or excision type 1 or 2 (see the text)
Biopsy and evaluation of the endocervical canal (see the text)
Evaluation of the endocervical canal
ASC-H, HSIL or cancer?
Yes
No
Specific conduct
Repeat cytopathology test and colposcopy in 6 months or triennial screening follow-up (see the text)
Specific conduct
Excision type 3
Repeat cytopathology test and colposcopy in 6 months (see the text)

Figure 2: The flow diagram for the recommended clinical approach for women with cytopathological test results of ASC-H. This flowchart is a translated version of the original flowchart provided by the Brazilian guidelines for cervical screening.

3.5 Data warehouse

The data warehouse (DW) is an integrated repository of data that employs a suite of tools to transform raw data into meaningful business information. This generated information is used as a foundation for decision-making. There are two main approaches to data warehouse design: one introduced by William Inmon in 1990 and another introduced by Ralph Kimball in 1996 (KIMBALL *et al.*, 2011; VELICANU *et al.*, 2007). Debates on which one is better and more effective have been going on for years. Neither is right or wrong, as they represent different data warehousing philosophies.

The philosophy of Inmon (INMON, 2005) is to design a complex database for the whole enterprise and then extract simple and separate databases (called data marts) from the complex database. An example of Inmon philosophy is the designing of a database for a company. The complex database for the whole company is designed first, and then from this complex database, data marts for each department (sales, finance, employee, etc) are extracted. On the other hand, Kimball's philosophy (KIMBALL *et al.*, 2011) is based on designing a simple and separate database for each section of the enterprise.

In this work, which is already very focused on a problem area, Kimball's approach was selected since it is more simple and focuses on ease of end-user accessibility (CHUCK *et al.*, 1998; KIMBALL *et al.*, 2011). The basic components of the Kimball's data warehouse approach are: i) operational source systems; ii) data staging area; iii) data presentation area; and iv) data access tools. Figure 3 shows a simplified diagram of a Kimball data warehouse.

Operational Source Systems

The operational source systems represent the systems that contain the source data. They are designed for data entry purposes and are not well suited for online queries and analysis.

Data Staging Area

The data staging area of the DW is both a storage area and a set of processes commonly referred to as extract-transform-load (ETL). Extracting means reading and understanding the source data and copying the data needed for the DW into the staging area for further manipulation.

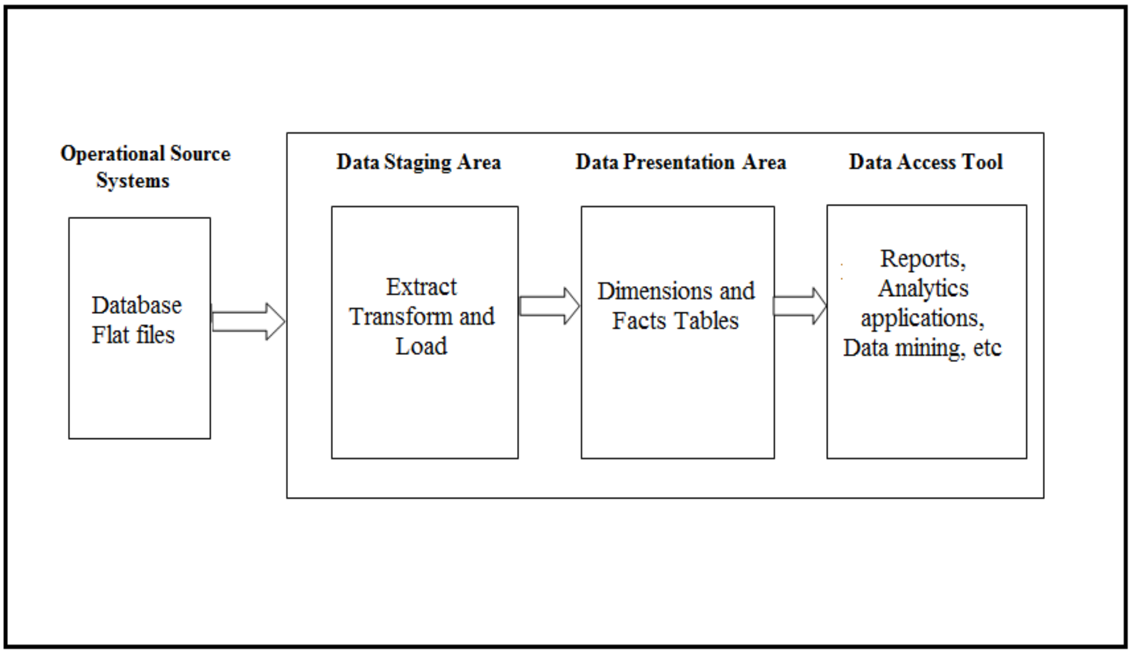

Operational Source Systems
Database Flat files
Data Staging Area
Extract Transform and Load
Data Presentation Area
Dimensions and Facts Tables
Data Access Tool
Reports, Analytics applications, Data mining, etc

Figure 3: Basic components of a data warehouse.

Transformation is a process of converting the extracted data into the desired form by applying many potential processes such as cleansing the data, correcting misspellings, resolving domain conflicts, dealing with missing elements, parsing into standard formats, combining data from multiple sources (linkage), deduplicating data, and assigning warehouse keys. The last process of the ETL is the loading of the transformed data into the DW presentation area. This loading process is commonly called "populating the DW". The data staging area can only accessed by business users and does not provide query and presentation services.

Data Presentation area

The data presentation area is an area where data is organized, stored, and made available for direct querying by users. Unlike the backroom staging area, the users can access the data presentation area via data access tools. There are two types of data modeling that are used for data presentation: The Entity Relationship (ER) modeling and the dimensional modeling types.

ER modeling produces a data model of the specific area of interest, using two basic concepts: the entities, which are objects or concepts that can have data about them, and the relationship between those entities. An entity is represented on the diagram as a rectangular box, while the relationship is represented by polygons. Attributes of entities are represented by oval or circular shapes. Figure 4 shows an example of an ER model.

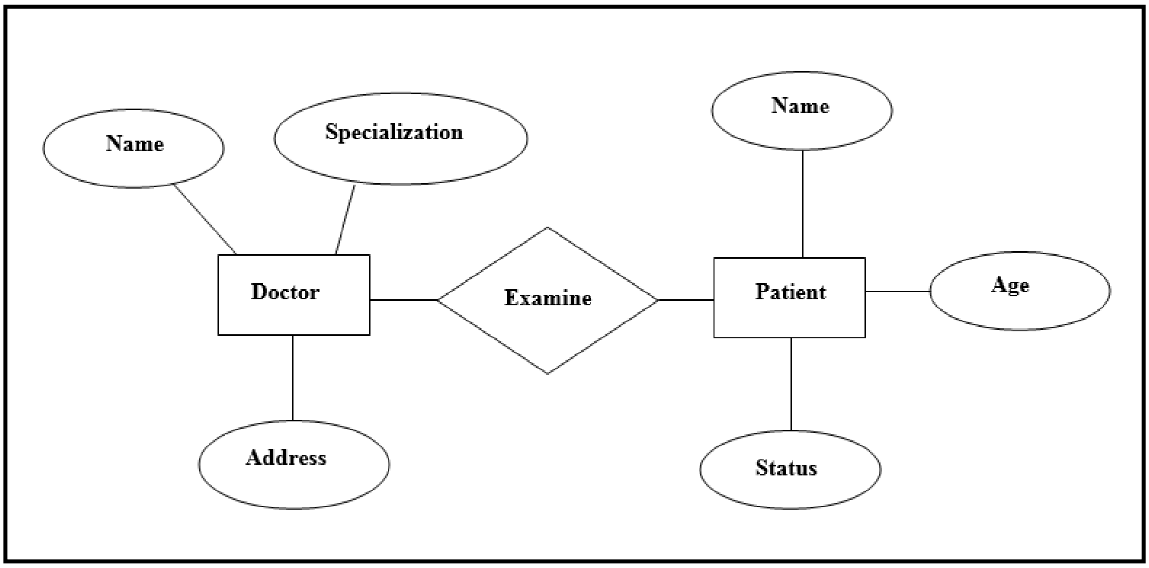


Figure 4: Example of an ER model.

Dimensional modeling is simpler, more expressive, and easier to understand than the ER modeling. It is especially useful for summarizing, rearranging, and presenting views of the data to support data analysis. Dimensional modeling focuses on numeric data, such as values, counts, weights, balances, and occurrences. The two main basic concepts of dimensional modeling are facts and dimensions.

A fact is a collection of related data items, consisting of measures and context data. Each fact represents an event that can be used in analyzing the business processes. For example, in the case of the evaluation of cervical cancer screening programs, one of the facts could be the number of inadequate specimens.

A dimension is a collection of elements used to constrain, group, or browse the facts. In the previous example of evaluation of screening programs, the dimensions could be a time when the specimen had been analyzed, or the laboratory that made the analysis, or any other element used to browse the fact.

There are two different models that can be used in dimensional modeling: i) star schema and ii) and snowflake schema. The star schema (Figure 5) has a simple structure with relatively few tables and well defined join paths. This simple design provides fast query response time and a simple schema. The snowflake schema (Figures 6) is a more complex data warehouse model than a star schema, where the dimension data are grouped into multiple tables instead of one large table. This complex structure results in queries that are more complex and reduces the query performance. The common element for both models is the fact table, which in both models holds the facts and keys to the dimension tables. Each dimension table is assigned a primary key, which is unique for each element of the dimension table. The primary key is replicated in a fact table where it is referred to as a foreign key. For both models, dimension tables may have many columns or attributes. These attributes describe the rows in the dimension table (ex. attributes of the time dimension could be year and month).

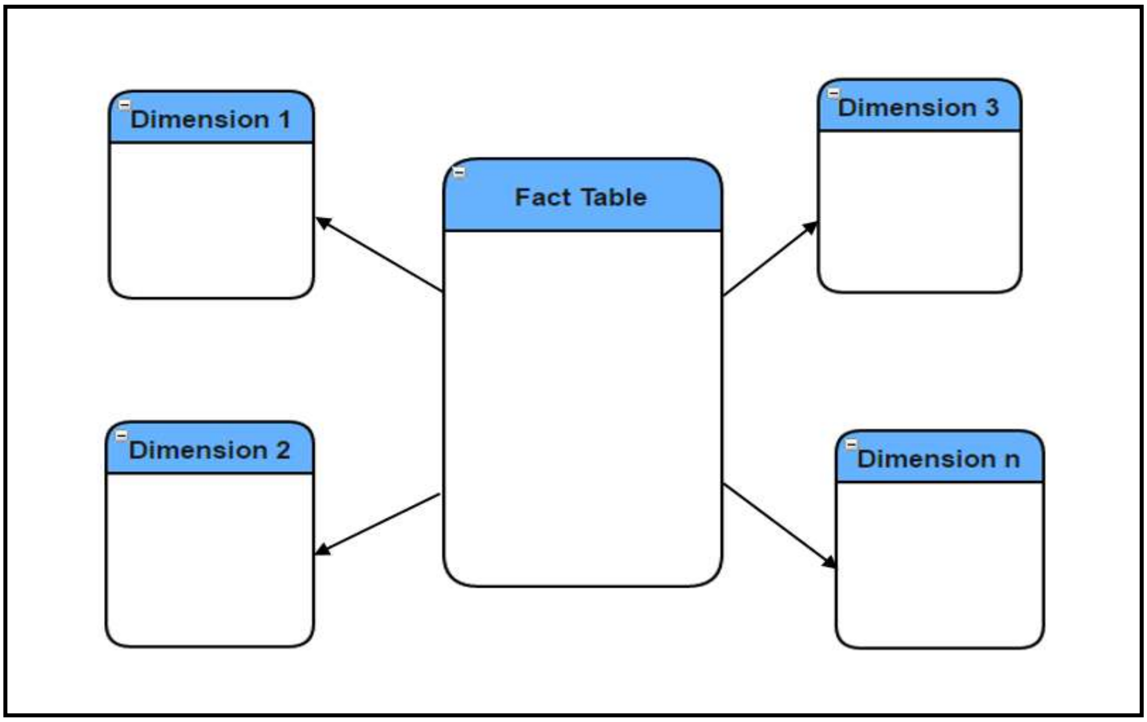


Figure 5: Example of a star schema of a data warehouse.

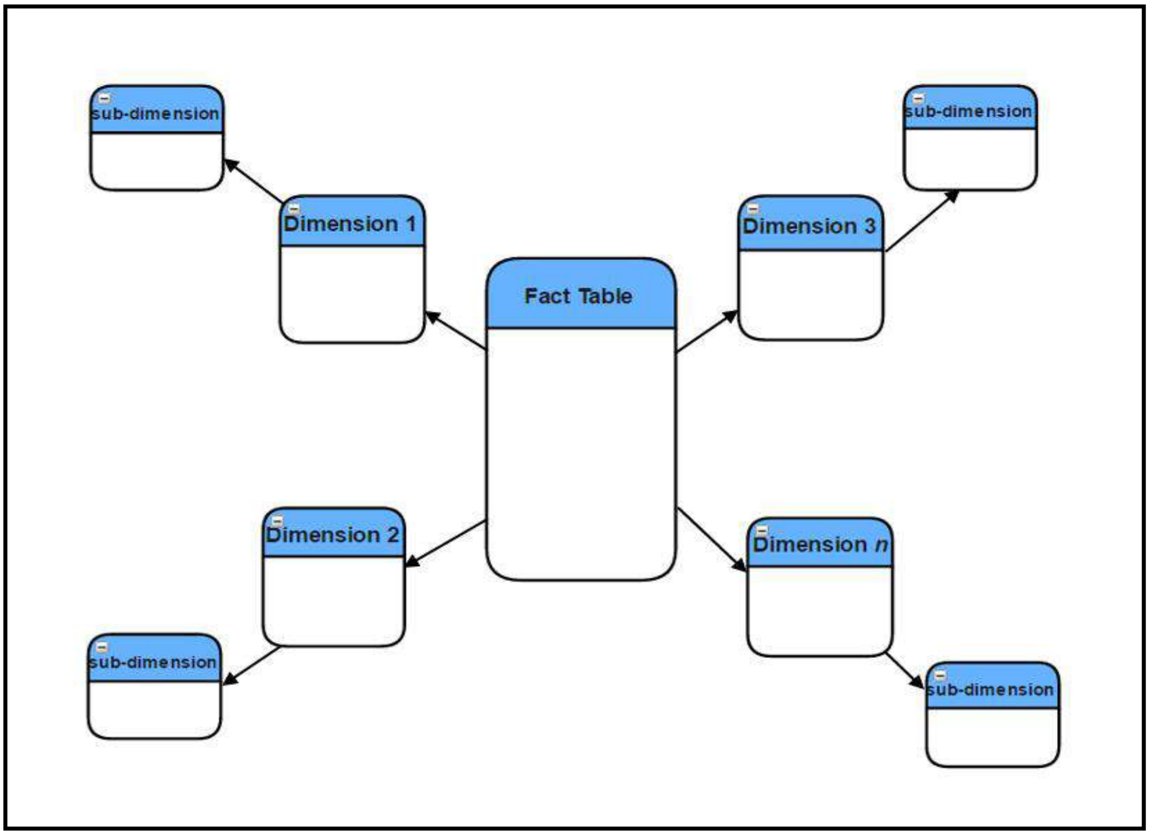


Figure 6: Example of a snowflake schema of a data warehouse.

Another important concept in dimensional modeling is granularity. Granularity refers to the level of detail or summarization of the units of data in the data warehouse. The lowest level of granularity is understood as raw, non-aggregated data. With increasing level of summarization, granularity increases (KIMBALL *et al.*, 2011). In the previous example of the time dimension, the lower level granularity is represented by month while the higher-level granularity is represented by year.

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 projection of a point in each axis corresponds to the dimension values associated to that fact. Figure 7 shows an example of a data cube for the dimensional model used for organizing the referral of patients to a cancer treatment center. In the example there is one fact table surrounded by three dimensions tables, which are time, treatment and hospital that referred the patient. Two metrics are shown in the fact table: number of patients and number of visits. The point in the cube browses the fact "number of patients" using the dimensions hospital, treatment and time respectively.

Usually a data warehouse provides tools that allows the visualization of the facts in several ways through charts, tables and maps. The visualization of a fact using only one dimension is a slicing (e.g. number of patients for year 2016 and for each combination of the values of the other dimensions), and using two or more dimensions is a dicing (e.g. number of patients for the (year 2016 or 2015) and (hospital "ABC" or hospital "CDE")). The roll-up is a navigation from a lower level of the dimension's hierarchy to a higher level of the hierarchy (e.g. municipality to state), while the drill down is performed by the navigation from a higher level to a lower level of the hierarchy. These operations give more interactivity to the process of data analysis/visualization, when compared to more traditional static query tools.

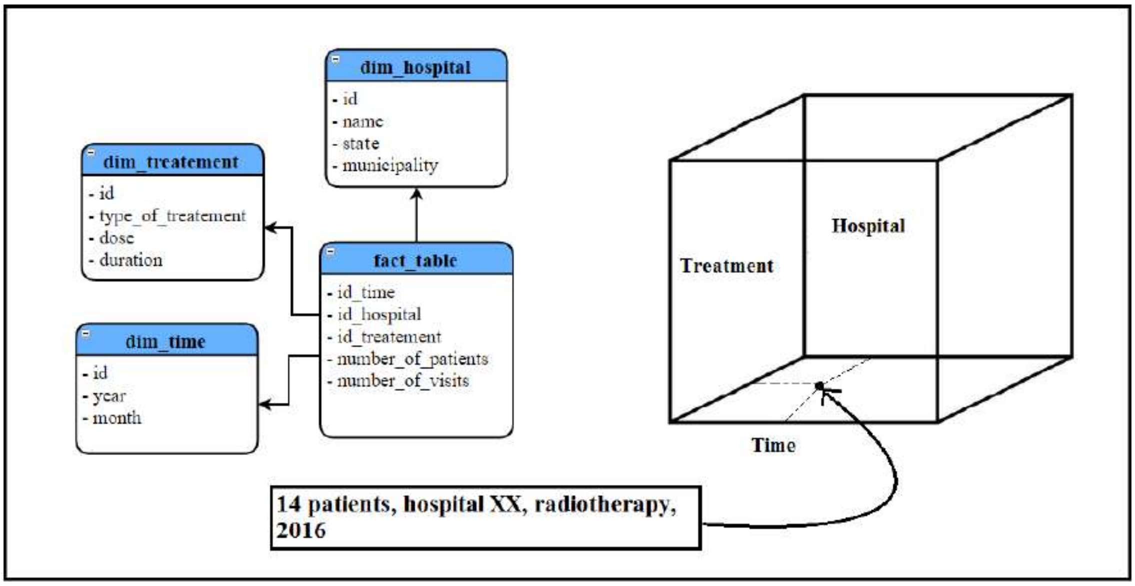


Figure 7: Example of a data cube.

Data Access Tools (DA)

The final major component of the data warehouse environment is the data access tools (DA). DA Tools allow end users to perform data analysis for decision making, by building structured query language (SQL) queries through pointing and clicking on the list of tables and fields in the data warehouse. DA Tools make the tasks of database administrators a lot easier, especially if the database being managed is large. DA Tools provide users with a friendly graphical user interface to the database, avoiding users the need for work directly with the query languages, which may look cryptic for them on the command line interface.

3.6 Record Linkage of Data

3.6.1 The record linkage process

The term record linkage, also referred to as data cleaning or object identification, is simply the bringing together of information from two different records that are believed to belong to the same person, family or entity (WINKLER, 2006). This might involve the linking of records within a single database file to identify the duplicate records, or records from many data files to identify the records that belong to the same entity (GILL, 2001). The record linkage of data is performed in four stages: preparation of the database; blocking of the records to be linked; matching or linking the records; and validation of the linkage process (HERZOG *et al.*, 2007).

Preparation of the database

The preparation of data is a process of understanding the data, identifying the variables, cleaning the data, and analyzing its consistency. The results of this process provide information for the choice of a set of variables to be used in the linkage process. Such variables must pass a set of procedures called "standardization of variables". The objective of the standardization is to facilitate the comparison between the variables contents (CHURCHES *et al.*, 2002). The standardization process involves the following steps:

- Standardization of the alphabetic characters which are involved, turning them into capital letters, removing accents and deleting the special characters (*, #,?, $, etc.).
- Partition/parsing of the character string into its component parts. This partitioning facilitates the cleaning of the fields, allowing the removal of unwanted information between the characters such as prepositions (*de, do, da,* etc.) and enabling the use of certain phonetic changes in certain fragments of the variable.
- Formatting of the variables by setting a standard format, for instance, the same date format for date variables.
- Application of phonetic codes, which are based on the phonetic similarity of words. This step is important in reducing errors due to variations in spelling as well as to perform the blocking of records.

Blocking of records

The blocking of records is a process of splitting a large dataset into a smaller dataset of individuals which has at least one common characteristic, such as geographic region or a specific clinical condition (DUSETZINA *et al.*, 2014). The objective of blocking is to reduce as much as possible the large number of potential comparisons between the records, by comparing only records that belong to the same subset (NEWCOMBE, 1967). This technique consists of separating the data files into logical blocks of mutually exclusive records, through the indexing of the files to be related. This indexing is performed using a blocking key composed of one or a set of variables chosen for this purpose. The blocking key is constructed from the available variables already standardized, where all records that have the same value recorded in this key will be entered into the same block (e.g. records with first name Smith will be in the same block if first name is the blocking key). The blocking step could be performed according to two strategies: single-step and multiple steps. In the single-step strategy, the blocking variables are selected, and then the records to be compared are separated according to the blocking variables. The comparison step is performed by comparing the records of the same block in the compared files. In the multistep strategy, different blocking variables may be used in sequential steps, where the unpaired records in the first phase of blocking are used in the second phase of blocking and so on.

Linking the records

The linking of records is the stage where the potential matches are brought together so that they may be compared. This comparison could be carried out using deterministic or probabilistic linkage. The deterministic record linkage is performed in the case of the existence of a common identifier(s), which permits the identification unambiguously (e.g. National Security Number). The probabilistic record linkage is performed in the case of the absence of the common identifier(s). In the probabilistic linkage, statistical models are used in order to sort the records as true pairs (that belong to the same entity) or false pairs (that belong to different entities) (FELLEGI *et al.*, 1969; NEWCOMBE, 1967; NEWCOMBE, 1988).

The first practical application of the probabilistic record linkage was performed by Howard Newcombe who used records of vital data such as name, birth date, address and other available information to find hereditary diseases (GILL, 2001; NEWCOMBE, 1967). In 1969, Fellegi and Sunter (1969) provided a formal mathematical model for ideas that had been introduced by Newcombe *et al*. In their approach, scores are associated to each pair of records based on weights attributed to the comparison of identifying variables such as name, parent's name, birthdate, sex, and address. Weights are used to measure the contribution of each variable to the probability of making an accurate classification, and are constructed from concepts widely used among epidemiologists in assessing the accuracy of diagnostic tests (FELLEGI *et al.*, 1969). For each variable i, let:

m_i = P(variable values agree|true pair) (3.1)

u_i= P(variable values agree|false pair) (3.2)

1− m_i =P (variable values disagree|true pair) (3.3)

1−u_i= P(variable values disagree|false pair) (3.4)

For each pair of records and each variable to be compared, one of these two weights applies: one weight for the situation of agreement of the values of the compared variables and another for the situation of disagreement. The weight for agreement is given by:

$$w_{ai} = log_2 \frac{m_i}{u_1} \quad (3.5)$$

The weight for disagreement is given by:

$$\omega_{\mathrm{di}} = \log_2 \left[\frac{1-\mathrm{m_i}}{1-\mathrm{u_1}}\right] \quad (3.6)$$

The total score of a pair is obtained from the sum of the weights for each variable used in the matching process. Since m_i is usually greater than u_i, the agreement weight contributes positively to the score, while the disagreement weight contributes negatively. By ordering all pairs according to their scores and by choosing appropriate cutoff points p_1 and p_2, with p_2 greater than p_1, it is possible to split the score space into three regions: i) those pairs above p_2, which are definitely classified as links; ii) those pairs below p_1, which are definitely not classified as pairs; and iii) those pairs between p1 and p_2, which form a grey zone, requiring further investigation.

Validation of the linkage process

The validation of the linkage is a process of estimating the linkage accuracy. Sensitivity and specificity are classical measures of the accuracy. Sensitivity is the proportion of true pairs that were classified as matched pairs in the linkage process. On the other hand, specificity is the proportion of the pairs that were classified as nonmatched among the pairs that are really false pairs. The linkage accuracy is related to several factors, such as the quality of information of the identifying variables, the linking strategy adopted and the correct specification of concordant, not concordant and inconclusive pairs (SILVEIRA *et al.*, 2009).

3.6.2. The software for implementation of record linkage process

From the 1950s through the early 1980s, researchers and practitioners undertaking large record linkage projects had to develop their own software. They often faced the choice of using less accurate methods or expending dozens of staff years to create proprietary systems (HERZOG *et al.*, 2007). Currently, a variety of commercial record linkage and deduplication systems are available. For many applications, the record linkage process is quite complex and requires a significant amount of customization or additional programming, which is not possible for the commercial systems. Additionally, a large amount of these softwares is specialized to a certain domain, which limited its use only to that domain. Several smaller record linkage systems are available free or at affordable prices. However, they are commonly limited in their ability to deal with different types of data, contain a limited amount of functionality (for example, they implement only a small number of commonly used string comparison functions), or they can only link small data sets (CHRISTEN, 2009).

In Brazil, the first software for record linkage was called RecLink, which was developed by Camargo *et al* (2000). Although RecLink has been used in several projects, it does not allow the choice of comparison and phonetic algorithms, and performs poorly in handling large databases (BASTOS, 2011). Aiming to overcome the limitations of the RecLink (especially dealing with large databases), software program called VincReg was developed by the medical informatics group at the Universidade do Estado do Rio de Janeiro (FREIRE *et al.*, 2010). VincReg was developed in Java and allows the user to configure many of the steps of the record linkage process. In addition, VincReg has obtained excellent results in terms of sensitivity and specificity when used to link databases of the Brazilian healthcare system such as SISCOLO, Oncology Module of the Outpatient Information System (APAC_ONCO, in Portuguese), the Hospital Information System (SIH, in Portuguese) and the Mortality Information System (SIM, in Portuguese) (CABRAL, 2010; FREIRE *et al.*, 2012; FREIRE *et al.*, 2015).

CHAPTER 4 – LITERATURE REVIEW

The evaluation of cervical cancer screening programs requires continuous monitoring of the performance indicators. Before this process of monitoring, the indicators should be identified and constructed. This chapter focuses on the identification of the performance indicators used for the evaluation of the screening programs worldwide; and types of performance indicators used in Brazil and the Brazilian available environments/tools for the visualization of these indicators.

4.1 The performance indicators used worldwide

A systematic review to identify the performance indicators used for the evaluation of the screening programs worldwide was carried out. The search was performed in *Biblioteca Virtual em Saúde* (BVS, in Portuguese), following the search strategy described in Box 2. In addition to it, a Google search for the gray literature was done with no language requirements. All studies about cost-effectiveness of the cervical cancer screening programs were excluded as well as duplicate articles.

Box 2: Search strategy for the literature review of the performance indicators used worldwide

[cervi$ cancer screening OR pap smear OR Papanicolaou OR cervical cytology OR uterine screening Not cost$effectiv$ AND (quality improvement OR quality indicators OR healthcare quality indicators OR practice guidelines OR quality of service OR effectiveness OR performance OR assessment)]

The review identified a total of 89 performance indicators distributed among 21 countries/regions from all the continents, except Africa, where the majority of countries neither have screening programs nor publish the articles in their own languages (AHMED *et al.*, 2014a). These indicators were classified into quantitative and qualitative (Figure 8).

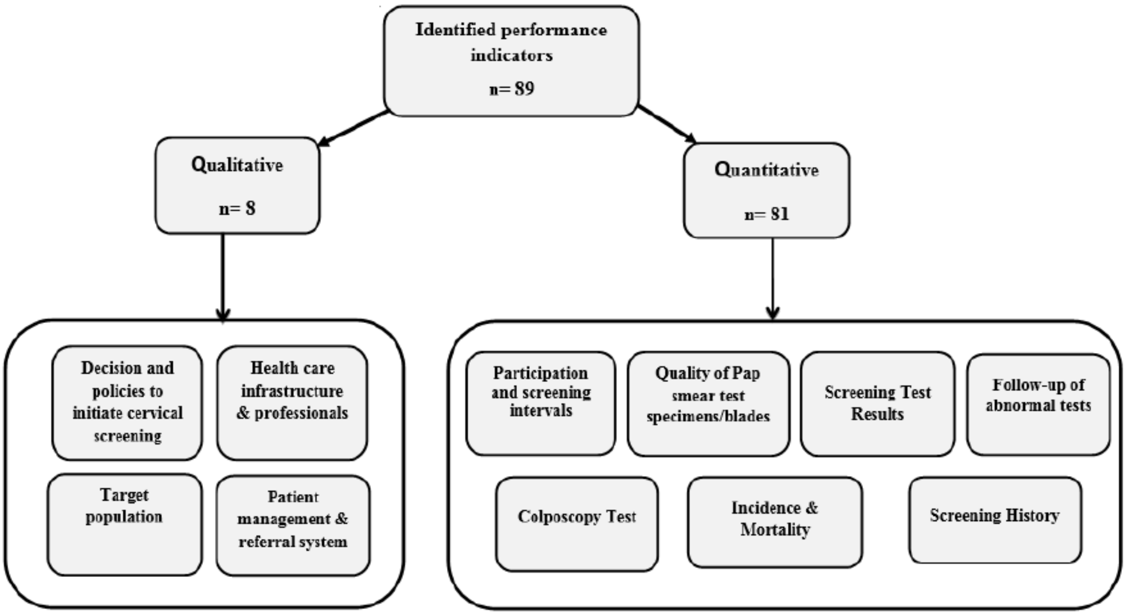

Identified performance indicators
n= 89
Qualitative
n= 8
Quantitative
n= 81
Decision and policies to initiate cervical screening
Health care infrastructure & professionals
Target population
Patient management & referral system
Participation and screening intervals
Quality of Pap smear test specimens/blades
Screening Test Results
Follow-up of abnormal tests
Colposcopy Test
Incidence & Mortality
Screening History

Figure 8: Classification of the identified performance indicators.

Out of the 89 indicators, 8 were qualitative, two per each group, and 81 were quantitative. The latter ones were distributed among seven groups. The description of each group is presented below:

Adequacy of the specimens/slides: Indicators related to the adequacy of the specimens/ slides, as well as the presence of the material of the Transformation Zone in the specimens.

Screening test results: Indicators related to all possible results of the cytopathology test.

Follow-up of abnormal tests: Indicators related to the follow-up of all abnormal cytopathology test results that need to undergo further evaluation to determine whether the women need to be treated. Indicators of this group are mainly concerned with the number of histopathology tests and the treatment of intraepithelial lesions.

Colposcopy exam: Indicators related to the colposcopy exam.

Participation and screening intervals: Indicators related to the women's participation in the program as well as the time interval between each screening test.

Screening history: Indicators about the screening history of women's diagnosis with invasive cancer. Those indicators represent a retrospective summary (up to 10 years) of screening prior to the diagnosis.

Incidence and mortality: Indicators related to the incidence and mortality from cervical cancer.

Decision and polices to initiate cervical screening: This group includes indicators related to the informed decisions to initiate cervical screening in the context of a National Cancer Control Program and the construction of program policies.

Healthcare infrastructure and professionals: This group includes indicators related to the availability of adequate infrastructure as well as the training of medical personnel.

Patient management and referral system: This group includes indicators related to the existence of referral system for women with an abnormality and the mechanisms to ensure that those women were attended for diagnosis and treatment.

Target population: This group includes indicators related to the definition of the target women and the means to invite them for screening.

The most frequent groups of indicators were: participation and screening intervals; screening test results and incidence and mortality. In the case of Brazil, the less reported indicators were related to the follow-up of abnormal tests, colposcopy exam, participation and screening intervals and screening history (AHMED *et al.*, 2014a).

4.2 The performance indicators used in Brazil

Another literature review was carried out to identify the used indicators in Brazil. The search was done in BVS and Web of Science, using the search strategy described in Box 3. The search was performed for the studies that only used SISCOLO or SISCAN web. Duplicated studies and studies related to cost-effectiveness were excluded.

Box 3: Search strategy for the literature review of the performance indicators used for the evaluation of Brazilian Cervical Cancer Screening Program

Brazi* cervi* cancer screening AND (coverage OR performance indicators OR quality control OR effectiveness OR evaluation OR monitoring OR SISCOLO OR SISCAN) Not cost-effectiveness.

Following the search strategy in Box 3, a total of 155 performance indicators were identified. The identified indicators were reported in studies that used SISCOLO. However, the scopes of the identified studies that used SISCAN web were out of the scope of the current study. Those indicators are distributed among 22 studies. The reported indicators of each study are presented in Box 4. The majority of the identified studies were regional studies focusing on test-based indicators. Five national studies were identified. The first one is a time series study that built a set of test-based indicators from the year 2006 to 2013 (COSTA *et al*., 2015). This type of time series studies is useful for the longitudinal monitoring of the program and permits the comparison between different regions. Two other national studies evaluated the Brazilian cervical cancer screening program by analyzing the distribution of the test-based indicators from 2002 until 2006 (DIAS *et al*., 2010) and from 2006 until 2009 (SANTOS *et al.*, 2012). The results showed that the program is still below the target in some states where there is a need to optimize the resources and provide access to the women who are not included in the screening program. Regional differences showed some failures on women screening, population coverage and adequacy of specimens.

The last two national studies used sets of test-based indicators to access the profile of the Brazilian cytopathology laboratories in the year 2002 (THULER et al., 2007) and the year 2010 (BORTOLON *et al*., 2012). In the study of Thuler *et al*.(2007), in addition to the test-based indicators, a set of six indicators related to the administration and the medical personnel of the laboratories was used. The test-based indicators in these studies are used for the internal quality monitoring (MIQ, in Portuguese) of the cytopathological examination which have standard values established by the Brazilian Ministry of Health (BRAZIL, 2013d). The MIQ indicators were also identified in other four regional studies (ARAUJO JR *et al*., 2015; ÁZARA *et al.*, 2014; PLEWKA *et al*., 2014; TOBIAS *et al.*, 2016). The results of both national and regional studies showed that those indicators could be excellent measures to access the quality of the cytopathology laboratories.

Studies that evaluate the actions of screening programs considering multiple aspects are essential to provide information for decision-making and better allocation of resources, as well as to permit the understanding of the effect of the indicators on one another. In this review, three studies were identified. The first study was carried out by Feitosa *et al.* (2007) to establish the screening profile of the 850 municipalities in Minas Gerais State. The study identified the municipalities with problems in implementing the program, such as the collection, the fixation and the analysis of the Pap smear specimens. The second study was carried out to establish the profile of primary healthcare units in three regions of the State of Rio de Janeiro. In order to carry out the study, a production of the cytopathologic tests during 2007 was obtained in a sample of 535 of these units. The units were classified according to four test-based indicators. These indicators permit the establishing of the units' profiles and the identification of the factors that affect the detection of the atypical cells, which is the main objective of the program (AHMED *et al.*, 2014b). The third study established the screening profile of two Brazilian cities (DISCACCIATI *et al.*, 2014). In this study, the prevalence of the cytopathological results according to women's age and time since the last examination was calculated. The results showed that the prevalence of the altered cytopathological results vary significantly with the region. This variation is due to the cytopathology test quality, which is a procedure that depends on the actions of health care professionals and, therefore, shows high performance variability.

The indicators related to the biopsy procedures are reported in one study (BRITO-SILVA *et al.*, 2014). In this study, the screening program in the municipality of São Paulo/Brazil was evaluated through the calculation of the number of the conducted biopsies, proxy of coverage and the results of cytopathology test.

Box 4: The performance indicators of studies identified in Brazil

Indicators			THULER *et al.*, 2007	FEITOSA *et al.*, 2007	CABRAL *et al.*, 2008	NOBRE *et al.*, 2009	UCHIMURA *et al.*, 2009	DIAS *et al.*, 2010	BASTOS, 2011
Test-based		Adequacy of slides/specimens	X			X		X	X
		Presence of TZ		X	X				X
		Percentages of cytopathology test results	X	X	X		X	X	X
		Ratio between low-grade and high-grade cytopathological findings	X	X	X	X	X		
		Percentages of histopathology test results							X
		Proxy of coverage				X	X	X	X
Woman-based	Coverage	Real							X
	Follow-up of women	According to the cytopathological findings							X
		Time interval between tests							X
		Missed women between tests							X
Indicators related to the administration of the laboratories			X						
Scope of Study			National	Regional	Regional	Regional	Regional	National	Regional

Box 4 (cont): The performance indicators of studies identified in Brazil

Indicators			SANTOS *et al.*, 2012	BORTOLON *et al.*, 2012	ÁZARA *et al.*, 2014	PLEWKA *et al.*, 2014	DISCACCIATI *et al.*, 2014	AHMED *et al.*, 2014a	DA SILVA *et al.*, 2014
Test-based		Adequacy of slides/specimens	X				X		
		Presence of TZ						X	
		Percentages cytopathology test results	X	X	X	X	X	X	X
		Ratio between low-grade and high-grade cytopathological findings	X	X	X	X			
		Percentages of histopathology test results							
		Proxy of coverage	X						X
Woman-	Coverage	Real							
	Follow-up of women	According to the cytopathological findings							
		Time interval between tests							
		Missed women between tests							
Indicators related to the administration of the laboratories									
Scope of Study			National	National	Regional	Regional	Regional	Regional	Regional

Box 4: (cont): The performance indicators of studies identified in Brazil

Indicators			BRITO-SILVA *et al.*, 2014	NASCIMENTO *et al.*, 2015	COSTA *et al.*, 2015	MORAES *et al.*, 2015	ASSUNÇÃO *et al.*, 2015	ARAUJO JR *et al.*, 2015	TOBIAS *et al.*, 2016	RODRIGUES *et al.*, 2016
Test-based		Adequacy of slides/specimens			X	X	X			X
		Presence of TZ			X					
		Percentages/quantities of cytopathology test results	X		X	X	X	X	X	X
		Ratio between low-grade and high-grade cytopathological findings			X			X	X	
		Percentages/quantities of histopathology test results	X							
		Proxy of coverage	X	X	X	X	X			X
Woman-based	Coverage	Real								
	Follow-up of women	According to the cytopathological findings								
		Time interval between tests								
		Missed women between tests								
Indicators related to the administration of the laboratories										
Scope of Study			Regional	Regional	National	Regional	Regional	Regional	Regional	Regional

In spite of the importance of the woman-based indicators, these are seldom used for the evaluation of the Brazilian Screening Program. Bastos et al. (2011) used SISCOLO data, from the year 2006 until 2009, to estimate the effectiveness of the screening program for the state of Rio de Janeiro as well as to construct a set of woman-based and test-based indicators. A probabilistic record linkage was performed and then the number of women identified in each annual interval was used to calculate the coverage of the cervical cancer screening. For the women identified in the first interval with a test result not indicating a cancer, a follow-up analysis of the adherence to the Brazilian screening standard approach was performed by monitoring three consecutive tests. In order to do this analysis, these women were categorized into three groups according to the result of the first identified test. Group 1 contains women with test result of normal and benign changes; group 2 contains women with test results of low-risk atypical findings; and group 3 contains women with test results of high-risk atypical findings. For the groups 1 and 2, the following indicators were evaluated: the presence or absence of subsequent tests; the average time to perform the tests; and results of the subsequent tests. The number of women who were followed up and not followed up during the observation period of three tests was calculated. The effectiveness was estimated by comparing the number of detected atypias at high risk for cancer and the number of potential atypias at high risk that could be detected when there were no failures in the detection process and in the follow-up of women. The results showed that the program detected 6,397 atypias at high risk and failed to detect 12,730 due to loss of follow-up of the eligible women and failures in the detection process due to the absence of specimens of cells from the TZ. Thus, its effectiveness was estimated as 33.4%.

In summary, the review shows that some indicators are more frequent than others (Figure 9). The most frequent indicators used in Brazil are: percentages of test results; a proxy of coverage; ratio between low-grade and high-grade cytological findings; and adequacy of Pap smear specimens. In spite of the importance of the indicators related to the presence of cells of TZ in the specimens, they were reported only in five studies. The less used indicators are the woman-based indicators related to the follow-up of women and real coverage, which were reported in only one study (BASTOS, 2011).

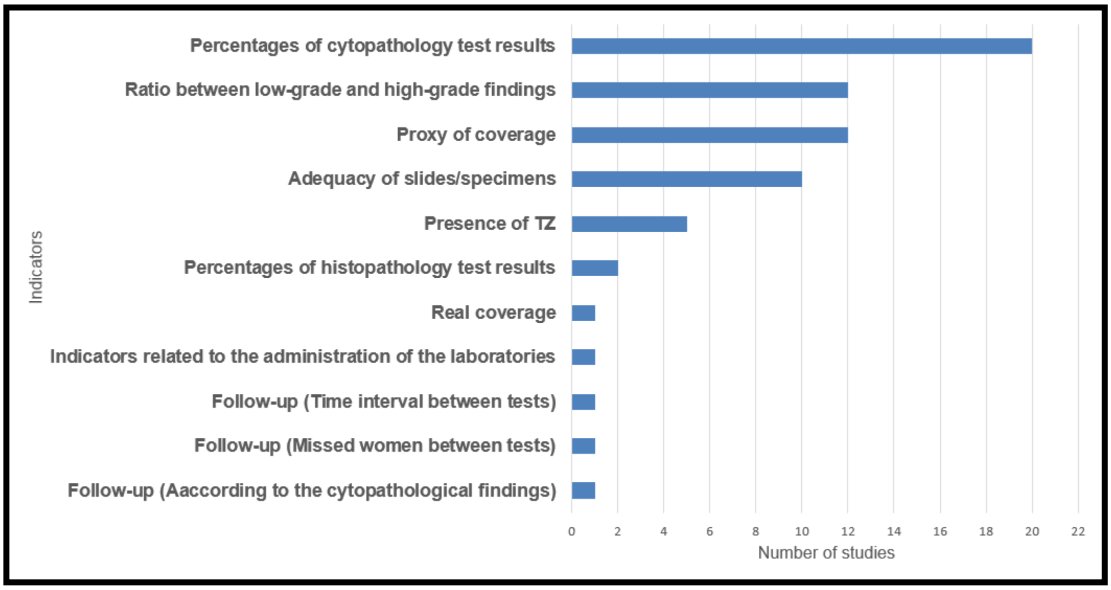


Figure 9: Distribution of the number of studies according to the identified indicators.

4.3 Visualization environments of performance indicators

The goal of visualization is to help user to easily understand and interpret huge and complex sets of information (KHAN *et al.*, 2011; SCHIRNDING, 2002). Five visualization environments of the performance indicators for the evaluation of the Brazilian Cervical Cancer Screening Program were identified. These environments were selected by searching the websites of the Brazilian Ministry of Health and its related institutions.

The first and largely used environment is an online tabulation tool developed by DATASUS called *Tabulador para Internet* (TabNet, in Portuguese). TabNet uses information from different data sources including SISCOLO to construct the indicators. TabNet permits the user to make a cross tabulation to visualize a set of test-based indicators related to the cytopathology and histopathology as well as colposcopy exam (BRAZIL, 2015b). Those indicators are visualized as quantities rather than percentages which requires an extra effort if they are to be used for comparative purposes.

The second tool is called *Tabulador para Windows* (TabWin, in Portuguese) which is a standalone tool offered by the Brazilian Ministry of Health (BRAZIL, 2016b). TabWin uses information from different data sources including SISCOLO to construct the test-based indicators. Using TabWin requires the downloading of the monthly

SISCOLO files and the indicators must be manually configured. This is a tedious and time-consuming process, which requires some skills in data management, which is usually not the case. On the other hand, TabWin has the advantage of allowing geographic maps to visualize the indicators, which helps the manager to monitor the indicators in the municipalities.

INCA offers an online indicators panel that used SISCOLO data to visualize a time series of five test-based indicators. These indicators are related to proxy of coverage, percentage of test results, adequacy of Pap smear specimens and ratio between low-grade and high-grade atypical findings (BRAZIL, 2015c).

Another online environment offered by the Ministry of Health is called Strategic Management Supporting Hall (SAGE, in Portuguese). SAGE uses data from the Outpatient Information System (SIA, in Portuguese) to visualize the quantities of performed Pap smear tests for all the Brazilian states in 2011 and 2012, as well as the quantities of performed tests in Brazil from 2011 until 2014 (BRAZIL, 2015d).

The last identified environment is called SISCAN web which is the web based platform of the Cancer Information System (SISCAN, in Portuguese), which integrates the data of the National Program for Control of Breast Cancer (SISMAMA, in Portuguese) and the SISCOLO data (BRAZIL, 2013a). SISCAN web also uses other data sources such as the National Register of Health Facilities (CNES, in Portuguese) and the National Register of Users of the Unified Health System (*Cartão* SUS, in Portuguese). In the SISCAN web, the women's data are registered online by both primary healthcare units and laboratories. One of the main advantages of the platform is to identify the women by means of the *Cartão* SUS as a unique identifier. Taking the advantages of this unique identifier, SISCAN web creates a follow-up module for women with altered test results. The women start to be followed-up from the time of the first altered test results. The follow-up module permits monitoring the history of each woman individually as well as monitorization of her future test results.

Unlike the other environments, the access to SISCAN web is not public. In order to access the platform, users need to log in, which is available only to the State, regional and municipality program coordinators as well as to the managers of primary healthcare units and laboratories (BRAZIL, 2013a).

Although SISCAN web was deployed in 2011, it is still in a phase of implementation and adjustment, and up to now, it is not completely deployed in many Brazilian States, including Rio de Janeiro State. Also, the health care units without

Internet access are still using the old paper-based system (i.e. they manually fill in the requisition forms and send them to the laboratory with the slides).

Box 5 summaries the indicators provided by the five mentioned environments and their visualization features. It is clear that the woman-based indicators are less applied in those environments. The only environment that seems to overcome this limitation is SISCAN web, considering the information reported in the manual (BRAZIL, 2013a). On the other hand, TabNet offers public access with the greatest number of test-based indicators.

Beside the advantages of the available environment, some limitations were observed:

- The indicators used in Brazil are constructed in a fragmented form.
- A limited number of the woman-based indicators is available.
- Managers need to consult many environments in order to evaluate the program.
- Environments present a limited visualization of the indicators.

The construction of a user-friendly environment that overcomes the above limitations is required in order to improve the evaluation of the screening program and to facilitate access to all indicators in the same interface.

Box 5: Indicators and features of the available visualization environments

Indicators and Visualization Features			**TabNet**	**INCA Indicators Panel**	**TabWin**	***SAGE***	**SISCANweb***
Indicators	**Cytopathology test results**	Percentages/quantities of test results	Y	L	Y	L	L
		Ratio between low-grade and high-grade findings	N	L	Y	N	N
	Coverage	Proxy	N	Y	Y	N	Y
		Real	N	N	N	N	Y
	Adequacy of slides/specimens	Adequacy of specimens	Y	Y	Y	N	Y
		Presence of TZ	Y	N	Y	N	Y
	Follow-up of women	According to the cytopathological findings	N	N	N	N	Y
		Time interval between tests	N	N	N	N	Y
		Missed women between tests	L	N	N	N	Y
Visualization Features		Does the environment have a public access?	Y	Y	Y	Y	N
		Does the environment provide graphical visualization of the indicators	L	L	Y	L	N
		Does the environment permit the visualization in many dimensions (ex. time, primary healthcare unit)?	Y	L	Y	N	Y
		Does the environment provide tools for report generation?	N	N	N	N	Y
		Does the environment permits saving of indicators information in data files?.	L	N	Y	L	N
		Does the environment have statistical analysis tools	N	N	Y	N	N

Legend: L=Limited; N= No and Y= Yes

* The provided information is from (BRAZIL, 2013a)

CHAPTER 5 – MATERIALS AND METHODS

5.1 Data warehouse building methodology

The building of the proposed DW model involved several steps. These steps started with the selection of the performance indicators to be constructed for the DW and ended with the visualization of those indicators in a user-friendly graphical user interface (GUI). Figure 10 illustrates the flowchart for the steps of the building process.

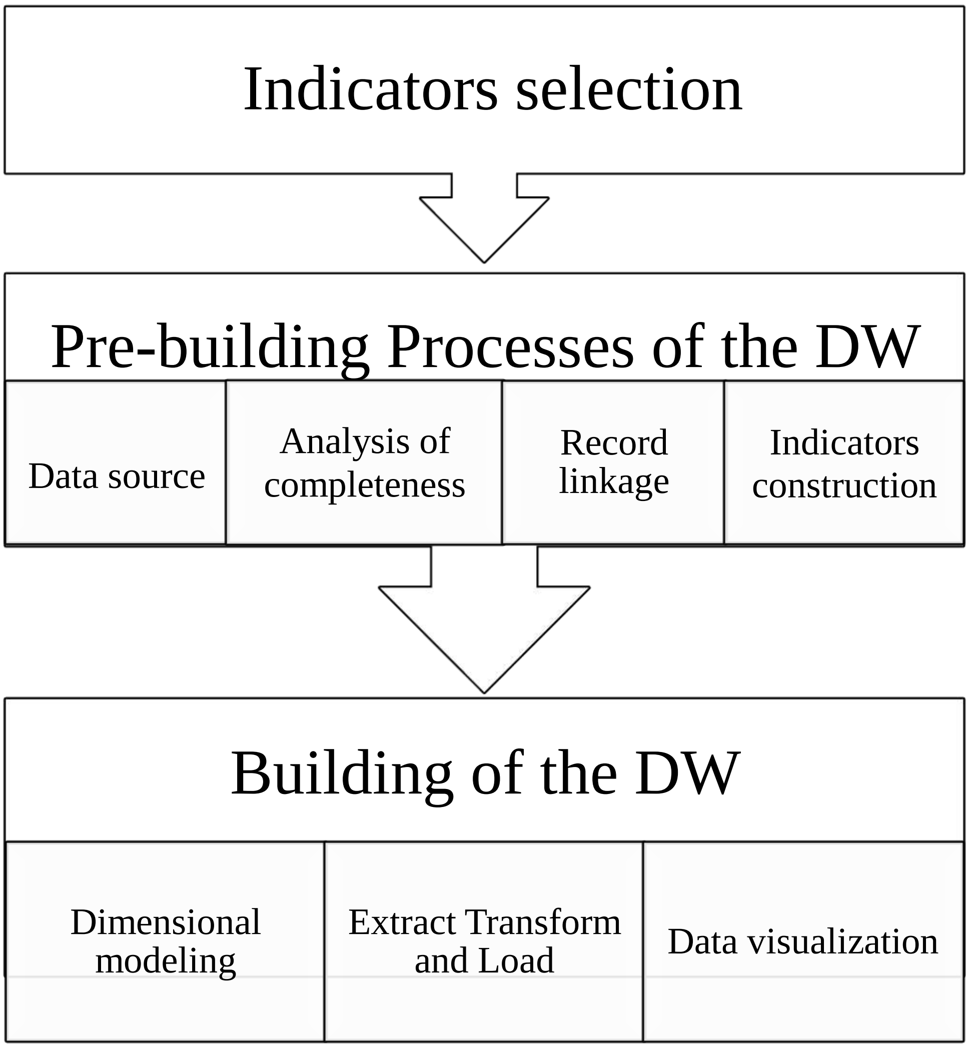


Figure 10: Flowchart of the data warehouse building process.

5.2 Indicators selection

The first step before the building of the proposed DW was the selection of the performance indicators to be obtained from the DW. The indicators were selected from the identified indicators from both international and national literature reviews, as described in sections 4.1 and 4.2. After the exclusion of the redundant/duplicate indicators in both reviews, two inclusion criteria were applied: availability of data in the data source; and data consistency. Applying these criteria, a set of 115 indicators was selected. The selected set of indicators was classified into two main classes: test-based and woman-based, and then each class was divided into groups (Figure 11).

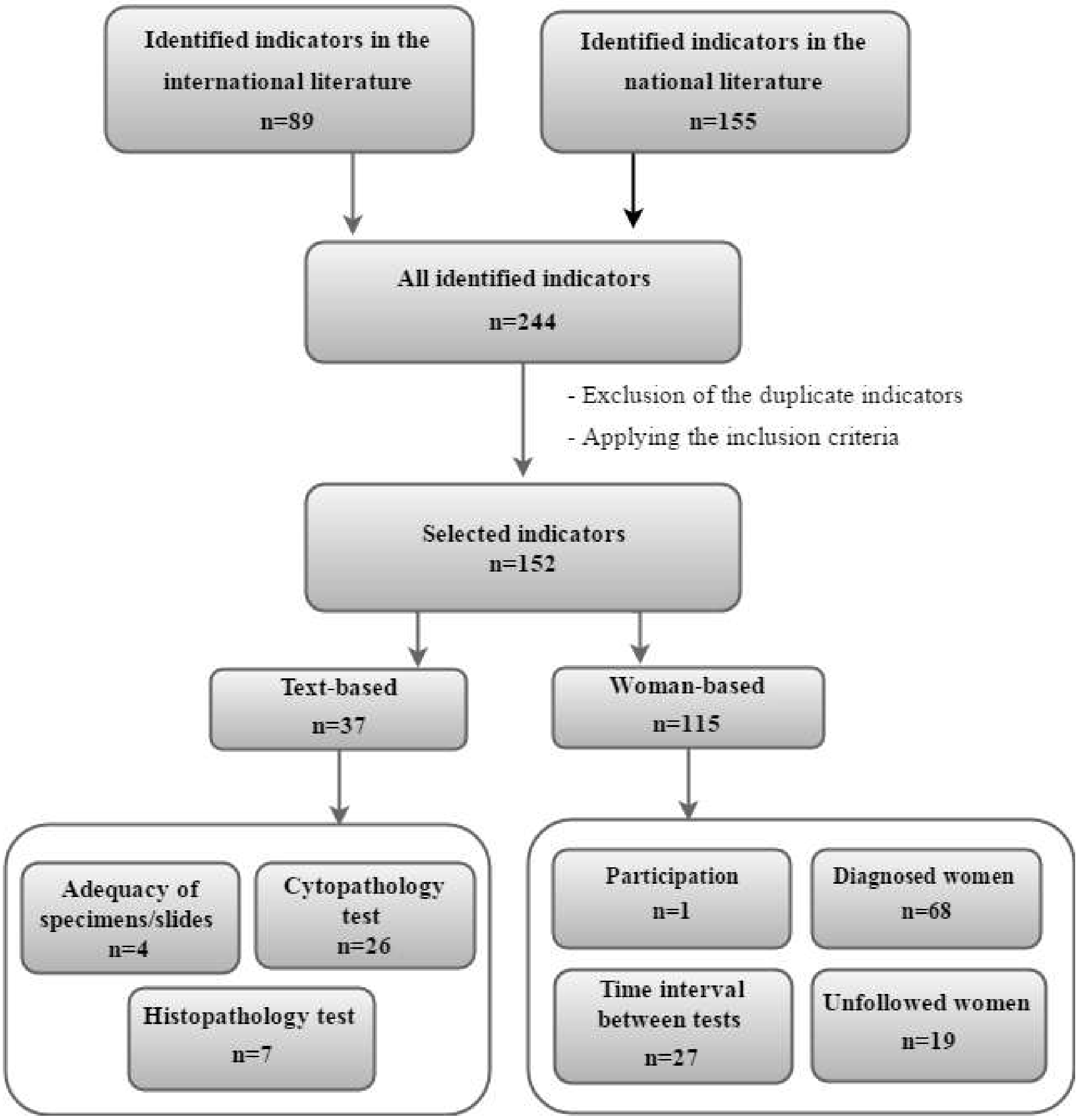

Identified indicators in the international literature
n=89
Identified indicators in the national literature
n=155
All identified indicators
n=244
- Exclusion of the duplicate indicators
- Applying the inclusion criteria
Selected indicators
n=152
Text-based
n=37
Woman-based
n=115
Adequacy of specimens/slides
n=4
Cytopathology test
n=26
Histopathology test
n=7
Participation
n=1
Diagnosed women
n=68
Time interval between tests
n=27
Unfollowed women
n=19

Figure 11: The selection process of the identified indicators.

In the following section, a description of each group will be presented where the first 3 groups represent the groups of the test-based indicators and the other groups represent the groups of the woman-based indicators.

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.

Cytopathology test: Indicators related to the results of the cytopathology test.

Histopathology test: Indicators related to the results of the histopathology test.

Participation: Indicator related to the participation of women (within the recommended age of the program) in the program.

Diagnosed women: Indicators related to the quantity of women classified by their age and the results of the cytopathology/histopathology tests.

Interval between tests: Indicators related to the average time interval between tests.

Unfollowed/missed women: Indicators related to the quantity of women who did not undergo the next recommended test.

The three groups of woman-based indicators (diagnosed women, interval between tests and unfollowed women) 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 (BRAZIL, 2006a).

5.3 Pre-building processes of the data warehouse

5.3.1 Data source

This study used the database of the National Cervical Cancer Information System (SISCOLO) for the municipality of Rio de Janeiro, from January 2012 to December 2014. The database has two tables: histopathology and cytopathology, which respectively contain the results of histopathologic (4,413 records) and cytopathologic (1,109,731 records) examinations of women. Originally, the data was organized into monthly files in Data Base File (DBF) format. These files were consolidated into two tables, which from now on will be called Histology and Cytology.

5.3.2 Analysis of completeness and consistency of the SISCOLO variables

An exploratory analysis was performed in order to assess the completeness and consistency of the SISCOLO variables. In order to perform this analysis, auxiliary layouts/dictionaries offered by DATASUS were used (BRAZIL, 2011c). These layouts contain a description of the variables in both Cytology and Histology tables. Once again, the values of those variables were obtained from the information in the cytopathology/histopathology acquisition forms. Although these acquisition forms were updated in the year 2012, the SISCOLO database remains the same (without inclusion/exclusion of some variables according to the updated acquisition forms). In this work, the frequencies of the existent SISCOLO variables were used to obtain the percentages of completeness. The following groups of fields were considered for the analysis:

Cytology table:

- Identification of the primary healthcare unit and laboratory;
- identification of woman;
- clinical information of woman (anamnesis and clinical examination);
- results of the test.

Histology table:

- identification of the primary healthcare unit and the laboratory;
- identification of woman;
- colposcopy exam;
- results of the test.

5.3.3 Linkage process of the SISCOLO data

The linkage process was performed using the VincReg software following the same strategy used by Freire *et al*. (2012). Figure 12 shows the flow diagram of the performed steps.

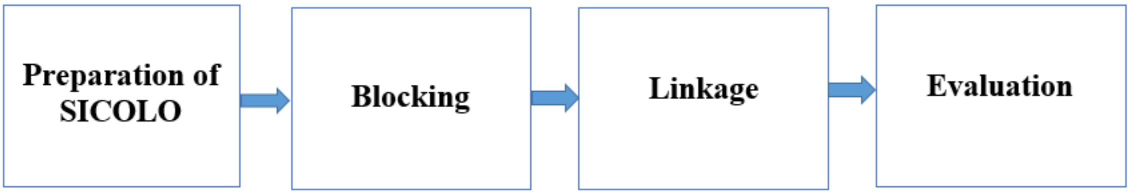


Figure 12: Flow diagram of the of the record linkage steps.

Preparation of SISCOLO data

The preparation of the SISCOLO data consisted of the analysis of completeness and consistency as described in section 5.3.2 and the processes of data cleansing, standardization and parsing. Both Cytology and Histology tables were imported into the MySQL database manager (MYSQL, 2015b). The cleansing process was performed using SQL queries where the woman's name, mother's name and address fields were processed as follows:

- conversion of all alphabetic characters to upper case;
- removal of blank spaces;
- removal of all accentuation, special characters and prepositions;
- removal of expressions such as "doutor" (doctor), "ignorado" (ignored), "rua" (road);
- removal of the annexes of the name such as "filha", "filho","neto".

The standardization process was performed with VincReg and the following steps were performed:

- breaking down of the birthdate into day, month and year of birth;
- breaking down of the name into parts as follows: first name, last name, middle names initials (if any), the person's second name (if any);
- coding of the first name and last name by applying a phonetic algorithm called Soundex with some modifications proposed for the Portuguese language (CAMARGO *et al.*, 2002).

Blocking

The blocking process was performed only in the Cytology table due to its large size. A five-step blocking strategy was performed where the blocking variables in each step were as follows: i) "phonetized first name of the woman", "phonetized first name of the mother", and "age group"; ii) "phonetized last name of the woman", "phonetized last name of the mother" and "age group"; iii) "phonetized first name of the woman", "phonetized last name of the mother", and "age group"; iv) "phonetized last name of the woman", "phonetized first name of the mother" and "age group"; v) "birth month" and "birth day". The "age group" variable was created from the woman's year of birth considering intervals of ten years. In order to reduce the size of the blocks in the blocking process, the first letter of the second name was added to the end of the common first names “Maria”, “Ana”, “Adriana” and “Marcia” before phonetization. Similarly, the first letter of the second last name was added to the beginning of the common surnames “Silva”, “Souza”, “Santos” and “Oliveira” before phonetization.

Linkage

In this step, three linkage processes were performed: internal linkage of the Histology table with itself, internal linkage of the Cytology table with itself, and external linkage of both Histology and Cytology tables. The internal linkages were performed to identify the same woman in a single table, while the external linkage was performed in order to identify the same woman in the both tables.

Both internal and external linkages were performed by applying a probabilistic linkage. In order to perform the linkage process, the VincReg software generated a table containing the pairs (links), their respective scores, and the values of the variables (name of the woman and mother, date of birth, residence, healthcare unit, and, when present, Individual Identification Number (CPF, in Portuguese), *Cartão* SUS, identity and telephone number). In order to define the values of the scores that established the cut-off points of the region of true pairs and false pairs, the generated table was sorted by the score value (decreasing order), and then a visual analysis was carried out.

The detailed description of all the steps of the linkage process can be found in (FREIRE *et al.*, 2012).

Evaluation

Freire *et al*. (2012) applied the same linkage strategy to a previous series of SISCOLO data and obtained excellent results in terms of sensitivity and specificity. Therefore, a formal evaluation of the linkage process was not performed in this study. However all records of women in the Cytology table with more than nine tests were manually inspected and corrected, in case a visual inspection indicated that one or more records did not belong to a specific woman. This manual inspection was also performed in all records of women with more than two tests in the Histology table.

5.3.4 Indicators construction methodology

This section will present the details of the construction methodology for both test-based and woman-based indicators described in section 5.2.

Construction of the test-based indicators

The test-based indicators were calculated as percentages or ratios. For example, the indicator “Percentage of exams with low grade squamous intraepithelial lesion (HSIL)” was calculated by the equation:

$$\frac{\text{Total number of tests with result HSIL}}{\text{Total number of performed tests}} * 100 \tag{5.1}$$

where: the number of performed tests represents the tests of the slides with satisfactory specimens.

A list of the all the test-based indicators and theirs formulas of calculation is presented in the Appendix 1.

Construction of the woman-based indicators

The following paragraphs will present the details of the construction methodology of both the participation and follow-up indicators. The participation was calculated by the formula:

$$\frac{\text{Total number of the screened women (25-64 years) in a specific year}}{\text{Female population (25-64 year) in the same year}} * 100 \tag{5.2}$$

In order to construct the follow-up indicators, the identified indicators in the work of (BASTOS, 2011) were recalculated in a more detailed form, permitting the follow-up of women according to the recommended approaches of the Brazilian Guidelines for Cervical Cancer Screening for the year 2016 (BRAZIL, 2016d). 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 DW should be up to date to meet the recent and future management needs.

In order to construct the follow-up indicators, 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 were analyzed with the aid of a specialist in pathology and another in the coordination of the Program of Prevention and Early Detection of Cancer (personal communication)[1]. 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 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 hah a result of ASC-US and 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 performed a cytopathology test and their clinical recommendation to perform 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. As mentioned before, this information is not available in the SISCOLO data in the observation period.
- Due to the dismissal of the colposcopy results, if a woman had undergo a histology test after the first one, this test was selected. Otherwise, if the next test was a cytopathology test, this test was selected.
- The time interval between tests was calculated by the difference between the date of release of the report of the result and the date of collection of the next specimen/biopsy.

[1] Information obtained by personal communication with Tereza Maria Piccinini Feitosa and Lucilia Maria Garbo Zarado.

- 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 cytopathological test results of 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 is to perform a cytopathology test within six months were followed until the fourth test, since the period of observation allowed the follow-up until the fourth test.
- In the cases when the guidelines do 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 performing a cytopathology test within 3 year 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 groups of women were formed. Box 6 shows each group and its corresponding entry year. The women (≥ 30 years) with initial results ASC-US and women (≥ 25 years) with initial results LSIL were classified into the same group, since they had the same recommended approaches until the fourth test (considering the absence of the colposcopy results).

The women (<25 years) with initial test results of LSIL and women (<25 years) with initial test results of ASC-US were classified into the same group since their follow-up required a long series of data (more than 3 years), which is not the case here.

However, in order to provide the program managers with a general view of the condition of women of this group, the corresponding flow diagram of this group was prepared in a way that allowed the managers to check if the women underwent the next test, and the results of this test.

Box 6: The 11 groups of women according to the initial cytopathology test result, age, and the corresponding initial clinical approach, and entry year in the database.

Groups	Initial clinical approach	Entry year
LSIL (women < 25 years) and ASC-US (women< 25 years)	Repeat cytopathology test in 3 years	2012
Normal findings and benign changes	Repeat cytopathology test in 12 months	2012/2013
ASC-US (women between 25 a 29 years)		
ASC-US (women ≥ 30) and LSIL (women ≥ 25 years)	Repeat cytopathology test in 6 months	
ASC-H	Perform colposcopy exam	2012/2013/ 2014*
AOI		
AGC		
HSIL (women≥ 25)		
HSIL (women ≤ 24 years)		
High-grade intraepithelial lesions, cannot exclude micro-invasion and squamous cell carcinoma		
Adenocarcinoma (AIS) in situ and invasive		

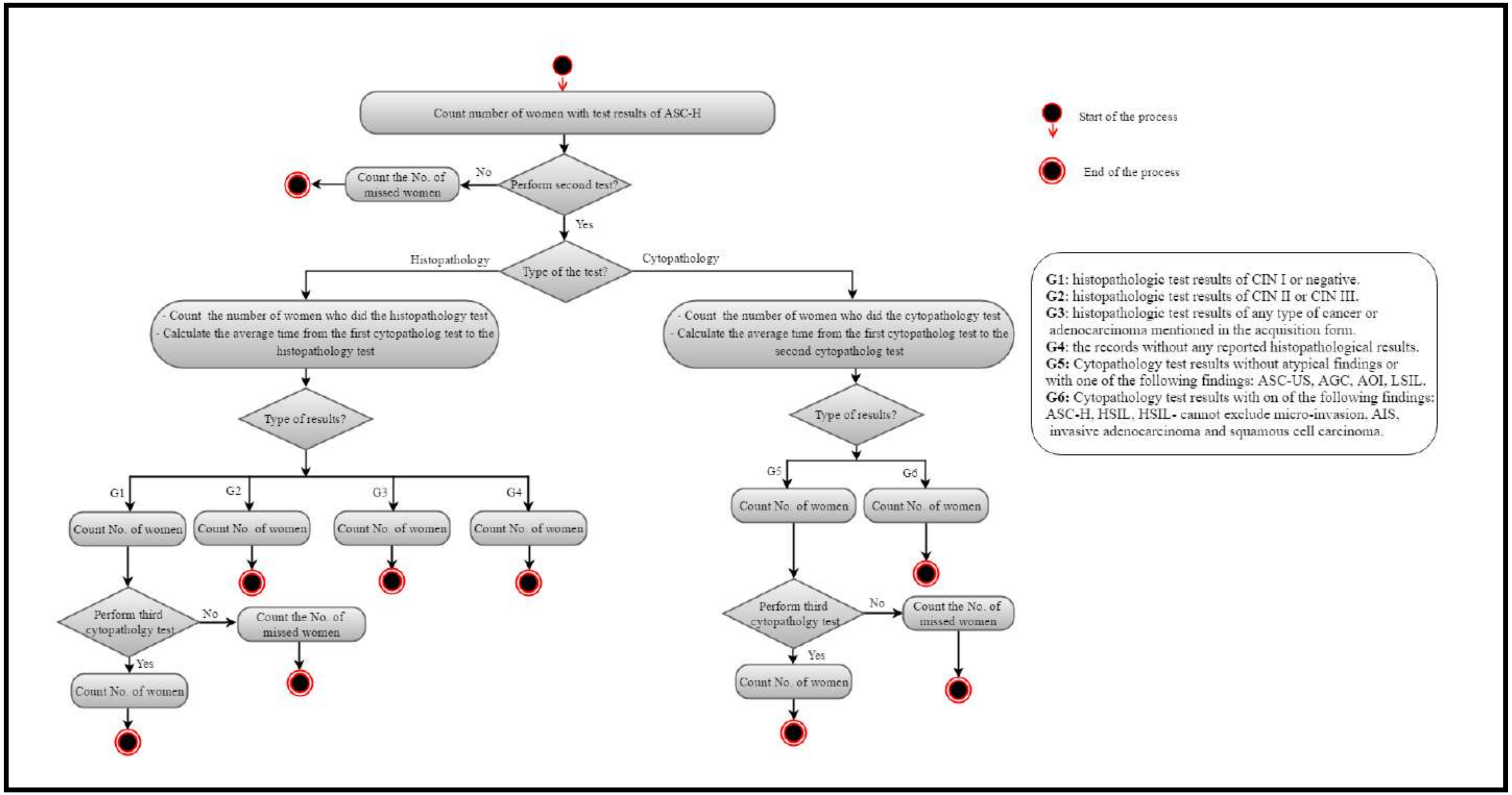
*For the women who performed the first test in 2014, it is necessary to have at least another cytopathological or histopathological test in the database.

In order to demonstrate how the clinical approaches were simplified, the following paragraphs will present in detail an example of the simplification of the recommended clinical approach for women with an initial test result of ASC-H (Figure 13). The original flow diagram of this case was presented in Figure 2, section 3.4. Considering the absence of the colposcopy information, the results of the colposcopy exam were ignored and the next test of the woman was considered to be the histopathology test, if present, or the next cytopathology test. This was a general rule for the construction of the simplified clinical approach.

For the women who performed a histopathology test, the results of this test were classified into four groups: G1) negative and CIN I; G2) CIN II/III; G3) all types of cancer and adenocarcinoma present in the acquisition form; and G4. The negative results were considered as those results with “lesions of benign characters”. The group labeled as G4 contains the women without any reported histopathological results. The women in group G1 were followed until the third test (cytopathology), while the follow-up of women in groups G2 and G3 was terminated at the second test, where the next recommended approach was to switch to another approach. Although groups G2 and G3 are classified as one group in the Brazilian Guidelines, they were separated in order to provide the manager with a more detailed view, which could help in the evaluation process.

For women who performed a cytopathology test, the result of this test was classified into two groups: G5 with ASC-H HSIL, HSIL cannot exclude micro-invasion, squamous cell carcinoma, AIS and invasive adenocarcinoma; and G6 with ASC_US, AOI, AGC, LSIL and negative test results. The negative results represented results without any atypical findings. Women in group G5 were followed until the third cytopathology test, while the follow-up of women in group G6 was terminated at the second test since the next recommended approach is to perform an excision procedure.

The simplified flow diagrams for the other 10 group are presented in Appendix 2.


Count number of women with test results of ASC-H
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
G1
G2
G3
G4
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
Type of results?
G5
G6
Count No. of women
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
Start of the process
End of the process

G1: histopathologic test results of CIN I or negative.
G2: histopathologic test results of CIN II or CIN III.
G3: histopathologic test results of any type of cancer or adenocarcinoma mentioned in the acquisition form.
G4: the records without any reported histopathological results.
G5: Cytopathology test results without atypical findings or with one of the following findings: ASC-US, AGC, AOI, LSIL.
G6: Cytopathology test results with on of the following findings: ASC-H, HSIL, HSIL- cannot exclude micro-invasion, AIS, invasive adenocarcinoma and squamous cell carcinoma.

Figure 13: The simplified flow diagram for the recommended clinical approach for the women with diagnosis of ASC-H.

The ideal calculation of the unfollowed (missed) women between the first and second tests should consider two groups of women (according to their colposcopy results), which are:

- the group of women whose clinical approach after the colposcopy test was to have a histopathology test subtracted from the total number of women who had a histopathology test; and
- the group of women whose clinical approach after the colposcopy test was to undergo a cytopathology test, subtracted from the total number of women who underwent a second cytopathology test.

Due to the absence of the information about the colposcopy results, the missed women were calculated as the total number of the women who had the first test subtracted from the total number of women who had a second test (which could be either a histopathology or a cytopathology test).

Formulation of the follow-up algorithms

In this step, the simplified flow diagrams were used to write the algorithms to calculate the indicators in all the 11 diagrams. An example of the algorithm for follow-up of women with the initial result of ASC-H is presented in Appendix 3.

5.4 Building of the data warehouse

The DW was implemented following Kimball's dimensional model. Figure 14 illustrates the steps for building the proposed DW, which are detailed below.

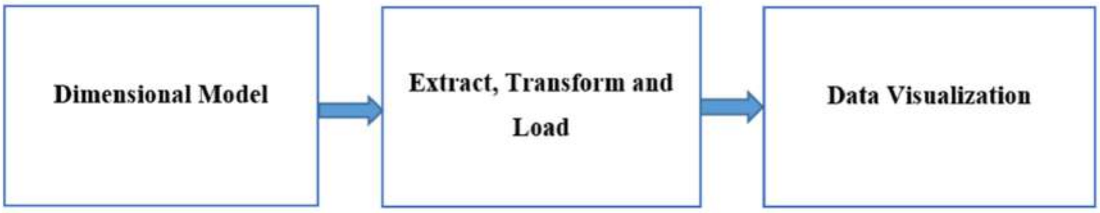


Figure 14: Flowchart of the data warehouse building process.

5.4.1 Dimensional modeling

In order to generate the dimensional model, the Cytology and Histology tables were first imported into the MySQL database manager (MYSQL, 2015b). After that, a star schema was generated using a MySQL workbench, which is a visual tool for modeling databases and managing MySQL servers (MYSQL, 2015a). The fact table for the schema contains the set of indicators that was selected in section 5.2. The schema has five dimensions, which are:

- Time;
- Municipality: represents residence municipality of screened women;
- Collection unit: represents the primary healthcare unit where the specimens were collected;
- Laboratory in which the specimens were analyzed;
- Age group: represents the age group of women at the time of specimen's collection. Here the age group is divided into intervals of 5 years.

These five dimensions were selected since they enable managers to browse the indicators from different views, which could enhance the decision making process. Although this study uses the SISCOLO data from the municipality of Rio de Janeiro, the dimension for the municipality of residence was included in the dimensional model because many women from other municipalities undergo the screening test in the municipality of Rio de Janeiro. In addition, many primary healthcare units in other municipalities sent their specimens to be analyzed in the municipality of Rio de Janeiro.

The attributes of the dimensions were selected in a way that made it easy to visualize the indicators at different levels/views of granularity. For example, the time dimension has year and month as attributes, so that any indicator could be monitored at the year level (higher granularity) or at the month level (lower granularity). Box 7 shows the attributes for each one of the five dimensions.

Box 7: Attributes of the proposed dimensions of the DW model

Dimension	Attributes
Time	Year and Month
Municipality	Name, code of the municipality, name of the health region, code of the heath 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

* National Register of Legal Entities (CNPJ, in Portuguese)

5.4.2 Extract, transform and load (ETL) process

As mentioned before, 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 (KIMBALL *et al.*, 2011). In this study, the data was provided by the following sources:

- The linked SISCOLO data as described in section 5.3.3.
- 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 (BRAZIL, 2015e; BRAZIL, 2015d).
- Data about the female population of the state of Rio de Janeiro according to the national census of the year 2010. This data was obtained from the site of the DATASUS (BRAZIL, 2015f).
- The data for the time and age group dimensions was obtained from the transformation of the corresponding variables in the SISCOLO database.

The ETL was implemented using MySQL as a database manager and Pentaho Data Integration (PDI) and the Java programming language (JAVA, 2016; MYSQL, 2015b; PENTAHO, 2016). PDI is one of the tools offered by the Pentaho BI Suite Business Intelligence (PBIS) platform community edition version 6.1.0.1-196 (PENTAHO, 2016).

PBIS 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. The following sections will describe how these tools were used in the ETL process for populating the dimensional tables, and the test-based and woman-based facts.

ETL of the dimensional tables

In this step, the PDI was used to extract, transform and load the data into the dimensional tables. The laboratory dimension contains eight attributes: name of the laboratory, CNES, CNPJ, code and name of the laboratory municipality, code and name of the laboratory health region and name of the federal unit. Since there is no unique table that has all of the mentioned attributes, data from five different tables (in Excel format) was used to load the data of the laboratory dimension (Figure 15). Names, attributes and sources of those tables are listed in Box 8.

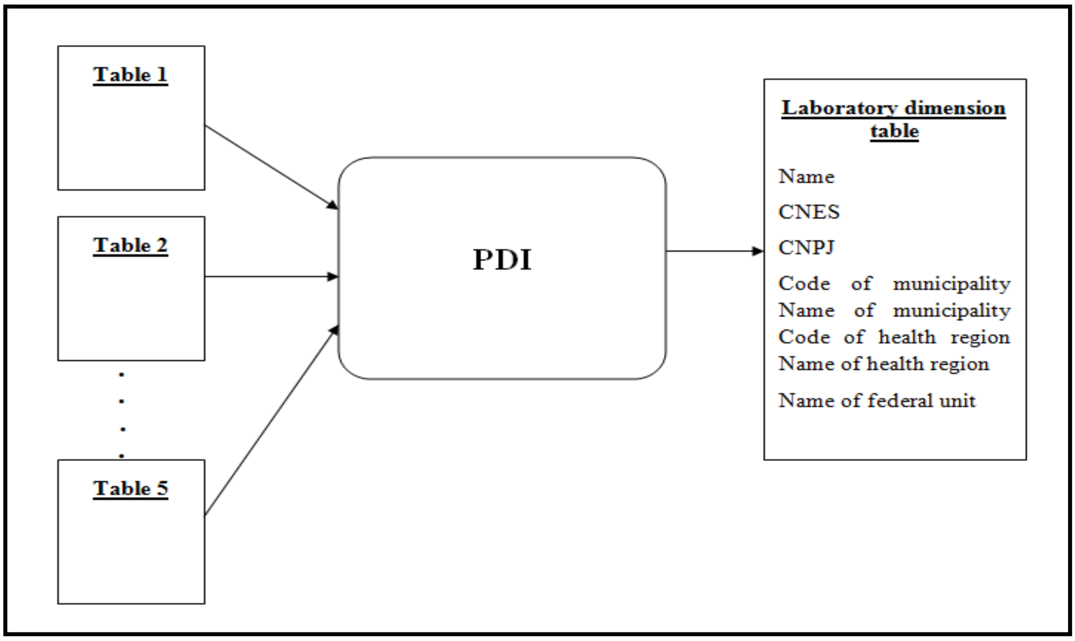


Figure 15: ETL process of the laboratory dimensional table.

Box 8: The tables used for the ETL process of the laboratory dimension

Table name	Attributes	Source
Lab	CNES CNPJ Laboratory name Name of the municipality	BRAZIL, 2015e
Municipality	Name of the municipality Code of the municipality Code of the federal unit	BRAZIL, 2015f
Health region	Code of the health region Name of the health region	BRAZIL, 2015f
Municipality_Health region	Code of the municipality Code of health region	BRAZIL, 2015f
Federal Unit	Code of federal unit Acronym of the federal unit	This table was created manually by the author

PDI was used to perform the steps below:

- Select the following attributes:
  - CNES, CNPJ, laboratory name and name of the municipality from the Lab table.
  - Municipality code, municipality name, and code of the federal unit from the municipality table.
  - Health region code and name from the health region table.
- Insert all the above attributes into the laboratory dimension table taking into account the assignment of the correct value for each attribute. The following example explains what is meant by assigning the correct value of attribute. The table Lab has “name of the municipality” as an attribute but do not have the “code of the municipality”. In order to add the attribute “code of the municipality”, we need to take the later one from the table “Municipality”. To ensure that each municipality will be assigned its correct code, the following condition was applied: “assign the code for the municipality if it has the same name in the two tables “Lab” and “Municipality” or, in the other words, obtain the code for the municipality from the join of the two tables on the attribute “Municipality Name”. Boxes 9, 10 and 11, illustrate that by joining the two tables on the attribute “Municipality Name” the correct code of municipality is assigned to both “Rio de Janeiro” and “Angra dos Reis” (Box 11).

Box 9: Example of records in the table “Lab”

CNES	CNPJ	Laboratory Name	Municipality Name
2271443	6092***00	Casa de Saude Sao Jose	Rio de Janeiro
2281295	2917***00	Laboratorio Sao Dimas	Angra dos Reis

Box 10: Example of records in the table “Municipality”

Name	Municipality Code	Federal unit
Rio de Janeiro	330455	33
Angra dos Reis	330010	33

Box 11: Example of records in the laboratory dimensional table

CNES	CNPJ	Laboratory Name	Municipality Name	Municipality Code
2271443	6092***00	Casa de Saude Sao Jose	Rio de Janeiro	330455
2281295	2917***00	Laboratorio Sao Dimas	Angra dos Reis	330010

Box 12 shows all the conditions that were used for loading the data into the laboratory dimension. In this table, in order to assign the value to any derived attribute, the attributes in the column “attribute” should be the same in the columns “first table” and “second table’, or only in the first table, when the second one is not stated.

Box 12: Conditions for assigning values to attributes of laboratory dimension

Derived Attribute	Attribute	First Table	Second Table
Code of the municipality	Name of the municipality	Municipality	Lab
Code of the health region	Code of the municipality	Municipality	Municipality_ Health region
Name of the Health region	Code of the health region	Health region	-
Acronym of the federal unit	Code of the federal unit	Federal unit	-

ETL of the test-based indicators

The construction of many test-based indicators required more than one fact. For example, the indicator "percentage of tests with result HSIL" is calculated by the following formula:

$$\frac{\text{Total number of tests with result HSIL}}{\text{Total number of performed tests}} * 100 \qquad (5.3)$$

In formula 5.3, the numerator and denominator represents two different facts. In order to perform the ETL process, a series of SQL queries were used to count the "total number of tests with results HSIL" and the "total number of performed tests" for each combination of the primary key values of the dimension tables. Then the resulting values from the queries were inserted in a temporary table in the MySQL database manager. The PDI was used to access the temporary table, obtain the stored values and insert them into the fact table. The processes of division and multiplication in equation 5.3 were performed as part of the visualization step, which will be explained in the visualization section.

ETL of the woman-based indicators

In order to perform the ETL process of the woman-based indicators, firstly both Cytology and Histology tables were prepared by the creation of all required variables for the follow-up process. Those variables were created according to the simplified flow diagrams described in section 5.3.4. Secondly, the created algorithms for all the 11 groups were translated into a JAVA program which was executed in order to load the fact table. The implementation and execution of the JAVA routines was performed by a JAVA expert who is one of the project team members.

5.4.3 Data visualization

The data visualization step consisted of three processes, which are: creation of the data cubes, preparation of the pre-defined reports; and the final visualization of the indicators in the graphical user interface. The following paragraphs present a description of each process.

Creation of the data cubes

After loading the facts and dimensions, five data cubes were generated using one tool of the PBIS called Pentaho Schema Workbench (PSW). These data cubes represent cubes for the visualization of the test-based indicators (adequacy of slides/specimens, cytopathology and histopathology), participation indicator and the last cube contains the indicators related to the follow-up of women in the first and second tests. The fact and dimensional tables were accessed by the PSW, and then the data cubes were generated in a semi-automatic way, using the graphical user interface of PSW, where the user only needs to define the facts, the dimensions, the granularity and the formula for the indicator calculation.

Preparation of the pre-defined reports

In order to visualize the follow-up indicators, another tool of PBIS called Pentaho Report Designer (PRD) was used to create the pre-defined reports in a standard and organized format. The PRD reports were prepared in a way that made it easy for the manager to monitor the indicators. The flow diagrams in section 5.3.4 were used to design simple/compact diagrams for the visualization of the indicators. Figure 16 presents the proposed flow diagram for the visualization of the follow-up process for the group of women with initial results of ASC-H. In Figure 16, the blue part of the boxes contains the information about the test (type of test, quantity of women who had the test and average time they took to undergo the next test). The white part of the boxes contains the quantities of women according to their test results where G1 to G6 correspond respectively to G1 to G6 explained in Figure 13 of section 5.3.4.

The graphical user interface

The created cubes and predefined reports were exported to the Pentaho BI-server, which is a web server that can be accessed via a web interface called User Console. The User Console of the BI-server has many tools that allow the user to visualize the data cubes and predefined reports. In this work, the pre-defined report of the follow-up indicators were visualized using the main interface of the User Console, while the test-based and participation, and the indicators related to the follow-up of women in the first and second tests, were visualized through an open-source plugin of the User Console called Saiku Analytics (SAIKU, 2015). The graphical interface of Saiku Analytics 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, the GUI provides interactive functionalities to perform the decision-support queries that managers normally need to address. Such functionalities involve the drill down, roll up and slicing/dicing of data. The Saiku GUI allows 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, PNG and JPEG.

5.5 Ethics Statement

This study was approved by the Research Ethics Committees of the Hospital Universitário Clementino Fraga Filho of the Universidade Federal do Rio de Janeiro/Brazil (CEP/HCFF-UFRJ – CAAE: 39106514.0.0000.5257) and the 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 unfeasible for all patients. For the purposes of this study, the information about the patients was anonymized and de-identified prior to the analysis, through the removal from the database of all details 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 was used only by authorized persons.

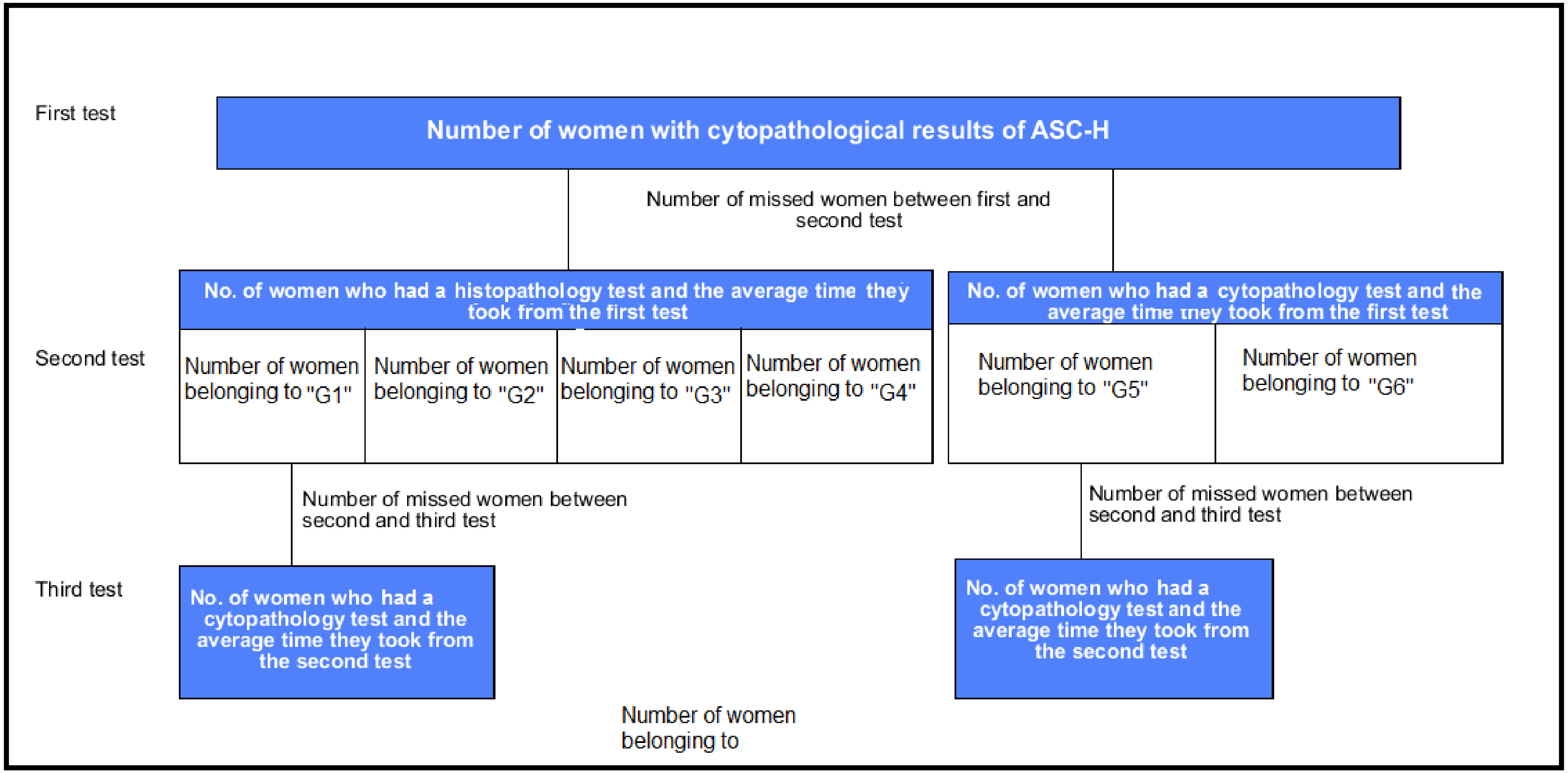

First test
Number of women with cytopathological results of ASC-H
Number of missed women between first and second test
Second test
No. of women who had a histopathology test and the average time they took from the first test
Number of women belonging to "G1"
Number of women belonging to "G2"
Number of women belonging to "G3"
Number of women belonging to "G4"
No. of women who had a cytopathology test and the average time they took from the first test
Number of women belonging to "G5"
Number of women belonging to "G6"
Number of missed women between second and third test
Number of missed women between second and third test
Third test
No. of women who had a cytopathology test and the average time they took from the second test
No. of women who had a cytopathology test and the average time they took from the second test
Number of women belonging to

Figure 16: The proposed flow diagram for reporting the indicators of follow-up of women with ASC-H initial result.

CHAPTER 6 – RESULTS

6.1 The completeness of the SISCOLO variables

In this section, the results of the exploratory analysis for both the Cytology (1,109,731 records) and Histology tables (4413 records) are presented. In the Cytology table, the percentages of completeness of the analyzed variables are shown in Table 1. In the group of variables “identification of the attended woman by the program”, the variables with the lowest percentages of completeness are: CPF, *Cartão* SUS and identity number. In addition, some inconsistencies were observed in the free text variables (woman's name, mother's name, and address). As examples of these consistencies can be cited: the name of the medical professional or nurse, instead of the name of the woman or mother, and numerical sequences in the names fields and presence of special characters. Regarding the group of variables “women's clinical information”, the variables with the lowest percentages of completeness are the date of the last menstruation and the previous cytopathology test. In this group, some inconsistencies were observed in the field “date of last menstruation”, for instance: women older than 85 years old and the year of last menstruation is the same year on the collection of the specimens.

In the Histology table, the percentages of completeness of the analyzed variables are presented in Table 2. In the group of variables “identification of the attended woman by the program”, the variables with the lowest percentages of completeness are CPF, *Cartão* SUS and identity number. The degree of completeness of the field “results of the test” is lower than in the Cytology table. The analysis of the reported histopathology results (Table 3) shows that the records without reported results represent 47.25 % of the records. Among these, 6.52% had results reported as a free text in the field “other malignant neoplasia”. The analysis of these latter records shows that they belong to 15 healthcare units/hospitals, with the majority of them in the municipality of Rio de Janeiro (Table 4). It was observed that the majority of the reported results of these units/hospitals were cancers or atypical findings with a high risk for cancer. As an example, the hospital with the highest number of records had 53.6%, 26.1% and 20.0% of the record reported as cancer, adenocarcinoma, and other findings, respectively.

Table 1: Percentages of completeness of the variables of the Cytology table.

Description of the variables	**Percentage of completeness**
Identification of primary healthcare unit and laboratory	
Code of medical record	93.20
Other variables	100.00
Identification of the attended woman by the program	
Cartão SUS	11.22
CPF	2.99
Identity number	27.89
Degree of education	100.00
Postal address code (CEP, in Portuguese) of residence	63.97
Residence district	91.17
Woman's name	100.00
Mother's name	99.27
Women's clinical information	
Date of last menstruation/rule	64.53
Are you using an Intra Uterine Device?	93.78
Are you pregnant?	93.78
Do you use contraceptive pills?	93.78
Do you use hormone/ remedy to treat menopause?	93.78
Have you ever had radiotherapy treatment?	93.78
Do you have any bleeding after sex?	93.78
Do you have any bleeding after menopause?	93.78
Inspection of woman's cervix	93.78
Signs suggestive of sexually transmitted diseases?	93.78
Year of the last preventive test*	61.19
Have you ever had a Pap smear test?	86.95
Adequacy of the sample	
All the variables	99.35
Results of test	
All the variables	100.00

*Percentage calculated based on the total answers of YES for the variable “Have you ever had a Pap smear test?”

Table 2: Percentages of completeness of the variables of the Histology table.

Description of the variables	Percentage of completeness
Identification of primary healthcare unit and laboratory	
Code of medical record	54.96
Other variables	100.00
Identification of the attended woman by the program	
Cartão SUS	15.40
Individual Identification Number (CPF, in Portuguese)	2.73
Identity number	27.37
Degree of education	100.00
Postal address code (CEP, in Portuguese) of residence	42.20
Residence district	93.06
Woman's name	100.00
Mother's name	98.79
Women's clinical information	
Variables about the referred cytopathological results	99.30
Colposcopy of the cervix - Result of colposcopy	73.90
Colposcopy of the cervix - Performed procedure	94.90
Results of test	
All the variables	83.70*

*This percentage include the reported results in the field “other malignant neoplasia”

Table 3: Classification of the reported results in the Histology table.

Description of the test	Number of Records	Percentage of records
Records with histopathology results reported as codes	2328	52.75
Records without reported histopathology results	2085	47.25
Total	4413	100.00

Table 4: Distribution of the tests reported as free text in the field “other malignant neoplasia” among the primary healthcare units/hospitals.

Name of the primary healthcare unit	**Municipality**	**Number of tests reported as free text**
Instituto de Ginecologia da UFRJ	Rio de Janeiro	69.00
Hospital Mario Kroeff	Rio de Janeiro	30.00
Fiotec Instituto Fernandes Figueira	Rio de Janeiro	8.00
SMDC Hospital Piedade	Rio de Janeiro	7.00
Casa da Mulher	Resende	4.00
Hospital Federal de Ipanema	Rio de Janeiro	3.00
Hospital Geral de Bonsucesso	Rio de Janeiro	3.00
Centro de Saude de Rio das Ostras	Rio das Ostras	2.00
Centro de Saude Coletiva Professor Manoel Jose Ferreira	Petrópolis	2.00
Hospital Maternidade Fernando Magalhaes	Rio de Janeiro	2.00
Policlinica Walter Gomes Francklin	Tres Rio	2.00
Cemes Centro Mun Espec	Itaguai	1.00
Policlinica Manoel Guilherme da Silveira	Rio de Janeiro	1.00
SMS Rio Hospital da Piedade	Rio de Janeiro	1.00
SMS Rio Policlinica Manoel Guilherme da S. Filho	Rio de Janeiro	1.00
Total		136.00

6.2 Record linkage of SISCOLO

6.2.1 Internal linkage of the Cytology table

The probabilistic record linkage of the Cytology table resulted in the identification of 845,651 women. Table 5 shows the distribution of the frequency of tests in the Cytology table. It can be observed that 99% of the identified women performed 1 to 3 tests in the observation period.

Table 5: Distribution of the frequency of tests in the Cytology table.

Frequency of tests	Number of women	Percentage
1	638,158	75.46
2	160,552	18.98
3	38,901	4.60
4	6,692	0.79
5	1,133	0.13
6	180	0.02
7	28	0.00
8	6	0.00
9	1	0.00

6.2.2 Internal linkage of the Histology table

The probabilistic record linkage of the cytology table resulted in identification of 3,822 women. Table 6 shows the distribution of the frequency of tests in the Histology table. It can be observed that 98% of the identified women performed 1 or 2 tests in the observation period.

Table 6: Distribution of the frequency of tests in the Histology table.

Frequency of tests	Number of women	Percentage
1	3,315	86.73
2	438	11.45
3	56	1.46
4	11	0.28
5	2	0.05

6.2.3 External linkage of the Cytology and the Histology tables

The probabilistic record linkage of the cytology and histology tables resulted in the identification of **2,599** women who had records in both tables. This linkage table consist of two elements, where the first element is the woman identification in the Histology table and the second element is her identification in the Cytology table.

6.3 The dimensional model

This section will present the DW dimensional model. As mentioned before, the DW was developed in Portuguese, so all the illustrated figures and terms will be presented in Portuguese.

The DW dimensional model is shown in Figure 17, with the fact table (indicadores_desempenho) connected to the dimensional tables "*tempo, unidade_coleta, faixa_etaria, laboratorio and municipio_residencia"*, which correspond to the dimensions "time, primary health care unit, age group, laboratory and residence municipality) respectively. Those dimensions are connected to the fact table by the following foreign keys: *tempo_id, faixa_etaria_id, municipio_residencia_id, u_coleta_id* and *laboratorio_id"* respectively. In addition to the foreign keys, the fact table contains the measured facts. As an example, the first four facts are *laminas_insatisfatorias, cito_cancer, diaplasia_moderada and n1_mulheres_hsil.* These facts represent the "number of unsatisfactory specimens" "number of cytopathological test results with cancer", "number of histopathological test results with moderate dysplasia" and "number of women with cytopathological test results of HSIL in the first test" respectively, for each combination of the primary key values of the dimensional tables.

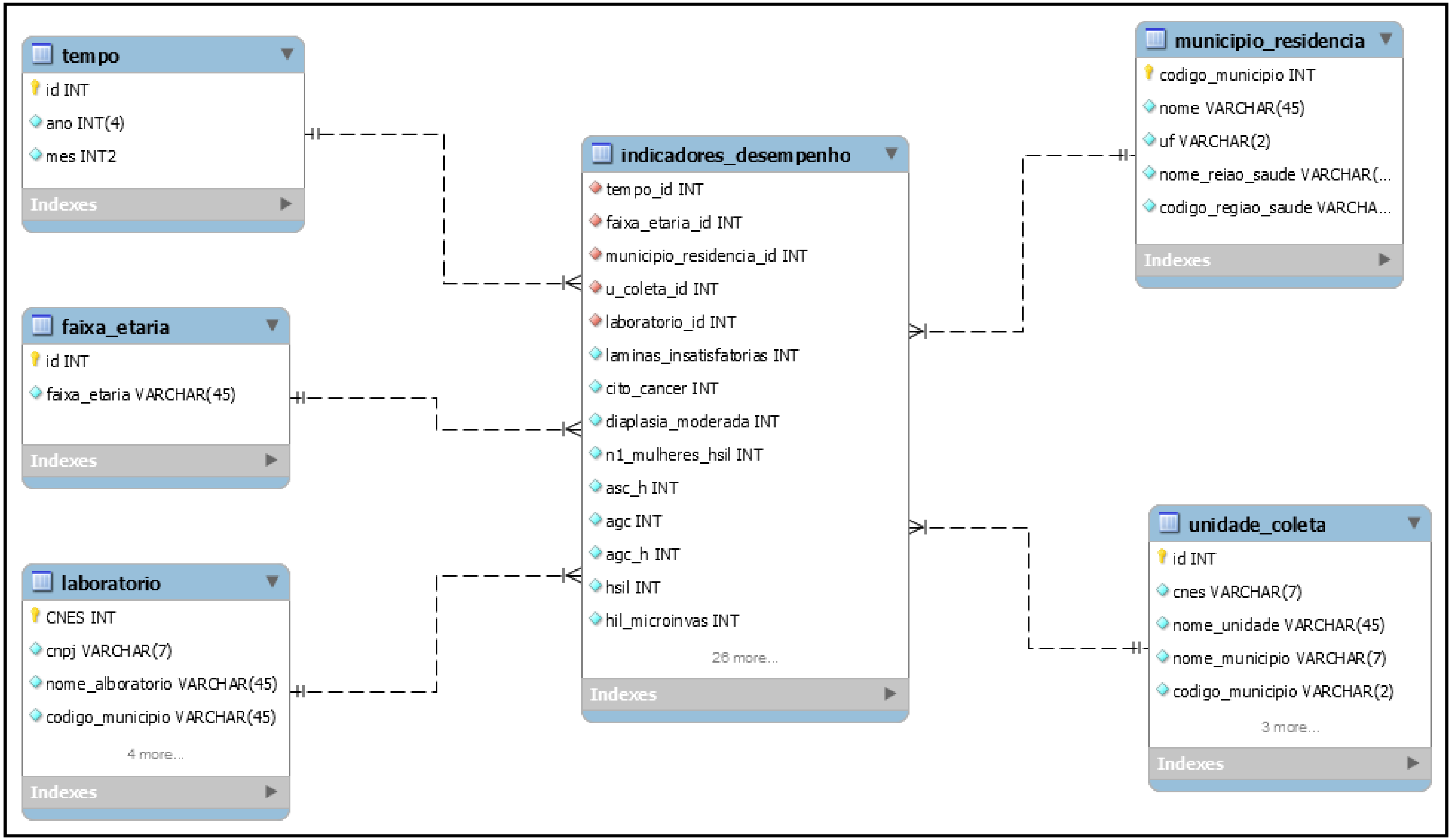


Figure 17: Dimensional model of the data warehouse.

6.4 The data visualization

As mentioned before, two GUIs were used for the indicators' visualization: the Pentaho User Console; and Saiku Analytics as a plugin to the User Console. In order to demonstrate the main features of these GUIs and how they could help managers in the decision-making process, four different examples are presented in this section. The first example presents the Pentaho User Console interface and the process of access to Saiku Analytics, as well as an example of a visualization of the indicator "percentage of the histopathological test results of CIN III". The second example shows the monitoring process of the indicator "percentages of the specimens with TZ". The third example presents the monitoring of the indicators related to the internal quality of the cytopathological examination. The last example presents the process of access to the pre-define report through the User Console as well as six examples of the visualization of the woman-based indicators.

6.4.1 The graphical interfaces (GUI)

The Pentaho User Console and the process of access to Saiku Analytics

The main interface of Pentaho User Console is illustrated in Figure 18 where the user could customize its appearance by adding the login of his company/department.

The functions of the main icons are presented below:

Procurar arquivo (search file, in English): used to access the predefined reports.

Novo (new, in English): used to select the available GUIs offered by the User Console.

Gerenciar Fonte de Dados (manage data source, in English): used to add/removed data sources.

Documentação (documentation, in English): used to access the documentation about the current version of Pentaho.

In order to access the interface of Saiku Analytics, the user should select the icon *Novo*, the list of the available GUIs will appear as illustrated in part (i) of Figure 19, and then Saiku Analytics should be selected. After the selection of Saiku Analytics, another screen (**ii**) will appear where the user should click on the icon "create a new query" for the access to the Saiku Analytics main GUI (**iii**). In this interface, the created data cubes should be selected from *cubo* (cube, in English) (indicated by **A** in Figure 19iii), then the indicators of the selected cube will be appear in the *medidas* (measures, in English) (**B** in

Figure 19iii). The dimensions are selected from *dimensões* (dimensions, in English) (**C** in the Figure 19iii). For each dimension, the hierarchical levels are shown below in decreasing order of detail:

- “tempo”: month (“mes”) and year (“ano”);
- “laboratorio”: laboratory name (“nome_laboratorio”), municipality (“nome_municipio”), health region (“regiao_saude_codigo”) and state (“uf”);
- “unidade_coleta”: unit name (“nome_unidade”), municipality (“nome_municipio”), region (“regiao_saude_codigo”) and state (“uf”);
- “municipio_residencia”: municipality (“nome_municipio”), region (“regiao_saude_codigo”) and state (“uf”).

An indicator’s value may be built and visualized independently for any level of each hierarchy, so that the user can go up (drill up) or down (drill down) any hierarchy.

After the selection of both indicator(s) and dimensions the result of the query is visualized in region **D** in Figure 19iii**.** Saiku Analytics offers a GUI where the user could visualized the requested queries in a cross-table (**E** in Figure 19iii) or use various types of graphs (**F, G** and **H** in Figure 19iii) as well as perform basic statistical analysis (**I** in Figure 19iii). In addition, the user can save the results of the queries in different formats such as PDF and XLS (**J** in Figure 19iii). In addition, the GUI offers other functionalities such as zoom, save/edit the query and drill across the cells (**K** in Figure 19iii).

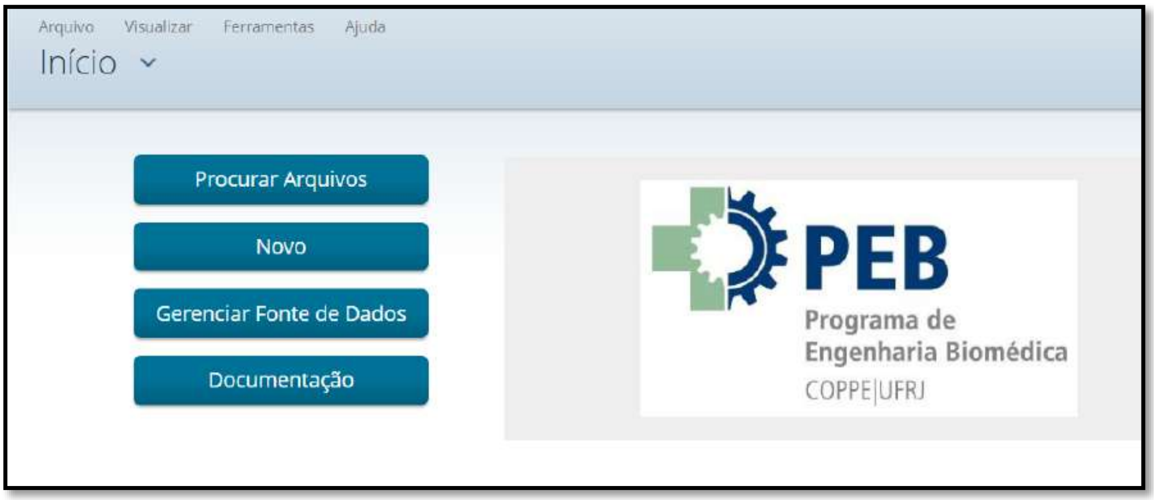
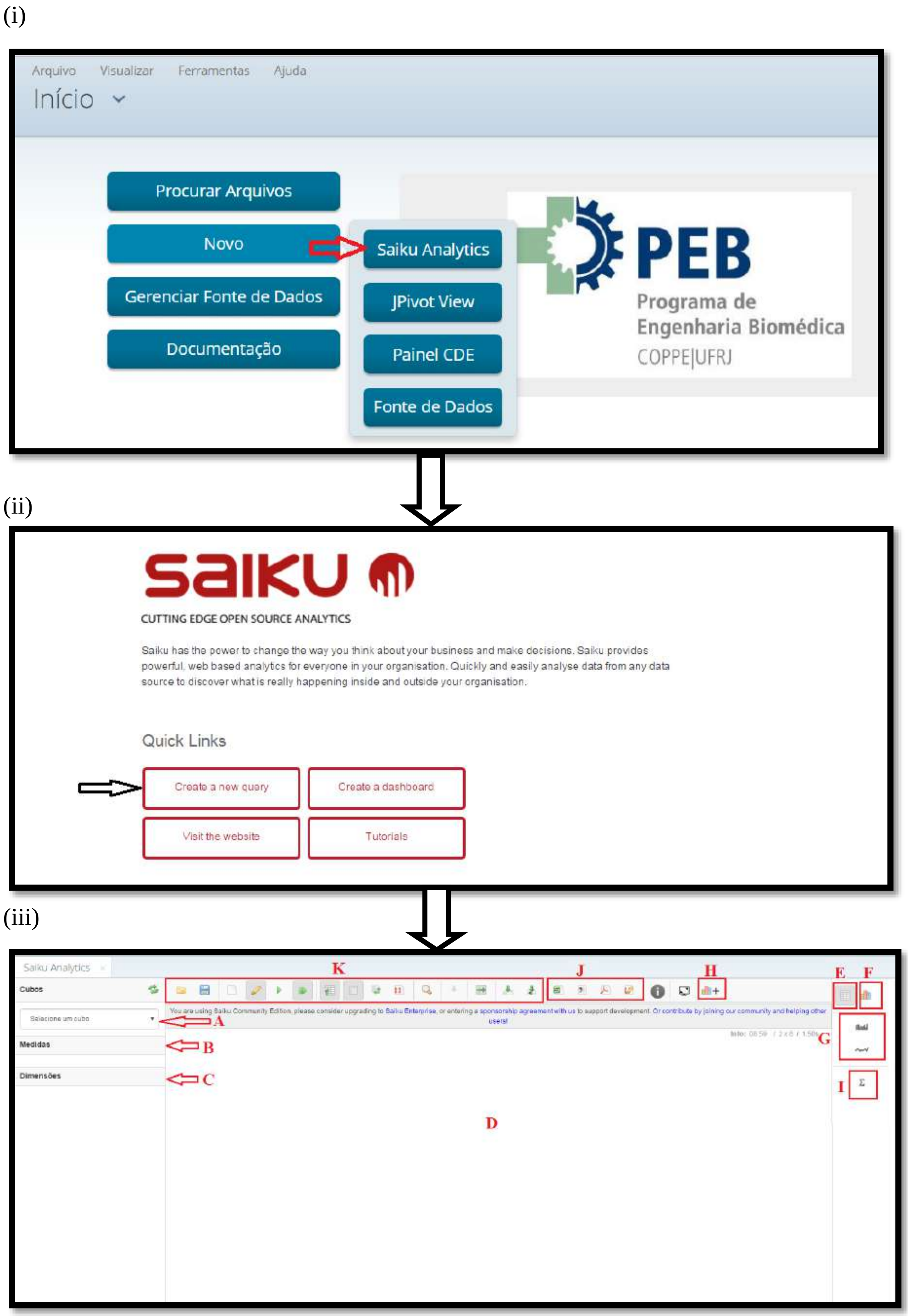


Figure 18: Snap shot of the graphical interface of Pentaho User Console.



Saiku has the power to change the way you think about your business and make decisions. Saiku provides powerful, web based analytics for everyone in your organisation. Quickly and easily analyse data from any data source to discover what is really happening inside and outside your organisation.

You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

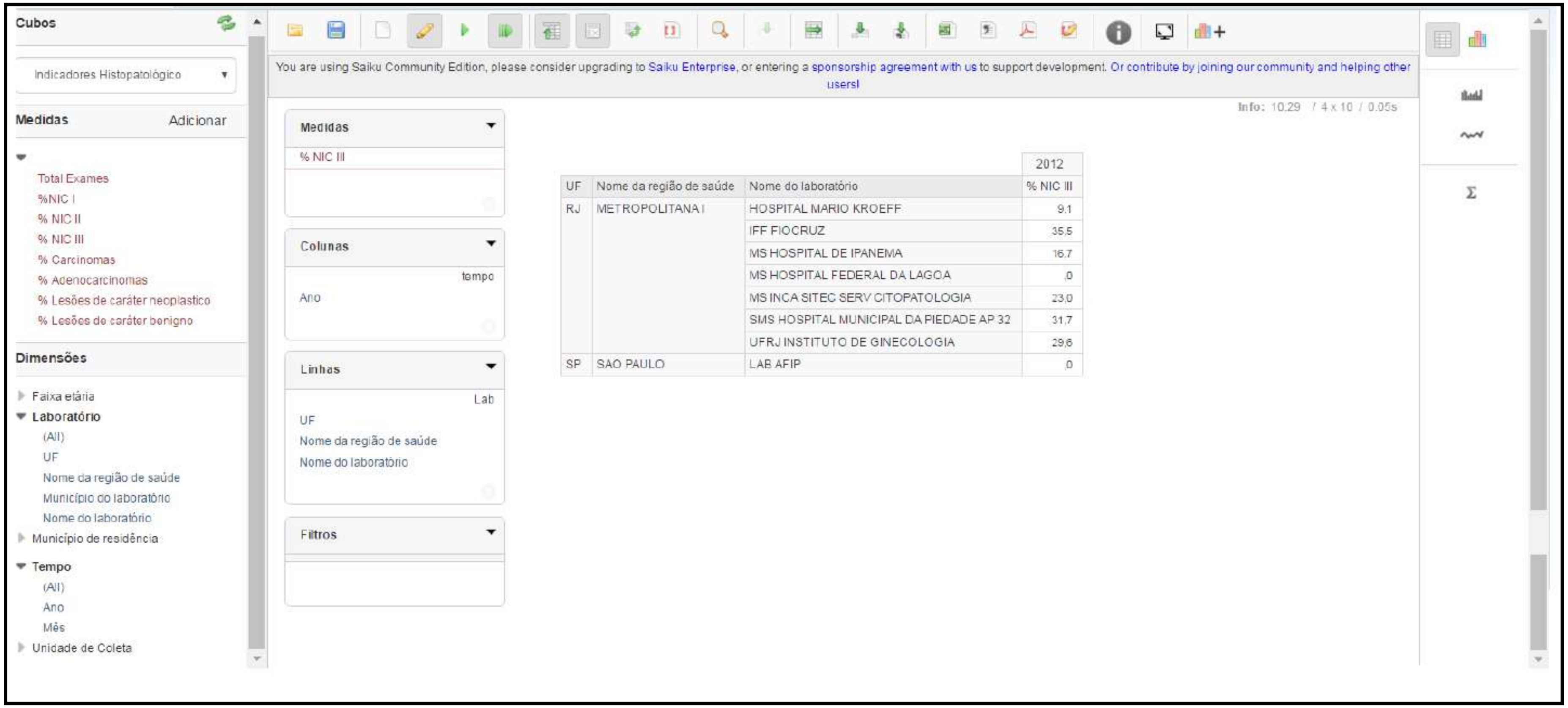
Figure 19: The process of access to the GUI of Saiku Analytics.

Visualization of the indicator "percentage of the histopathological test results of CIN II"

In this example, the visualization of the indicator "percentage of the histopathological test results of CIN III" is presented for the municipality of Rio de Janeiro in the year 2012.

In order to obtain the snapshot of figure 20, the following steps were applied:

- select the cube "*indicadores histopatológicos*" from the *cubo* region;
- select the indicator "% NIC III*"* from the *Medidas* region;
- click on the dimension *laboratório* and then select the attributes *nome do laboratório, nome da região de saúde and UF* and drag them to the *linhas* region;
- select the dimension *tempo* and drag the attribute *Ano* into the c*olunas* region and then click on the dragged attribute, the snapshot in Figure 21 will appear, and then select the year 2012.



Cubos

Indicadores Histopatológico

Medidas Adicionar

Total Exames
%NIC I
% NIC II
% NIC III
% Carcinomas
% Adenocarcinomas
% Lesões de caráter neoplastico
% Lesões de caráter benigno

Dimensões

Faixa etária
Laboratório
(All)
UF
Nome da região de saúde
Município do laboratório
Nome do laboratório
Município de residência
Tempo
(All)
Ano
Mês
Unidade de Coleta

You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Info: 10.29 / 4 x 10 / 0.05s

Medidas
% NIC III

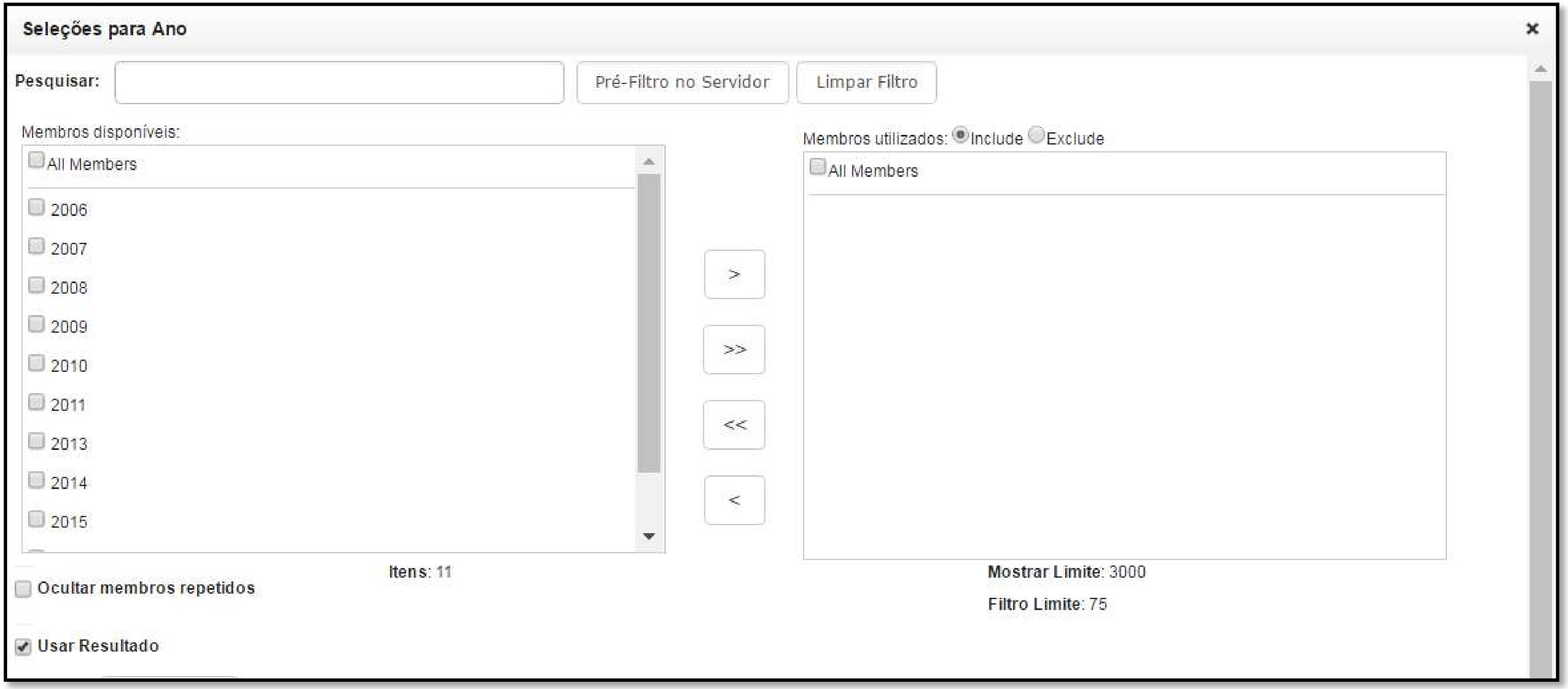
Colunas
tempo
Ano

Linhas
Lab
UF
Nome da região de saúde
Nome do laboratório

Filtros

			2012
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 20: Output screen for the query "percentages of histopathologic test results with NIC III" in the year 2012 per laboratory.



Seleções para Ano

Pesquisar:

Membros disponíveis:

- □ All Members
- □ 2006
- □ 2007
- □ 2008
- □ 2009
- □ 2010
- □ 2011
- □ 2013
- □ 2014
- □ 2015

Itens: 11

Membros utilizados: ● Include ○ Exclude

- □ All Members

Mostrar Limite: 3000

Filtro Limite: 75

□ Ocultar membros repetidos

☑ Usar Resultado

Figure 21: Screen to select the year of the analysis of the specimens with result NIC III in the laboratory.

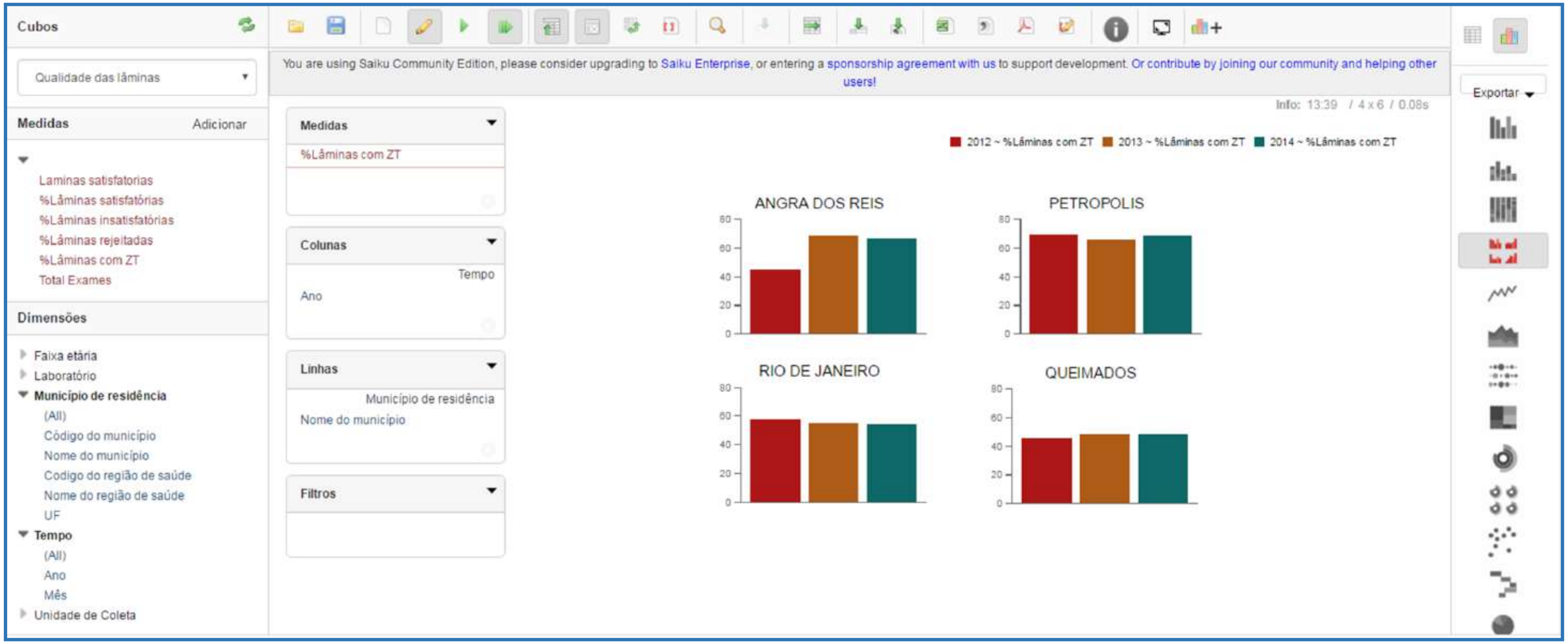
6.4.2 Monitoring of the indicators related to the adequacy of specimens/slides

In this example, the visualization of the indicator “percentage of the specimens with representation of the materials of the transformation zone” will be presented. The program managers should monitor this indicator since the absence of the cells of the transformation zone is one of the factors that are associated with the increased presence of false negative results (GAUZA *et al.*, 2010; ROBERTS *et al.*, 2016; SHIRATA *et al.*, 1998). Figure 22 shows a Saiku Analytics screen where the indicator is presented in a bar chart using the *Ano* attribute of the time dimension and the *Nome* attribute of the primary healthcare unit dimension. The indicator was visualized for four municipalities: Angra dos Reis, Petropolis, Rio de Janeiro and Queimados. The value of the indicator varied among these municipalities. It increased in Angra dos Reis and remained the same (around 60%) in both Rio de Janeiro and Petropolis. For the municipality of Queimados, the indicator remained the same but with low values (around 40%).

In order to identify the reason for the low values of the indicator in Queimados, the manager needs to make a detailed investigation of the primary healthcare units of the municipality. As an example, Figure 23 shows the indicator “percentages of specimens with TZ” for the 15 primary healthcare units of Queimados in 2013. It is visualized using the attribute “year of the time dimension” and the attribute “name of the primary healthcare unit dimension”. The indicator value in five of the primary healthcare units (marked by an asterisk “*”) is less than the recommended value (20%) by both the Brazilian Cervical Cancer Screening Program and the Pan American Health Organization (OPS, 1990).

The manager may want to know the contribution of the five units with slide problems to the municipality production. Figure 24 shows the pie chart of the “total number of the performed exams” indicator for all the 15 primary healthcare units. The names of the primary healthcare units with values less than 20% are surrounded by red triangles. The five primary healthcare units contribute approximately 24.4 % to the whole production, which could affect the value of the indicator in the municipality.

Depending on the information from the above analysis, the manager should make the appropriate decision to enhance the collection skills of the medical personnel of these primary healthcare units.



You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Info: 13:39 / 4 x 6 / 0.08s

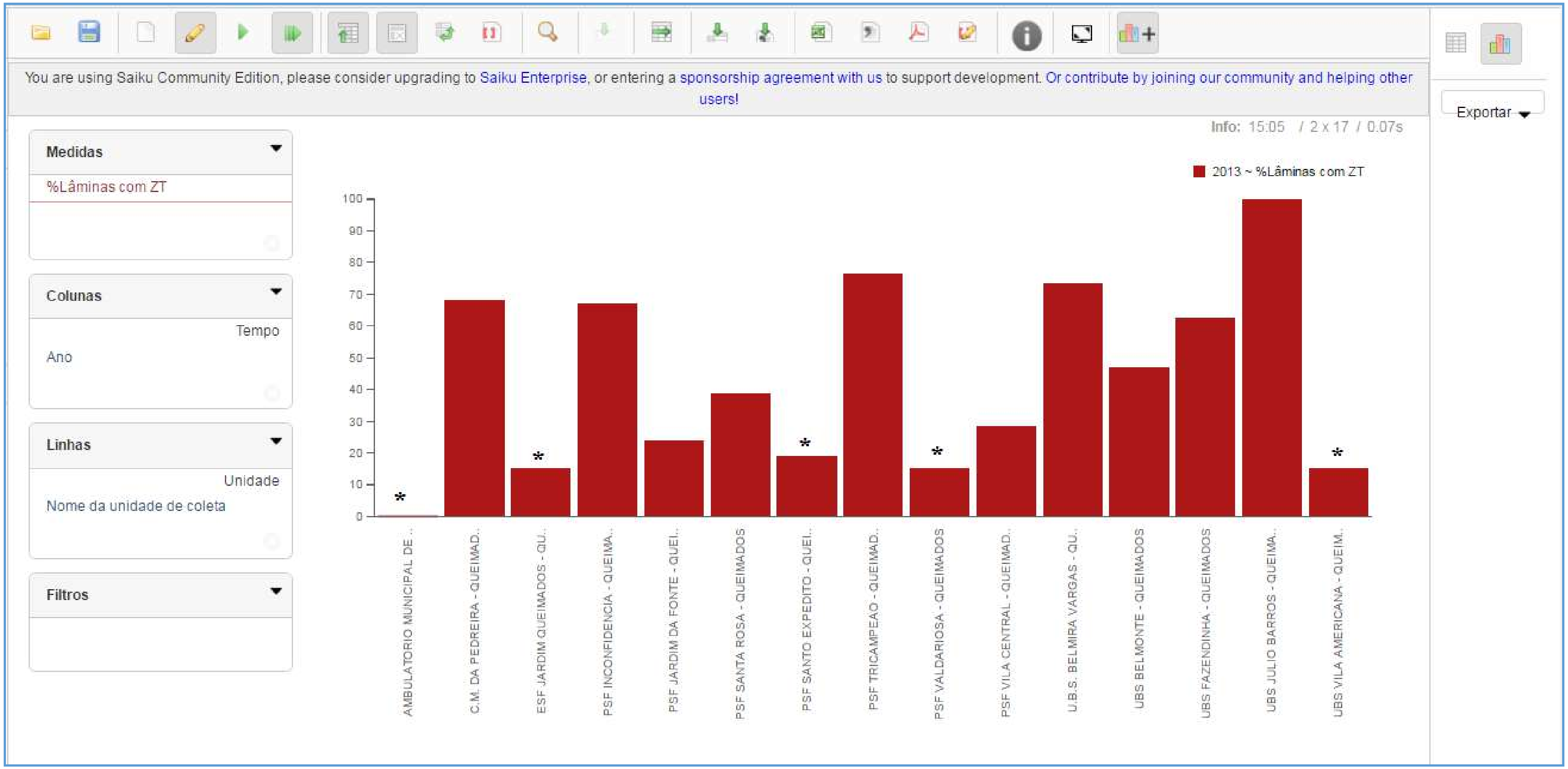
Medidas
%Lâminas com ZT

Colunas
Tempo
Ano

Linhas
Município de residência
Nome do município

Filtros

Figure 22: Output screen for the bar chart of “percentage of the specimens with representation of the materials of the transformation zone” in four municipalities in three different years.



You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Medidas

%Lâminas com ZT

Colunas

Tempo

Ano

Linhas

Unidade

Nome da unidade de coleta

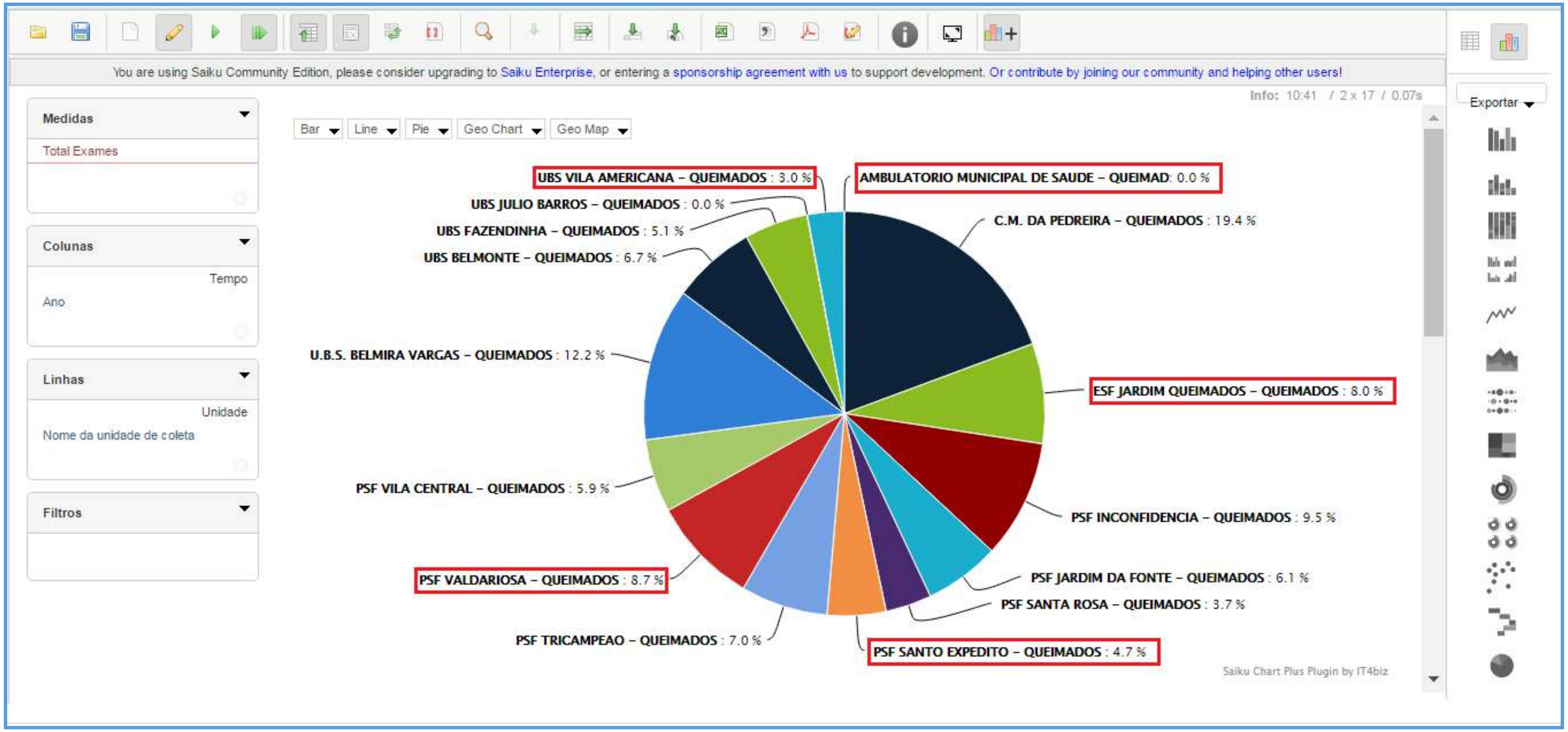
Filtros

Info: 15:05 / 2 x 17 / 0.07s

Exportar

🟥 2013 ~ %Lâminas com ZT
100
90
80
70
60
50
40
30
20
10
0
*
*
*
*
*
AMBULATORIO MUNICIPAL DE ..
C.M. DA PEDREIRA - QUEIMAD..
ESF JARDIM QUEIMADOS - QU..
PSF INCONFIDENCIA - QUEIMA..
PSF JARDIM DA FONTE - QUEI..
PSF SANTA ROSA - QUEIMADOS
PSF SANTO EXPEDITO - QUEI..
PSF TRICAMPEAO - QUEIMAD..
PSF VALDARIOSA - QUEIMADOS
PSF VILA CENTRAL - QUEIMAD..
U.B.S. BELMIRA VARGAS - QU..
UBS BELMONTE - QUEIMADOS
UBS FAZENDINHA - QUEIMADOS
UBS JULIO BARROS - QUEIMA..
UBS VILA AMERICANA - QUEIM..

Figure 23: Output screen for the bar chart of “percentage of the specimens with representation of the materials of the transformation zone” in 15 primary healthcare units of Queimados in 2013.



You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Info: 10:41 / 2 x 17 / 0.07s

Medidas

Total Exames

Colunas

Tempo

Ano

Linhas

Unidade

Nome da unidade de coleta

Filtros

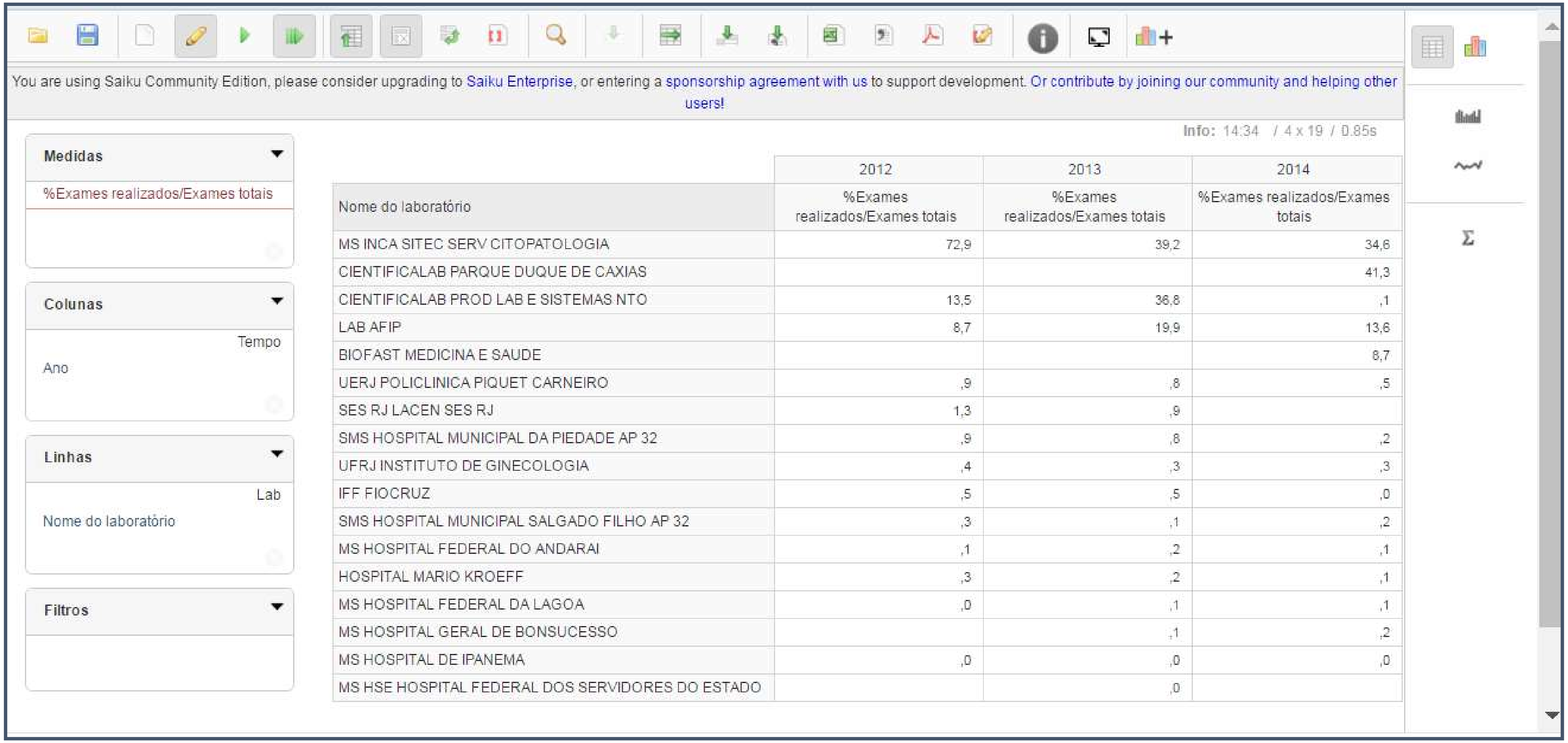
Bar
Line
Pie
Geo Chart
Geo Map
Exportar
UBS VILA AMERICANA – QUEIMADOS : 3.0 %
AMBULATORIO MUNICIPAL DE SAUDE – QUEIMAD: 0.0 %
UBS JULIO BARROS – QUEIMADOS : 0.0 %
UBS FAZENDINHA – QUEIMADOS : 5.1 %
C.M. DA PEDREIRA – QUEIMADOS : 19.4 %
UBS BELMONTE – QUEIMADOS : 6.7 %
U.B.S. BELMIRA VARGAS – QUEIMADOS : 12.2 %
ESF JARDIM QUEIMADOS – QUEIMADOS : 8.0 %
PSF VILA CENTRAL – QUEIMADOS : 5.9 %
PSF INCONFIDENCIA – QUEIMADOS : 9.5 %
PSF VALDARIOSA – QUEIMADOS : 8.7 %
PSF JARDIM DA FONTE – QUEIMADOS : 6.1 %
PSF SANTA ROSA – QUEIMADOS : 3.7 %
PSF TRICAMPEAO – QUEIMADOS : 7.0 %
PSF SANTO EXPEDITO – QUEIMADOS : 4.7 %
Saiku Chart Plus Plugin by IT4biz

Figure 24: Output screen for the bar chart of “quantities of the performed tests” in 15 primary healthcare units of Queimados in 2013.

6.4.3 Monitoring of the indicators related to the internal quality of the cytopathological tests

Figure 25 shows the percentages of the performed tests per laboratory in relation to all performed tests in the municipality of Rio de Janeiro. The blank space in the indicator values appears when the laboratory did not analyze any specimens in the specific year. One can observe that approximately 95% of the specimens were analyzed in three laboratories in the year 2012, which are “MS INCA SITEC SERV CITOPATOLOGIA”, “CIENTIFICALAB PROD LAB E SISTEMAS NTO” and “LAB AFIP”.

In order to monitor the internal quality of the cytopathological examination of those laboratories, five indicators were displayed in Figure 26. They are: percentage of the altered test results among the satisfactory tests (positive index); percentage of atypical squamous cells (ASC) among the satisfactory tests; percentage of ASC cells among the altered tests; ratio of atypical squamous cells (ASC-US and ASC-H) to squamous intraepithelial lesions (LSIL and HSIL) and percentages of tests with HSIL results. One can see in Figure 26 that the laboratory “LAB AFIP” presented two indicators (marked by red squares) with values outside the standard values recommended by the Brazilian Ministry of Health (BRAZIL, 2016c), while the laboratory “CIENTIFICALAB PROD LAB E SISTEMAS NTO” presented four indicators with values outside the recommended standards.

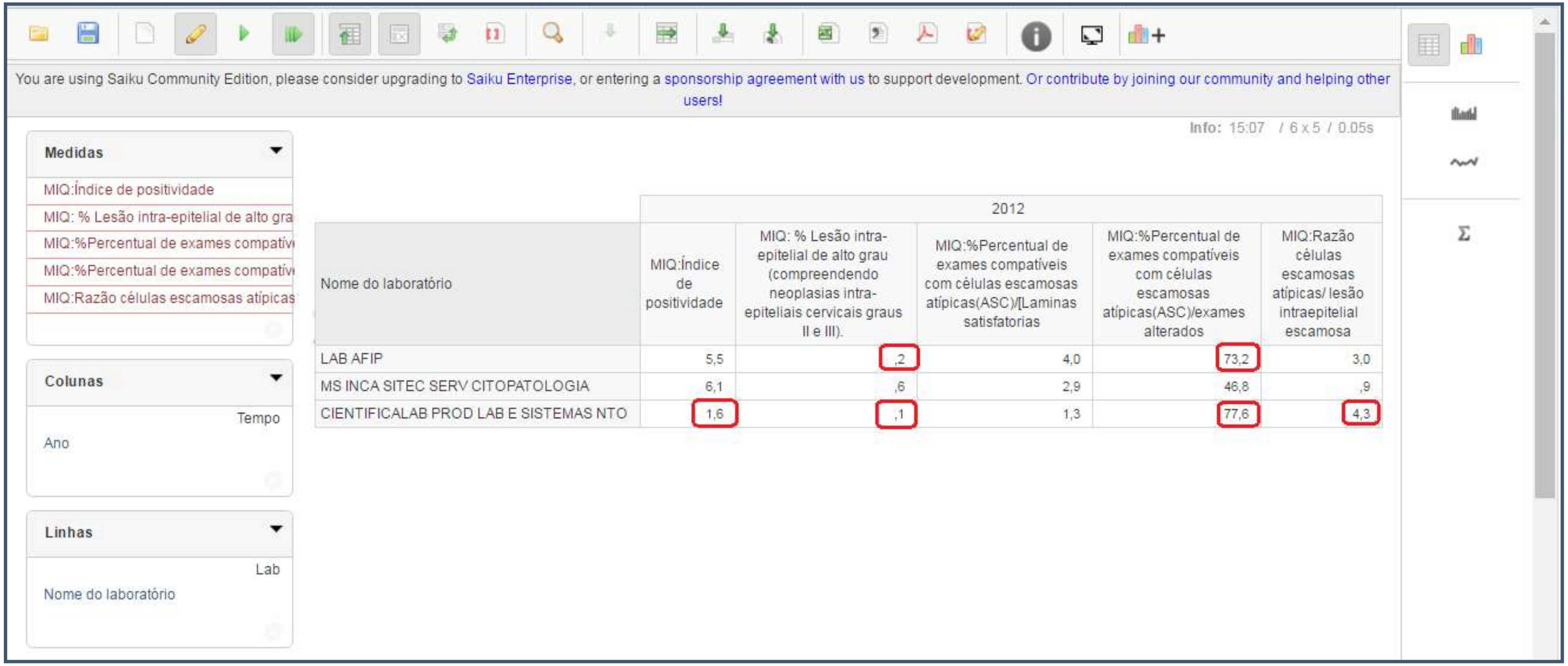


You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Info: 14:34 / 4 x 19 / 0.85s

	2012	2013	2014
Nome do laboratório	%Exames realizados/Exames totais	%Exames realizados/Exames totais	%Exames realizados/Exames totais
MS INCA SITEC SERV CITOPATOLOGIA	72,9	39,2	34,6
CIENTIFICALAB PARQUE DUQUE DE CAXIAS			41,3
CIENTIFICALAB PROD LAB E SISTEMAS NTO	13,5	36,8	,1
LAB AFIP	8,7	19,9	13,6
BIOFAST MEDICINA E SAUDE			8,7
UERJ POLICLINICA PIQUET CARNEIRO	,9	,8	,5
SES RJ LACEN SES RJ	1,3	,9	
SMS HOSPITAL MUNICIPAL DA PIEDADE AP 32	,9	,8	,2
UFRJ INSTITUTO DE GINECOLOGIA	,4	,3	,3
IFF FIOCRUZ	,5	,5	,0
SMS HOSPITAL MUNICIPAL SALGADO FILHO AP 32	,3	,1	,2
MS HOSPITAL FEDERAL DO ANDARAI	,1	,2	,1
HOSPITAL MARIO KROEFF	,3	,2	,1
MS HOSPITAL FEDERAL DA LAGOA	,0	,1	,1
MS HOSPITAL GERAL DE BONSUCESSO		,1	,2
MS HOSPITAL DE IPANEMA	,0	,0	,0
MS HSE HOSPITAL FEDERAL DOS SERVIDORES DO ESTADO		,0	

Figure 25: Output screen for the query "percentages of the performed tests by the laboratories in relation to all performed tests in the municipality of Rio de Janeiro" per name of the laboratory and year.



You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Info: 15:07 / 6 x 5 / 0.05s

Medidas

- MIQ:Índice de positividade
- MIQ: % Lesão intra-epitelial de alto gra
- MIQ:%Percentual de exames compativ
- MIQ:%Percentual de exames compativ
- MIQ:Razão células escamosas atípicas

Colunas

Tempo

Ano

Linhas

Lab

Nome do laboratório

	2012				
Nome do laboratório	MIQ:Índice de positividade	MIQ: % Lesão intra-epitelial de alto grau (compreendendo neoplasias intra-epiteliais cervicais graus II e III).	MIQ:%Percentual de exames compatíveis com células escamosas atípicas(ASC)/[Laminas satisfatorias	MIQ:%Percentual de exames compatíveis com células escamosas atípicas(ASC)/exames alterados	MIQ:Razão células escamosas atípicas/ lesão intraepitelial escamosa
LAB AFIP	5,5	,2	4,0	73,2	3,0
MS INCA SITEC SERV CITOPATOLOGIA	6,1	,6	2,9	46,8	,9
CIENTIFICALAB PROD LAB E SISTEMAS NTO	1,6	,1	1,3	77,6	4,3

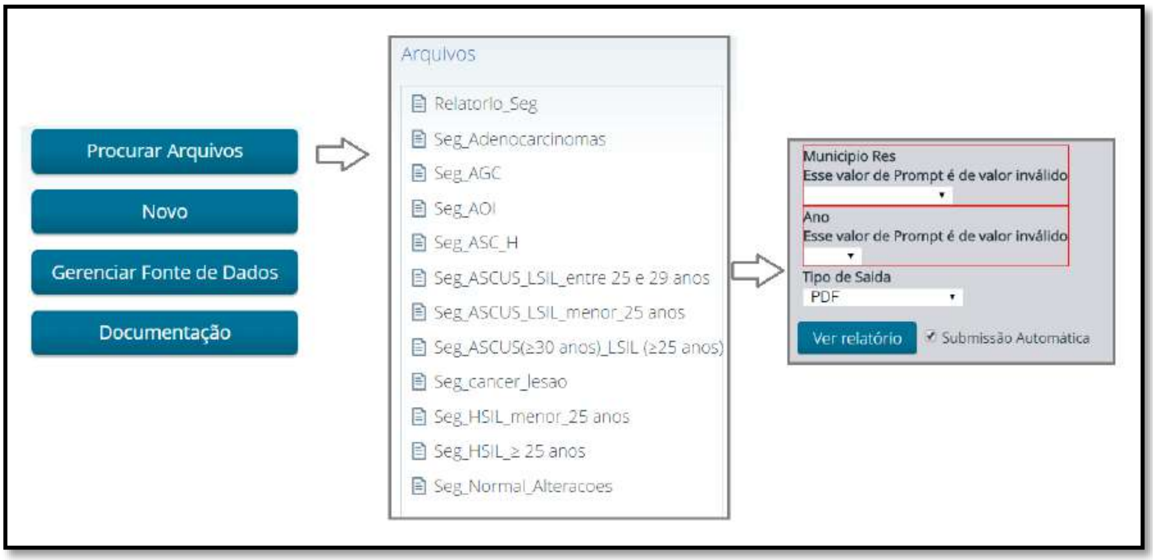
Figure 26: Output screen for the visualization of the indicators of the internal quality monitoring of the cytopathological examination per name of the laboratory and year.

6.4.4 Visualization of the woman-based indicators

In this section, the process of accessing the pre-defined report as well as six examples of the visualization of the woman-based indicators will be presented.

Process of accessing the pre-defined reports

Figure 27 shows the process of accessing and visualizing the pre-defined reports. The user should click on the icon *procurar arquivos*, and then a screen with the pre-defined reports will appear. In this screen 12 reports are shown. The first one is a general report about the follow-up process, and the other 11 reports are for the follow-up of each one of the created groups of women. The user should select the report of interest, then a screen will appear for the selection of the dimensions. The indicators in the pre-defined reports are visualized using the attribute "*nome do municipio*" of the dimension residence municipality and the attribute "*Ano*" of the dimension time. The user should select the desired municipality and year, and click on the button *ver relatório* (view report, in English) or mark the option *submissão automatic* (automatic submission, in English) and then select *Tipo de saida* (output type, in English), which could be PDF, XLS, CSV, RTF, TXT or HTML.



Municipio Res
Esse valor de Prompt é de valor inválido

Ano
Esse valor de Prompt é de valor inválido

Tipo de Saida
PDF

☑ Submissão Automática

Figure 27: The followed process in order to access and visualize the pre-defined report.

Examples of the visualization of the woman-based indicators

Figure 28 shows a snapshot of a pre-defined report that provides a general view of the participation of the target population in the screening program, and the percentages of the screened women classified by their initial cytopathological test results. It can be observed that women with test results without atypical findings represent 94.83% of all screened women, where the complement 5.17% presented the women with altered test results.

In order to provide information of the follow-up of each of the 11 group of women, according to the initial test results and age, a pre-defined report for each group was prepared. Figures 29-32 show four examples of the pre-defined reports for the follow-up of women with initial results of ASC-H; HSIL (women ≥ 25 years old); ASC-US (women ≥30 years old) and LSIL (women ≥25 years old); and without atypical findings.

The follow-up of the women with the initial cytopathological result of ASC-H is presented in Figure 29. It can be observed that 68.6% of the women were lost from follow-up. Among those women, 62.0 % were lost between the first and second test and 38% were lost between the second and third test. Considering that the initial clinical approach for this group is to perform an immediate colposcopy, it was observed that the group who performed the histopathology test took a relatively long time (4.8 months). Regarding the average time between cytopathology tests, it can be observed that, on average, the time until the next test is 10.2 months, which is 4 months more than the recommended time.

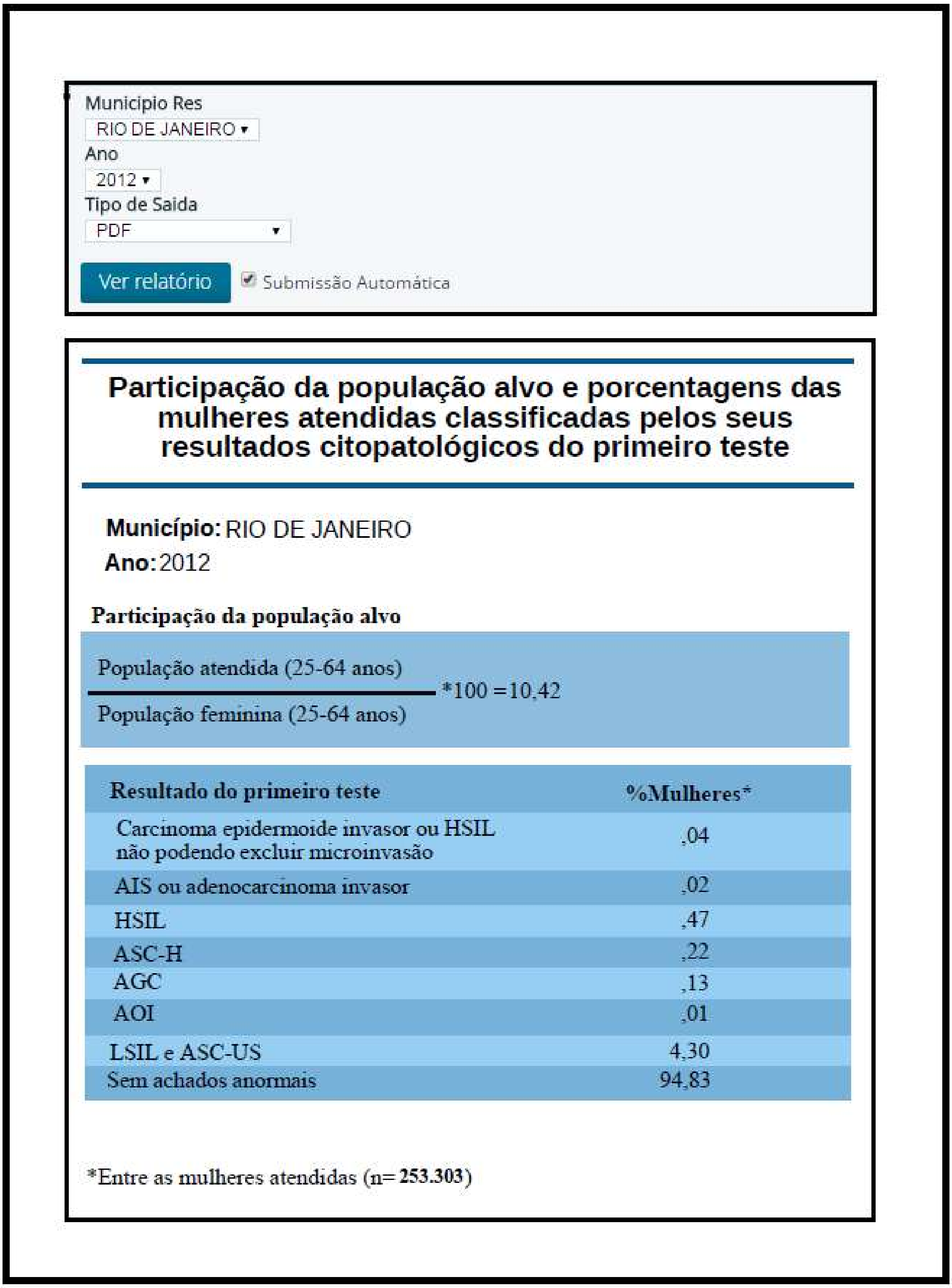
The follow-up of the women with the initial cytopathological result of HSIL is presented in Figure 30. It can be observed that 56.5% of the women were lost from follow-up. Among these women, 61.4% were lost between the first and second test and 38.4% were lost between the second and third test. The diagnosis of 65.2% of the women who performed a histopathology test, confirmed the initial results of HSIL. It can be observed that the average time between the first and the second cytopathology test was 12.4 months, and 5 months until the next histopathology test.

The follow-up of the women (≥ 30 years old) with the initial result of ASC-US and women (≥ 25 years old) with the initial result of LSIL is presented in Figure 31. It can be observed that 77.4% of the women were lost from follow-up. Among these women, 55.6% were lost between the first and second test, 42.5% were lost between the second and third test, and 1.9% were lost between the third and fourth test. Regarding the average time until the next test, it can be observed that only one group of women adhered to the recommendation of the screening program. This group contains the women with positive test results in the third cytopathology test and performed a fourth cytopathology test within (6.2 months), which is approximately equal to the recommended time by the screening program (6 months).

The follow-up of women with cytopathological results without atypical findings is presented in Figure 32. It can be observed that 61.8% of the women were not identified in the SISCOLO data in the period of observation, which. The majority of these women (98.2%) were not identified between the first and second test and 1.8 % were not identified between the second and third test. For the women who performed second test, the average time was 12 months.

The output screens of the pre-defined reports for the follow-up of the other five groups of women by their initial test are presented in the Appendix 4.

Figure 33 shows the output screen of the annual participation of the target population (25-64 years old) in the screening program in the municipality of Rio de Janeiro in the years 2012, 2013 and 2014. The figure shows that the values of annual participation are approximately the same for the three years.



Municipio Res

RIO DE JANEIRO ▾

Ano

2012 ▾

Tipo de Saida

PDF ▾

Participação da população alvo e porcentagens das mulheres atendidas classificadas pelos seus resultados citopatológicos do primeiro teste

Município: RIO DE JANEIRO

Ano: 2012

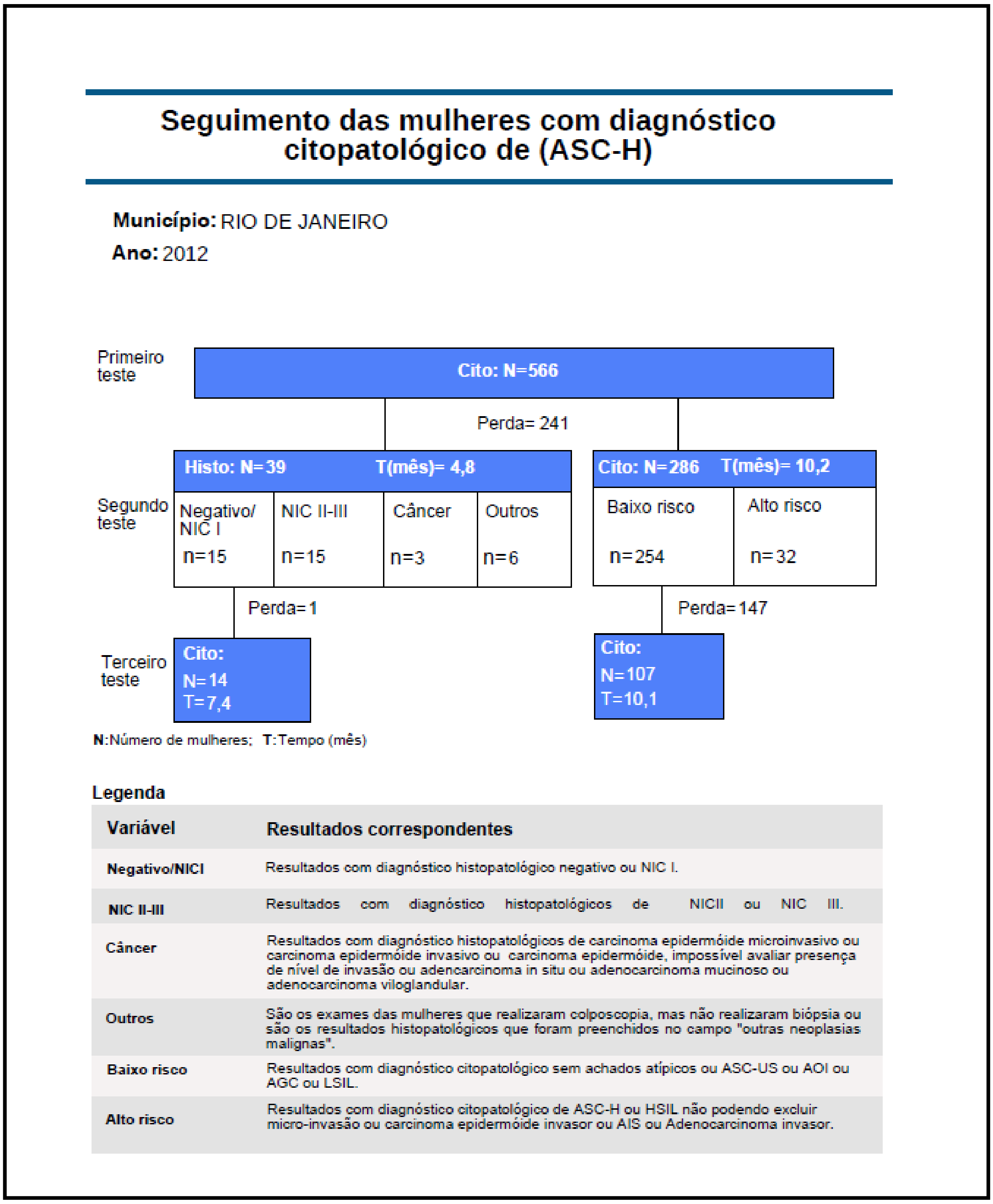
Participação da população alvo

$$\frac{\text{População atendida (25-64 anos)}}{\text{População feminina (25-64 anos)}} *100 = 10,42$$

Resultado do primeiro teste	%Mulheres*
Carcinoma epidermoide invasor ou HSIL não podendo excluir microinvasão	,04
AIS ou adenocarcinoma invasor	,02
HSIL	,47
ASC-H	,22
AGC	,13
AOI	,01
LSIL e ASC-US	4,30
Sem achados anormais	94,83

*Entre as mulheres atendidas (n= **253.303**)

Figure 28: Output screen for a pre-defined report about the participation of the target population in the screening program and the percentages of the screened women classified by their initial cytopathological test results.



Seguimento das mulheres com diagnóstico citopatológico de (ASC-H)

Município: RIO DE JANEIRO

Ano: 2012

Primeiro teste

Cito: N=566

Perda= 241

Segundo teste

Histo: N=39 T(mês)= 4,8

Negativo/ NIC I n=15

NIC II-III n=15

Câncer n=3

Outros n=6

Cito: N=286 T(mês)= 10,2

Baixo risco n=254

Alto risco n=32

Perda=1

Perda=147

Terceiro teste

Cito: N=14 T=7,4

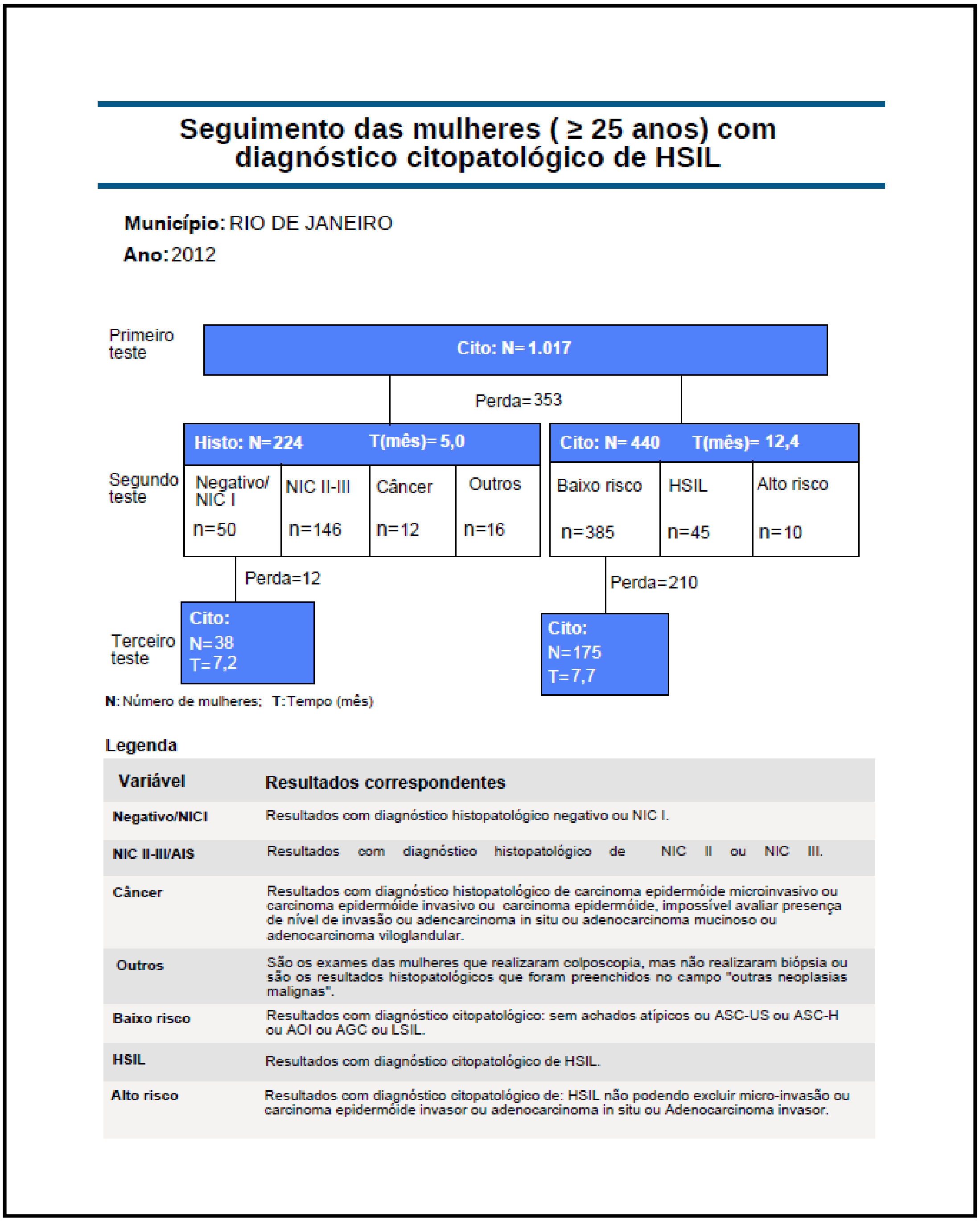
Cito: N=107 T=10,1

N:Número de mulheres; **T**:Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
NIC II-III	Resultados com diagnóstico histopatológicos de NICII ou NIC III.
Câncer	Resultados com diagnóstico histopatológicos de carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adencarcinoma in situ ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas".
Baixo risco	Resultados com diagnóstico citopatológico sem achados atípicos ou ASC-US ou AOI ou AGC ou LSIL.
Alto risco	Resultados com diagnóstico citopatológico de ASC-H ou HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou AIS ou Adenocarcinoma invasor.

Figure 29: Output screen for the follow-up indicators for women with initial cytopathological test result of ASC-H in the municipality of Rio de Janeiro in the year 2012.



Seguimento das mulheres (≥ 25 anos) com diagnóstico citopatológico de HSIL

Município: RIO DE JANEIRO

Ano: 2012

Primeiro teste

Cito: N= 1.017

Perda=353

Segundo teste

Histo: N=224 T(mês)= 5,0

Negativo/ NIC I	NIC II-III	Câncer	Outros
n=50	n=146	n=12	n=16

Cito: N= 440 T(mês)= 12,4

Baixo risco	HSIL	Alto risco
n=385	n=45	n=10

Perda=12

Perda=210

Terceiro teste

Cito: N=38 T=7,2

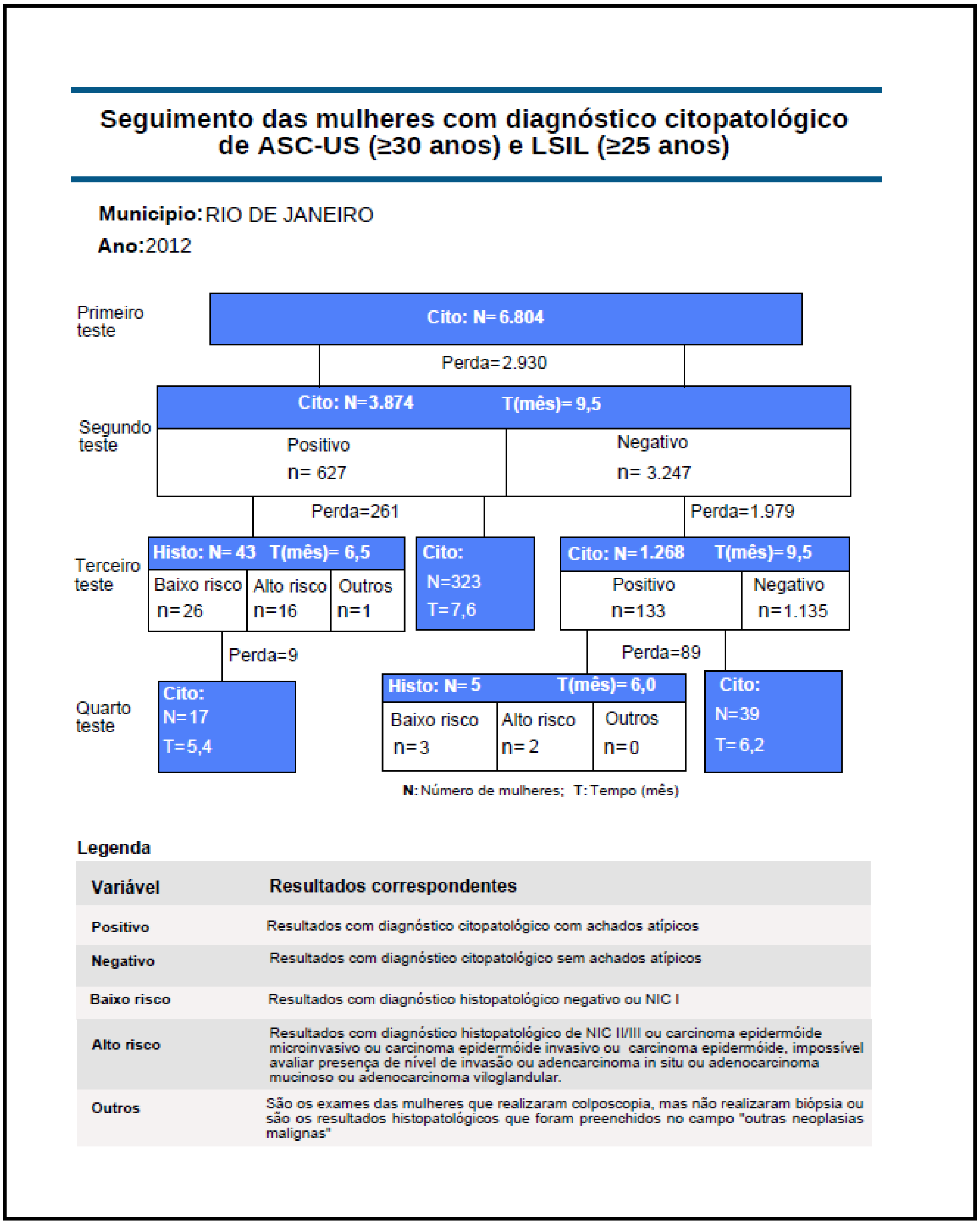
Cito: N=175 T=7,7

N: Número de mulheres; **T:** Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
NIC II-III/AIS	Resultados com diagnóstico histopatológico de NIC II ou NIC III.
Câncer	Resultados com diagnóstico histopatológico de carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adencarcinoma in situ ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas".
Baixo risco	Resultados com diagnóstico citopatológico: sem achados atípicos ou ASC-US ou ASC-H ou AOI ou AGC ou LSIL.
HSIL	Resultados com diagnóstico citopatológico de HSIL.
Alto risco	Resultados com diagnóstico citopatológico de: HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou adenocarcinoma in situ ou Adenocarcinoma invasor.

Figure 30: Output screen for the follow-up indicators for women (≥25 years old) with initial cytopathological test result of HSIL in the municipality of Rio de Janeiro in the year 2012.



Seguimento das mulheres com diagnóstico citopatológico de ASC-US (≥30 anos) e LSIL (≥25 anos)

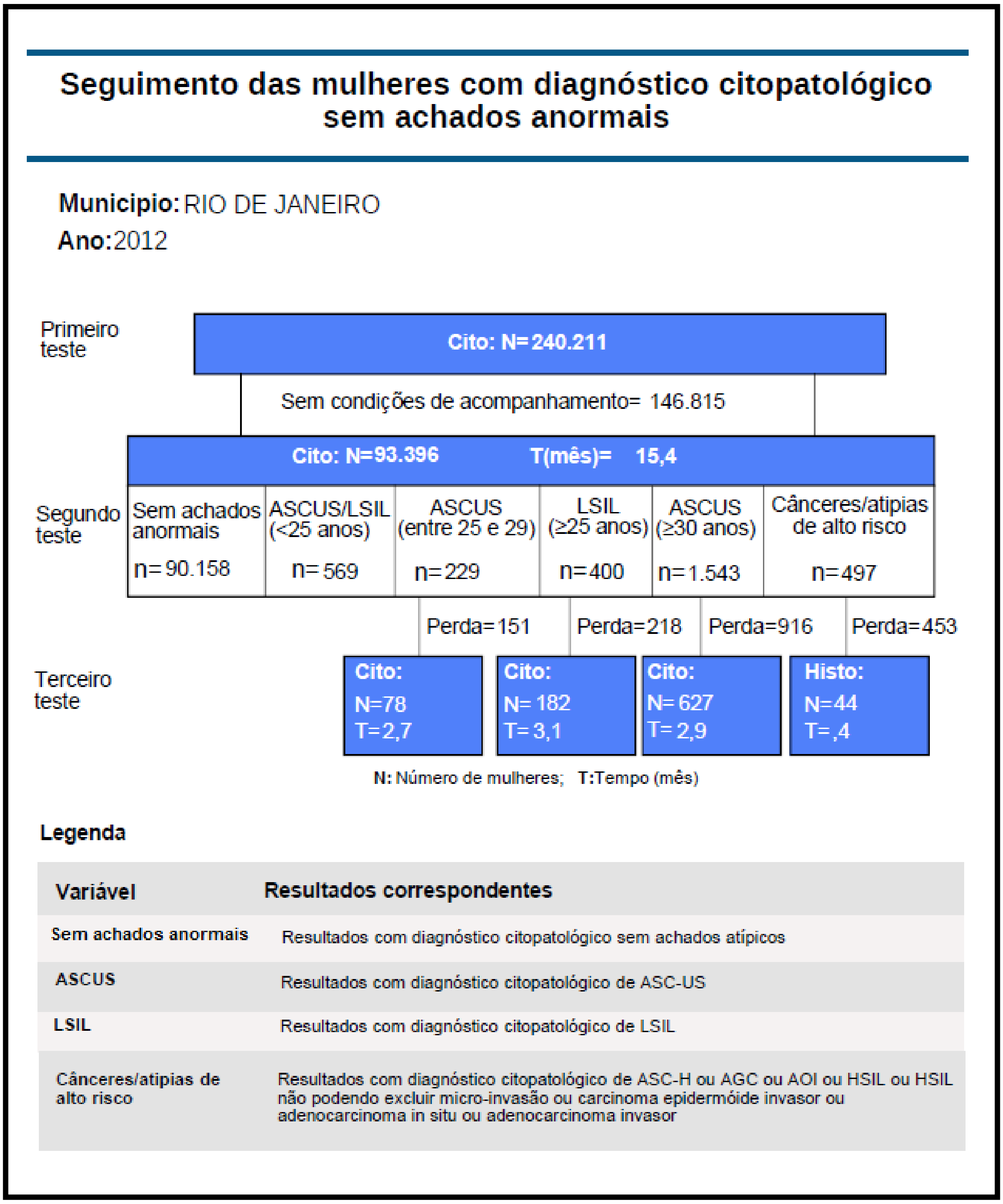
Municipio: RIO DE JANEIRO

Ano: 2012

Legenda

Variável	Resultados correspondentes
Positivo	Resultados com diagnóstico citopatológico com achados atípicos
Negativo	Resultados com diagnóstico citopatológico sem achados atípicos
Baixo risco	Resultados com diagnóstico histopatológico negativo ou NIC I
Alto risco	Resultados com diagnóstico histopatológico de NIC II/III ou carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adencarcinoma in situ ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas"

Figure 31: Output screen for the follow-up indicators for women (≥ 30 years old) with initial cytopathological test result of ASC-US and women (≥ 25 years old) with the initial cytopathological test result of LSIL, in the municipality of Rio de Janeiro in the year 2012.



Seguimento das mulheres com diagnóstico citopatológico sem achados anormais

Municipio: RIO DE JANEIRO
Ano:2012

Primeiro teste

Cito: N=240.211

Sem condições de acompanhamento= 146.815

Cito: N=93.396 T(mês)= 15,4

Segundo teste

Sem achados anormais	ASCUS/LSIL (<25 anos)	ASCUS (entre 25 e 29)	LSIL (≥25 anos)	ASCUS (≥30 anos)	Cânceres/atipias de alto risco
n=90.158	n=569	n=229	n=400	n=1.543	n=497

Perda=151 Perda=218 Perda=916 Perda=453

Terceiro teste

Cito: N=78 T=2,7

Cito: N= 182 T= 3,1

Cito: N= 627 T= 2,9

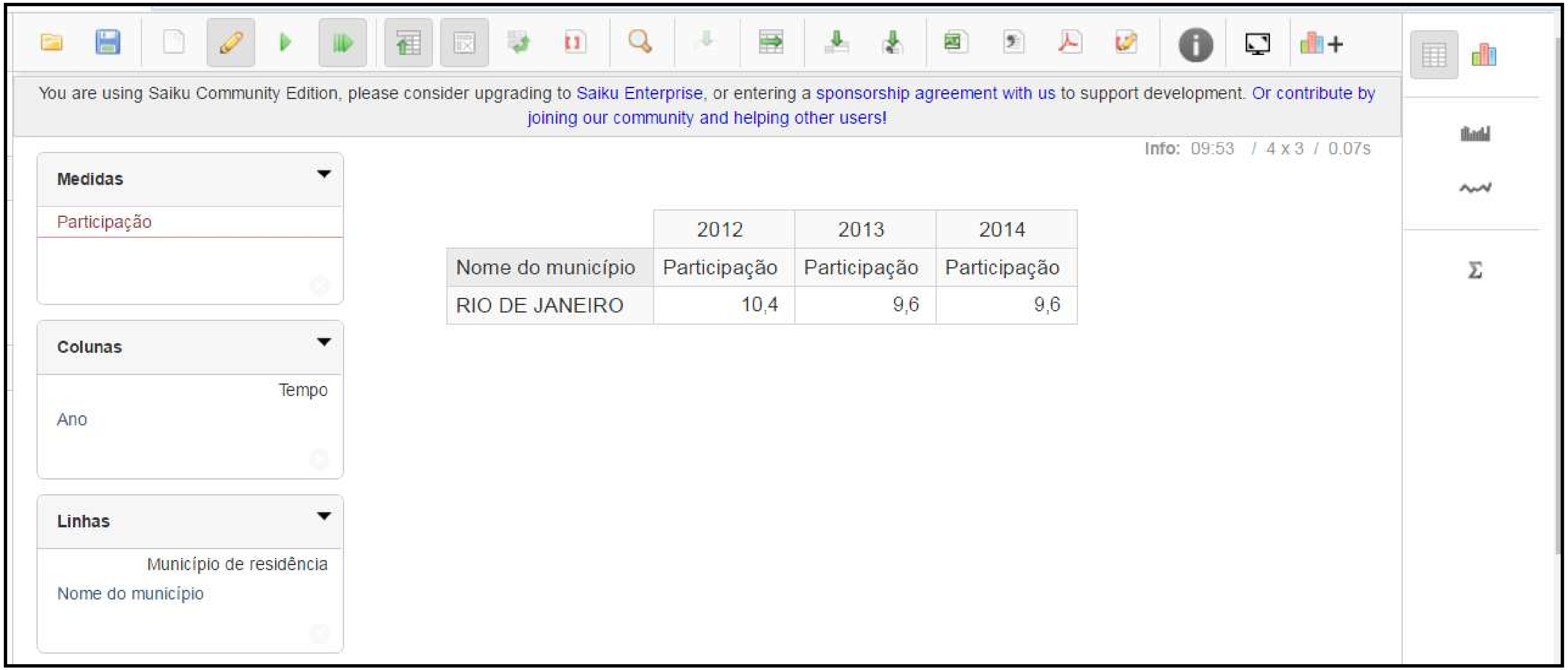
Histo: N=44 T= ,4

N: Número de mulheres; **T:**Tempo (mês)

Legenda

Variável	Resultados correspondentes
Sem achados anormais	Resultados com diagnóstico citopatológico sem achados atípicos
ASCUS	Resultados com diagnóstico citopatológico de ASC-US
LSIL	Resultados com diagnóstico citopatológico de LSIL
Cânceres/atipias de alto risco	Resultados com diagnóstico citopatológico de ASC-H ou AGC ou AOI ou HSIL ou HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou adenocarcinoma in situ ou adenocarcinoma invasor

Figure 32: Output screen for the follow-up indicators for women with the initial cytopathological test result without atypical findings, in the municipality of Rio de Janeiro in the year 2012.



You are using Saiku Community Edition, please consider upgrading to Saiku Enterprise, or entering a sponsorship agreement with us to support development. Or contribute by joining our community and helping other users!

Info: 09:53 / 4 x 3 / 0.07s

Medidas

Participação

Colunas

Tempo

Ano

Linhas

Município de residência

Nome do município

	2012	2013	2014
Nome do município	Participação	Participação	Participação
RIO DE JANEIRO	10,4	9,6	9,6

Figure 33: Output screen for the annual participation of the target population (25-64 years old) in the screening program in the municipality of Rio de Janeiro in the years 2012, 2013 and 2014.

CHAPTER 7 – DISCUSSION

The management of any cervical cancer screening program consists of many processes. Among them is the monitoring and evaluation of both test-based and woman-based indicators (WHO, 2002). According to the International Agency for Research on Cancer, the integration of test-based and woman-based indicators is essential for the evaluation of the program overall effectiveness (IARC, 2005). Several national and international studies and guidelines highlighted the essential role of those indicators for the detection and control of cancer and cancer precursor lesions (BASTOS, 2011; CCPCN, 2010; FEITOSA *et al.*, 2007; GAGE *et al.*, 2003; NCCSPRI, 2014).

The DW implemented in this study offers a user interface that allows the visualization of a set of test-based and woman-based 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 processes' failures. The visualization of the follow-up indicators by group of women, according to their test result and age, is the major contributions of this DW due to the absence of this information on the current available environments. The monitoring of the follow-up indicators is one of the essential actions in the screening process since it ensures the adherence of women to the program's clinical approaches (IARC, 2005). This follow-up should be performed systematically for all the women who enrolled in the program, not only for women with positive test results (WHO, 1988; WHO, 2014).

The developed DW also provides the user with a great flexibility for the visualization of the test-based indicators in a single user interface, where one could create one's own cross-tabulation report using any combination of the five proposed dimensions. This cross-tabulation is performed in an automatic way by just dragging the dimensions and facts into the specified position. Considering that the dataset used to develop the DW is limited to the production of tests performed in the municipality of Rio de Janeiro, 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 order to analyze the whole production of the municipalities, the inclusion of SISCOLO data for the State should be considered in the future development of the DW.

Many cross-section 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 JR *et al.*, 2015; ÁZARA *et al.*, 2014; BORTOLON *et al.*, 2012; THULER *et al.*, 2007; TOBIAS *et al.*, 2016). Those indicators could be obtained from the available environments indirectly, where the managers have to perform manual work in order to calculate them. On the other hand, the developed DW enables the user to obtain them in a direct form that reduces both effort and time as well as enables managers to continuously monitor and evaluate the performance of the laboratories that provide services to the municipality. Through this evaluation, an appropriate decision could be taken to enhance both the overall performance and the staff's performance.

To the best of our knowledge, we only identified in the international literature studies that described tools for the clinical decision regarding the diagnosis or treatment of the premalignant and early stage malignant conditions (AHRQ, 2013). Therefore, the proposed tool will be compared with the tools provided by the Brazilian Ministry of Health. Although both TabNet and TabWin allow the visualization of the test-based indicator in many dimensions, the developed DW managed to overcome some of the limitations of these environments. The DW provides the indicators in percentages, while TabNet provides them in quantities, which requires an extra effort if they are to be used for comparative purposes. In the case of TabWin, the user has to configure the indicators manually, which is a tedious and time-consuming process. On the other hand, TabWin has the advantage of allowing the visualization of the indicators on geographic maps. The inclusion of additional visualization techniques such as geographic maps is considered in the future development of the DW.

In this work, a probabilistic record linkage was performed on the SISCOLO data in order to allow the construction of the woman-based indicators. These indicators are visualized in an easy way and give the manager a global profile of the screened women and their adherence to the program's clinical approaches. This profile provides the manager with measurable indicators about the coverage of the program, the number of examined women, their age, the test results, the proportion of women who underwent a confirmatory histopathology test, the average time the women took in order to perform the next recommended test, and the number of missed women who did not perform the next recommended test. Those indicators are visualized in a compact flow diagram that follows the logic used in the follow-up diagrams of the Brazilian Guidelines for Cervical Cancer Screening (BRAZIL, 2016d). The motivation for this compact visualization was to facilitate the analysis of the follow-up process of women for nontechnical personnel, and the development and implementation of up-to-date algorithms for the indicators' calculation. The flow diagrams used in this study were based on the guidelines review in 2016. Therefore, the findings in this study should be interpreted with caution since the DW data set covers a period previous to 2016, where the clinical approaches for the problem had some differences.

The findings of the woman-based indicators cannot be directly compared with other findings in the literature, but it is possible to make an indirect comparison with the findings presented by Bastos (2011), related to these indicators for the State of Rio de Janeiro. Regarding the participation of the target population, it could be observed that 10 % of the target population is screened annually, which is similar to the obtained value by Bastos (2011). It could be observed that both values are a third of the established goal by the Brazilian Ministry of Health (BRAZIL, 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 (ANS, 2015). The percentage of loss to follow-up for women eligible to repeat a cytopathology test in six months was 43.1% in the municipality of Rio de Janeiro and 63.7% in the State of Rio de Janeiro (BASTOS, 2011). 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 cytopathology test was 6.5 months, which was longer than the 4 months observed for the State.

It is worth mentioning that among the group of women without atypical findings in the first test, there was a subgroup who presented a second test with atypia with high risk for cancer or even cancer within a period of 15 months. This is an unexpected situation considering the natural history of the disease (FORSMO *et al*., 1997). This is a type of finding which should be seen as an alert, and needs to be further investigated by the local manager. It was also observed in this group that the majority of women who underwent the first test did not have a second test in the SISCOLO data 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.

The analysis of the groups with the recommendation to repeat the cytopathology test after 1 or 3 years was limited due to the size of the dataset. To overcome this limitation, there is a need to complement the dataset in order to have a longer period to monitor these recommendations. However, the update of the dataset to be loaded into the DW is not a trivial process, as was explained in the sessions 5.3 and 5.4. The performing of these processes could be an obstacle for the user without the support of technical personnel, and a strategy for the DW management should be considered before its implementation. This is also a topic for further development of the DW.

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). A low percentage of completeness of the *Cartão* SUS was also observed (15.40%), although it was higher than that observed by Bastos (0.4%). The *Cartão* SUS is a key field that could be used as a unique identifier for performing the record linkage process, which is essential in the construction of the woman-based indicators. Another problem related to the consistency of the dataset was the absence of the histopathology test results for 47.24% of the records, as well as the absence of data on the colposcopy results (93.9%). In addition, the reporting of 6.52% of the records as text rather than code hampered the analysis, since the inclusion of these records requires the application of text processing algorithms, which is beyond the scope of this thesis. These problems hindered the analysis of adherence for the women whose recommended clinical approach is to perform an immediate colposcopy. The colposcopy results are important for checking the eligibility of women to undergo histopathology/cytopathology tests, and the time of diagnosis of the atypical findings with cancer or high risk for cancer. This colposcopy information is most likely to be encountered in other SUS information systems. However, the SISCOLO does not have integration with the other information systems. The observed low values in the percentages of completeness of the SISCOLO data could be overcome through the training of the healthcare personnel, as well as raising their level of awareness about the impact of incompleteness and data inconsistency on the management of the program actions.

The developed DW provides flexible visualization features for both test-based and woman-based 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.

The developed DW 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.

REFERENCES

ACS, 2015, American Cancer Society (ACS), Cervical Cancer Overview. Available from: https://old.cancer.org/acs/groups/cid/documents/webcontent/003042-pdf.pdf. Accessed in: Feb-2017.

AHMED, S. Y. M and ALMEIDA, R. T, "A review of performance indicators for evaluation of cervical cancer screening programs". *XXIV Congresso Brasileiro de Engenharia Biomédica,* 207, Uberlandia, Brazil, 2014a.

AHMED, S. Y. M, COSTA, J. F . G, SILVA, M .G .P, *et al.*, "Perfil de unidades básicas de saúde quanto às ações de rastreamento do câncer do colo do útero no rj". *XXIV Congresso Brasileiro de Engenharia Biomédica,* 273, Uberlandia, Brazil, 2014b.

Agency for Healthcare Research and Quality (AHRQ). Decision Support Tools for Screening and Treatment Decisions in Early Cancer. Evidence-based Practice Center Systematic Review Protocol, 2013. Available from: http://effectIvehea lthcare.ahrq. gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1749&pageaction=displayproduct.

ANS, 2015, Agência Nacional de Saúde Suplementar (ANS), Beneficiários de planos privados de saúde, por cobertura assistencial (Brasil - 2003-2011). Available from: <http://www.ans.gov.br/images/stories/Materiais_para_pesquisa/Perfil_setor/Caderno_informacao_saude_suplementar/2015_mes12_caderno_informacao.pdf>. Accessed in: Feb-2017.

ARAUJO JR, MARIO LUCIO C, SANTANA, DANIELA A, ALMEIDA, LÍVIA B, *et al.* "Quality in cytopathology: an analysis of the internal quality monitoring indicators of the Instituto Nacional de Câncer", **Jornal Brasileiro de Patologia e Medicina Laboratorial**. 51, 2, pp. 102-107, 2015.

ASSUNÇÃO, J. R. GOMES, ARAÚJO, D. D. O, ARAÚJO, D. V, *et al.* "Avaliação de indicadores para câncer de colo do útero no período de 2008 a 2012", **Revista Ciência Plural**. 1, 3, pp. 38-50, 2015.

ÁZARA, C. Z. S, MANRIQUE, E. J. C, TAVARES, S. B. N, *et al.* "Internal quality control indicators of cervical cytopathology exams performed in laboratories monitored by the External Quality Control Laboratory", **Revista Brasileira de Ginecologia e Obstetrícia**. 36, 9, pp. 398-403, 2014.

BASTOS, E. A. *Estimativa da efetividade do programa de rastreamento do câncer do colo do útero no estado do Rio de Janeiro*. MSc. dissertation, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, COPPE/UFRJ, Programa de Engenharia Biomédica, Rio de Janeiro, 2011.

BORTOLON, P. C, SILVA, M. A. F, CORRÊA, F. M, *et al.* "Avaliação da qualidade dos laboratórios de citopatologia do colo do útero no Brasil", **Revista Brasileira de Cancerologia**. 58, 3, pp. 435-444, 2012.

BRAZIL, 2005, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Portaria 2439. Política Nacional de Atenção Oncológica. Available from: <http://www1.inca.gov.br/inca/Arquivos/Legislacao/portaria_2439.pdf>. Accessed in: Feb-2017.

BRAZIL, 2006a, Ministério da Saúde, Secretaria de Atenção à Saúde, Instituto Nacional de Câncer (INCA), Coordenação de Prevenção e Vigilância, Nomenclatura Brasileira para Laudos Cervicais e Condutas Preconizadas: recomendações para profissionais de saúde 2ª ed., Rio de Janeiro. Available from: <http://bvsms.saude.gov.br/bvs/publicacoes/Nomenclaturas_2_1705.pdf>. Accessed in: Feb - 2017.

BRAZIL, 2006b, Ministério da Saúde, Gabinete do Ministro, Portaria 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. Available from: <http://bvsms.saude.gov.br/bvs/saudelegis/gm/2006/prt0399_22_02_2006.html>. Accessed in: Feb-2017.

BRAZIL, 2010a, Ministério da Saúde, Gabinete do Ministro, Portaria nº 310 de 10 de Fevereiro de 2010. Available from: <http://bvsms.saude.gov.br /bvs/saudelegis/gm/2010/prt0310_10_02_2010.html>. Accessed in: Feb-2017.

BRAZIL, 2010b, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Plano de ação para redução da incidência e mortalidade por câncer do colo do útero: sumário executivo. Available from: <http://www1.inca.gov.br /inca/Arquivos/ sumario_colo _utero_versao_2011.pdf>. Accessed in: Feb-2017.

BRAZIL, 2011a, Ministério da Saúde, Instituto Nacional de Câncer, Sistemas de informação do controle do câncer de mama (sismama) e do câncer do colo do útero (siscolo): manual gerencial. Available from: <http://bvsms.saude.gov.br /bvs/publicacoes/inca/Sistema_de_informacao_do_controle_do_cancer_de_mama.pdf>. Accessed in: Feb-2017.

BRAZIL, 2011b, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Coordenação Geral de Ações Estratégicas. Divisão de Apoio à Rede de Atenção Oncológica. Diretrizes brasileiras para o astreamento do câncer do colo do útero. Available from: <http://www1.inca.gov.br/inca/Arquivos/ Diretrizes_rastreamento _ cancer_colo_utero.pdf>. Accessed in: Dec-2016.

BRAZIL, 2011c, Ministério da Saúde, Departamento de Informática do SUS, Sistemas de Informação do Controle do Câncer de Mama (SISMAMA) e do Câncer do Colo do Útero (SISCOLO), Siscolo Lay-Out de Tabelas. Available from: <ftp://ftp.datasus.gov.br/siscam/siscolo/ArqSISCAM_IEV400.zip>. Accessed in: Nov-2016.

BRAZIL, 2012a, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Manual de Gestão da Qualidade para Laboratórios de Citopatologia. Available from: <http://www1.inca.gov.br/inca/Arquivos/publicacoes/manual_gestao_qualidade_laboratorio_citopatologia.pdf>. Accessed in: Feb-2017.

BRAZIL, 2012b, Ministério da Saúde, Secretaria de Atenção à Saúde, Instituto Nacional de Câncer (INCA), Coordenação de Prevenção e Vigilância, Nomenclatura Brasileira para Laudos Citopatológicos Cervicais 3ª ed., Rio de Janeiro. Available from: <http://bvsms.saude.gov.br/bvs/publicacoes/inca/nomenclatura_brasileira_laudos_citopatologicos.pdf>. Accessed in: Feb-2017.

BRAZIL, 2013a, Ministério da Saúde, Instituto Nacional de Câncer, Sistema de informação do câncer: manual preliminar para apoio à implantação. Available from: <http://bvsms.saude.gov.br/bvs/publicacoes/inca/siscan_manual_preliminar.pdf>. Accessed in: Feb-2017.

BRAZIL, 2013b, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Portaria nº 3.394, de 30 de dezembro de 2013, Institui o Sistema de Informação de Câncer (SICAN) no âmbito do Sistema ùnico de Saúde (SUS). Available from: <http://bvsms.saude.gov.br/bvs/saudelegis/gm/2013/prt3394_30_12_2013.html>. Accessed in: Feb-2017.

BRAZIL, 2013c, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Portaria nº 3.388, de 30 de dezembro de 2013, Redefine a Qualificação Nacional em Citopatologia na prevenção do câncer do colo do útero (QualiCito), no âmbito da Rede de Atenção à Saúde das Pessoas com Doenças Crônicas. Available from: <http://bvsms.saude.gov.br/bvs/saudelegis/gm/2013/prt3388_30_12_2013.html>. Accessed in: Feb-2017.

BRAZIL, 2013d, Ministério da Saúde, Gabinete do Ministro, Portaria nº 3.388, de 30 de dezembro de 2013, Redefine a Qualificação Nacional em Citopatologia na prevenção do câncer do colo do útero (QualiCito), no âmbito da Rede de Atenção à Saúde das Pessoas com Doenças Crônicas. Available from: <http://bvsms.saude.gov.br/bvs/saudelegis/gm/2013/prt3388_30_12_2013.html>. Accessed in: Feb-2017.

BRAZIL, 2015a, Ministério da Saúde, Instituto Nacional de Câncer, câncer do colo do útero. Available from: <http://www2.inca.gov.br/wps/wcm/ connect/tiposdecancer/site/home++/colo_utero/definicao>. Accessed in: Jan-2017.

BRAZIL, 2015b, Ministério da Saúde, Departamento de Informação e Informática do SUS. Sistema de informação do câncer do colo do útero e mama, Informações Estatísticas Available from: <http://w3.datasus.gov.br/siscam/index.php?area=0401>. Accessed in: Aug-2016.

BRAZIL, 2015c, Ministério da Saúde, Instituto Nacional de Câncer, Painel de indicadores do câncer do colo de útero. Available from: <http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_nacional_controle_cancer_colo_utero/indicadores>. Accessed in: Nov-20 16.

BRAZIL, 2015d, Ministério da Saúde, Sala de apoio a gestão estratégica, Prevenção e tratamento de câncer de colo e mama. Available from: <http://189.28.128.178/sage/>. Accessed in: Aug-2016.

BRAZIL, 2015e, Ministério da Saúde, Departamento de Informática do SUS, Cadastro Nacional de Estabelecimentos de Saúde. Available from: <http://cnes.datasus.gov.br/>. Accessed in: May-2016.

BRAZIL, 2015f, Ministério da Saúde, Departamento de Informática do SUS, Tabelas Nacionais. Available from: <ftp://ftp.datasus.gov.br/territorio/tabelas>. Accessed in: Mar-2015.

BRAZIL, 2016a, Ministério da Saúde, Instituto Nacional de Câncer, Atlas de Mortalidade por Câncer. Available from: <http://www1.inca.gov.br/vigilancia/mortalidade.asp>. Accessed in: Feb-2017.

BRAZIL, 2016b, Departamento de Informática do SUS. TAB para Windows (TABWIN). Available from: <http://datasus.saude.gov.br/informacoes-de-saude/ferramentas/tabwin> . Accessed in: Feb-2017.

BRAZIL, 2016c, Ministério da saúde, Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), Manual de gestão da qualidade para laboratório de citopatologia.

Available from: <http://www1.inca.gov.br/inca/Arquivos/livro _completo_ manual_ citopatologia.pdf>. Accessed in: Jan-2016.

BRAZIL, 2016d, Ministério da Saúde, Instituto Nacional de Câncer (INCA), Coordenação de Prevenção e Vigilância (Conprev), Divisão de Detecção Precoce e Apoio à Organização de Rede. Diretrizes brasileiras para o astreamento do câncer do colo do útero. 2nd ed. Available from: <http://www1.inca.gov.br/inca/Arquivos/D Diretrizes_para_o_Rastreamento_do_cancer_do_colo_do_utero_2016_corrigido.pd f>. Accessed in: Feb-2017.

BRAZIL, 2017, Ministério da Saúde, Instituto Nacional de Câncer, controle do câncer do colo doútero, histórico das ações. Available from: <http://www2.inca.gov.br/wps/ wcm/connect/acoes_programas/site/home/nobrasil/programa_nacional_controle_ca ncer_colo_utero/historico_acoes>. Accessed in: Jan-2017.

BRITO-SILVA, K, BEZERRA, A. F. B., CHAVES, L. D. P, *et al.* "Integrality in cervical cancer care: evaluation of access", **Revista de Saude Publica**. 48, 2, pp. 240-248, 2014.

BURD, E. M "Human papillomavirus and cervical cancer", **Clinical Microbiology Reviews**. 16, 1, pp. 1-17, 2003.

CABRAL, M. D. B. *Proposta de relacionamento probabilístico dos registros da base de dados do programa de rastreamento do câncer do colo do útero.* DSc. thesis, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia, COPPE/UFRJ, Programa de Engenharia Biomédica, Rio de Janeiro, 2010.

CABRAL, M. D. B, FEITOSA, T.M.P, FIGUEIREDO, R.M, *et al.*, "Análise do rastreamento do câncer do colo do útero no Estado do Rio de Janeiro". In: proceeding of *XXI Congresso Brasileiro de Engenharia Biomédica*, pp. 323-326, Salvador/Brazil, 2008.

CAMARGO, JR. K. R and COELI, C. M "Evaluation of different blocking strategies in probabilistic record linkage", **Revista Brasileira de Epidemiologia**. 5, 2, pp. 185-196, 2002.

CCPCN, 2010, Government of Canada, Public Health Agency of Canada. Report from the Screening Performance Indicators Working Group, Cervical Cancer Prevention and Control Network (CCPCN): Performance Monitoring for Cervical Cancer Screening Programs in Canada. Available from: <http://www.phac-aspc.gc.ca/cd-mc/cancer/pmccspc-srpdccuc/index-eng.php>. Accessed in: Feb-2017.

CHRISTEN, PETER "Development and user experiences of an open source data cleaning, deduplication and record linkage system", **ACM SIGKDD Explorations Newsletter**. 11, 1, pp. 39-48, 2009.

CHUCK, B, DIRK, H, DON, S*, et al.*, **Data modeling techniques for data warehousing**. 1st ed. United State, Colorado, IBM, 1998.

CHURCHES, TIM, CHRISTEN, PETER, LIM, KIM, *et al.* "Preparation of name and address data for record linkage using hidden Markov models", **BMC Medical Informatics and Decision Making**. 2, 1, pp. 1, 2002.

CLARK, I, WHITING, P, TWIN, J, *et al.*, **Pathology Handbook**. 5 Ed. Australia, Christina Vett-Joice, 2012.

COSTA, R. A, LONGATTO-FILHO, A, PINHEIRO, C*, et al.* "Historical Analysis of the Brazilian Cervical Cancer Screening Program from 2006 to 2013: A Time for Reflection", **PLoS ONE**. 10, 9, pp. e0138945, 2015.

CPAC, 2011, Canadian Partnership Against Cancer (CPAC), Cervical Cancer Screening in Canada Monitoring Program Performance 2006 –2008. Available from: <https://content.cancerview.ca/download/cv/prevention_and_screening/cccic_microsite/documents/cccicervicalcsreportpdf?attachment=0>. Accessed in: Feb-2017.

CRUK, 2015, Cancer Research UK, What is cervical cancer. Available from: < http://www.cancerresearchuk.org/about-cancer/cervical-cancer/about>. Accessed in: May-2016

DIAS, MARIA BEATRIZ KNEIPP, GLÁUCIA, JEANE and ASSIS, TOMAZELLI MÔNICA "Rastreamento do câncer de colo do útero no Brasil: análise de dados do Siscolo no período de 2002 a 2006", **Epidemiologia e Serviços de Saúde**. 19, 3, pp. 293-306, 2010.

DISCACCIATI, M. G, BARBOZA, B. M. S and ZEFERINO, L. C "Por que a prevalência de resultados citopatológicos do rastreamento do câncer do colo do útero pode variar significativamente entre duas regiões do Brasil?", **Revista Brasileira de Ginecologia e Obstetrícia**. 36, 05, pp. 192-197, 2014.

DUSETZINA, S. B, TYREE, S, MEYER, A. M, *et al.* **Linking data for health services research: a framework and instructional guide**. Report 14-EHC033-EF, Agency for Healthcare Research and Quality (US), Rockville (MD). 2014.

FEITOSA, T. M. P and ALMEIDA, R. T "Perfil de produção do exame citopatológico para controle do câncer do colo do útero em Minas Gerais, Brasil, em 2002 Pap smear screening for the control of cervical cancer in Minas Gerais State, Brazil, 2002", **Cadernos de Saúde Pública**. 23, 4, pp. 907-917, 2007.

FELLEGI, I. P and SUNTER, A. B "A theory for record linkage", **Journal of the American Statistical Association**. 64, 328, pp. 1183-1210, 1969.

FORBES, C. A, JEPSON, R. G and MARTIN-HIRSCH, P. P "Interventions targeted at women to encourage the uptake of cervical screening", **The Cochrane Library**. pp. 1999.

FORSMO, S, BUHAUG, H, SKJELDESTAD, F. E, *et al.* "Treatment of pre-invasive conditions during opportunistic screening and its effectiveness on cervical cancer incidence in one Norwegian county", **International Journal of Cancer**. 71, 1, pp. 4-8, 1997.

FREIRE, S. M., ALMEIDA, R.T., CABRAL, M. D. B., *et al.* "A record linkage process of a cervical cancer screening database", **Computer Methods and Programs in Biomedicine**. 108, 1, pp. 90-101, 2012.

FREIRE, S. M., SOUZA, R. C. and ALMEIDA, R. T. **Avaliação de técnicas para vinculação de registros de base de dados e desenvolvimento de um framework para aplicação dessas técnicas**. Report CNPq/ANS – No25/2007, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasília. 2010.

FREIRE, S. M., SOUZA, R.C. and ALMEIDA, R. T. "Integrating Brazilian health information systems in order to support the building of data warehouses", **Research on Biomedical Engineering**. 31, 3, pp. 0-0, 2015.

GAGE, J. C, FERRECCIO, C., GONZALES, M., *et al.* "Follow-up care of women with an abnormal cytology in a low-resource setting", **Cancer Detection and Prevention**. 27, 6, pp. 466-471, 2003.

GAKIDOU, EMMANUELA, NORDHAGEN, STELLA and OBERMEYER, ZIAD "Coverage of cervical cancer screening in 57 countries: low average levels and large inequalities", **PLoS Med**. 5, 6, pp. e132, 2008.

GAUZA, JOSÉ EDUARDO, POPE, LEONORA ZOZULA BLIND, POSSAMAI, DIMITRI SAUFER, *et al.* "The magnitude of appropriate cytologic sample in the detec-tion of precedent wounds of the uterine cervical cancer", **Arquivos Catarinenses de Medicina**. 39, 4, pp. 2010.

GILL, L., "Methods for automatic record matching and linking and their use in national statistics". London: Office for National Statistics (National Statistics Methodological, Series 25). 2001.

GIRIANELLI, V. R, THULER, L. C. S and SILVA, G. A "Qualidade do sistema de informação do câncer do colo do útero no estado do Rio de Janeiro", **Revista de Saúde Pública**. 43, 4, pp. 580-588, 2009.

HERZOG, THOMAS N, SCHEUREN, FRITZ J and WINKLER, WILLIAM E, **Data quality and record linkage techniques**. United State, New York, 1st ed, 2007.

IARC, **Cervix cancer screening: IARC Handbook of Cancer Prevention**. 1st ed. France, Lyon, IARC Press, 2005.

IARC, 2015, International Agency for Research on Cancer (IARC). Cervical cancer. Available from: <http://screening.iarc.fr/cervicalindex.php>. Accessed in: Sept-2016.

INMON, WILLIAM H, **Building the data warehouse**. 4 th ed. United States, John Wiley &amp;amp Sons, 2005.

JAVA, 2016, ORACLE Corporation. JAVA. Version 7.

KHAN, M and KHAN, S "Data and information visualization methods, and interactive mechanisms: A survey", **International Journal of Computer Applications**. 34, 1, pp. 1-14, 2011.

KIMBALL, R and ROSS, M, **The data warehouse toolkit: the complete guide to dimensional modeling**. 2nd. United States, John Wiley &amp;amp and Sons, 2011.

LAZCANO-PONCE, E. C, BUIATTI, E, NÁJERA-AGUILAR, P, *et al.* "Evaluation model of the Mexican national program for early cervical cancer detection and proposals for a new approach", **Cancer Causes & Control**. 9, 3, pp. 241-251, 1998.

MAYOCLINIC, 2015, Diseases and conditions-cancer. Available from: <http://www.mayoclinic.org/diseases-conditions/cancer/basics/risk-factors/con-20032378>. Accessed in: June-2015.

MILLER, A. B, **Cervical cancer screening programmes: managerial guidelines**. 1 st. Switzerland, Geneva, World Health Organaization, 1992.

MORAES, M, N and JERÔNIMO, C, G, F "Análise Dos Resultados De Exames Citopatológicos Do Colo Uterino", **Revista de Enfermagem UFPE On Line**. pp. 2015.

MYSQL, Oracle Corporation. MySQL Work Bench. Ver 5.5. 2015a.

MYSQL, Oracle Corporation. MySQL server Ver 6.3. 2015b.

NASCIMENTO, G. W. C, PEREIRA, C. C. A., NASCIMENTO, D. I. C, *et al.* "Cervical cancer screening coverage in the state of Minas Gerais, Brazil between 2000-2010: a study using data from the Cervical Cancer Information System (SISCOLO)", **Cadernos Saúde Coletiva**. 23, 3, pp. 253-260, 2015.

NAYAR, R and WILBUR, D. C "The Pap Test and Bethesda 2014", **The Journal of Clinical Cytology and Cytopathology**. pp. 121-132, 2015.

NCCSPRI, 2014, The National Cervical Screening Programme of Republic of Ireland (NCCSPRI), Guidelines for quality assurance in cervical screening. Available from: <http://www.cervicalcheck.ie/_fileupload/QualityAssurance/NCSS-PUB-Q-1%20 Guidelines%20for%20Quality%20Assurance%20in%20Cervical%20Screening.pdf >. Accessed in: Feb-2017.

NEWCOMBE, H. B "Record linking: the design of efficient systems for linking records into individual and family histories", **American Journal of Human Genetics**. 19, 3 Pt 1, pp. 335, 1967.

NEWCOMBE, H. B, **Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration, and Business**. 1st ed. Oxford, Oxford University Press, 1988.

NOBRE, A, ALVES, J. C and NETO, D.L "Avaliação de indicadores de rastreamento do câncer do colo do útero no Amazonas, Norte do Brasil, de 2001 a 2005", **Revista Brasileira de Cancerologia**. 55, 3, pp. 213-220, 2009.

OPS, 1990, Organización Panamericana de la Salud (OPS), Manual de normas y procedimientos para el control del cáncer de cuello uterino. Available from: <http://apps.who.int/iris/bitstream/10665/173976/1/Manual%20de%20normas%20 y%20procedimientos%20para%20el%20control%20del%20cancer%20de%20cuell o%20uterino.pdf >. Accessed in: Feb-2017.

PENTAHO, Hitachi group company. Pentaho Suite Business Intelligence. Ver 6.1.0.1-196. 2016.

PLEWKA, J, TURKIEWICZ, M, DUARTE, B. F, *et al.* "Avaliação dos indicadores de qualidade de laboratórios de citopatologia cervical", **Revista do Instituto Adolfo Lutz**. 73, 2, pp. 140-147, 2014.

ROBERTS, JENNIFER M., JIN, FENGYI, THURLOE, JULIA K., *et al.* "The value of a transformation zone component in anal cytology to detect HSIL", **Cancer Cytopathology**. 124, 8, pp. 596-601, 2016.

ROBLES, S. C "Is a once-in-a-lifetime pap smear the best option for low-resourced settings?", **International Journal of Cancer**. 111, 1, pp. 160-161, 2004.

RODRIGUES, JOSENIRA FREITAS, MOREIRA, BEATRIZ AMARAL, ALVES, TAMARA GABRIELA SILVA, *et al.* "Rastreamento do câncer do colo do útero na região ampliada oeste de Minas Gerais", **Revista de Enfermagem do Centro-Oeste Mineiro**. 6, 2, pp. 2016.

SAIKU, Meteorite.bi Company. Ver 3.3.2. 2015.

SANTIAGO, S. MARIA and ANDRADE, M. G. G "Avaliação de um programa de controle do câncer cérvico-uterino em rede local", **Cadernos de Saúde Pública**. 19, 2, pp. 571-578, 2003.

SANTOS, R. S, MELO, E. C. P and SANTOS, K.M "Análise espacial dos indicadores pactuados para o rastreamento do câncer do colo do útero no Brasil", **Texto & Contexto Enfermagem**. 21, 4, pp. 800-810, 2012.

SCHIRNDING, Y, **Health in sustainable development planning: the role of indicators** 1 Ed. Geneva, World Health Organization, 2002.

SELLORS, J.W and SANKARANARAYANAN, R, 2003, Colposcopy and Treatment of Cervical Intraepithelial Neoplasia: a Beginners' Manual. Available from: <http://screening.iarc.fr/doc/Colposcopymanual.pdf>. Accessed in: Feb-2017.

SHIRATA, NEUZA KASUMI, PEREIRA, SONIA MARIA MIRANDA, CAVALIERE, MARIA JOSÉ, *et al.* "Celularidade dos esfregaços cérvicovaginais: importância em programas de garantia de qualidade em citopatologia", **Jornal Brasileiro de Ginecologia**. 108, 3, pp. 63-66, 1998.

SILVEIRA, DANIELE PINTO DA and ARTMANN, ELIZABETH "Accuracy of probabilistic record linkage applied to health databases: systematic review", **Revista de Saúde Pública**. 43, 5, pp. 875-882, 2009.

THULER, L. C . S, ZARDO, L . M and ZEFERINO, L . C "Perfil dos laboratórios de citopatologia do Sistema Único de Saúde", **O Jornal Brasileiro de Patologia e Medicina Laboratorial**. 43, 2, pp. 103-114, 2007.

TOBIAS, A. H. G, AMARAL, R. G, DINIZ, E. M, *et al.* "Quality Indicators of Cervical Cytopathology Tests in the Public Service in Minas Gerais, Brazil", **Revista Brasileira de Ginecologia e Obstetrícia/RBGO Gynecology and Obstetrics**. 38, 02, pp. 065-070, 2016.

UCHIMURA, N. SHOZO, NAKANO, K, NAKANO, L.C. G, *et al.* "Quality and performance of pap smears in the cervical cancer screening program in a city of southern Brazil", **Revista da Associação Médica Brasileira**. 55, 5, pp. 569-574, 2009.

VELICANU, M and MATEI, G "Building a Data Warehouse step by step", **Economic Informatics, Forthcoming**. 42, 2, pp. 83-89, 2007.

WHO, 1988, World Health Organization, Cytological screening in the control of cervical cancer: Technical guidelines. Available from: <http://apps.who.int/iris/bitstream/ 10665/39794/1/9241542195.pdf>. Accessed in: March-2016.

WHO, 2002, World Health Organization, Cervical cancer screening in developing countries. Available from: <http://www.who.int/cancer/media/en/ cancer_ cervical_ 37321.pdf>. Accessed in: Jul-2016.

WHO, **World Health Organization, Planning and implementing cervical cancer prevention and control programs: a Manual for Managers**. 1st ed. Seattle, Alliance for Cervical Cancer Prevention, 2004.

WHO, 2007, World Health Organization, Cancer Control, WHO Guide for Effective Programmes: Early Detection. Available from: <http://www.who.int/cancer/ modules/Early%20Detection%20Module%203.pdf>. Accessed in: Jun-2015.

WHO, 2012, World Health Organization , Screening for cervical cancer. Available from: <http://www.who.int/cancer/detection/cervical_cancer_screening/en/>. Accessed in: Nov-2016.

WHO, 2013, World Health Organization , Monitoring national cervical cancer prevention and control programmes. Available from: <http://apps.who.int/iris/bitstream/ 10665/79316/1/9789241505260_eng.pdf?ua=1>. Accessed in: Feb-2017.

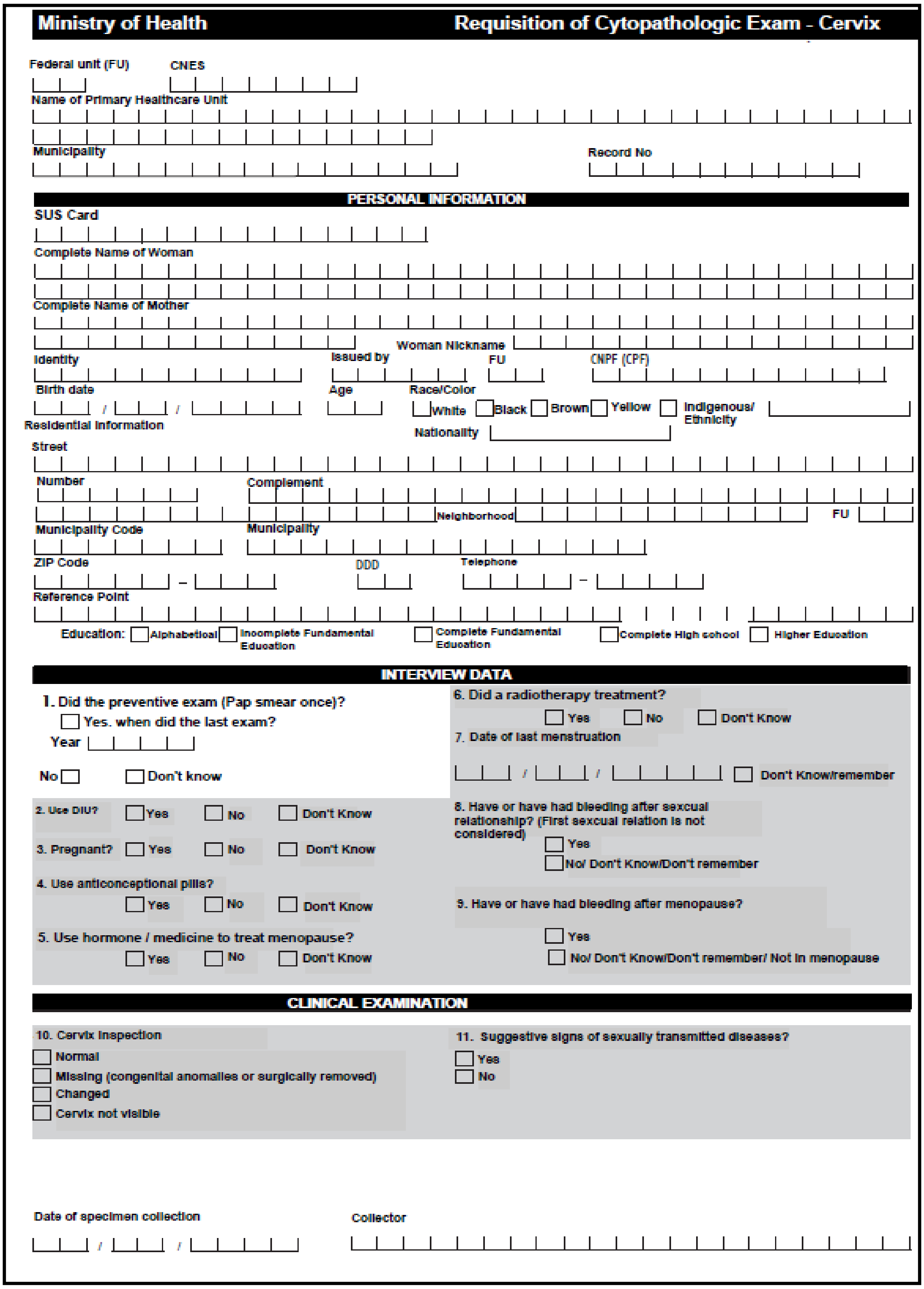
WHO, 2014, World Health Organization, Comprehensive cervical cancer control, a guide to essential practice. Available from: <http://apps.who.int/iris/bitstream/10665/ 144785/1/9789241548953_eng.pdf>. Accessed in: Feb-2017.

WHO, 2016, World Health Organization . Screening as well as vaccination is essential in the fight against cervical cancer. Available from: <http://www.who.int/ reproductivehealth/topics/cancers/fight-cervical-cancer/en/>. Accessed in: Jan-2017.

WINKLER, W. E., 2006, Data Quality: Automated Edit/Imputation and Record Linkage, Washington, DC: Statistical Research Division, U.S. Bureau of the Census. Available from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi= 10.1.1.79.1519 >. Accessed in: Feb-2017.

ANNEX 1: Cytopathologic Requisition Form

A 1.1: Translated form



Ministry of Health **Requisition of Cytopathologic Exam - Cervix**

Federal unit (FU) CNES

Name of Primary Healthcare Unit

Municipality Record No

PERSONAL INFORMATION

SUS Card

Complete Name of Woman

Complete Name of Mother

Woman Nickname

Identity Issued by FU CNPF (CPF)

Birth date / /

Age Race/Color □ White □ Black □ Brown □ Yellow □ Indigenous/ Ethnicity

Residential information Nationality

Street

Number Complement

Neighborhood FU

Municipality Code Municipality

ZIP Code - DDD Telephone -

Reference Point

Education: □ Alphabetical □ Incomplete Fundamental Education □ Complete Fundamental Education □ Complete High school □ Higher Education

INTERVIEW DATA

1. Did the preventive exam (Pap smear once)?
□ Yes. when did the last exam?
Year
No □ □ Don't know

2. Use DIU? □ Yes □ No □ Don't Know

3. Pregnant? □ Yes □ No □ Don't Know

4. Use anticonceptional pills?
□ Yes □ No □ Don't Know

5. Use hormone / medicine to treat menopause?
□ Yes □ No □ Don't Know

6. Did a radiotherapy treatment?
□ Yes □ No □ Don't Know

7. Date of last menstruation
/ / □ Don't Know/remember

8. Have or have had bleeding after sexcual relationship? (First sexcual relation is not considered)
□ Yes
□ No/ Don't Know/Don't remember

9. Have or have had bleeding after menopause?
□ Yes
□ No/ Don't Know/Don't remember/ Not in menopause

CLINICAL EXAMINATION

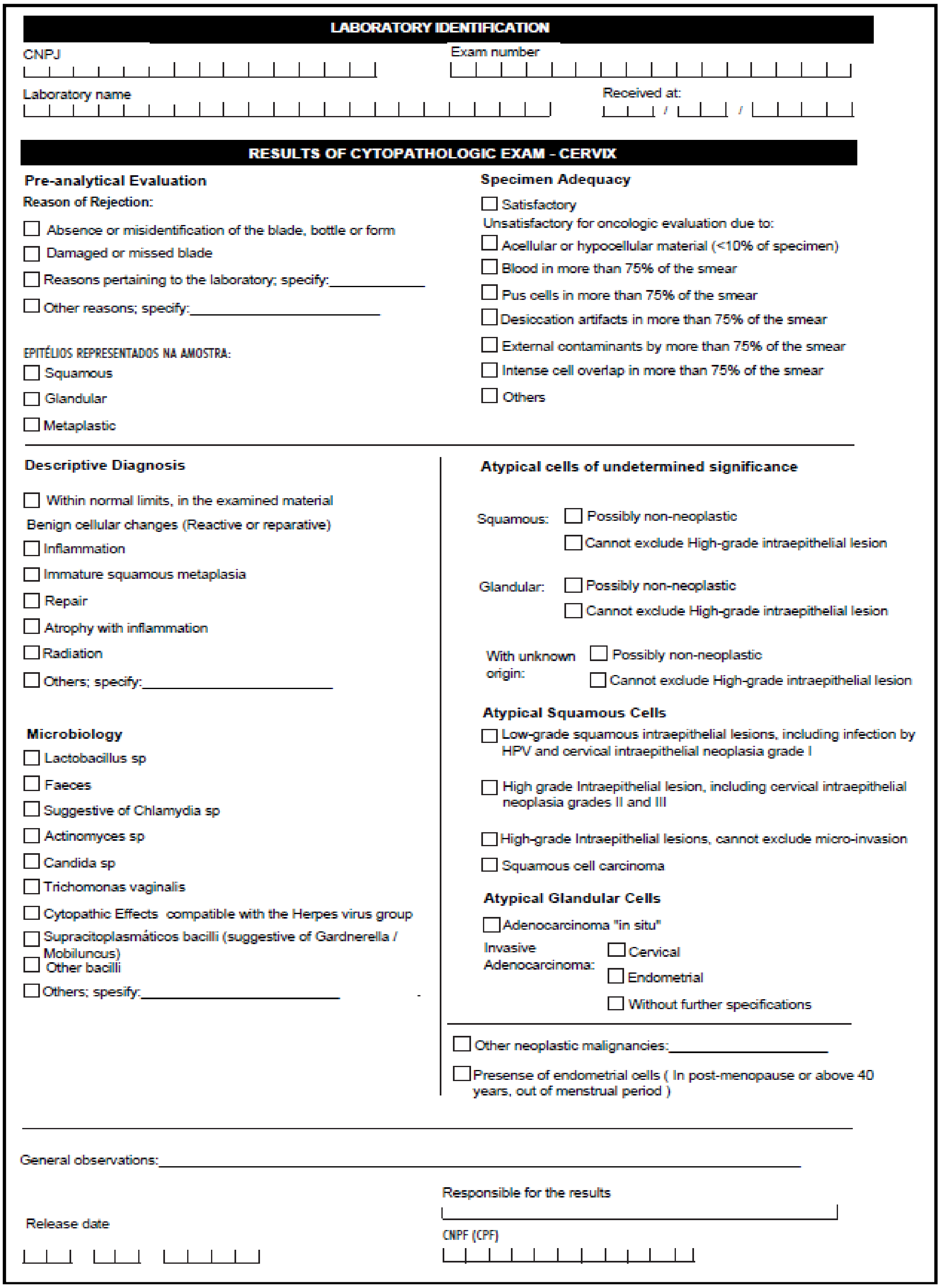
10. Cervix inspection
□ Normal
□ Missing (congenital anomalies or surgically removed)
□ Changed
□ Cervix not visible

11. Suggestive signs of sexually transmitted diseases?
□ Yes
□ No

Date of specimen collection / /

Collector

A 1.2: Translated form



LABORATORY IDENTIFICATION

CNPJ |_|_|_|_|_|_|_|_|_|_|_|_|_|_|

Exam number |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|

Laboratory name |_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|_|

Received at: |_|_| / |_|_| / |_|_|_|_|

RESULTS OF CYTOPATHOLOGIC EXAM - CERVIX

Pre-analytical Evaluation

Reason of Rejection:

□ Absence or misidentification of the blade, bottle or form

□ Damaged or missed blade

□ Reasons pertaining to the laboratory; specify:____________

□ Other reasons; specify:________________________

EPITÉLIOS REPRESENTADOS NA AMOSTRA:

□ Squamous

□ Glandular

□ Metaplastic

Specimen Adequacy

□ Satisfactory

Unsatisfactory for oncologic evaluation due to:

□ Acellular or hypocellular material (<10% of specimen)

□ Blood in more than 75% of the smear

□ Pus cells in more than 75% of the smear

□ Desiccation artifacts in more than 75% of the smear

□ External contaminants by more than 75% of the smear

□ Intense cell overlap in more than 75% of the smear

□ Others

Descriptive Diagnosis

□ Within normal limits, in the examined material

Benign cellular changes (Reactive or reparative)

□ Inflammation

□ Immature squamous metaplasia

□ Repair

□ Atrophy with inflammation

□ Radiation

□ Others; specify:________________________

Microbiology

□ Lactobacillus sp

□ Faeces

□ Suggestive of Chlamydia sp

□ Actinomyces sp

□ Candida sp

□ Trichomonas vaginalis

□ Cytopathic Effects compatible with the Herpes virus group

□ Supracitoplasmáticos bacilli (suggestive of Gardnerella / Mobiluncus)

□ Other bacilli

□ Others; spesify:_________________________

Atypical cells of undetermined significance

Squamous: □ Possibly non-neoplastic

□ Cannot exclude High-grade intraepithelial lesion

Glandular: □ Possibly non-neoplastic

□ Cannot exclude High-grade intraepithelial lesion

With unknown origin: □ Possibly non-neoplastic

□ Cannot exclude High-grade intraepithelial lesion

Atypical Squamous Cells

□ Low-grade squamous intraepithelial lesions, including infection by HPV and cervical intraepithelial neoplasia grade I

□ High grade Intraepithelial lesion, including cervical intraepithelial neoplasia grades II and III

□ High-grade Intraepithelial lesions, cannot exclude micro-invasion

□ Squamous cell carcinoma

Atypical Glandular Cells

□ Adenocarcinoma "in situ"

Invasive Adenocarcinoma: □ Cervical

□ Endometrial

□ Without further specifications

□ Other neoplastic malignancies:____________________

□ Presense of endometrial cells (In post-menopause or above 40 years, out of menstrual period)

General observations:____

Responsible for the results

|______________________________________|

Release date

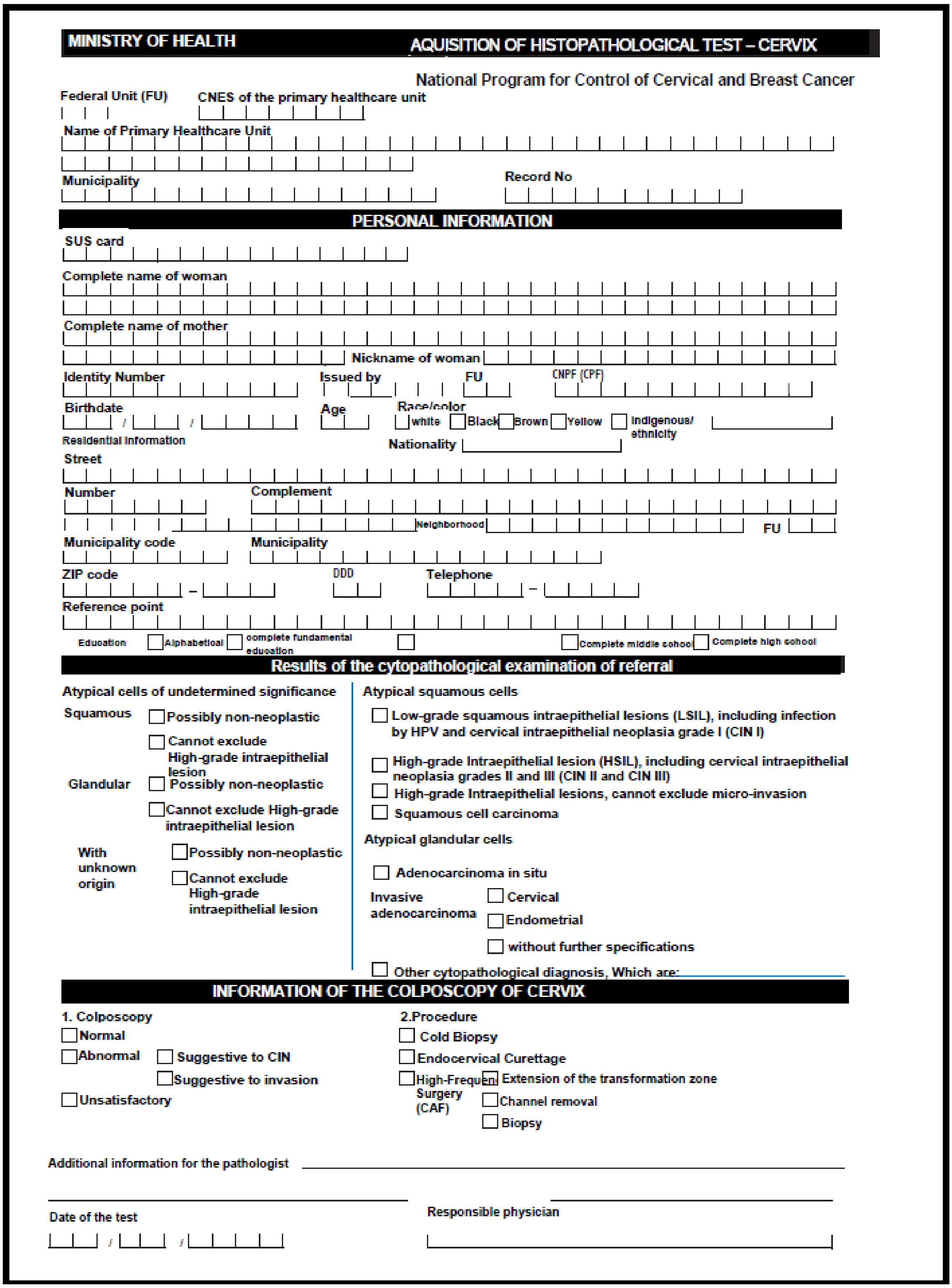
|_|_| |_|_| |_|_|_|_|

CNPF (CPF)

|_|_|_|_|_|_|_|_|_|_|_|

ANNEX 2: Histopathologic Requisition Form

A 2.1: Translated form



MINISTRY OF HEALTH **AQUISITION OF HISTOPATHOLOGICAL TEST – CERVIX**

National Program for Control of Cervical and Breast Cancer

Federal Unit (FU) CNES of the primary healthcare unit

Name of Primary Healthcare Unit

Municipality Record No

PERSONAL INFORMATION

SUS card

Complete name of woman

Complete name of mother

Nickname of woman

Identity Number Issued by FU CNPF (CPF)

Birthdate / / Age Race/color □ white □ Black □ Brown □ Yellow □ Indigenous/ ethnicity

Residential information Nationality

Street

Number Complement

Neighborhood FU

Municipality code Municipality

ZIP code – DDD Telephone –

Reference point

Eduoation □ Alphabetioal □ oomplete fundamental eduoation □ □ Complete middle sohool □ Complete high sohool

Results of the cytopathological examination of referral

Atypical cells of undetermined significance

Squamous □ Possibly non-neoplastic
□ Cannot exclude High-grade intraepithelial lesion

Glandular □ Possibly non-neoplastic
□ Cannot exclude High-grade intraepithelial lesion

With unknown origin □ Possibly non-neoplastic
□ Cannot exclude High-grade intraepithelial lesion

Atypical squamous cells

□ Low-grade squamous intraepithelial lesions (LSIL), including infection by HPV and cervical intraepithelial neoplasia grade I (CIN I)

□ High-grade Intraepithelial lesion (HSIL), including cervical intraepithelial neoplasia grades II and III (CIN II and CIN III)

□ High-grade Intraepithelial lesions, cannot exclude micro-invasion

□ Squamous cell carcinoma

Atypical glandular cells

□ Adenocarcinoma in situ

Invasive adenocarcinoma □ Cervical
□ Endometrial
□ without further specifications

□ Other cytopathological diagnosis, Which are:

INFORMATION OF THE COLPOSCOPY OF CERVIX

1. Colposcopy
□ Normal
□ Abnormal □ Suggestive to CIN
□ Suggestive to invasion
□ Unsatisfactory

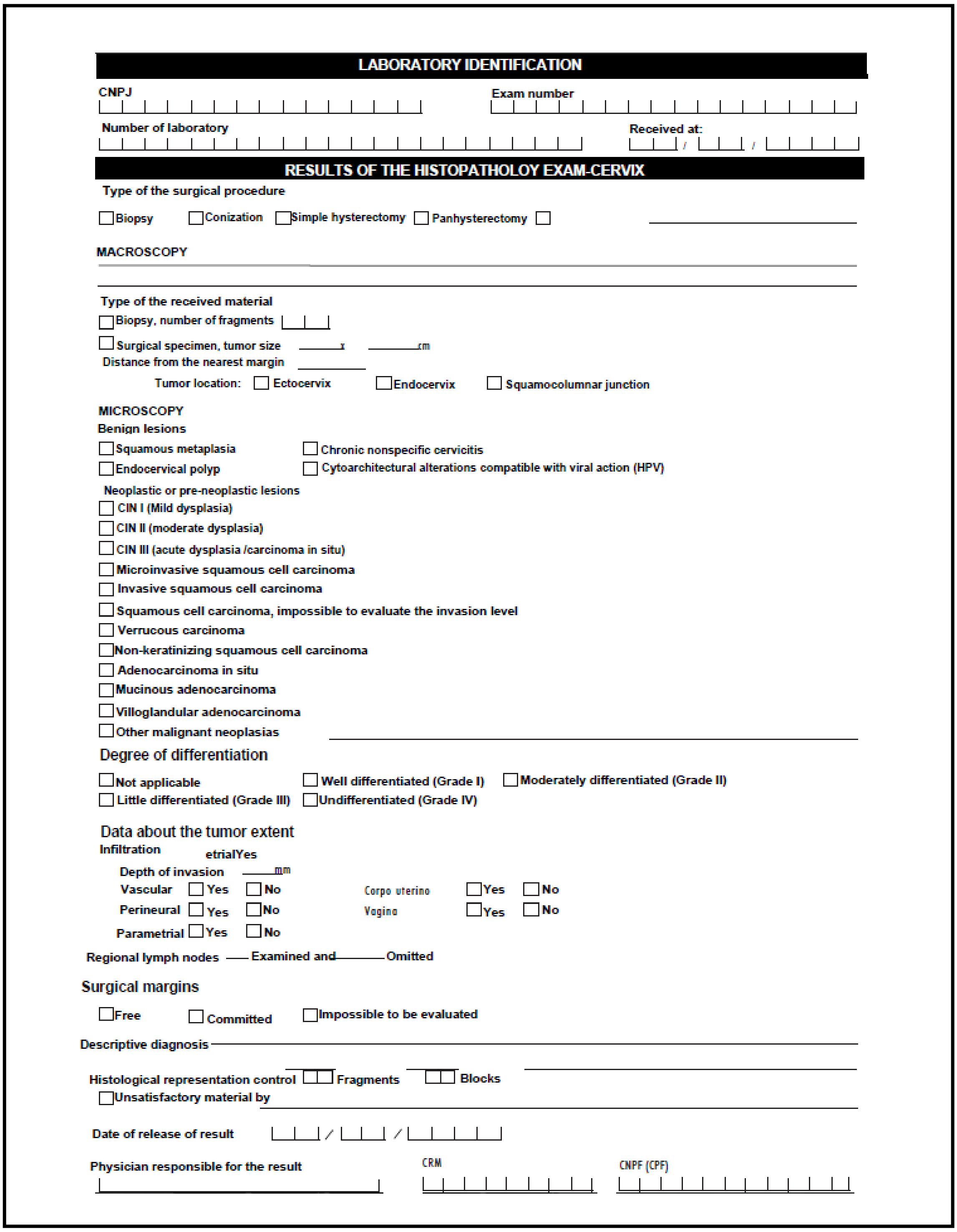
2.Procedure
□ Cold Biopsy
□ Endocervical Curettage
□ High-Frequen Surgery (CAF) □ Extension of the transformation zone
□ Channel removal
□ Biopsy

Additional information for the pathologist

Date of the test / /

Responsible physician

A 2.2: Translated form



LABORATORY IDENTIFICATION

CNPJ

Exam number

Number of laboratory

Received at: / /

RESULTS OF THE HISTOPATHOLOY EXAM-CERVIX

Type of the surgical procedure

□Biopsy □Conization □Simple hysterectomy □ Panhysterectomy □

MACROSCOPY

Type of the received material

□Biopsy, number of fragments

□Surgical specimen, tumor size x cm

Distance from the nearest margin

Tumor location: □ Ectocervix □Endocervix □ Squamocolumnar junction

MICROSCOPY

Benign lesions

□Squamous metaplasia □Chronic nonspecific cervicitis

□Endocervical polyp □ Cytoarchitectural alterations compatible with viral action (HPV)

Neoplastic or pre-neoplastic lesions

□ CIN I (Mild dysplasia)

□CIN II (moderate dysplasia)

□CIN III (acute dysplasia /carcinoma in situ)

□Microinvasive squamous cell carcinoma

□ Invasive squamous cell carcinoma

□Squamous cell carcinoma, impossible to evaluate the invasion level

□ Verrucous carcinoma

□Non-keratinizing squamous cell carcinoma

□ Adenocarcinoma in situ

□Mucinous adenocarcinoma

□Villoglandular adenocarcinoma

□Other malignant neoplasias

Degree of differentiation

□Not applicable □ Well differentiated (Grade I) □Moderately differentiated (Grade II)

□ Little differentiated (Grade III) □Undifferentiated (Grade IV)

Data about the tumor extent

Infiltration etrialYes

Depth of invasion mm

Vascular □ Yes □ No Corpo uterino □Yes □No

Perineural □ Yes □No Vagina □Yes □No

Parametrial □Yes □ No

Regional lymph nodes Examined and Omitted

Surgical margins

□Free □ Committed □Impossible to be evaluated

Descriptive diagnosis

Histological representation control □□ Fragments □□ Blocks

□Unsatisfactory material by

Date of release of result / /

Physician responsible for the result

CRM

CNPF (CPF)

APPENDIX 1: List of performance indicators

Each indicator is obtained by the following formula:

Indicator (%) = (Numerator /Denominator)*100, Ratio = (Numerator /Denominator)

A. Adequacy of Pap smear specimens

No	Indicator	Numerator	Denominator
1	Percentage of rejected slides	Number of rejected exams	Total number of tests
2	Percentage of unsatisfactory specimens	Number of tests with adequacy "un satisfactory"	Total number of tests
3	Percentage of satisfactory specimens	Number of tests with adequacy "satisfactory"	Total number of tests
4	Percentage of specimens with presence of TZ cells	Number of tests with presence of TZ cells	Total number of performed tests

*Performed tests represent all tests with adequacy of specimens “satisfactory”

B. Cytopathology test results

No	Indicator	Numerator	Denominator
1	Percentage of tests without changes	Number of tests with results " without changes	Total number of performed tests
2	Percentage of tests with low risk atypical cells	Number of tests with result ”ASC-US and LSIL“	Total number of performed tests
3	Percentage of tests with high risk atypical cells	Number of tests with result “ASC_H, AGC, AOI, HSIL, HSIL- cannot exclude micro invasion, AIS, invasive adenocarcinoma and cancer”	Total number of performed tests
4	Percentage of tests with cancer	Number of tests with result “cancer”	Total number of performed tests
5	Percentage of tests within normal limits	Number of tests with result “within normal limits”	Total number of performed tests
6	Percentage of tests with reactivate or reparative benign changes	Number of tests with result “reactivate or reparative benign changes”	Total number of performed tests
7	Percentage of tests with ASC-US	Number of tests with result “ASC-US”	Total number of performed tests
8	Percentage of tests with ASC-H	Number of tests with result “ASC-H”	Total number of performed tests

9	Percentage of tests with AGC-Possibly non-neoplastic	Number of tests with result “AGC-Possibly non-neoplastic”	Total number of performed tests
10	Percentage of tests with AGC- H	Number of tests with result “AGC-H”	Total number of performed tests
11	Percentage of tests with AOI- non-neoplastic	Number of tests with result “AOI- non-neoplastic”	Total number of performed tests
12	Percentage of tests with AOI-H	Number of tests with result “AOI-H”	Total number of performed tests
13	Percentage of tests with LSIL	Number of tests with result “LSIL”	Total number of performed tests
14	Percentage of tests with HSIL, including HSIL	Number of tests with result “HSIL”	Total number of performed tests
15	Percentage of tests with High-grade Intraepithelial lesions, cannot exclude micro-invasion	Number of tests with result “high-grade Intraepithelial lesions, cannot exclude micro-invasion”	Total number of performed tests
16	Percentage of tests with Squamous cell carcinoma	Number of tests with result “squamous cell carcinoma”	Total number of performed tests
17	Percentage of tests with Adenocarcinoma in situ	Number of tests with result “adenocarcinoma in situ”	Total number of performed tests
18	Percentage of tests with Invasive cervical adenocarcinoma	Number of tests with result “invasive cervical adenocarcinoma”	Total number of performed tests
19	Percentage of tests with endometrial adenocarcinoma	Number of tests with result “endometrial adenocarcinoma”	Total number of performed tests
20	Percentage of tests with Invasive adenocarcinoma/ without further specifications	Number of tests with result “invasive adenocarcinoma/ without further specifications”	Total number of performed tests
21	Ratio between low grade lesions and high grade lesions	Number of tests with result “LSIL”	Number of tests with result “HSIL”
22	Ratio between high grade lesions and cancer	Number of tests with result “ HSIL ”	Number of tests with result “ squamous cell carcinoma”
23	Percentage of altered tests (Positive index)	Number of tests with ASC-US; ASC-H; AOI, AGC, LSIL; HSIL; HSIL cannot exclude micro invasion cancer; AIS and invasive adenocarcinoma;	Total number of performed tests

24	Percentage of atypical squamous cell (ASC) among the satisfactory tests	Number of tests with ASC-US and ASC-H	Total number of performed tests
25	Percentage of atypical squamous cell (ASC) among the altered tests	Number of tests with ASC-US and ASC-H	Total number of altered tests
26	Ratio for atypical squamous cells/squamous intraepithelial lesion	Number of tests with ASC-US and ASC-H	Number of tests with LSIL and HSIL

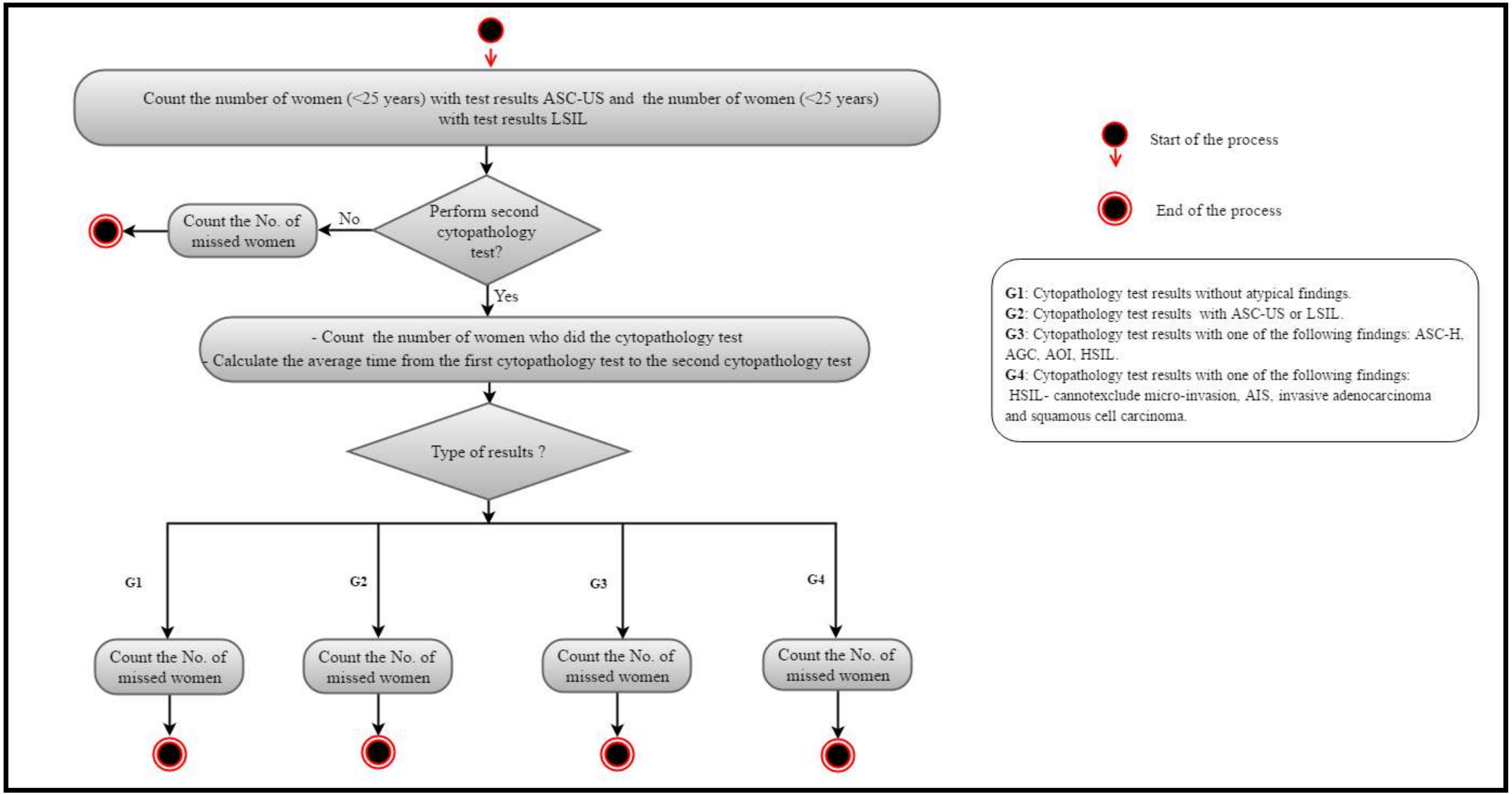
C. Histopathology test results

No	Indicator	Numerator	Denominator
1	Percentage of histological tests with benign lesions	histological tests with result “Benign lesions”	Total number of performed tests
2	Percentage of histological tests with neoplastic or pre-neoplastic lesions	histological tests with result “neoplastic or pre-neoplastic lesions”	Total number of performed tests
3	Percentage of histological tests with CIN I (mild dysplasia)	histological tests with result “CIN1”	Total number of performed tests
4	Percentage of histological tests with CIN II (moderate dysplasia)	histological tests with result “CIN II”	Total number of performed tests
5	Percentage of histological tests with CIN III (severe dysplasia /carcinoma *in situ*)	histological tests with result “CIN III”	Total number of performed tests
6	Percentage of histological tests with carcinomas	histological tests with result “carcinoma”	Total number of performed tests
7	Percentage of histological tests with adenocarcinomas	histological tests with result “adenocarcinomas”	Total number of performed tests

APPENDIX 2

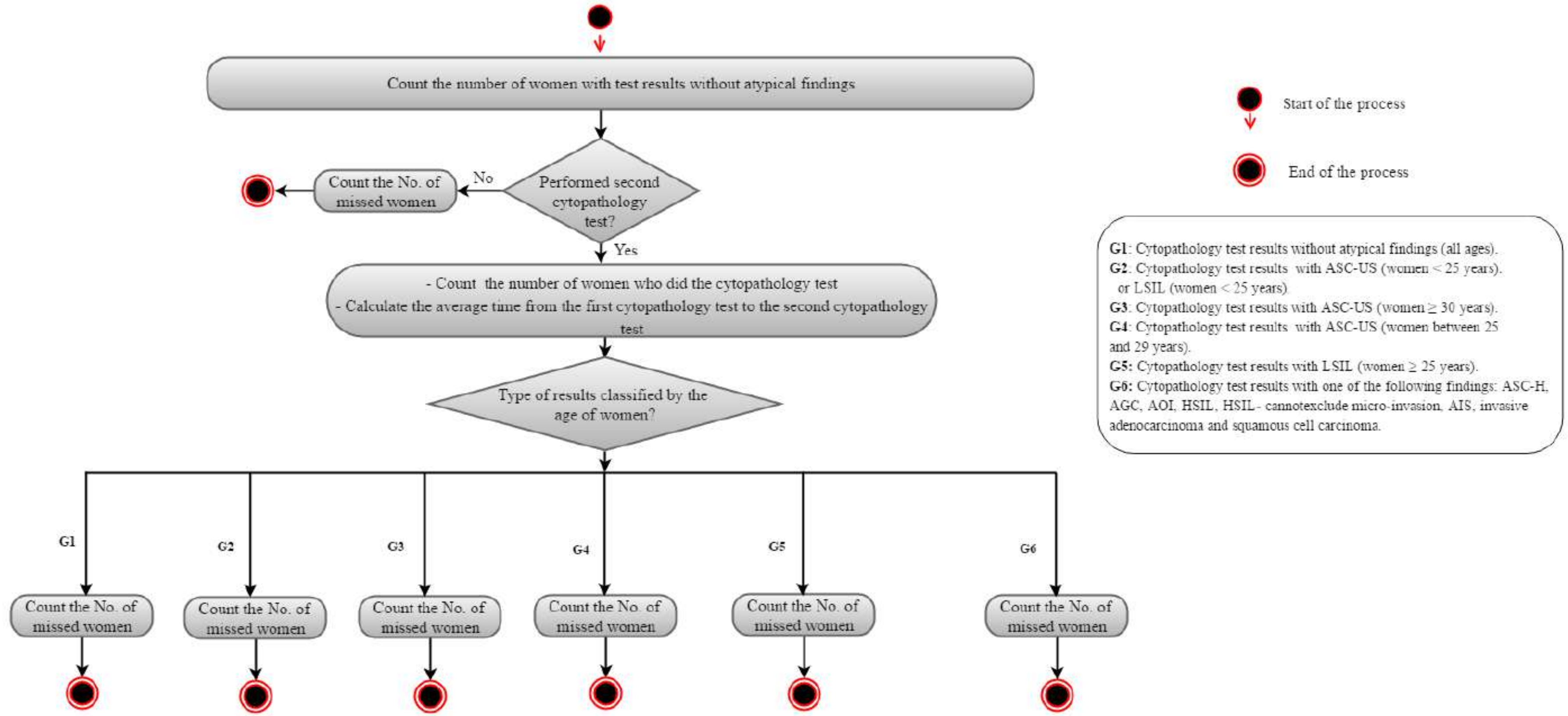
The simplified flow diagrams for the follow-up of the groups of women according to their age and test results

A 2.1: The simplified flow diagram for the recommended clinical approach for the women (<25 years) with diagnosis of ASC-US and women (<25 years) with initials diagnosis of LSIL.


Count the number of women (<25 years) with test results ASC-US and the number of women (<25 years) with test results LSIL
Perform second cytopathology test?
No
Count the No. of missed women
Yes
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopathology test to the second cytopathology test
Type of results ?
G1
G2
G3
G4
Count the No. of missed women
Count the No. of missed women
Count the No. of missed women
Count the No. of missed women
Start of the process
End of the process

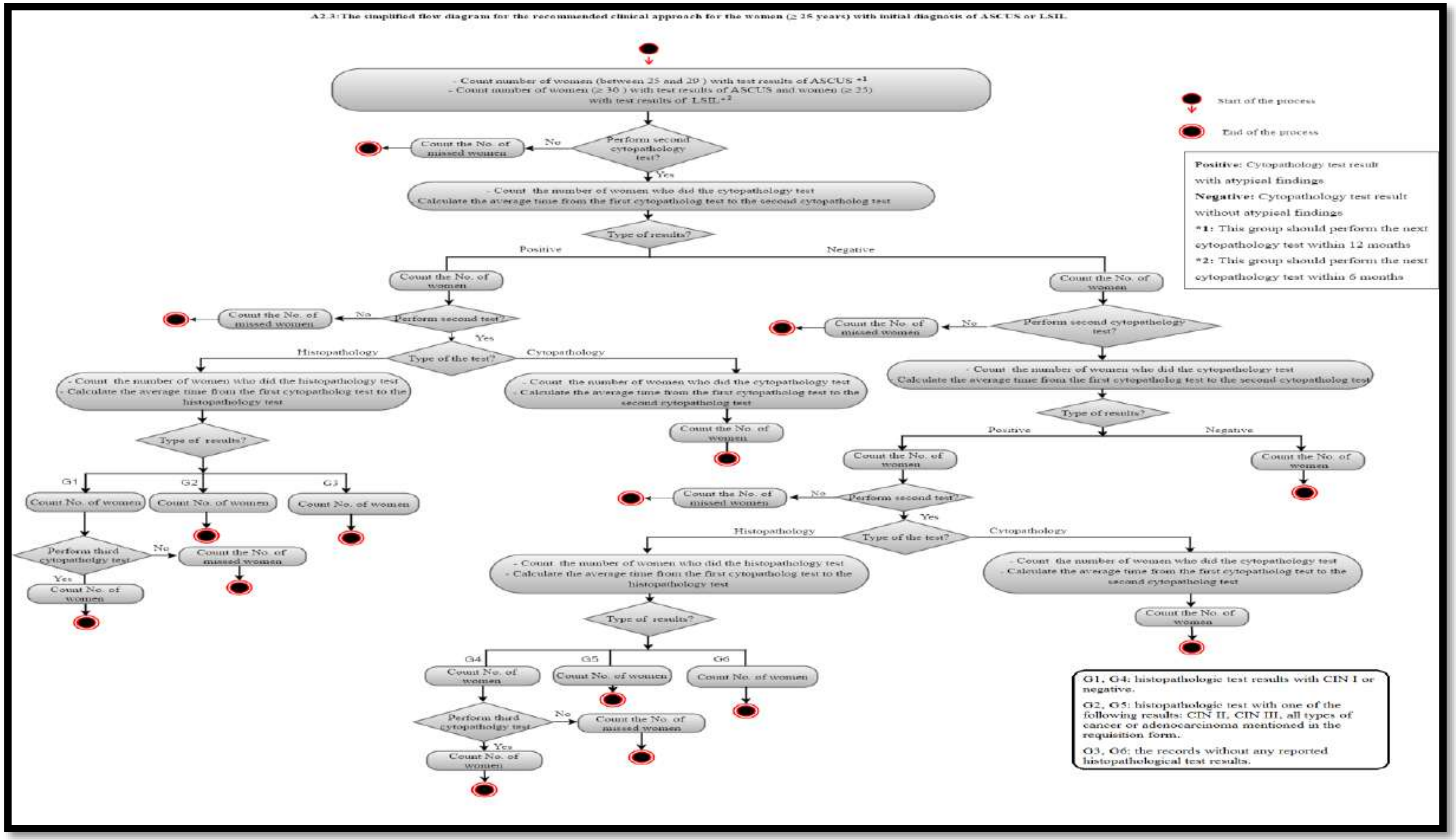
G1: Cytopathology test results without atypical findings.
G2: Cytopathology test results with ASC-US or LSIL.
G3: Cytopathology test results with one of the following findings: ASC-H, AGC, AOI, HSIL.
G4: Cytopathology test results with one of the following findings: HSIL- cannotexclude micro-invasion, AIS, invasive adenocarcinoma and squamous cell carcinoma.

A 2.2: The simplified flow diagram for the recommended clinical approach for the women with initial diagnosis without atypical findings.


Count the number of women with test results without atypical findings
Performed second cytopathology test?
No
Count the No. of missed women
Yes
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopathology test to the second cytopathology test
Type of results classified by the age of women?
G1
G2
G3
G4
G5
G6
Count the No. of missed women
Count the No. of missed women
Count the No. of missed women
Count the No. of missed women
Count the No. of missed women
Count the No. of missed women
Start of the process
End of the process

G1: Cytopathology test results without atypical findings (all ages).
G2. Cytopathology test results with ASC-US (women < 25 years).
or LSIL (women < 25 years)
G3: Cytopathology test results with ASC-US (women ≥ 30 years).
G4: Cytopathology test results with ASC-US (women between 25 and 29 years).
G5: Cytopathology test results with LSIL (women ≥ 25 years).
G6: Cytopathology test results with one of the following findings: ASC-H, AGC, AOI, HSIL, HSIL - cannotexclude micro-invasion, AIS, invasive adenocarcinoma and squamous cell carcinoma.

A 2.3: The simplified flow diagram for the recommended clinical approach for the women (≥ 25years) with initial diagnosis of ASC-US or LSIL.



A2.3:The simplified flow diagram for the recommended clinical approach for the women (≥ 25 years) with initial diagnosis of ASCUS or LSIL.

- Count number of women (between 25 and 29) with test results of ASCUS *1
- Count number of women (≥ 30) with test results of ASCUS and women (≥ 25) with test results of LSIL *2
Perform second cytopathology test?
No
Count the No. of missed women
Yes
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
Positive
Negative
Count the No. of women
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Count the No. of women
Type of results?
G1
G2
G3
Count No. of women
Count No. of women
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
Count the No. of women
Perform second cytopathology test?
No
Count the No. of missed women
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
Positive
Negative
Count the No. of women
Count the No. of women
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Count the No. of women
Type of results?
G4
G5
G6
Count No. of women
Count No. of women
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
Start of the process
End of the process

Positive: Cytopathology test result with atypical findings

Negative: Cytopathology test result without atypical findings

***1:** This group should perform the next cytopathology test within 12 months

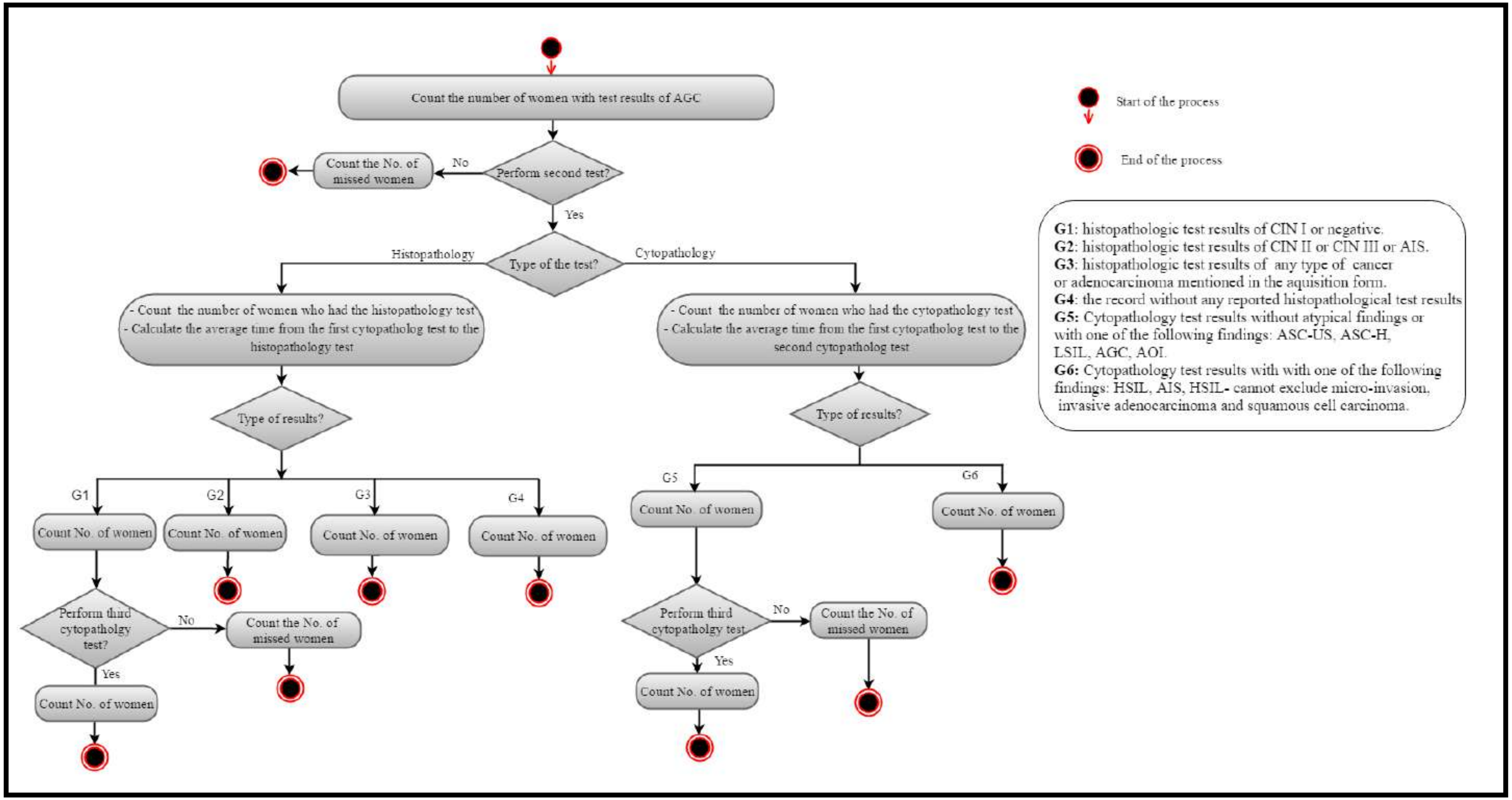
***2:** This group should perform the next cytopathology test within 6 months

G1, G4: histopathologic test results with CIN I or negative.

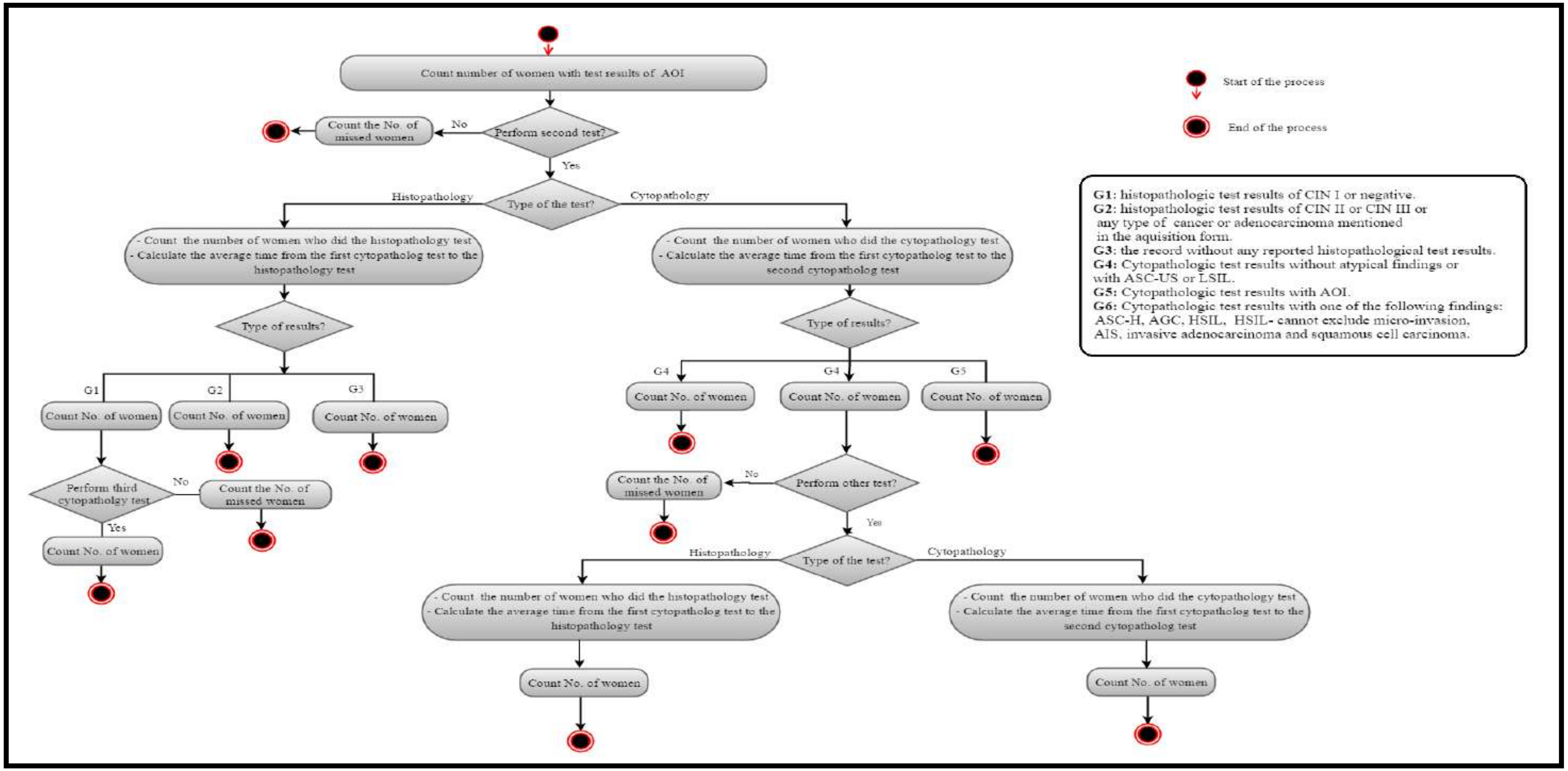
G2, G5: histopathologic test with one of the following results: CIN II, CIN III, all types of cancer or adenocarcinoma mentioned in the requisition form.

G3, G6: the records without any reported histopathological test results.

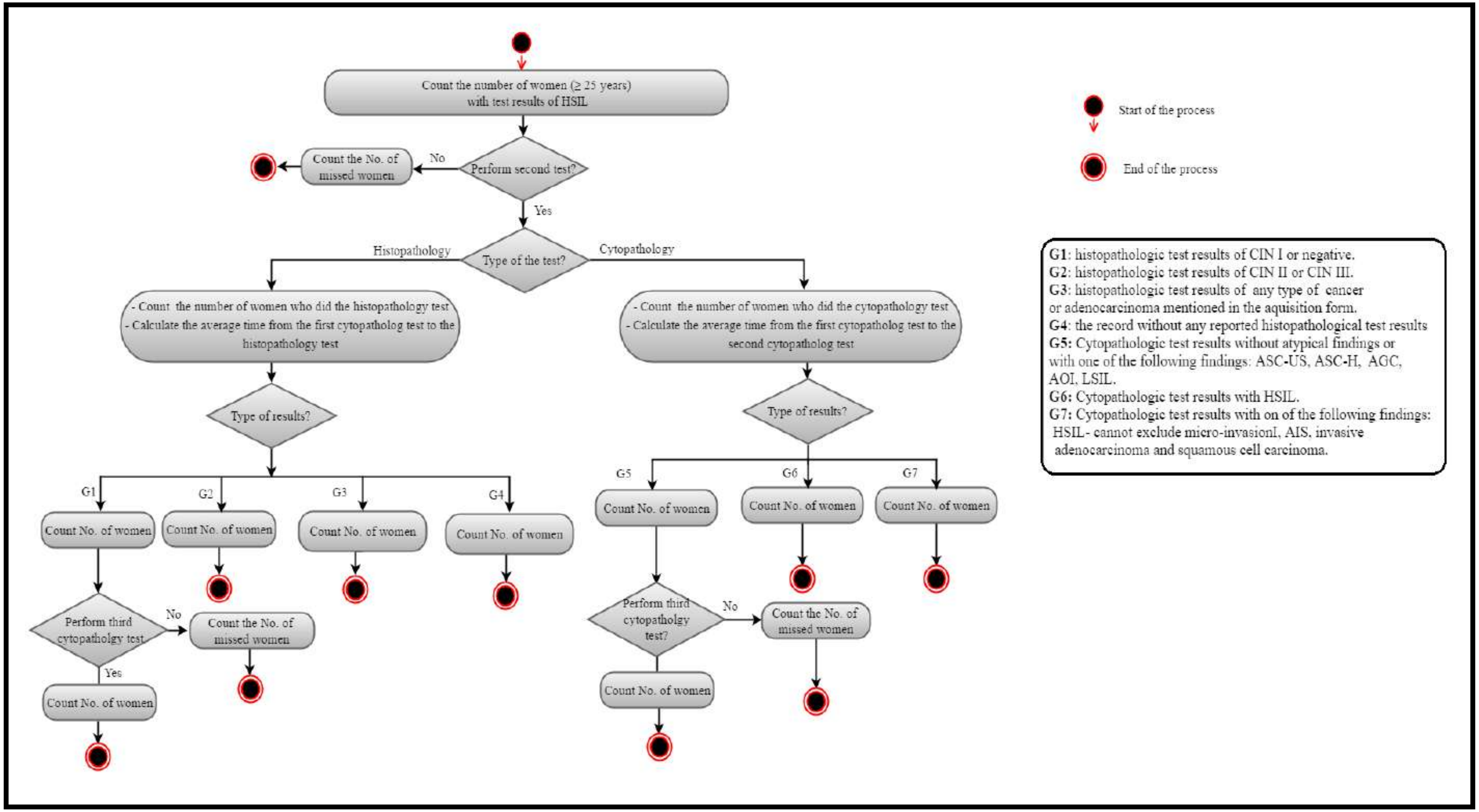
A 2.4: The simplified flow diagram for the recommended clinical approach for the women with initial diagnosis of AGC.


Count the number of women with test results of AGC
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who had the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who had the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
Type of results?
G1
G2
G3
G4
G5
G6
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Perform third cytopatholgy test?
No
Count the No. of missed women
Yes
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
Start of the process
End of the process
G1: histopathologic test results of CIN I or negative.
G2: histopathologic test results of CIN II or CIN III or AIS.
G3: histopathologic test results of any type of cancer or adenocarcinoma mentioned in the aquisition form.
G4: the record without any reported histopathological test results
G5: Cytopathology test results without atypical findings or with one of the following findings: ASC-US, ASC-H, LSIL, AGC, AOI.
G6: Cytopathology test results with with one of the following findings: HSIL, AIS, HSIL- cannot exclude micro-invasion, invasive adenocarcinoma and squamous cell carcinoma.

A 2.5: The simplified flow diagram for the recommended clinical approach for the women with initial diagnosis of AOI.

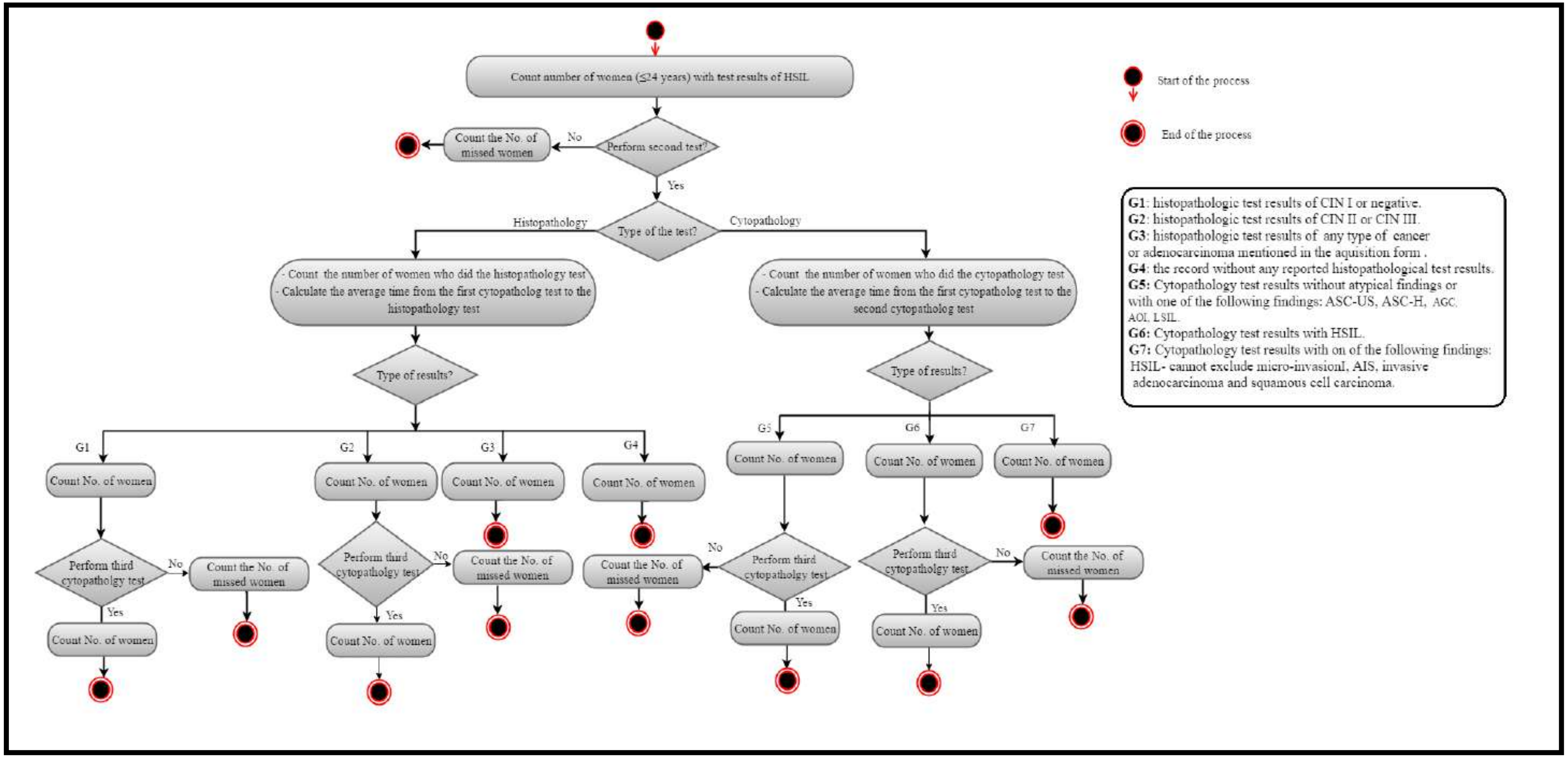

Start of the process
End of the process
Count number of women with test results of AOI
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
G1
G2
G3
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
Type of results?
G4
G4
G5
Count No. of women
Perform other test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Count No. of women
Count No. of women
G1: histopathologic test results of CIN I or negative.
G2: histopathologic test results of CIN II or CIN III or any type of cancer or adenocarcinoma mentioned in the aquisition form.
G3: the record without any reported histopathological test results.
G4: Cytopathologic test results without atypical findings or with ASC-US or LSIL.
G5: Cytopathologic test results with AOI.
G6: Cytopathologic test results with one of the following findings: ASC-H, AGC, HSIL, HSIL- cannot exclude micro-invasion, AIS, invasive adenocarcinoma and squamous cell carcinoma.

A 2.6: The simplified flow diagram for the recommended clinical approach for the women (≥25 years) with initial diagnosis of HSIL.


Count the number of women (≥ 25 years) with test results of HSIL
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
Type of results?
G1
G2
G3
G4
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Perform third cytopatholgy test
No
Count the No. of missed women
Yes
Count No. of women
G5
G6
G7
Count No. of women
Count No. of women
Count No. of women
Perform third cytopatholgy test?
No
Count the No. of missed women
Count No. of women
Start of the process
End of the process

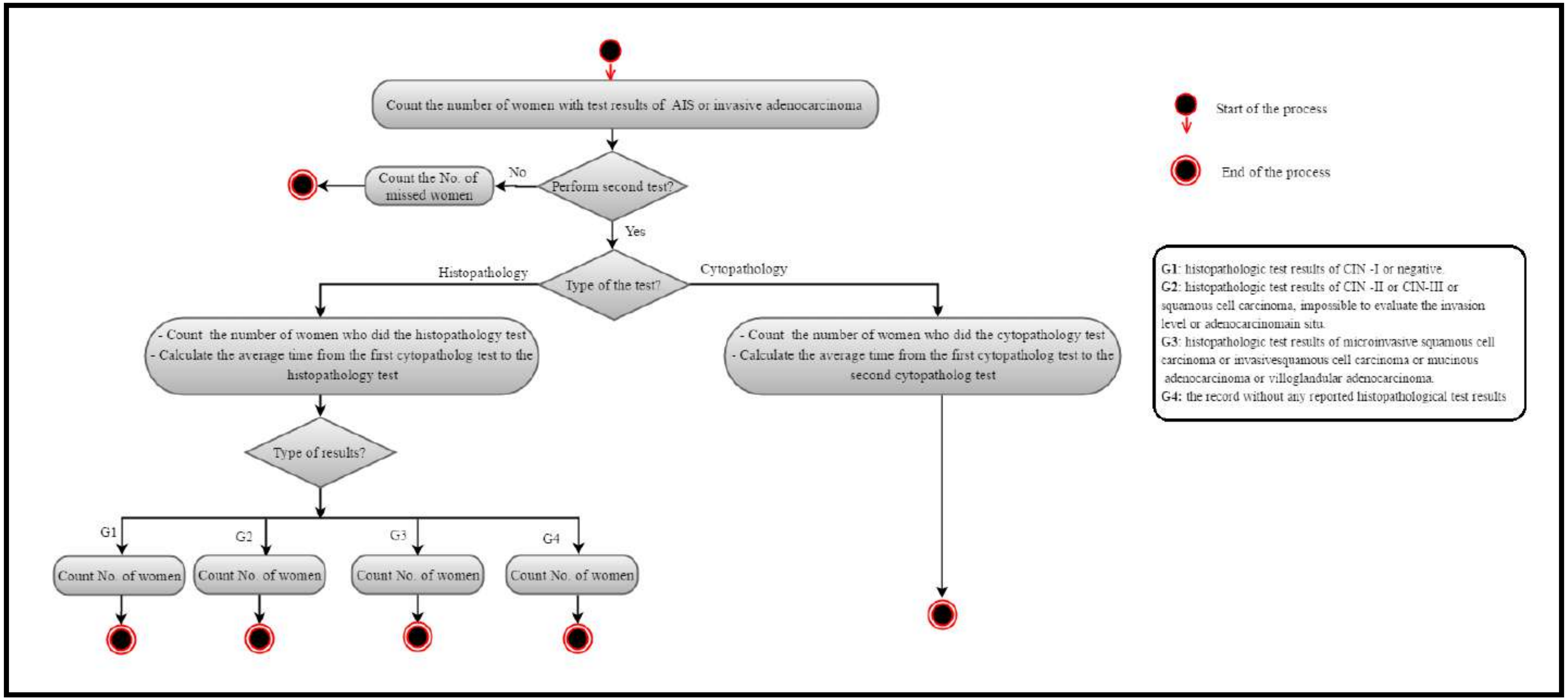
G1: histopathologic test results of CIN I or negative.
G2: histopathologic test results of CIN II or CIN III.
G3: histopathologic test results of any type of cancer or adenocarcinoma mentioned in the aquisition form.
G4: the record without any reported histopathological test results
G5: Cytopathologic test results without atypical findings or with one of the following findings: ASC-US, ASC-H, AGC, AOI, LSIL.
G6: Cytopathologic test results with HSIL.
G7: Cytopathologic test results with on of the following findings: HSIL- cannot exclude micro-invasionI, AIS, invasive adenocarcinoma and squamous cell carcinoma.

A 2.7: The simplified flow diagram for the recommended clinical approach for the women (≤24 years) with initial diagnosis of HSIL.

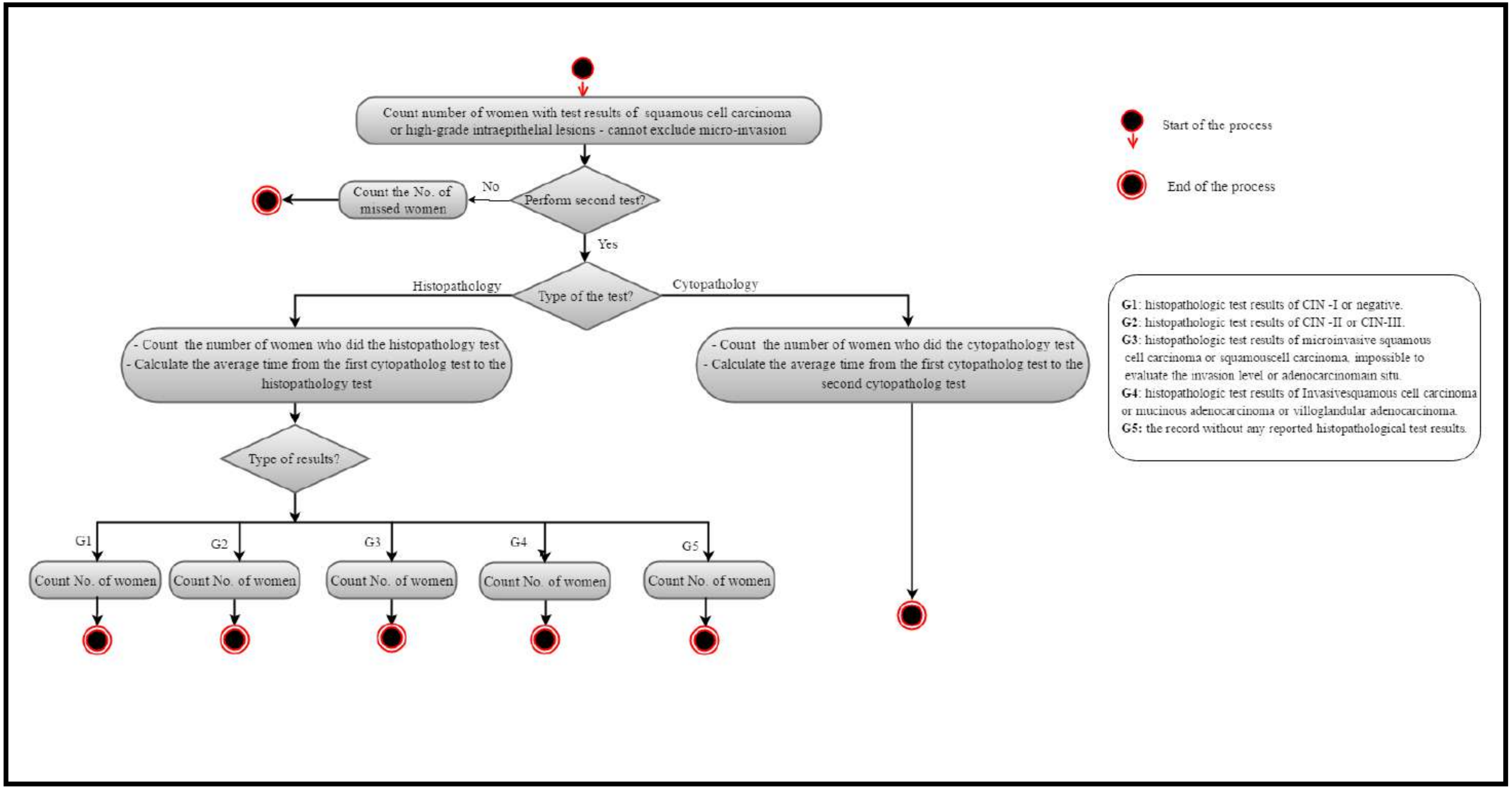

Count number of women (≤24 years) with test results of HSIL
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
Type of results?
G1
G2
G3
G4
G5
G6
G7
Count No. of women
Perform third cytopatholgy test
No
Yes
Count the No. of missed women
Start of the process
End of the process

G1: histopathologic test results of CIN I or negative.
G2: histopathologic test results of CIN II or CIN III.
G3: histopathologic test results of any type of cancer or adenocarcinoma mentioned in the aquisition form .
G4: the record without any reported histopathological test results.
G5: Cytopathology test results without atypical findings or with one of the following findings: ASC-US, ASC-H, AGC, AOI, LSIL.
G6: Cytopathology test results with HSIL.
G7: Cytopathology test results with on of the following findings: HSIL- cannot exclude micro-invasionl, AIS, invasive adenocarcinoma and squamous cell carcinoma.

A 2.8: The simplified flow diagram for the recommended clinical approach for the women with initial diagnosis of AIS or invasive adenocarcinoma.


Count the number of women with test results of AIS or invasive adenocarcinoma
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
G1
G2
G3
G4
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Start of the process
End of the process
G1: histopathologic test results of CIN -I or negative.
G2: histopathologic test results of CIN -II or CIN-III or squamous cell carcinoma, impossible to evaluate the invasion level or adenocarcinomain situ.
G3: histopathologic test results of microinvasive squamous cell carcinoma or invasivesquamous cell carcinoma or mucinous adenocarcinoma or villoglandular adenocarcinoma.
G4: the record without any reported histopathological test results

A 2.9: The simplified flow diagram for the recommended clinical approach for the women with initial diagnosis of squamous cell carcinoma or high-grade intraepithelial lesions - cannot exclude micro-invasion


Count number of women with test results of squamous cell carcinoma or high-grade intraepithelial lesions - cannot exclude micro-invasion
Perform second test?
No
Count the No. of missed women
Yes
Type of the test?
Histopathology
Cytopathology
- Count the number of women who did the histopathology test
- Calculate the average time from the first cytopatholog test to the histopathology test
- Count the number of women who did the cytopathology test
- Calculate the average time from the first cytopatholog test to the second cytopatholog test
Type of results?
G1
G2
G3
G4
G5
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Count No. of women
Start of the process
End of the process
G1: histopathologic test results of CIN -I or negative.
G2: histopathologic test results of CIN -II or CIN-III.
G3: histopathologic test results of microinvasive squamous cell carcinoma or squamouscell carcinoma, impossible to evaluate the invasion level or adenocarcinomain situ.
G4: histopathologic test results of Invasivesquamous cell carcinoma or mucinous adenocarcinoma or villoglandular adenocarcinoma.
G5: the record without any reported histopathological test results.

APPENDIX 3: The algorithms for the follow-up of the group of women with ASC-H initial results

This appendix contains the algorithm for the follow-up of the group of women with ASC-H initial results. This algorithm was written according to the illustrated flow diagram in the Figure 13. Below is the description of the used symbols in indicators representation in the algorithm:

N: number of women.

C: cytopathology test.

H: histopathology test.

T: average time interval between the release of the results of the former test and collection of the specimens of the latter test.

L: number of the lost women between the tests.

As an example, the indicator T1C2C means the time interval between the first cytopathology test and second cytopathology test. The following is the details of the algorithm:

1. Selection of (uid)s of women with test results of ASC-H in the first test (variable asc_h = 1 in the table Cytology) - (**Indicator: N1C**).

2. Within the N1C (uid)s, calculate how many performed histopathology test (**Indicator: N1H**) where the time of the histology test is after the cytology test (T1H> T1C).

3. For the N1H (uid)s calculate the time interval until the first cytopathology test - (**Indicator: T1C1H**).

4. Within the (uid)s of N1H how many have negative result or CIN I (variable negativa_nic1=1 in the table Histology) - (Indicator: NNEGNIC1).

5. Within the (uid)s of NNEGNIC1, calculate how many performed second cytopathology test (**Indicator: N2CB**), where the time of this test is after the histology test (T_2CB> T_1H).

6. For the (uid)s of N2CB, calculate the time interval to the histopathology test (uids of NNEGNIC1) - (**Indicator: T1H2CB**).

7. Calculate the number of the missed women between the histopathology test (uids of NNEGNIC1) until second cytopathology test (uids of N2CB) **-** (**Indicator: L1H2CB**) where: P1H2CB = NNEGNIC1 - N2CB.

8. Within the (uid)s of N1H how many have result of CIN I or CIN II (variable nic_23=1 in the table Histology) - (**Indicator: NNIC23**).

9. Within the (uid)s of N1H how many have result of adenocarcinoma or cancer (variable aden_cancer_histo=1 in the table Histology) - (**Indicator: NADENCANCER**).

10. Within the (uid)s of N1C, calculate how many have performed second cytopathology test (Indicator: N2C), where the time of this test is after the first cytology test (T2C> T1C).

11. For the (uid)s of N2C, calculate the time interval to the first cytology test (uids of N1C) - (**Indicator: T1C2C**).

12. Within the (uid)s of N2C how many have result of ASC-H or more severe (variable f_alto_risco =1 in the table Cytology) - (**Indicator: NALTORISCO**).

13. Within the (uid)s of N2C how many have result of low risk for cancer (variable f_baixo_risco =1 in the table Cytology) - (**Indicator: NBAIXORISCO**).

14. Within the (uid)s of NBAIXORISCO, calculate how many have perform third cytopathology test (**Indicator: N3CA**).

15. For the (uid)s of N3CA, calculate the time interval to cytopathology test (uids of NBAIXORISCO) - (**Indicator: T1C2C**).

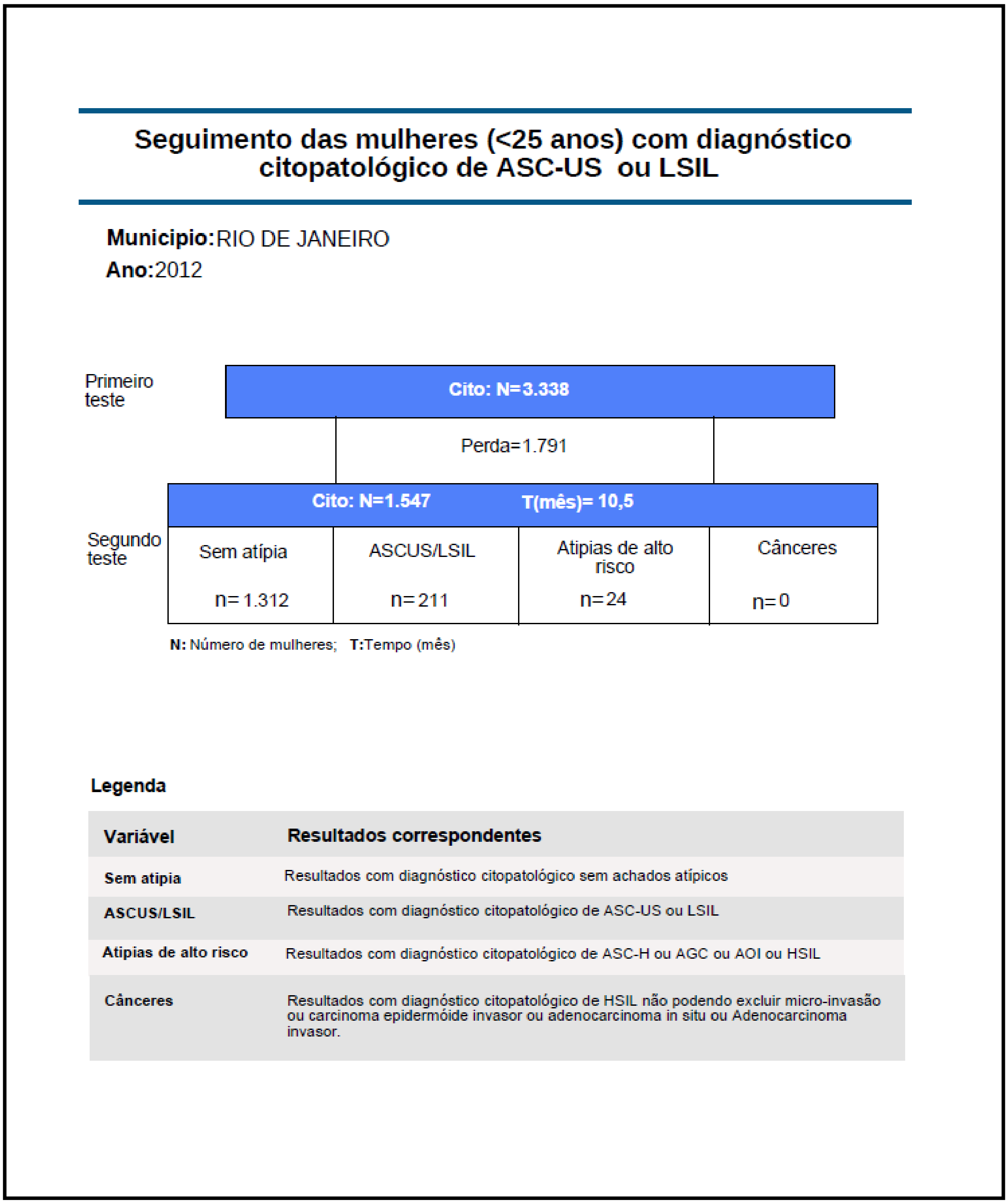
16. Calculate the loss of follow-up between the second cytopathology test (uids of NBAIXORISCO) and third cytology test (uids of N3CA) - (**Indicator: P2C3CA**), Where: L2C3CA = NBAIXORISCO - N3CA.

17. Calculate the loss of follow-up between the first cytopathology test and (second cytopathology test (uids of N2C) and first histopathology test (uids of N1H) **-** (**Indicator: L2C3CA**), Where: L1C2CIH= N1C – (N2C+ N1H).

APPENDIX 4

Output screens for additional follow-up indicators

A 4.1: Output screen for the follow-up indicators for women (<25 years) with initial cytopathological test result of ASC-US or LSIL in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres (<25 anos) com diagnóstico citopatológico de ASC-US ou LSIL

Municipio:RIO DE JANEIRO
Ano:2012

Primeiro teste

Cito: N=3.338

Perda=1.791

Segundo teste

Cito: N=1.547 T(mês)= 10,5

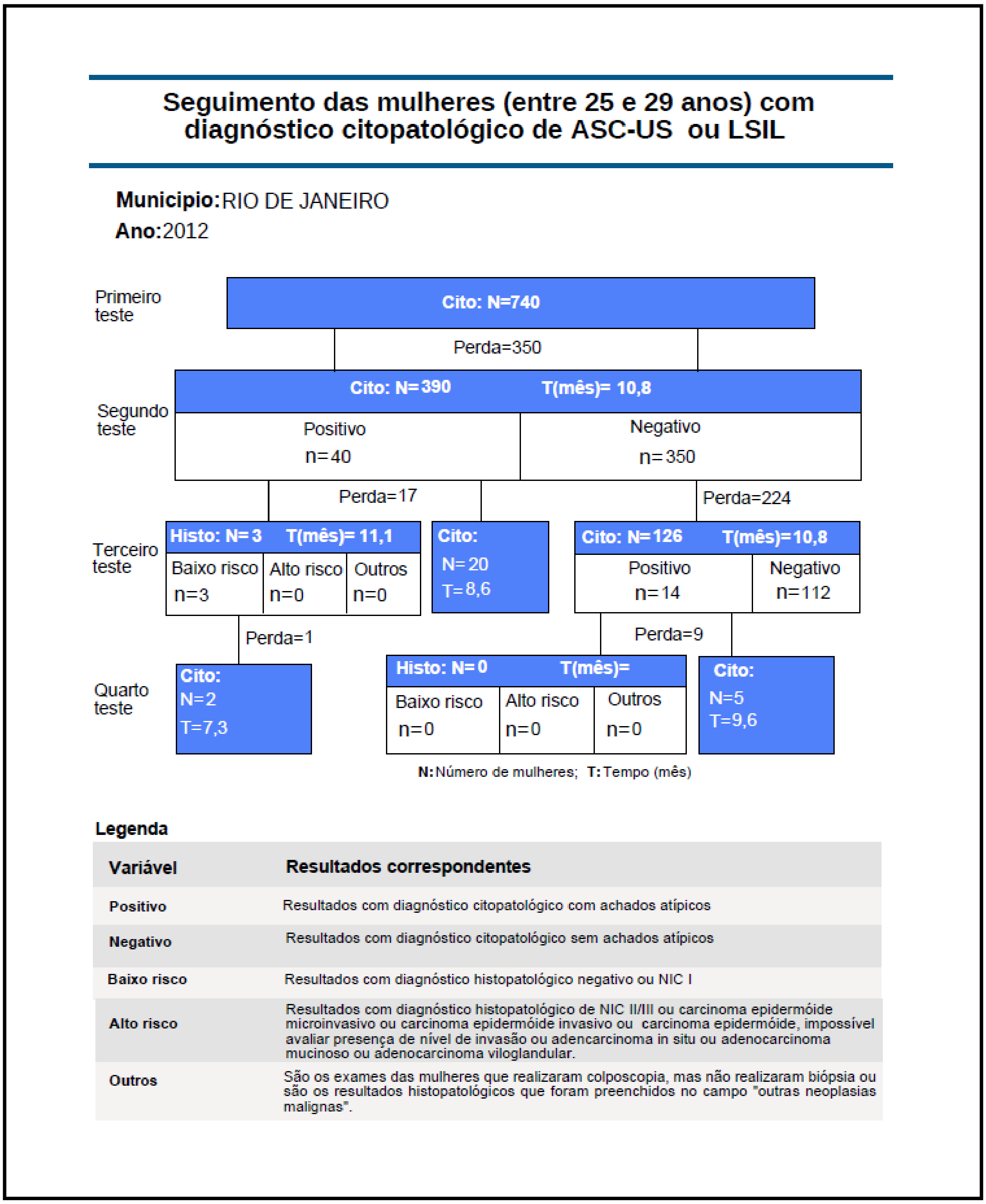
Sem atípia	ASCUS/LSIL	Atipias de alto risco	Cânceres
n= 1.312	n=211	n=24	n=0

N: Número de mulheres; **T:**Tempo (mês)

Legenda

Variável	**Resultados correspondentes**
Sem atipia	Resultados com diagnóstico citopatológico sem achados atípicos
ASCUS/LSIL	Resultados com diagnóstico citopatológico de ASC-US ou LSIL
Atipias de alto risco	Resultados com diagnóstico citopatológico de ASC-H ou AGC ou AOI ou HSIL
Cânceres	Resultados com diagnóstico citopatológico de HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou adenocarcinoma in situ ou Adenocarcinoma invasor.

A 4.2: Output screen for the follow-up indicators for women (between 25-29 years) with initial cytopathological test result of ASC-US or LSIL in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres (entre 25 e 29 anos) com diagnóstico citopatológico de ASC-US ou LSIL

Municipio:RIO DE JANEIRO

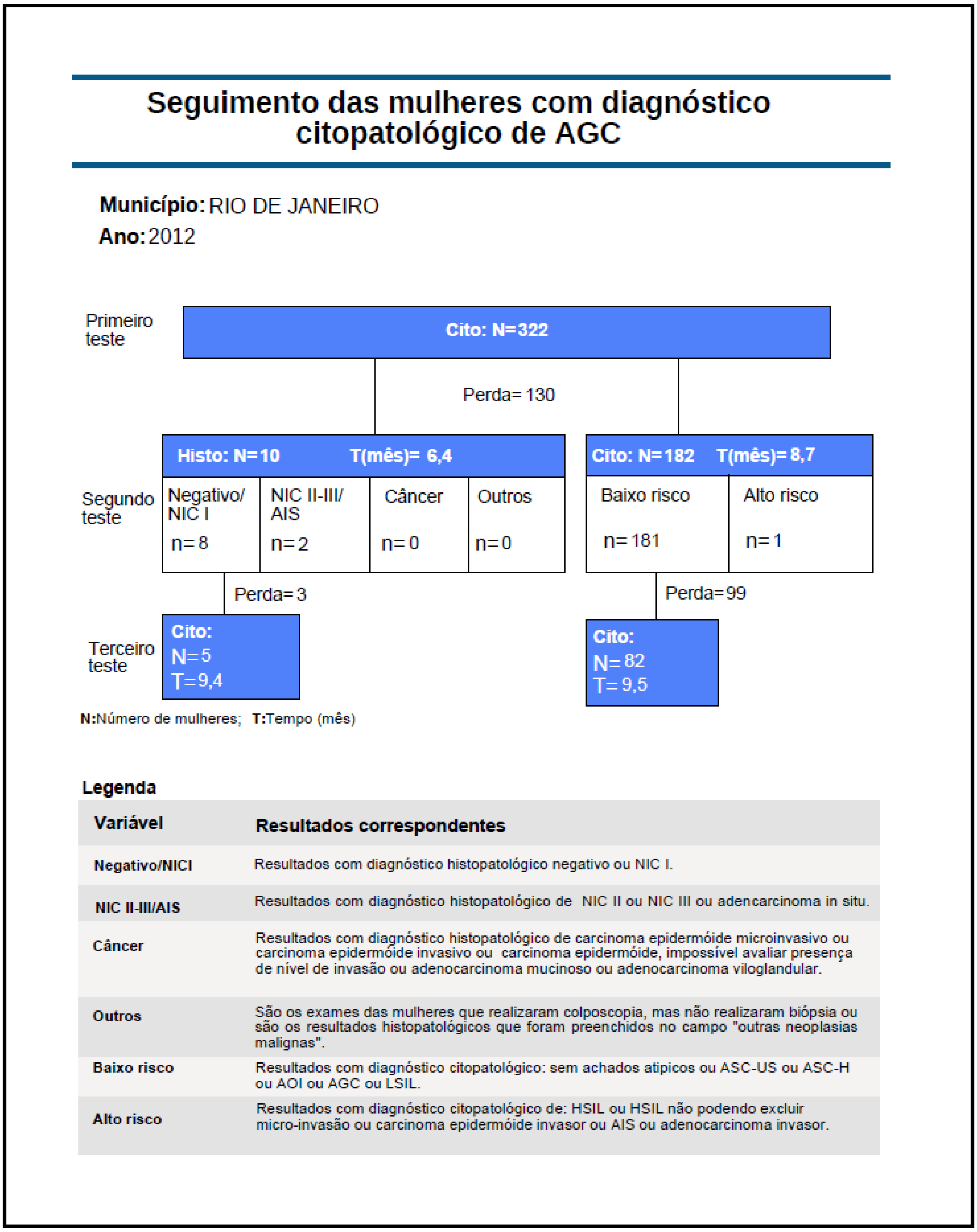
Ano:2012

N:Número de mulheres; **T:**Tempo (mês)

Legenda

Variável	Resultados correspondentes
Positivo	Resultados com diagnóstico citopatológico com achados atípicos
Negativo	Resultados com diagnóstico citopatológico sem achados atípicos
Baixo risco	Resultados com diagnóstico histopatológico negativo ou NIC I
Alto risco	Resultados com diagnóstico histopatológico de NIC II/III ou carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adencarcinoma in situ ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas".

A 4.3: Output screen for the follow-up indicators for women with initial cytopathological test result of AGC in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres com diagnóstico citopatológico de AGC

Município: RIO DE JANEIRO

Ano: 2012

Primeiro teste

Cito: N=322

Perda= 130

Segundo teste

Histo: N=10 T(mês)= 6,4

Negativo/ NIC I	NIC II-III/ AIS	Câncer	Outros
n=8	n=2	n=0	n=0

Cito: N=182 T(mês)= 8,7

Baixo risco	Alto risco
n= 181	n=1

Perda=3

Perda=99

Terceiro teste

Cito: N=5 T=9,4

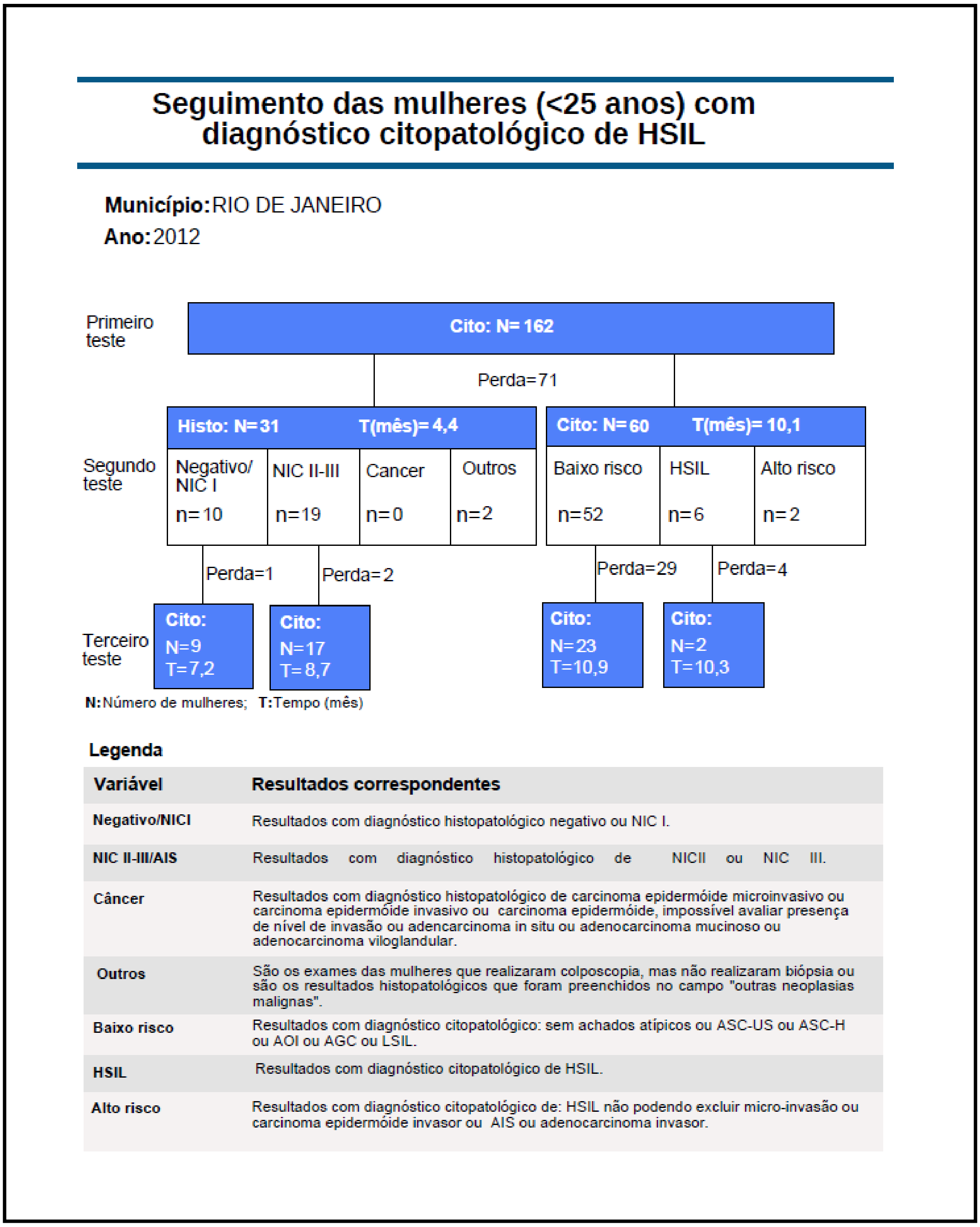
Cito: N= 82 T= 9,5

N:Número de mulheres; **T:**Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
NIC II-III/AIS	Resultados com diagnóstico histopatológico de NIC II ou NIC III ou adencarcinoma in situ.
Câncer	Resultados com diagnóstico histopatológico de carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas".
Baixo risco	Resultados com diagnóstico citopatológico: sem achados atipicos ou ASC-US ou ASC-H ou AOI ou AGC ou LSIL.
Alto risco	Resultados com diagnóstico citopatológico de: HSIL ou HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou AIS ou adenocarcinoma invasor.

A 4.4: Output screen for the follow-up indicators for women (<25 years old) with initial cytopathological test result of HSIL in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres (<25 anos) com diagnóstico citopatológico de HSIL

Município: RIO DE JANEIRO

Ano: 2012

Primeiro teste

Cito: N= 162

Perda=71

Segundo teste

Histo: N=31 T(mês)= 4,4

Negativo/ NIC I n=10

NIC II-III n=19

Cancer n=0

Outros n=2

Cito: N=60 T(mês)= 10,1

Baixo risco n=52

HSIL n=6

Alto risco n=2

Perda=1

Perda=2

Perda=29

Perda=4

Terceiro teste

Cito: N=9 T=7,2

Cito: N=17 T=8,7

Cito: N=23 T=10,9

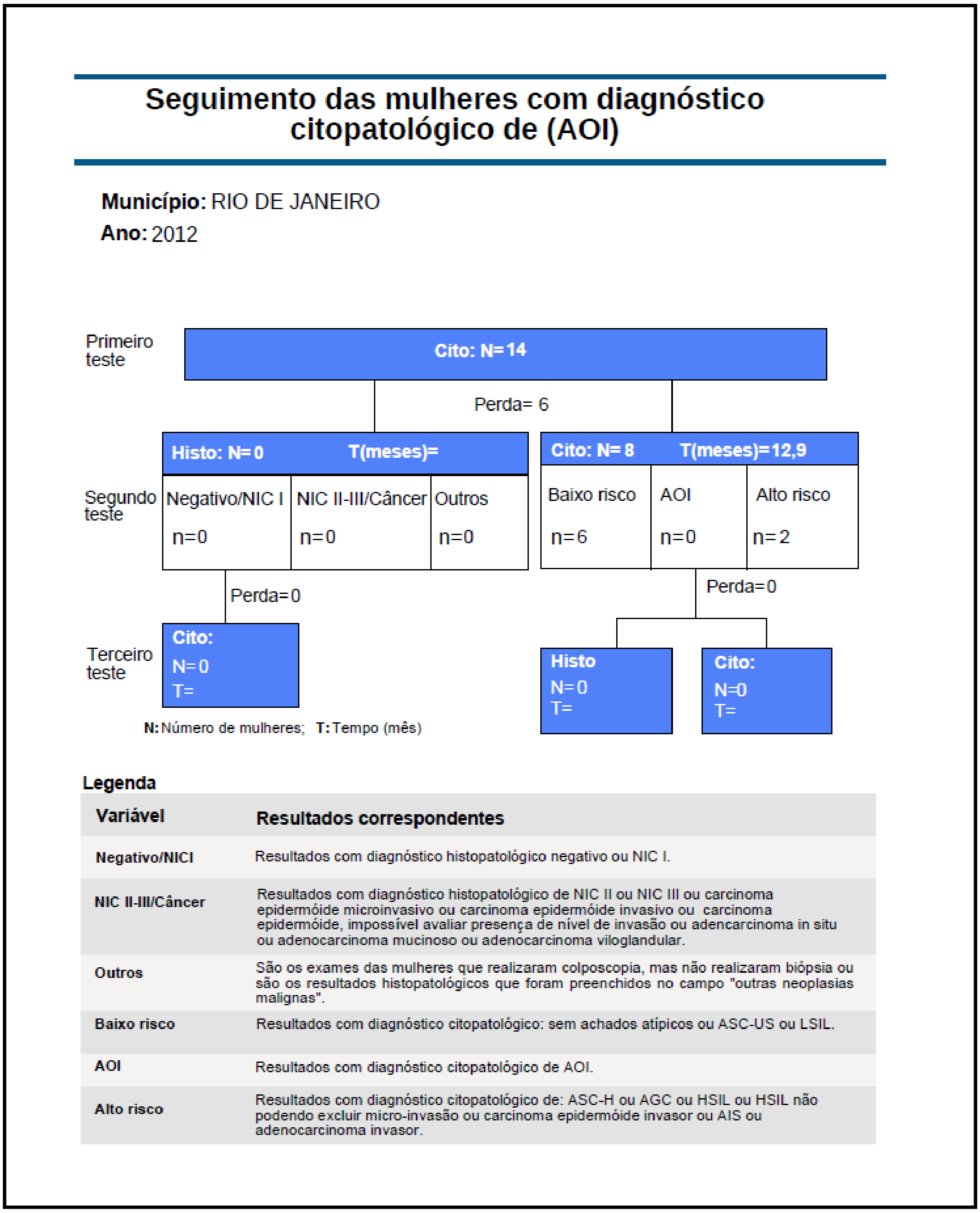
Cito: N=2 T=10,3

N: Número de mulheres; **T:** Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
NIC II-III/AIS	Resultados com diagnóstico histopatológico de NICII ou NIC III.
Câncer	Resultados com diagnóstico histopatológico de carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adencarcinoma in situ ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas".
Baixo risco	Resultados com diagnóstico citopatológico: sem achados atípicos ou ASC-US ou ASC-H ou AOI ou AGC ou LSIL.
HSIL	Resultados com diagnóstico citopatológico de HSIL.
Alto risco	Resultados com diagnóstico citopatológico de: HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou AIS ou adenocarcinoma invasor.

A 4.5: Output screen for the follow-up indicators for women with initial cytopathological test result of AOI in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres com diagnóstico citopatológico de (AOI)

Município: RIO DE JANEIRO

Ano: 2012

Primeiro teste

Cito: N=14

Perda= 6

Segundo teste

Histo: N=0 T(meses)=

Negativo/NIC I n=0

NIC II-III/Câncer n=0

Outros n=0

Cito: N=8 T(meses)=12,9

Baixo risco n=6

AOI n=0

Alto risco n=2

Perda=0

Perda=0

Terceiro teste

Cito: N=0 T=

Histo N=0 T=

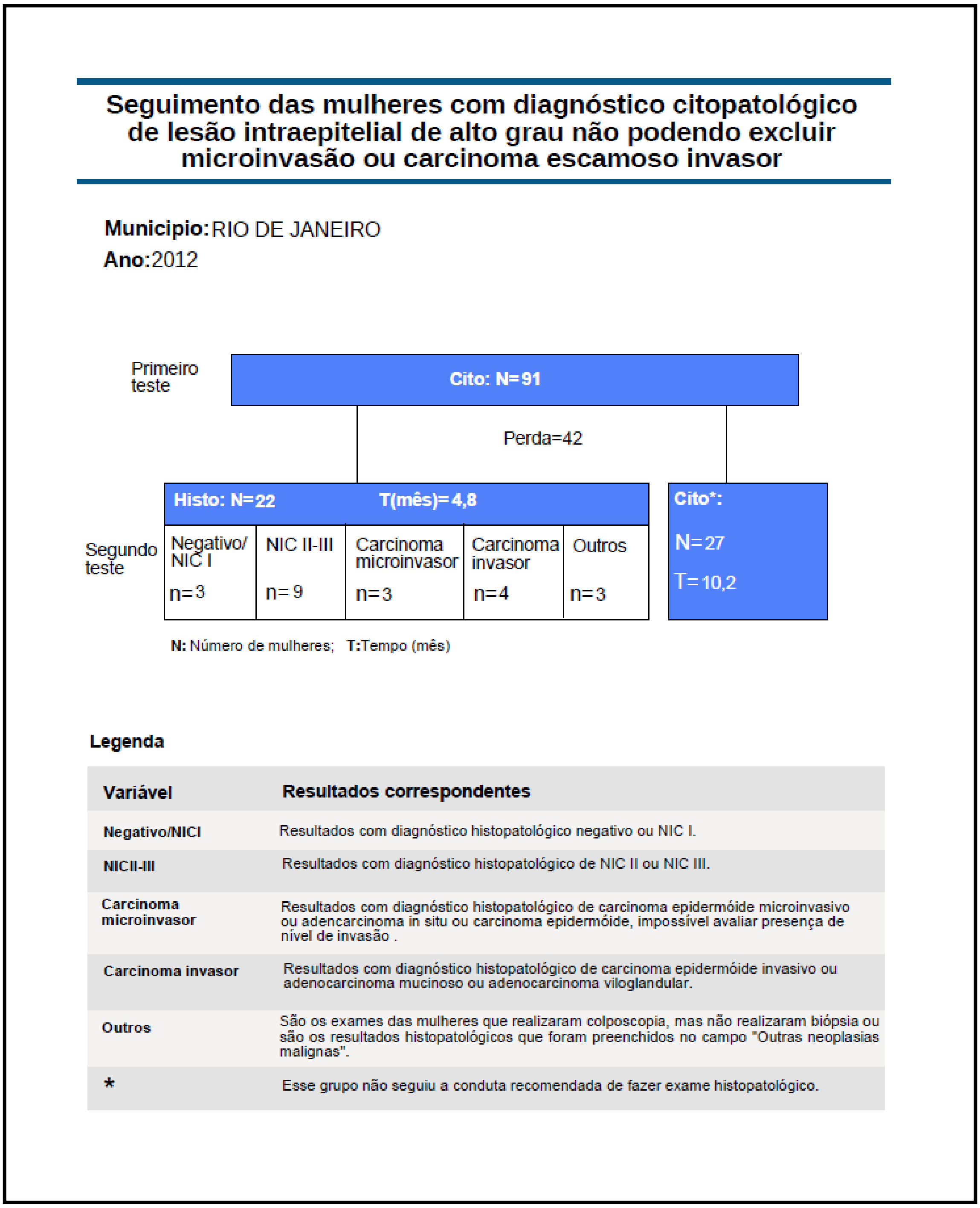
Cito: N=0 T=

N: Número de mulheres; **T:** Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
NIC II-III/Câncer	Resultados com diagnóstico histopatológico de NIC II ou NIC III ou carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou adencarcinoma in situ ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "outras neoplasias malignas".
Baixo risco	Resultados com diagnóstico citopatológico: sem achados atípicos ou ASC-US ou LSIL.
AOI	Resultados com diagnóstico citopatológico de AOI.
Alto risco	Resultados com diagnóstico citopatológico de: ASC-H ou AGC ou HSIL ou HSIL não podendo excluir micro-invasão ou carcinoma epidermóide invasor ou AIS ou adenocarcinoma invasor.

A 4.6: Output screen for the follow-up indicators for women with initial cytopathological test result of squamous cell carcinoma or high-grade intraepithelial lesions, cannot exclude micro-invasion in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres com diagnóstico citopatológico de lesão intraepitelial de alto grau não podendo excluir microinvasão ou carcinoma escamoso invasor

Municipio: RIO DE JANEIRO
Ano: 2012

Primeiro teste

Cito: N=91

Perda=42

Segundo teste

Histo: N=22 T(mês)=4,8

Negativo/ NIC I	NIC II-III	Carcinoma microinvasor	Carcinoma invasor	Outros
n=3	n=9	n=3	n=4	n=3

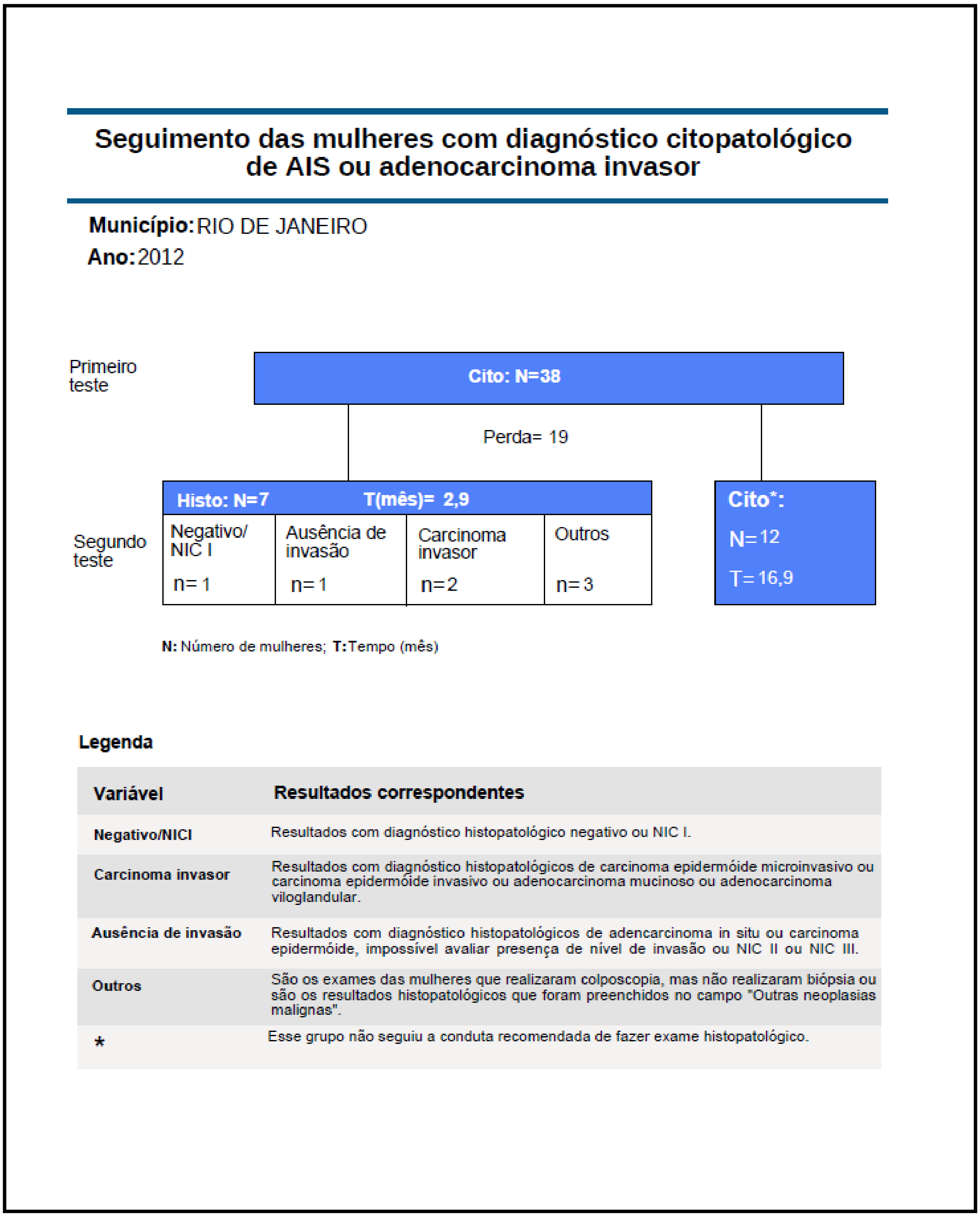
Cito*:
N=27
T=10,2

N: Número de mulheres; **T:**Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
NICII-III	Resultados com diagnóstico histopatológico de NIC II ou NIC III.
Carcinoma microinvasor	Resultados com diagnóstico histopatológico de carcinoma epidermóide microinvasivo ou adencarcinoma in situ ou carcinoma epidermóide, impossível avaliar presença de nível de invasão .
Carcinoma invasor	Resultados com diagnóstico histopatológico de carcinoma epidermóide invasivo ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "Outras neoplasias malignas".
*	Esse grupo não seguiu a conduta recomendada de fazer exame histopatológico.

A 4.7: Output screen for the follow-up indicators for women with initial cytopathological test result of adenocarcinoma in situ or invasive adenocarcinoma, in the municipality of Rio de Janeiro in year 2012.



Seguimento das mulheres com diagnóstico citopatológico de AIS ou adenocarcinoma invasor

Município: RIO DE JANEIRO

Ano: 2012

Primeiro teste

Cito: N=38

Perda= 19

Segundo teste

Histo: N=7 **T(mês)= 2,9**

Negativo/ NIC I	Ausência de invasão	Carcinoma invasor	Outros
n= 1	n= 1	n=2	n= 3

Cito*:
N= 12
T= 16,9

N: Número de mulheres; **T:** Tempo (mês)

Legenda

Variável	Resultados correspondentes
Negativo/NICI	Resultados com diagnóstico histopatológico negativo ou NIC I.
Carcinoma invasor	Resultados com diagnóstico histopatológicos de carcinoma epidermóide microinvasivo ou carcinoma epidermóide invasivo ou adenocarcinoma mucinoso ou adenocarcinoma viloglandular.
Ausência de invasão	Resultados com diagnóstico histopatológicos de adencarcinoma in situ ou carcinoma epidermóide, impossível avaliar presença de nível de invasão ou NIC II ou NIC III.
Outros	São os exames das mulheres que realizaram colposcopia, mas não realizaram biópsia ou são os resultados histopatológicos que foram preenchidos no campo "Outras neoplasias malignas".
*	Esse grupo não seguiu a conduta recomendada de fazer exame histopatológico.