

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Especialização em Business Intelligence

Ambiente de Business Intelligence para Informações
Gerenciais do Sistema de Rastreamento do Câncer do Colo do
Útero

Zina Reis Pinheiro

André Vargas Abs da Cruz



Zina Reis Pinheiro

**Ambiente de Business Intelligence para Informações
Gerenciais do Sistema de Rastreamento do Câncer do Colo
do Útero**

Monografia apresentada ao Departamento de Engenharia Elétrica da PUC/Rio como parte dos requisitos para a obtenção do título de Especialização em Business Intelligence. Aprovada pela Comissão Examinadora abaixo assinada.

André Vargas Abs da Cruz

Rio de Janeiro
29 de agosto de 2009



Dedicatória

A minha família Marcelo, Zeno e Juca por acreditarem que o sacrifício valeria a pena.

Agradecimentos

Aos colegas e amigos do INCA João Luiz Vianna, Levi Matos Marins, Gislane da Silva, Rafael Pereira e Zenas Reis da Silva pelo apoio técnico e disposição nos momentos necessários.

Ao Paulo Camanho por me ouvir.

Ao Gerson Locci por me acompanhar no aprendizado de BI.

Ao Marcus e meus pais pelo computador e suporte emocional.

Sumário

RESUMO.....	6
ABSTRACT.....	9
1. INTRODUÇÃO.....	10
1.1. Motivação	10
1.2. Objetivos do Trabalho	12
1.3. Descrição do Trabalho	12
1.4. Organização da Monografia.....	13
2. DESCRIÇÃO DO PROBLEMA.....	14
2.1. Câncer.....	14
2.1.1. Incidência e Mortalidade por Câncer.....	15
2.1.2. Câncer do Colo do Útero	19
2.2. Sistema de Informação na Saúde Pública	21
2.2.1. Departamento de Informática do SUS – DATASUS	21
2.2.2. Sistema de Informação do Câncer do Colo do Útero (SISCOLO)....	22
2.2.3. Exame Citológico ou Citopatológico	25
2.2.4. Exame Histopatológico	25
2.2.5. Seguimento	27
2.2.6. Monitoramento Externo da Qualidade (MEQ).....	28
3. METODOLOGIAS.....	29
3.1. Banco de dados multidimensional (Data Warehouse)	29
3.2. Processo de Carga de Dados (ETL)	32
3.3. On-Line Analytical Processing (OLAP)	33
4. ARQUITETURA DO SISTEMA PROPOSTO.....	34
4.1. Indicadores do SISCOLO.....	34
4.1.1. Exames Citopatológicos e Histopatológicos	36
4.1.2. Monitoramento Externo de Qualidade do Exame (MEQ).....	37
4.1.3. Seguimento	38
4.2. Data Warehouse do SISCOLO.....	38
4.2.1. Fato Citopatológico	39
4.2.2. Fato Histopatológico	40

4.2.3.	Fato Seguimento	41
4.2.4.	Fato População IBGE x Exames	41
4.3.	Processo de Carga de Dados – ETL.....	42
4.4.	Painel de Indicadores – Compartilhar Resultados.....	45
5.	RESULTADOS	46
6.	CONCLUSÕES E TRABALHOS FUTUROS.....	51
6.1.	Outros Data Marts	52
6.2.	Data Mining	53
	REFERÊNCIAS BIBLIOGRÁFICAS.....	54

Lista de Ilustrações

Figura 1: Data Marts e base de dados origem do Observatório do Câncer	11
Figura 2: Estimativa 2008 das taxas brutas de incidência por 100.000 e do número dos casos novos por câncer, segunda a região - Estimativa 2008 (INCA).	15
Figura 3: Estimativa 2008 de número dos casos novos por câncer, por região - Estimativa 2008 (INCA).	16
Figura 4: Mortalidade por câncer, Brasil, homens e mulheres – Atlas de Mortalidade por Câncer (INCA).	16
Figura 5: Mortalidade mais frequentes - Homens – Atlas de Mortalidade por Câncer (INCA)	17
Figura 6: Mortalidade mais freqüente – Mulheres - Atlas de Mortalidade por Câncer (INCA)	17
Figura 7: Mortalidade mais frequentes - Mulheres, Região Norte - Atlas de Mortalidade por Câncer (INCA).	18
Figura 8: Mortalidade mais frequentes - Mulheres, Região Nordeste - Atlas de Mortalidade por Câncer (INCA).	18
Figura 9: Tela Principal – SISCOLO.....	22
Figura 10: Fluxo SISCOLO - Tipo de Exame	23
Figura 11: Fluxo Operacional SISCOLO	24
Figura 12: Fluxo SISCOLO – Diagnóstico Citopatológico	25
Figura 13: Fluxo SISCOLO - Exame Histopatológico	26
Figura 14: Estratégia de seguimento do câncer do colo do útero.	27
Figura 15: Fluxo SISCOLO - Seguimento	28
Figura 16: Star Schema	31
Figura 17: Snowflake Scheme.....	31
Figura 18: Fact Constellation Scheme	32
Figura 19: Modelo Conceitual - Fato Citopatológico.....	40
Figura 20: Modelo Conceitual - Fato Histopatológico.....	40
Figura 21: Modelo Conceitual - Fato Seguimento	41
Figura 22: População Brasileira Contagem Populacional de 1996 – IBGE	42
Figura 23: Modelo Conceitual - Fato População x Exames.....	42
Figura 24: Arquitetura de Carga Fatos	43
Figura 25: Arquitetura de Carga Dimensões	44
Figura 26: Exemplo de relatório disponível para homologação do usuário.	45
Figura 27: Painel de Indicadores - Site do INCA.....	46
Figura 28: Exemplo Painel 01 - Brasil	48
Figura 29: Exemplo Painel 01 - por Região.....	48
Figura 30: Exemplo Painel 03 - Todas as Faixas Etárias.....	49
Figura 31: Exemplo Painel 01 - por município.....	50
Figura 32: Atlas de Mortalidade por Câncer - BI sem DW	51

Resumo

O uso de sistemas de informação em saúde no Brasil é relativamente recente¹ e não parece ter interferido (pelo menos ainda) em uma desejada e mais ampla propagação. Criou-se rapidamente uma extensa base de dados de informações nacionais da saúde, mas não existe uma política unificada de comunicação entre estes sistemas, tornando as informações dispersas, redundantes e, muitas vezes, impossibilitando até mesmo o seu uso.

Uma visão unificada das diferentes informações possibilitaria aos gestores do Sistema Único de Saúde (SUS) resultados claros e precisos para controlar, regular e avaliar a qualidade da assistência, resultando em elementos necessários para ações de prevenção e controle de doenças.

Considerando essa necessidade, o Instituto Nacional de Câncer (INCA) vem investindo nos últimos anos em uma plataforma de Informações Gerenciais em Câncer que possibilite a integração dos sistemas de informação em câncer e, ao mesmo tempo, auxilie os gestores da rede de atenção oncológica na utilização dessa informação para agilizar tomadas de decisão.

Iniciou-se em 2008 o mapeamento de alguns projetos que deverão situar-se nesse ambiente de integração de informações. A primeira fase dessa implantação refere-se às informações de câncer do colo do útero.

Estima-se que o câncer do colo do útero em 2008 seja o segundo câncer mais comum na população feminina brasileira, representando 19% de todas as neoplasias em mulheres (INSTITUTO NACIONAL DE CÂNCER 2008). O método mais efetivo de prevenção de câncer do colo do útero, com redução em até 80% da mortalidade, são os programas de rastreamento do câncer através do exame preventivo (Papanicolau). O Inca, em parceria com o DATASUS (empresa de informática do SUS), desenvolveu um sistema de rastreamento do câncer do colo do útero (SISCOLO) dotado de uma base de dados com mais de 61 milhões de registros, desde 2001.

Com a presente monografia, pretende-se descrever as técnicas de business intelligence (data warehouse e OLAP) utilizadas para resolver o problema para consolidar, analisar e gerar relatórios abrangentes a partir de um

¹ Em 1975 foi implantado o Sistema de Informação de Mortalidade – SIM, sistema pioneiro a empregar documento individualizado e padronizado para a coleta das informações sobre óbito.

único repositório de dados com regras consistentes às informações do SISCOLO.

Palavras chaves: business intelligence, data warehouse, OLAP, cancer.

Abstract

Although the use of health information systems in Brazil is relatively recent², its rapid expansion was not detained. Currently, there are extensive national health databases managed by both government and private institutions. However, there is no intercommunication policy between these systems, which justifies for its disbursement, redundancy and often, impossible use of its information.

A common repository of the different available information would enable health managers to better analyze health care quality and create more effective propositions to prevent and control diseases based on clear and precise results.

Taking this into consideration the Brazilian National Cancer Institute (INCA) has invested during the last years in a Cancer Information Management platform that allows the integration of cancer information systems in order to support oncology network managers in their decision-making.

In 2008 INCA began to lay out projects that should populate this environment with integrated cancer information. The first project to be implemented in this environment refers to the details of cervical cancer.

It is estimated that in 2008 cervical cancer is the second most common cancer in Brazilian women, this represents 19% of all female cancer tumors (INSTITUTO NACIONAL DE CÂNCER 2008). The most effective method of preventing cervical cancer, with up to 80% mortality reduction, is cancer screening programs using Pap tests (or Pap smears). A joint effort between INCA and DATASUS resulted in the development of a cervical cancer screening information system (SISCOLO) with more than 61 million records, since 2001.

This paper describes the use of Business Intelligence techniques (Data Warehouse and OLAP) to solve the problem of consolidating, analyzing and generating comprehensive reports from a single database with consistent rules of SISCOLO information.

Key words: Business Intelligence, data warehouse, OLAP, cancer.

² The Mortality Information System (SIM) was implemented in 1975 as the pioneer system to use individualized and standard forms to collect information on death certificates.

1. Introdução

1.1. Motivação

A Missão e Objetivo Estratégico do INCA justificam a criação de uma área onde todas as informações referentes à oncologia no Brasil sejam compartilhadas de maneira consolidada e de fácil acesso aos gestores de saúde.

Missão do Instituto Nacional de Câncer

Ações nacionais integradas para prevenção e controle do Câncer

Objetivo Estratégico

"Exercer plenamente o papel governamental na prevenção e controle do câncer, assegurando a implantação das ações correspondentes em todo o Brasil, e, assim, contribuir para a melhoria da qualidade de vida da população."

Com base nessa diretriz o INCA tem investido num observatório do câncer que gere consultas das diversas informações referentes à neoplasia (Figura 1).

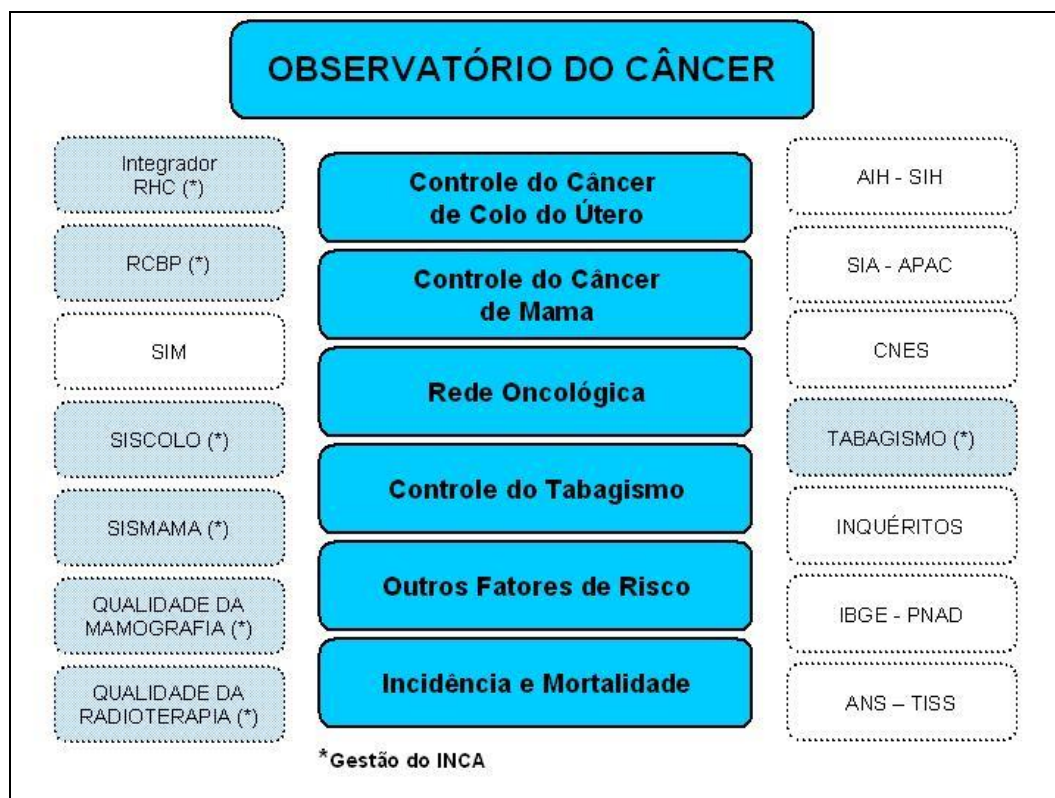


Figura 1: Data Marts e base de dados origem do Observatório do Câncer

Pretende-se criar nesta plataforma várias bases de dados dimensionais das informações nacionais do câncer. As bases hoje disponíveis para esse trabalho são os Registros de Câncer (registro de base populacional – *RCBP* e hospitalar – *RHC*), Rastreamento de Câncer (câncer do colo do útero – *SISCOLO* e câncer de mama – *SISMAMA*), Informações de Procedimentos (hospitalar – *SIH*, ambulatorial – *AIH* e alta complexidade – *APAC*), Sistema de Informações de Mortalidade (*SIM*) entre outros.

Em 2005, as articulações de ações dirigidas ao câncer de mama e do colo do útero estabeleceram a Política Nacional de Atenção Oncológica e a publicação da portaria GM nº. 2439 de 08/12/2005. Dentro das ações listadas o fortalecimento do sistema de informação e o desenvolvimento de pesquisas e mobilização social direcionaram o INCA a investir primeiramente na base de dados já existente e de maior direcionamento institucional.

Em 2008 iniciou-se o projeto do primeiro *DW* de câncer com as informações do sistema do câncer do colo do útero.

1.2. Objetivos do Trabalho

Os objetivos deste trabalho são:

Analisar os dados encontrados nas bases de dados do SISCOLO verificando a integridade das informações, com o objetivo de criar uma base dimensional com todas as informações necessárias para atender aos indicadores do câncer do colo do útero;

Definir os relatórios e gráficos que melhor expressam os indicadores do câncer do colo do útero e criá-los utilizando aplicativo OLAP.

O objetivo geral deste trabalho é estabelecer um processo de criação (passo a passo) de indicadores de câncer utilizando as ferramentas de *Business Intelligence (BI)* que possa servir no futuro para desenvolver os outros indicadores da plataforma de Informações Gerenciais em Câncer.

1.3. Descrição do Trabalho

Foram necessárias quatro etapas para implementar o ambiente de informações gerenciais: (1) identificação e análise dos indicadores, (2) criação da data warehouse, (3) construção dos processos de carga da data warehouse e (4) apresentação dos resultados em forma de relatórios gráficos nos painéis de indicadores.

Na primeira etapa os indicadores que constituem o Painel de Indicadores da Rede de Atenção Oncológica no Brasil foram analisados e foram identificadas as variáveis necessárias para cada indicador, assim como suas origens.

Na segunda etapa as variáveis identificadas para cada indicador foram mapeadas a um dos 4 fatos definidos - citopatológico, histopatológico, seguimento e população. O *data warehouse* (DW) foi modelado utilizando o modelo estrela.

Após modelar o *DW* iniciou-se a terceira etapa onde foi utilizada uma ferramenta de ETL para carregar o *staging* área, fatos e dimensões.

Na quarta e última etapa de implantação do ambiente de BI foi construído painéis que representam um ou mais indicadores cada, utilizando uma ferramenta OLAP.

1.4. Organização da Monografia

Esta monografia está dividida em cinco capítulos adicionais, descritos abaixo:

O capítulo 2 apresenta a descrição do problema e detalhamento da base de dados e sistema transacional da origem.

O terceiro capítulo é dedicado ao detalhamento da metodologia às premissas adotadas para o ambiente.

O quarto capítulo descreve a arquitetura e cada etapa da solução, as dificuldades encontradas em cada uma delas, assim como decisões tomadas e suas consequências.

O quinto e sexto capítulo é dedicado aos resultados e conclusões deste trabalho.

2. Descrição do Problema

Neste capítulo será apresentado o câncer do colo do útero, como doença com incidência e mortalidade crescente apesar de se apresentar como doença de fácil prevenção e tratamento. Acredita-se que o causa essa situação contraditória é a falta de informação ao gestor, portanto dificultando a criação de ações que possam diminuir essa incidência e consequentemente mortalidade. Técnicas de *BI* aplicadas na área da saúde, especificamente na Saúde Pública, apresentaram ganhos, porém algumas particularidades, como recursos escassos e dados incompletos e ausentes, dentre outros, no contexto brasileiro dificultam a aplicação direta destas ferramentas.

2.1. Câncer

Países que utilizaram um programa de rastreamento de câncer do colo do útero tiveram uma redução da incidência para menos de dez casos por 100.000 mulheres/ano (ZEFERINO 2008), independente do nível inicial de incidência. Baseado nessa informação, em agosto de 1998, o Ministério da Saúde (MS) iniciou uma ação nacional, convidando as mulheres de 35 a 49 anos de idade a comparecer aos serviços de saúde e realizar o exame citopatológico. Iniciou-se o programa de rastreamento do câncer do colo do útero no Brasil.

Mas fazer exame de rastreamento não é suficiente para garantir uma redução da mortalidade. É necessário que as mulheres que tenham um exame alterado recebam tratamento adequado. No Brasil, há evidências de que um percentual significativo de mulheres que fazem uma avaliação colposcópica não recebe qualquer tipo de acompanhamento.

Nessa sessão serão apresentados os últimos dados de incidência e mortalidade por câncer, as particularidades do câncer do colo do útero, a história do Sistema de Informação do câncer do colo do útero (SISCOLO), assim como as informações nela encontradas.

2.1.1. Incidência e Mortalidade por Câncer

É chamado de câncer o conjunto de mais de 100 doenças que tem em comum o crescimento desordenado (maligno) de células. Essas células podem invadir os tecidos e órgãos, espalhar-se (metástase) para outras regiões do corpo. A formação de um tumor é dada quando estas células agressivas e incontroláveis aglomeram-se (INSTITUTO NACIONAL DE CÂNCER 2009).

Estima-se que em 2008, e válido também para o ano de 2009, terão 466.730 casos novos de câncer masculino e feminino no Brasil (Figura 2). O câncer de pele do tipo não melanoma, também chamado de carcinoma basocelular e carcinoma epidermóide, é estimado o mais incidente (115 mil casos novos) na população brasileira, porém esse câncer apresenta altos índices de cura, principalmente devido à facilidade do diagnóstico precoce. Seguido ao câncer de pele, os cânceres mais prevalentes serão os cânceres de próstata (49 mil novos casos) e pulmão (18 mil novos casos) no sexo masculino e os câncer de mama (49 mil casos novos) e do colo do útero (19 mil novos casos) no sexo feminino (Figura 3). Esse é o mesmo perfil da magnitude observada nas estimativas mundiais.

Região	Estimativa dos Casos Novos		
	Masculino	Feminino	Total
Norte	8.470	9.150	17.620
Nordeste	36.600	42.360	78.960
Centro-Oeste	14.460	14.050	28.510
Sul	52.000	47.580	99.580
Sudeste	120.330	121.730	242.060
BRASIL	231.860	234.870	466.730

Figura 2: Estimativa 2008 das taxas brutas de incidência por 100.000 e do número dos casos novos por câncer, segunda a região - Estimativa 2008 (INCA).

Localização Primária	Norte	Nordeste	Centro-Oeste	Sul	Sudeste
Mama Feminina	1.210	7.630	2.630	9.500	28.430
Traquéia, Brônquio e Pulmão	1.020	3.630	1.690	7.210	13.720
Estômago	1.200	3.840	1.250	4.360	11.150
Próstata	1.750	9.820	3.200	9.500	25.260
Colo do Útero	1.700	4.720	1.350	3.470	7.440
Cólon e Reto	550	2.680	1.430	5.950	16.380
Esôfago	230	1.360	560	3.110	5.290
Leucemias	530	1.900	650	1.820	4.640
Cavidade Oral	400	2.500	750	2.500	8.010
Pele Melanoma	150	450	250	1.950	3.120
Outras Localizações	4.550	14.190	7.260	27.270	64.610
Subtotal	13.290	52.720	21.020	76.640	188.050
Pele não Melanoma	4.330	26.240	7.490	22.940	54.010
TOTAL	17.620	78.960	28.510	99.580	242.060

Figura 3: Estimativa 2008 de número dos casos novos por câncer, por região - Estimativa 2008 (INCA).

Em cinco anos, de 2002 a 2006, o índice de mortalidade por câncer aumentou de 12,8% para 14,5% no Brasil (Figura 4).

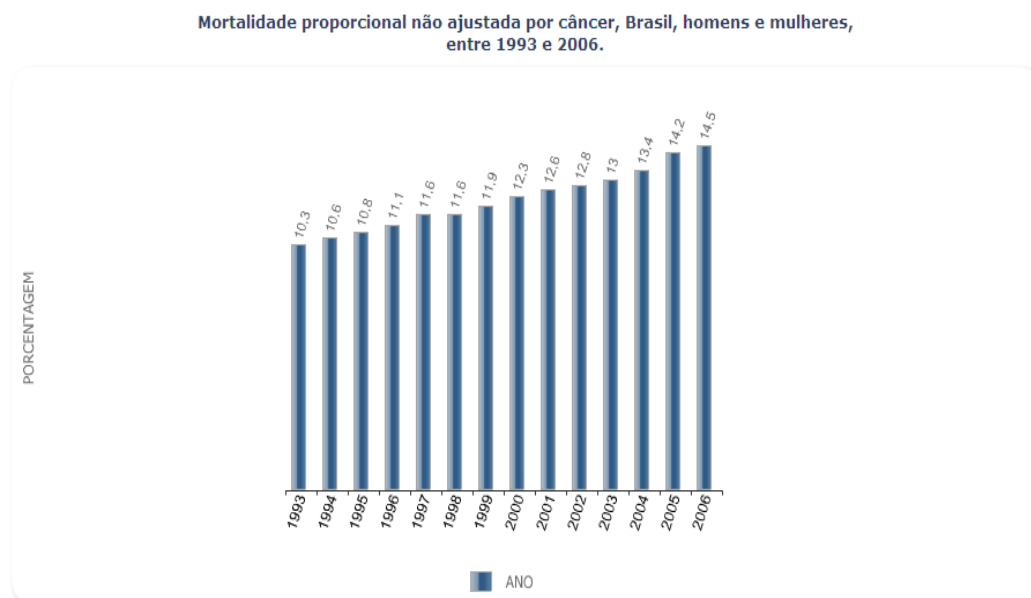


Figura 4: Mortalidade por câncer, Brasil, homens e mulheres – Atlas de Mortalidade por Câncer (INCA).

Nos homens, como visto na, a taxa de morte que mais cresceu nos últimos cinco anos foi a de câncer de próstata. Nas mulheres, o câncer de mama foi o câncer com maior índice de crescimento (Figura 5).

Taxas de mortalidade das 5 localizações primárias mais frequentes em 2006, ajustadas por idade, pela população mundial, por 100.000 Homens, Brasil, entre 2002 e 2006.



Figura 5: Mortalidade mais frequentes - Homens – Atlas de Mortalidade por Câncer (INCA)

Taxas de mortalidade das 5 localizações primárias mais frequentes em 2006, ajustadas por idade, pela população mundial, por 100.000 Mulheres, Brasil, entre 2002 e 2006.



Figura 6: Mortalidade mais freqüente – Mulheres - Atlas de Mortalidade por Câncer (INCA)

Mesmo sendo o câncer de mama aquele com maior taxa de mortalidade nacional, o câncer de colo de útero continua um dos 5 cânceres mais prevalentes no Brasil. Ao analisar a taxa de mortalidade individualmente por região, o câncer de colo de útero é a de maior taxa de mortalidade no Norte (Figura 7) e segunda maior taxa de mortalidade no Nordeste (Figura 8).

Taxas de mortalidade das 5 localizações primárias mais frequentes em 2006, ajustadas por idade, pela população mundial, por 100.000 Mulheres, Região Norte, entre 2002 e 2006.

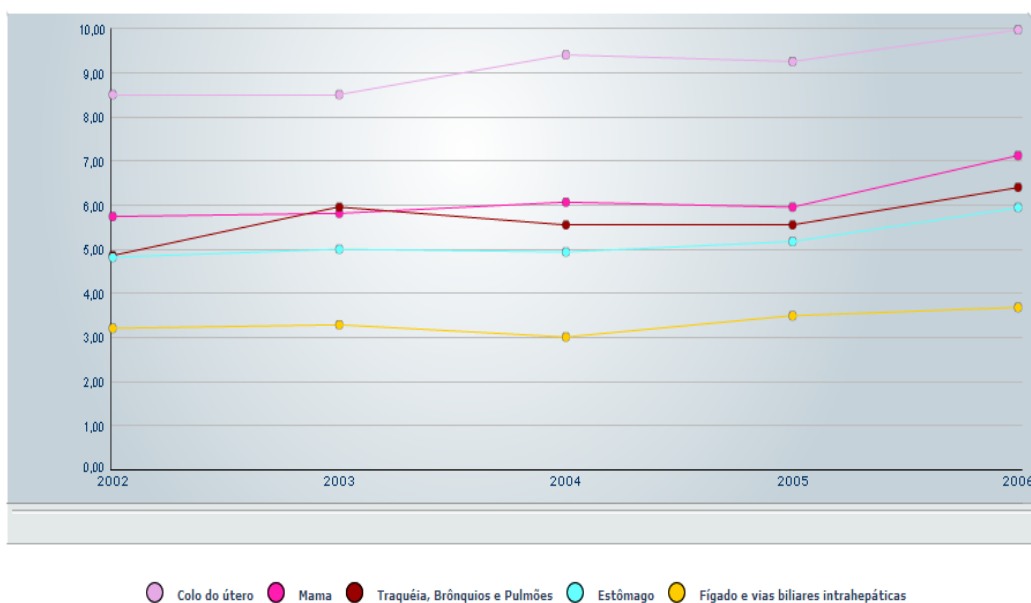


Figura 7: Mortalidade mais frequentes - Mulheres, Região Norte - Atlas de Mortalidade por Câncer (INCA).

Taxas de mortalidade das 5 localizações primárias mais frequentes em 2006, ajustadas por idade, pela população mundial, por 100.000 Mulheres, Região Nordeste, entre 2002 e 2006.

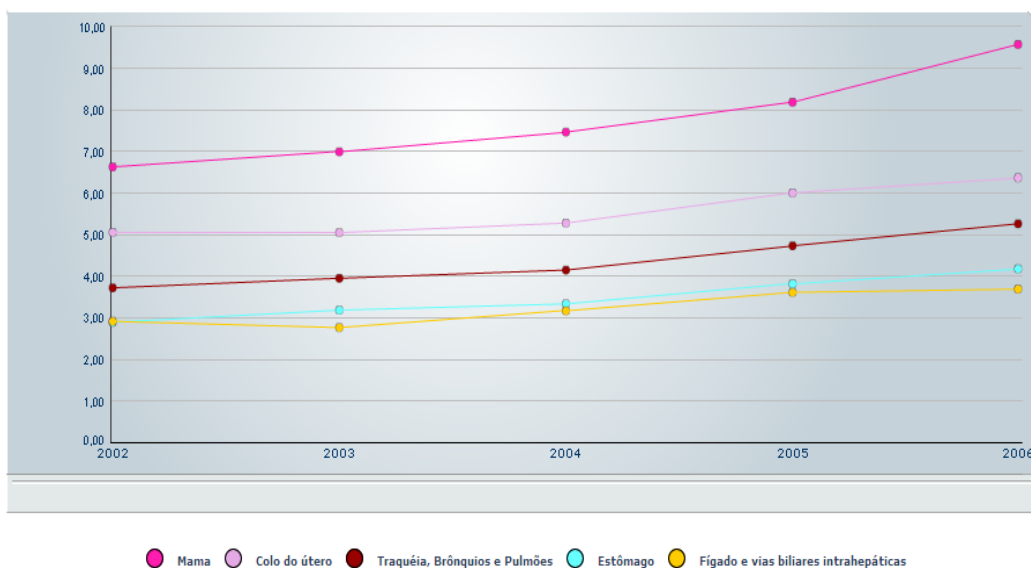


Figura 8: Mortalidade mais frequentes - Mulheres, Região Nordeste - Atlas de Mortalidade por Câncer (INCA).

Isso significa que apesar dos esforços do SUS para reduzir a mortalidade por câncer do colo do útero, este continua prevalente em todo país.

2.1.2. Câncer do Colo do Útero

O câncer do colo do útero se inicia no colo uterino da mulher, que é a parte do útero que fica no fundo da vagina. O útero é o órgão que envolve o bebê durante a gravidez e ao nascer de parto vaginal, o bebê passa pelo canal central do colo uterino. Este tumor maligno é de progressão relativamente lenta e que não manifesta sintomas na sua fase inicial. No início, estas anormalidades ainda não se caracterizam como um câncer e são denominadas displasias. Porém algumas dessas alterações ou displasias podem dar início a uma série de alterações que podem levar ao aparecimento do câncer do colo do útero.

Se células pré-cancerosas se transformam em células verdadeiramente tumorais e se espalham mais profundamente no colo uterino ou outros órgãos e tecidos, a doença é chamada de câncer de colo uterino ou cervical (vindo da palavra *cervix*, outro sinônimo para colo do útero).

Carcinoma de células escamosas – representam de 85% a 90% de todos os casos, e adenocarcinomas cerca de 10%.

Os fatores de risco abaixo agem direta ou indiretamente sobre o epitélio do colo uterino levando pequenas alterações que, com o passar do tempo, se não tratadas, podem levar ao câncer do colo do útero.

Fatores de Risco – Câncer do colo do útero
Multiplicidade de parceiros.
Monogamia com parceiros não monogâmicos.
Tabagismo.
Dieta.
Uso prolongado de anticoncepcional oral.
Doenças sexualmente transmissíveis em particular o HPV (Papiloma Vírus Humano).

Uso de drogas imunossupressoras.
Doenças que levam a diminuição da imunidade.

Nestas alterações iniciais a mulher não apresenta qualquer sintoma visível.

O HPV é o vírus sexualmente transmissível mais comum no mundo e alguns estudos comprovam que cerca de 20% dos adultos sexualmente ativos são infectados por um ou mais dos seus 120 tipos. O MS estima que cerca de 3% das mulheres portadoras do HPV podem desenvolver câncer de colo uterino. Mas a identificação e o tratamento das lesões precursoras previnem o desenvolvimento da doença maligna. O vírus do HPV está presente em mais de 90% dos casos de câncer do colo do útero.

Como o papilomavírus humano pode permanecer durante anos em estado de latência, suas manifestações podem aparecer ou reaparecer em qualquer momento da vida sem um motivo aparente. O HPV, porém não é sinônimo de promiscuidade, pois a maior parte das pessoas se contagia no início da vida sexual e não porque teve muitos parceiros.

2.1.2.1. Exame Preventivo do Câncer do Colo do Útero

A melhor maneira de detectar a lesão precursora de câncer do colo do útero é o exame Papanicolau que é feito a cada 1 ou 3 anos no exame ginecológico. Este exame preventivo consiste na coleta de material citológico do colo uterino. Para a coleta do material, é introduzido um espelho vaginal e procede-se à escamação ou esfoliação da superfície externa e interna do colo através de uma espátula de madeira e de uma escovinha endocervical.

Durante o exame o ginecologista pode perceber uma pequena alteração, que pode ser tratada como infecção e após tratamento repetir o exame. Quando uma alteração continua sendo identificada, a mulher é encaminhada para o exame de colposcopia onde o colo uterino pode ser checado, procurando áreas suspeitas. O colposcópio é o instrumento que é inserido na vagina para o exame. Este exame não é doloroso e não apresenta qualquer efeito colateral. O colposcópio dá uma visão aumentada e iluminada dos tecidos da vagina e colo do útero.

O próximo passo pode ser a realização de uma biópsia. A biópsia é um pequeno fragmento retirado de áreas suspeitas para exame microscópico. Se a lesão for pequena, o médico poderá tentar retirá-la totalmente durante a biópsia.

Se a biópsia confirma câncer do colo do útero, o paciente pode ser encaminhado para um especialista para tratamento. O especialista poderá pedir e fazer exames adicionais para verificar se o câncer está além do colo do útero.

Após o tratamento, a paciente é acompanhada para detectar lesões residuais ou recorrentes.

2.2. Sistema de Informação na Saúde Pública

2.2.1. Departamento de Informática do SUS – DATASUS

O Decreto 100 de 16/04/1991 criou a Fundação Nacional de Saúde - FNS, definindo o Departamento de Informática do SUS - DATASUS como órgão integrante da estrutura básica da FNS. Ao DATASUS compete "especificar, desenvolver, implantar e operar sistemas de informação relativos às atividades finalísticas do SUS, em consonância com as diretrizes do órgão setorial" (art.12).

DATASUS iniciou suas atividades em outubro 1991, a partir da incorporação dos recursos humanos, acervo técnico e equipamentos da DATAPREV relativos às atividades de informática do SUS, bem como quadros e recursos de informática da FSESP e SUCAM.

O objetivo inicial do Ministério da Saúde foi criar um Sistema Nacional de Informações em Saúde - SIS, integrado em todo o território nacional, abrangendo questões epidemiológicas e de prestação de serviços.

Em julho de 1992, em Belo Horizonte - MG, ocorreu o II Congresso Brasileiro de Epidemiologia, promovido pela Associação Brasileira de Pós-Graduação em Saúde Coletiva - ABRASCO. Como parte das atividades do Congresso, realizou-se a Oficina de Trabalho "Utilização de Grandes Bancos de Dados Nacionais".

2.2.2. Sistema de Informação do Câncer do Colo do Útero (SISCOLO)

O Ministério da Saúde iniciou o programa Nacional de Controle do Câncer do Colo do Útero em agosto de 1998 para reduzir a mortalidade e os efeitos físicos, psíquicos e sociais do câncer do colo do útero na mulher brasileira.

A partir de 1998, ocorreu à ampliação das ações de comunicação social, garantia do financiamento dos procedimentos ambulatoriais e criação de um software Sistema de informação do Câncer do Colo do Útero (SISCOLO) para melhor gerenciamento das informações do Programa Nacional de Controle do Câncer do Colo do Útero. Atrelando as informações ao sistema de faturamento, o Ministério da Saúde pretendia garantir o envio das informações das unidades de saúde que fizessem o exame preventivo do câncer do colo do útero.

O INCA em parceria com o DATASUS desenvolveu o SISCOLO (Figura 9) como o sistema de rastreamento de câncer do colo do útero. Atualmente a base de dados do SISCOLO possui mais de 61 milhões de registros.

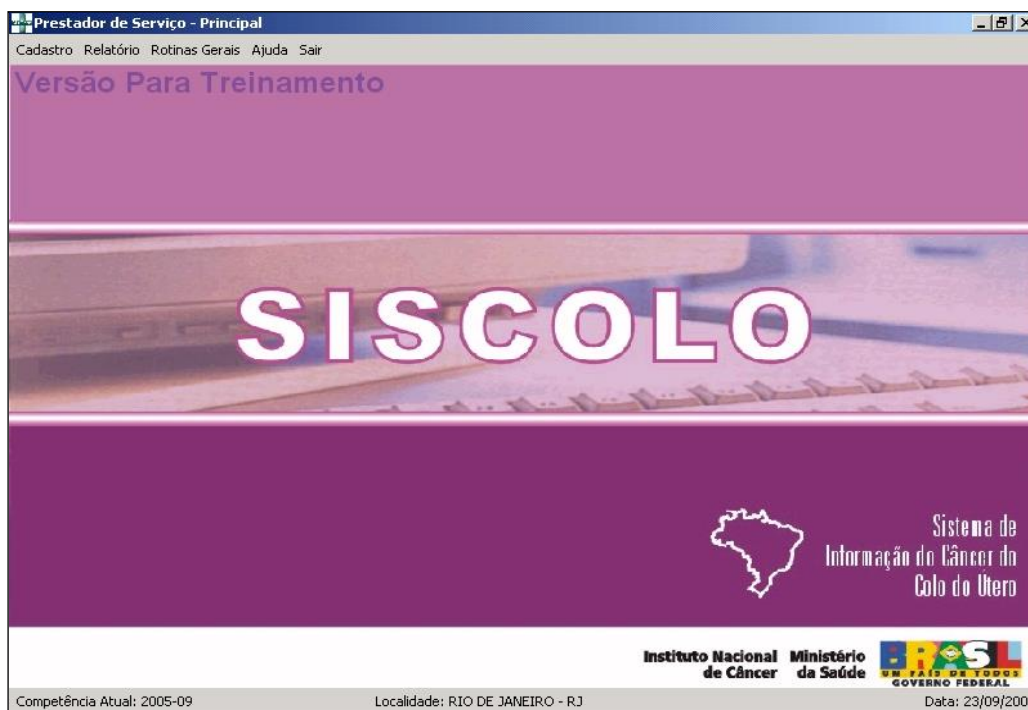


Figura 9: Tela Principal – SISCOLO

2.2.2.1. Fluxo do SISCOLO

Mensalmente os laboratórios conveniados pelo SUS digitam ou importam arquivos com os resultados dos exames citopatológico e/ou histopatológico coletados pelo serviço. Esses dois tipos de exames são coletados para o rastreamento de câncer do colo do útero (Figura 10).

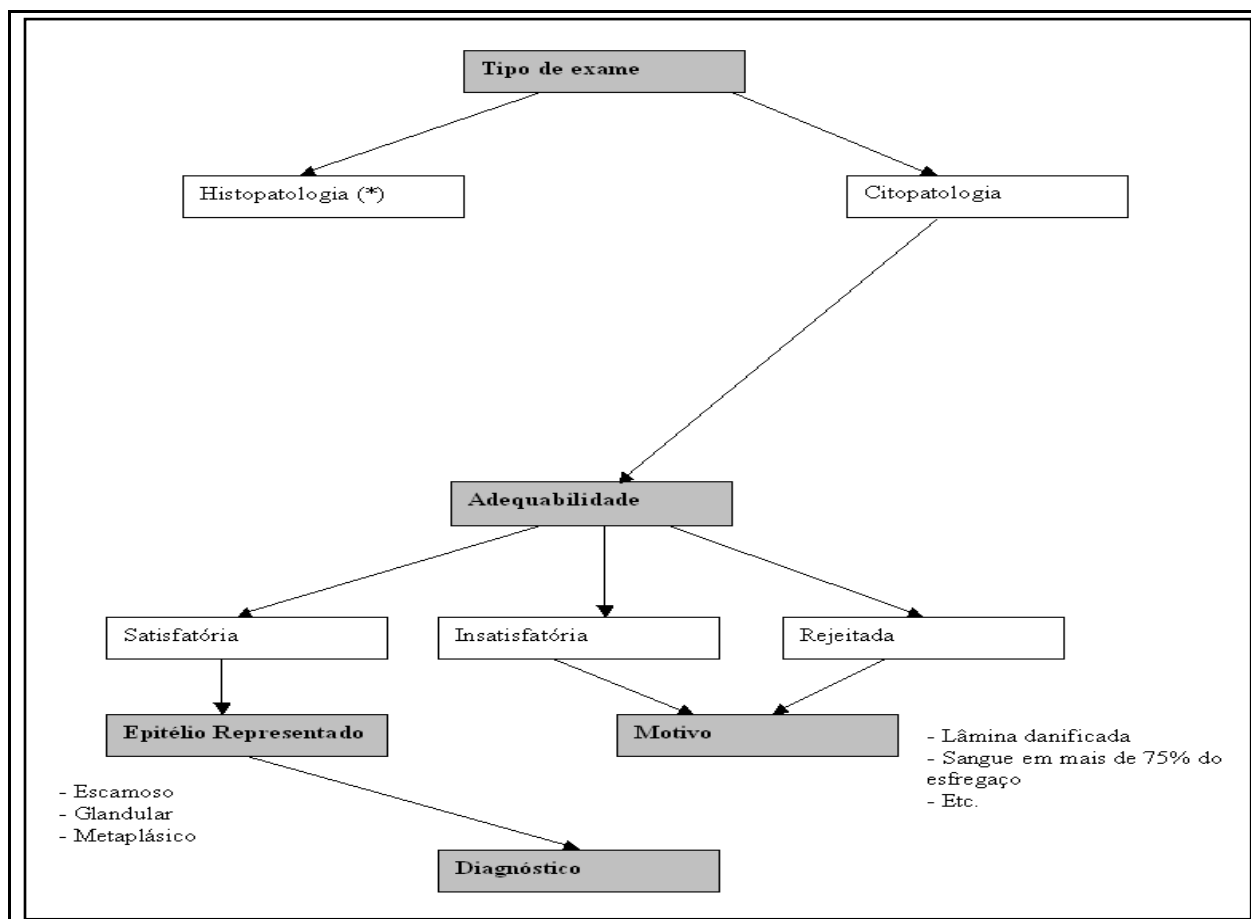


Figura 10: Fluxo SISCOLO - Tipo de Exame

No final de cada competência (mensal), esses dados são enviados para as secretarias estaduais e para o faturamento do SUS. O sistema que faz todo o faturamento de procedimentos do SUS chama-se Sistema de Informações Ambulatoriais (SIASUS).

A informação recebida pelo estado (dados do paciente, diagnóstico do exame e anamnese) será repassada para uma base central e descentralizada para os municípios para acompanhamento. Somente as mulheres que

apresentarem exames alterados são automaticamente direcionadas para o município para que seja iniciado o processo de seguimento.

Os serviços de saúde municipais devem realizar o acompanhamento das mulheres com exames citopatológicos alterados (atipias) e informar sobre os encaminhamentos.

Os dados enviados para o faturamento do SUS, permitirão que estes serviços recebam recursos relativos à quantidade de exames realizados.

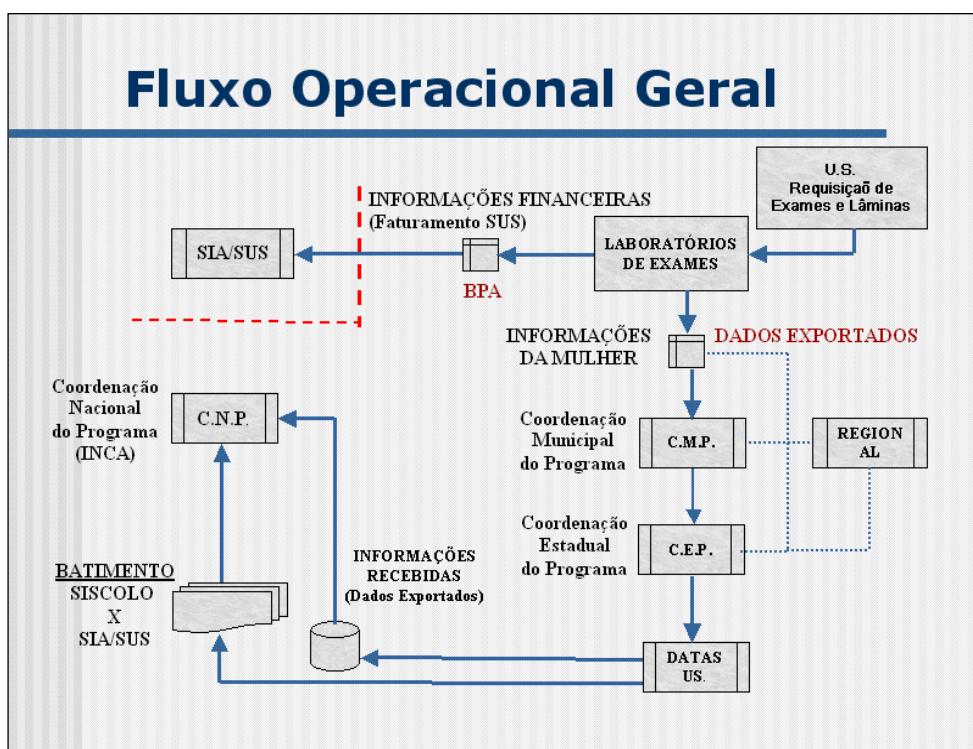


Figura 11: Fluxo Operacional SISCOLO

2.2.3. Exame Citológico ou Citopatológico

O exame citopatológico é feito para estudar as alterações celulares. Após a coleta e feita colocação do material na lâmina, deve-se proceder à sua fixação imediata em álcool e enviá-lo ao laboratório.

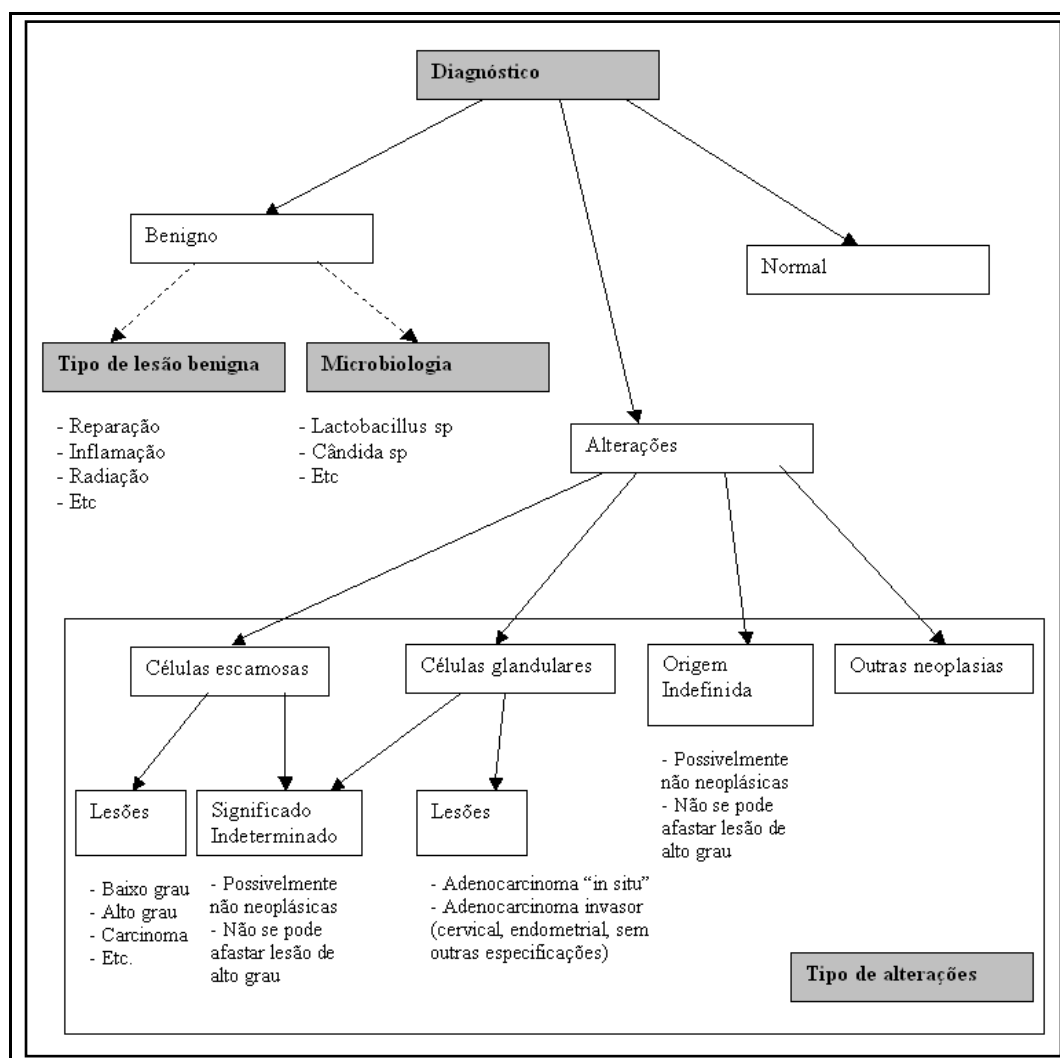


Figura 12: Fluxo SISCOLO – Diagnóstico Citopatológico

2.2.4. Exame Histopatológico

Este exame trata do estudo dos tecidos do organismo ao microscópio. O exame histopatológico permite afirmar seguramente a natureza de uma lesão.

No exame histopatológico, um fragmento do tecido é examinado, avaliando-se, então, toda a sua composição, sendo, portanto, mais preciso que o exame citológico. A obtenção do material pode ser feita por meio de uma pequena cirurgia, chamada biópsia. Há casos em que todo o nódulo ou tumor é retirado e há casos em que apenas uma parte é retirada.

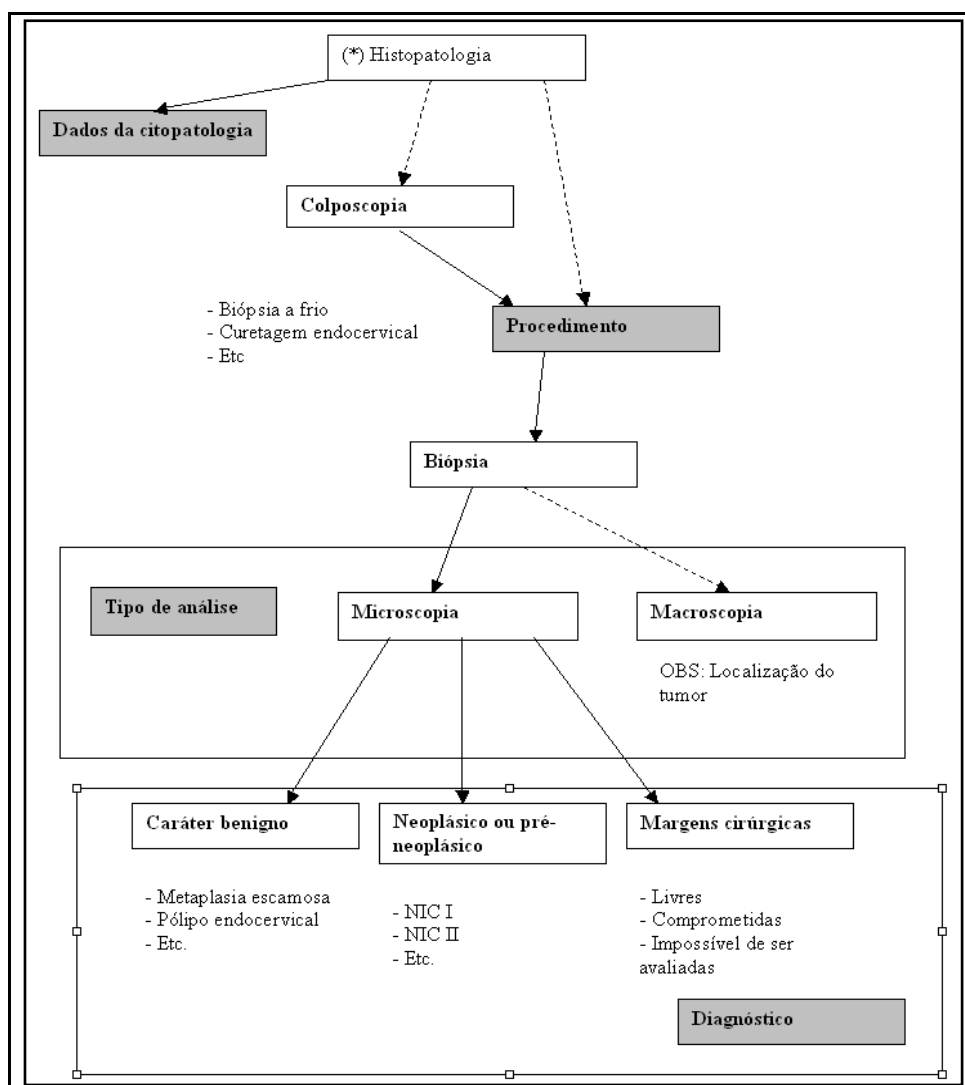


Figura 13: Fluxo SISCOLO - Exame Histopatológico

2.2.5. Seguimento

Esta informação é histórica, ou seja, serão armazenadas todas as informações dos pacientes nas diversas situações. As informações contidas neste módulo fazem com que os programas de prevenção do câncer sejam eficazes. Todos os procedimentos que conduzem a um diagnóstico completo, tratamento adequado do caso e observação dos diversos desfechos: cura ou óbito são registrados. O paciente diagnosticado ou não com alteração no exame preventivo deverá retornar a uma unidade de saúde para novo exame ou tratamento (Figura 14).

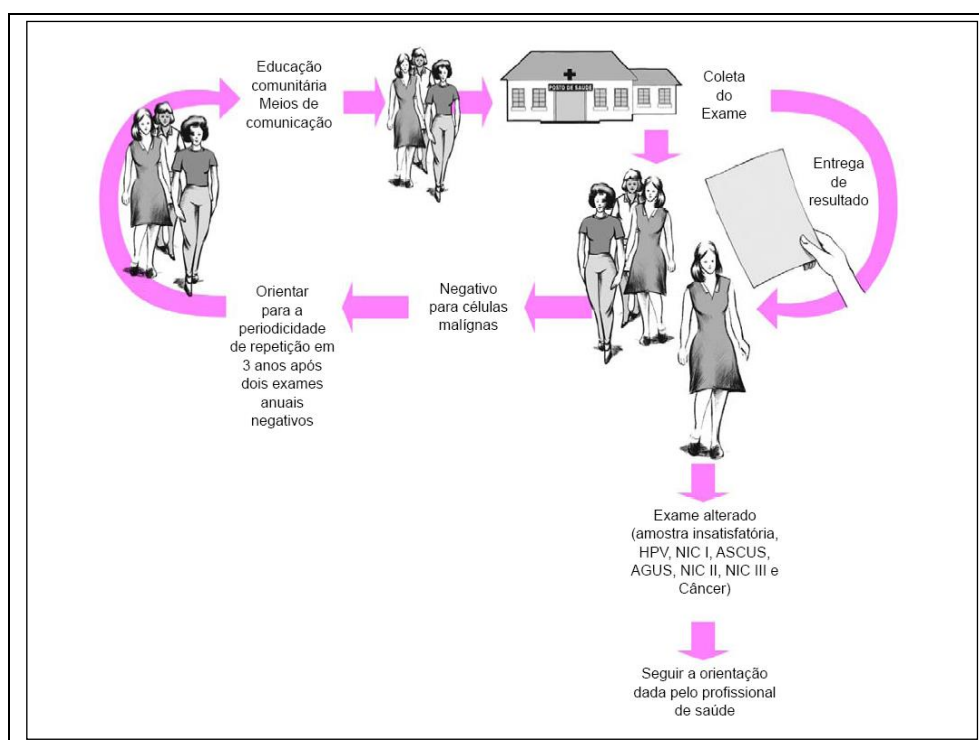


Figura 14: Estratégia de seguimento do câncer do colo do útero.

A Figura 15 detalha todas as informações registradas no seguimento em caso de exames alterados.

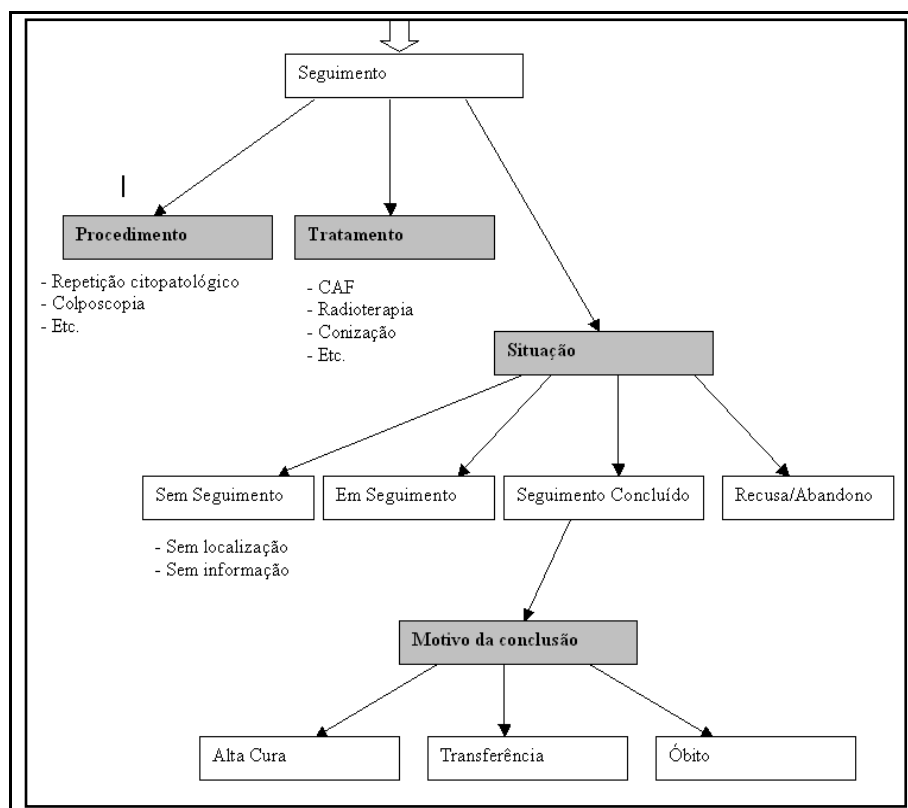


Figura 15: Fluxo SISCOLO - Seguimento

2.2.6. Monitoramento Externo da Qualidade (MEQ)

O monitoramento Externo da Qualidade envolve a nova leitura de lâminas do exame citopatológico por outro laboratório para que o resultado possa ser comparado. Caso haja algum tipo de discordância o exame deverá ser repetido e os laboratórios envolvidos avaliados pela a qualidade do seu diagnóstico. Esse controle de qualidade ocorre apenas para exames citopatológicos.

3. Metodologias

Antes de definir as ferramentas desta monografia é esclarecedor definir o conceito de *BI* que trata-se de um “termo guarda-chuva” que inclui todas as tecnologias que auxiliam a coleta, organização, análise, compartilhamento e monitoramento das informações. Algumas das tecnologias *BI* utilizadas neste trabalho são *DW*, *Data Mart*, *ETL* e *OLAP*.

Para se obter um bom ambiente de *BI*, um dos componentes básicos, e provavelmente o mais necessário, é um *DW* (ou um *Data Mart* para ambientes de *BI* com menor abrangência ou assunto específico). Neste capítulo, será apresentado a definição de *DW* e os métodos para sua construção. Será apresentado o processo de carga de dados de um *DW*, descrevendo os cuidados que devem ser tomados especificamente se tratando de um sistema de saúde. Por último será abordado o conceito de compartilhamento dos dados via uma ferramenta de *OLAP*.

Nesse capítulo são apresentadas as tecnologias de *BI* que podem ser relacionados à saúde apesar de constatar-se que são pouco utilizados, apesar de claramente identificado como útil para a melhor qualidade de decisões. Pressupõe-se que esse baixo uso das ferramentas de inteligência computacional deve-se ao fato do valor de investimento inicial, tanto de ferramentas como de recursos humanos, atualmente ser alto e o benefício do uso de *BI* não está estimado.

3.1. Banco de dados multidimensional (Data Warehouse)

Os sistemas que contêm dados operacionais de uma empresa apresentam informações essenciais para o gestor do negócio. Por exemplo, no caso das unidades de saúde, os sistemas que registram cada atendimento e o tratamento que o paciente recebe, armazenam dados dos pacientes (idade, sexo, estado físico, residência, etc.), dados da doença e tratamento, assim como informações da infra-estrutura da unidade de saúde e o seu corpo clínico.

Em meados dos anos 80 grande parte dos estudiosos de banco de dados conheceu os fundamentos de bancos relacionais através do livro de Chris Date's *An Introduction to Relational Databases* (DATE 2003). Ao longo dos anos esse

conceito foi ampliado e aperfeiçoado objetivando a criação de uma estrutura mais eficiente para armazenamento das informações sem redundância e robusta.

No entanto, alguns problemas poderão ocorrer quando os analistas acessarem esses dados operacionais diretamente:

- Desempenho no acesso às informações;
- Habilidade para criar consultas dessas informações necessitando da ajuda ou dependência de um profissional especializado;
- Informações faltantes e sem padronização causada por versionamento do sistema ou preenchimento falho.

A criação de um banco dimensional resolve esses problemas. *Data Warehousing* é o processo de consolidação, organização e uso de dados operacionais isoladas ou em uma única base de dados que objetiva a criação de banco dimensional ou DW. Todo banco dimensional é composto por uma tabela central chamado de tabela de fato e várias tabelas auxiliares chamadas tabelas dimensionais. A tabela fato é a estrutura que contem os valores quantitativos da sua base de dados. As informações de cada fato são armazenadas nas tabelas de dimensão. A menor granularidade de um *DW* é o atributo que são os itens dentro das dimensões.

As tabelas dimensionais do *DW* são conectadas diretamente a tabela de fatos por campos chaves. As estruturas mais comuns de um *DW* são *Star* (Figura 16) e *Snowflake Scheme* (Figura 17).

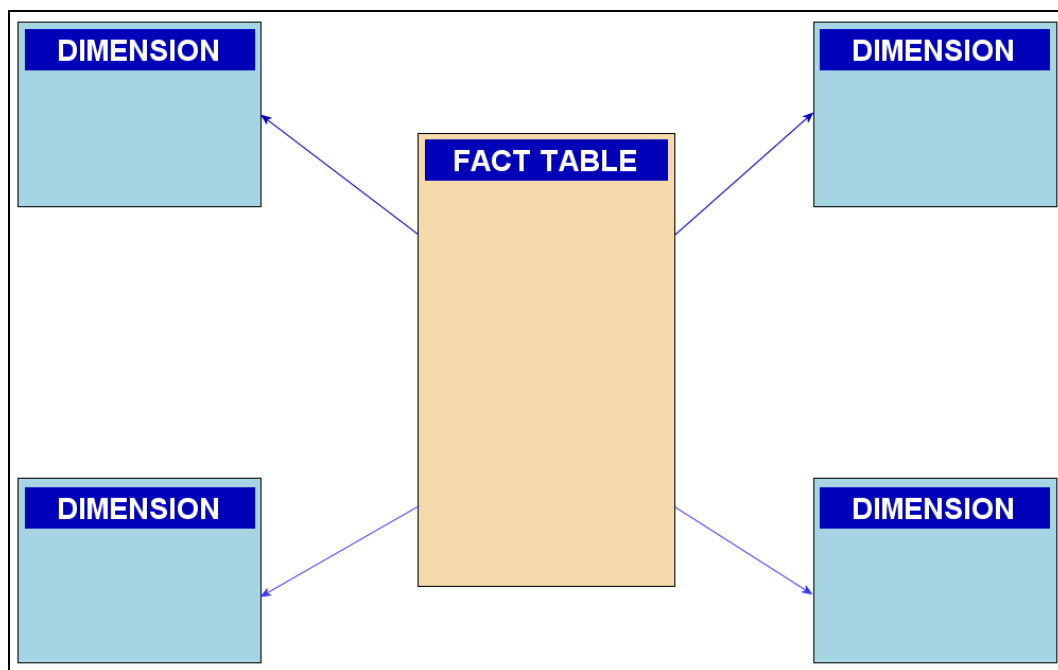


Figura 16: Star Schema

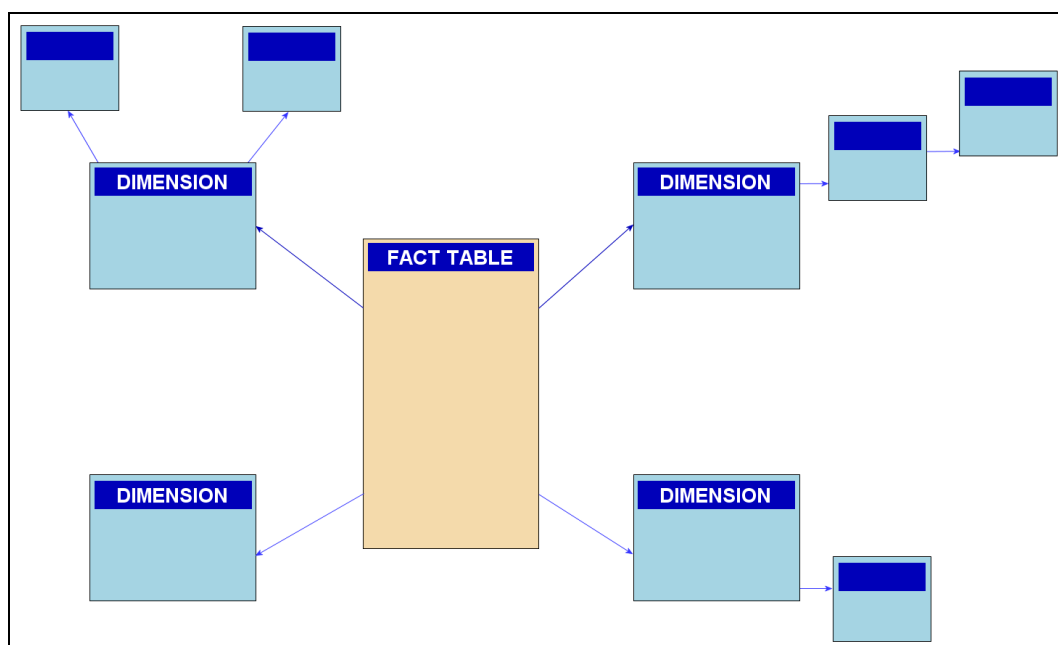


Figura 17: Snowflake Scheme

Existe ainda uma terceira e menos usada estrutura de base de dados dimensional chamada *Fact Constellation Scheme* (Figura 18).

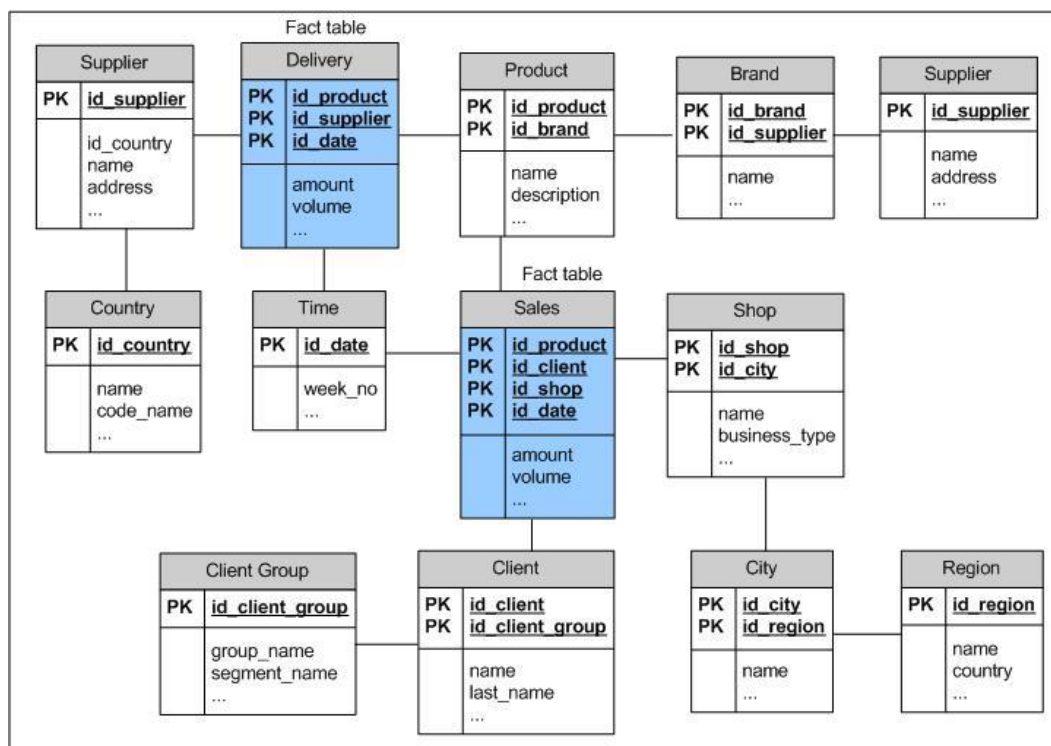


Figura 18: Fact Constellation Scheme

3.2. Processo de Carga de Dados (ETL)

A etapa mais crítica de um projeto de BI é a de carga de dados no DW, chamada Extração, Transformação e *Loading*/Carregamento (ETL). Essa etapa automatiza o processo de converter, re-formatar e integrar informações da base de dados relacional para uma base multidimensional. Aplicativos de ETL podem também ser usados para migração de dados quando esse necessita de correções e será executado com alguma frequência, reduzindo os riscos de processos manuais.

A etapa de ETL ocorre numa área de processamento temporária chamada de *Data Staging Área*. Nessa área ocorre a preparação dos dados que irão para o DW (ou *Data Mart*).

Lembrando que neste processo um grande volume de dados está sendo convertido um ambiente que consolida os registros, seguindo algum critério para sumariá-los, alguns passos devem ser identificados para assegurar a consistência dos dados nos relatórios finais. Apesar de alguns sistemas de ETL dedicarem os seus processos a alterar os dados quando esses estão incorretos, as melhores práticas de BI sugerem que se façam correções nos bancos de

dados de origem. Após a correção da informação encontrada incorreta ou ausente no processo de ETL o DW poderá ser atualizado.

3.3. On-Line Analytical Processing (OLAP)

A etapa de análise dos dados é feita através do processamento de *OLAP*. Este termo foi definido por CODD (1993) e ampliada ao longo dos anos e significa análise rápida de informações multidimensionais compartilhadas.

Usa-se nesse modelo representações de “cubos multidimensionais” que sumariza um conjunto de fatos e dimensões de um *DW*.

O uso de *OLAP* é encontrado em varias áreas comerciais e da ciência para identificar informações estratégicas de maneira ágil, precisa e consistente. Alguns exemplos são apoios à captação de doadores de sangue e sistema de acompanhamento da saúde das famílias de Novo Hamburgo. Apesar de cada vez mais estarmos encontrando exemplos de sucesso, na saúde pode-se considerar uma ferramenta inovadora.

Existem vários aplicativos *OLAP* no mercado, dentre elas, Oracle Discoverer, MicroStrategy, Business Objects, Cognos BI, SQL Server Analysis Services.

4. Arquitetura do Sistema Proposto

Este capítulo detalha os passos de criação do *DW*, processo de carga dos dados, bem como a criação de cada painel de indicadores finais.

Foi dividido da seguinte maneira, primeiramente enumerar os indicadores e sua interpretação. A seguir o detalhamento de cada fato dentro do *DW* e as dimensões, acompanhada de justificações para cada escolha. O terceiro passo é a carga dos dados com informações da limpeza de dados e arquitetura de ETL. Por último a criação dos painéis com exemplificações dos relatórios necessários para dar suporte a cada indicador e visão de cada painel.

São, ao todo, 61 milhões de registros desde 2001, inclusos nesses registros são 35 milhões de exames citopatológicos e o restante dos registros exames histopatológicos, informações de seguimento e monitoramento externo.

4.1. Indicadores do SISCOLO

Em outubro de 2006 o INCA organizou uma Oficina de Validação de indicadores que formarão um Painel de Indicadores da Rede de Atenção Oncológica no Brasil. Pretende-se ampliar o processo de construção de um modelo de avaliação e monitoramento da Política Nacional de Atenção Oncológica. Nessa lista de indicadores referentes aos exames citopatológicos e histopatológicos, monitoramento externo e seguimento de pacientes com exames alterados. A partir desses indicadores iniciou-se a especificação das informações necessárias no *DW* das informações do SISCOLO.

Cada indicador é construído através de uma variedade de origens como: IBGE, DATASUS e alguns sistemas e tabelas internas. Essas informações precisam ter algumas variáveis em comum para serem utilizadas como chave primária. Nesse caso foi utilizadas tabelas de localidade que contém códigos de município que é composto de 7 dígitos, sendo os dois primeiros referentes ao código do estado e os cinco últimos o município. Outra chave primária que distingue cada estabelecimento da saúde é o Cadastro Nacional de Estabelecimentos de Saúde - CNES. Esse código identifica cada estabelecimento de saúde e informa condições não só de endereço, nome de fantasia, como condições de infra-estrutura de funcionamento.

4.1.1. Exames Citopatológicos e Histopatológicos

- **Razão entre lesão de alto grau e carcinoma invasor de câncer do colo do útero**

Conceito: Indicador de desempenho do programa de controle de câncer do colo do útero, permitindo ao programa detectar precocemente as lesões pré-malignas.

Método de Cálculo:

$$\frac{\text{Nº. total de exames com diagnóstico de lesão de alto grau}}{\text{Nº. total de exames com diagnóstico de carcinoma invasor}}$$

OBS: antes de 2006 utilizar nomenclatura para NIC II (lesão de alto grau) e NIC III (carcinoma invasor) no numerador. Após 2006 o numerador é o descrito acima.

- **Percentual de amostras insatisfatórias (adequabilidade)**

Conceito: Número de amostras insatisfatórias elevadas está associado principalmente ao problema nas etapas de coleta e processamento das amostras e reflete a inadequabilidade das amostras para diagnóstico.

Método de Cálculo:

$$\frac{\text{Nº. de amostras insatisfatórias}}{\text{Nº. total de exames citopatológicos}}$$

Variação: Substituir as amostras insatisfatórias por satisfatórias, mas limitadas.

- **Cobertura de exames citopatológicos cérvico-vaginais**

Conceito: Indica a capacidade das ações de detecção para mobilizar e captar a população-alvo para rastreamento.

Método de Cálculo:

Nº. total de exames citopatológico (faixa etária 25 a 59 anos)

Nº. total de mulheres (faixa etária 25 a 59 anos)

OBS: este indicador irá utilizar a população referente ao último ano liberado pelo IBGE, ou seja, caso o ano atual não tenha sido liberado será usado o ano anterior.

• **Percentual captação de mulheres sem citopatológico anterior**

Conceito: Indica a capacidade das ações de detecção para mobilizar e captar a população-alvo para o rastreamento.

Método de Cálculo:

Nº. total de mulheres SEM citologia anterior (faixa etária 25 a 59 anos)

Nº. total de mulheres (faixa etária 25 a 59 anos)

Variação: Substituir SEM citologia anterior para COM citologia anterior e SEM INFORMAÇÃO.

4.1.2. Monitoramento Externo de Qualidade do Exame (MEQ)

• **Percentual de laboratórios que participam do Monitoramento Externo da Qualidade dos exames citopatológicos do colo do útero (Papanicolaou)**

Conceito: Percentual de laboratórios prestadores de serviços ao SUS que realizam exames citopatológicos do colo do útero (Papanicolaou) participantes do processo de Monitoramento Externo da Qualidade (MEQ).

Método de Cálculo:

Nº. de laboratórios que realizam exames citopatológicos do colo do útero e submetem suas lâminas a MEQ

Nº total de laboratórios que realizam exames citopatológicos do colo do útero

• **Percentual de exames citopatológicos submetidos à releitura pelo Monitoramento Externo da Qualidade.**

Conceito: Percentual de exames citopatológicos do colo do útero (Papanicolaou) submetidos a uma nova leitura feita por um novo laboratório.

Método de Cálculo:

Nº. de exames citopatológicos do colo do útero submetidos a uma nova leitura

Nº total de exames citopatológicos do colo do útero

• **Percentual de concordância observada entre diagnóstico:**

Método de Cálculo:

Nº. de exames citopatológicos do colo do útero concordantes entre laboratório principal e laboratório MEQ

Nº total de exames citopatológicos do colo do útero submetidos ao MEQ

4.1.3. Seguimento

Relatórios por diagnóstico, tratamento, situação do seguimento (óbito, abandono, alta) e tempo de seguimento do paciente.

4.2. Data Warehouse do SISCOLO

Dentro da etapa de construção do DW, foram modelados quatro fatos que armazenam as informações necessárias do SISCOLO:

1. Fato Citopatológico – referente aos exames de preventivo, identificando quantitativo de exames e os seus diagnósticos;

2. Fato Histopatológico – referente aos exames histológicos, identificando o quantitativo de exames e os seus diagnósticos;
3. Fato Seguimento – referente ao acompanhamento dos pacientes com exames alterados que precisarão de tratamento;
4. Fato População versus Exames – referente ao total de pacientes com granularidade de município, relacionado ao quantitativo de exames histológicos e citopatológicos.

Para armazenar as informações do monitoramento externo, o mesmo fato citopatológico diferenciando o laboratório principal e os laboratórios de monitoramentos externos com indicativos.

As premissas do DW para os exames citopatológicos e histopatológicos são:

- Serão carregadas informações a partir de janeiro/2000.
- O tipo de procedimento cirúrgico para os exames histopatológicos é sempre Biópsia.
- Não existe estadiamento para o SISCOLO (tanto para exame citopatológico e histopatológico).
- Não haverá utilização de informações do cartão do SUS (identificação do paciente) para associar um exame histopatológico com seu exame citopatológico. Isso significa que não haverá nesse DW informação que possibilite a identificação do paciente.
- Todos os fatos são carregados mensalmente com informações referentes a competência (mês) anterior.

4.2.1. Fato Citopatológico

O Fato Citopatológico armazena todas as informações do exame citológico do câncer do colo do útero como Paciente (local de residência, faixa etária, escolaridade), Unidade de Saúde que fez a solicitação de exame (localidade, tipo de prestador), Exame (período / tempo, diagnóstico, qualidade do exame) tanto para o laboratório principal como para o laboratório de monitoramento externo.

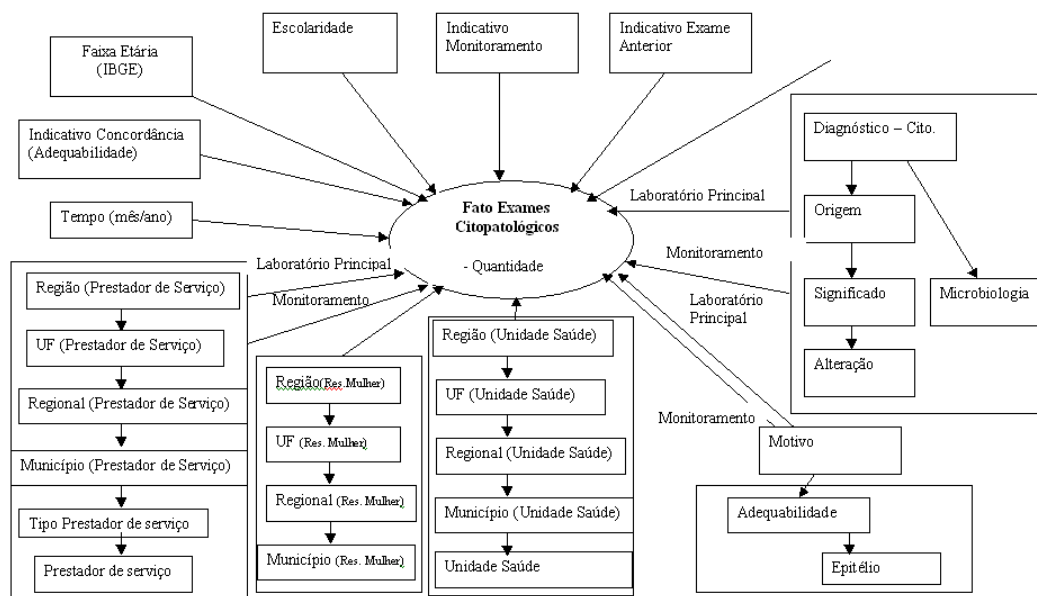


Figura 19: Modelo Conceitual - Fato Citopatológico

4.2.2. Fato Histopatológico

O Fato Histopatológico armazena todas as informações do exame histológico do câncer do colo do útero como Paciente (local de residência, faixa etária, escolaridade), Unidade de Saúde que fez a solicitação de exame (localidade, tipo de prestador), Exame (período / tempo, diagnóstico, qualidade do exame) (Figura 20).

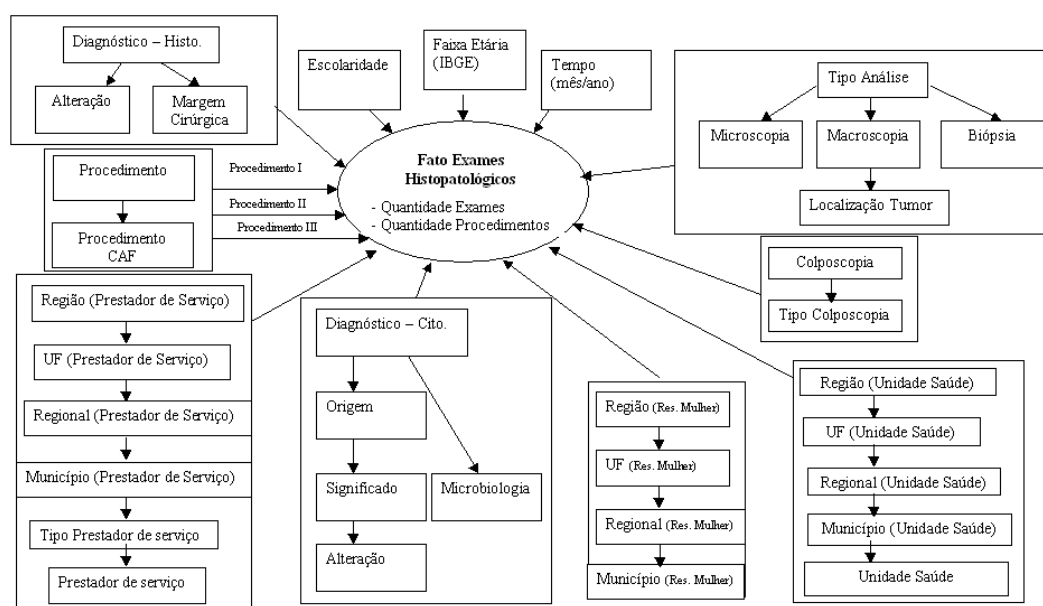


Figura 20: Modelo Conceitual - Fato Histopatológico

4.2.3. Fato Seguimento

Esse fato contém informação dos seguimentos dos pacientes com exame citopatológico alterado. As informações desse fato referem-se ao Paciente (local de residência, origem do paciente), Unidade de Saúde, tratamento (período), situação do seguimento (motivo de conclusão, se for o caso), Exame (período, tipo de exame, repetição de cito) (Figura 21).

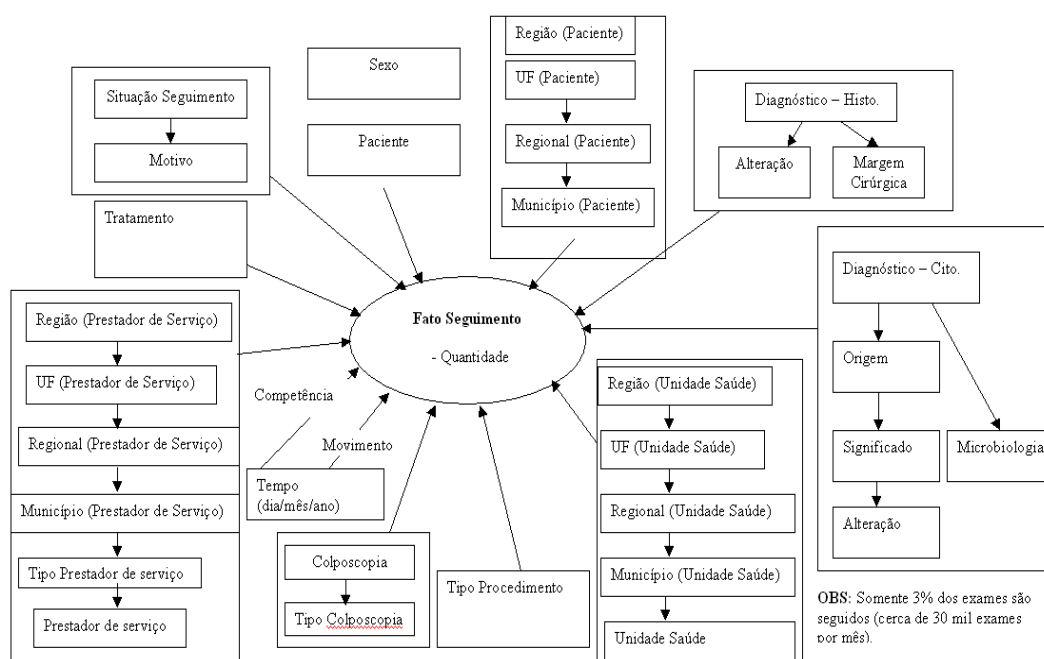


Figura 21: Modelo Conceitual - Fato Seguimento

4.2.4. Fato População IBGE x Exames

As estimativas populacionais, tanto censitárias quanto intercensitárias, utilizadas no DW do SISCOLO, foram fornecidas pelo Instituto Brasileiro de Geografia e Estatística – IBGE. As populações são classificadas por faixa etárias (Figura 22).

População Padrão Brasil		
Índice	Faixa etária	Total
1	00-04	15.666.747
2	05-09	16.441.227
3	10-14	17.568.584
4	15-19	16.723.387
5	20-24	14.445.732
6	25-29	13.000.784
7	30-34	12.422.671
8	35-39	10.911.407
9	40-44	9.261.885
10	45-49	7.463.915
11	50-54	5.882.787
12	55-59	4.854.733
13	60-64	4.002.121
14	65-69	3.238.057
15	70-74	2.271.773
16	75-79	1.457.689
17	80 ou +	1.465.664
Total		157.070.163

Figura 22: População Brasileira Contagem Populacional de 1996 – IBGE

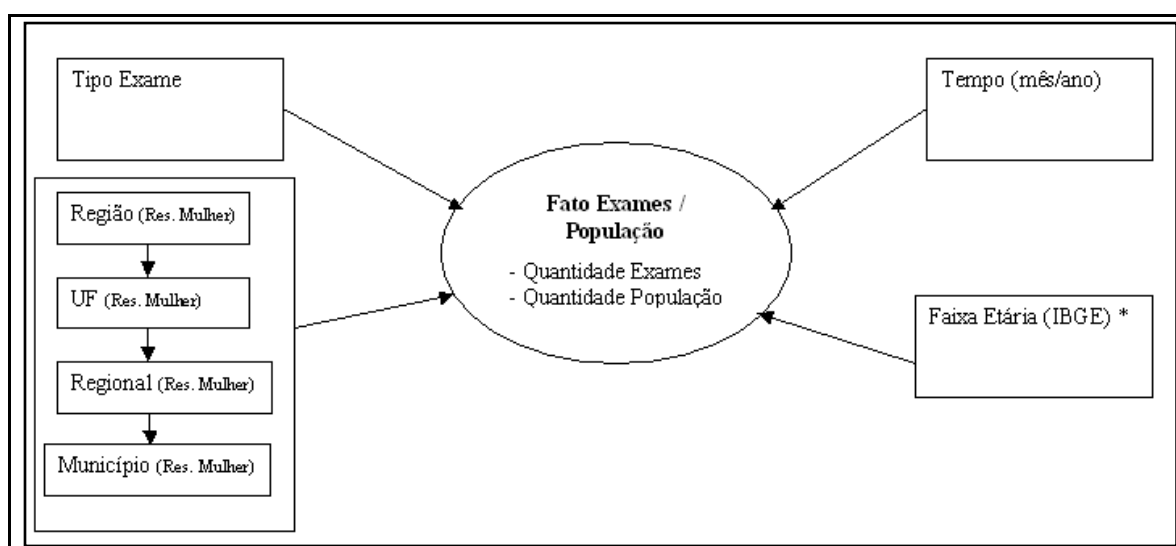


Figura 23: Modelo Conceitual - Fato População x Exames

4.3. Processo de Carga de Dados – ETL

Apesar do objetivo do Ministério da Saúde em criar o SISCOLO ser rastrear o câncer do colo do útero o maior interesse dos usuários dos laboratórios é o faturamento dos exames. Com isso a qualidade dos dados

epidemiológicos é prejudicada e o processo de ETL torna-se crucial para a limpeza e consistência dos dados.

No início do projeto de BI no INCA, grande parte dos recursos foi destinada à compra de OLAP e contratação de consultoria e funcionários que dominassem a nova tecnologia. Optou-se por não investir numa ferramenta ETL e utilizar a ferramenta da Oracle, OWB 10g que já era largamente utilizada para seus aplicativos transacionais.

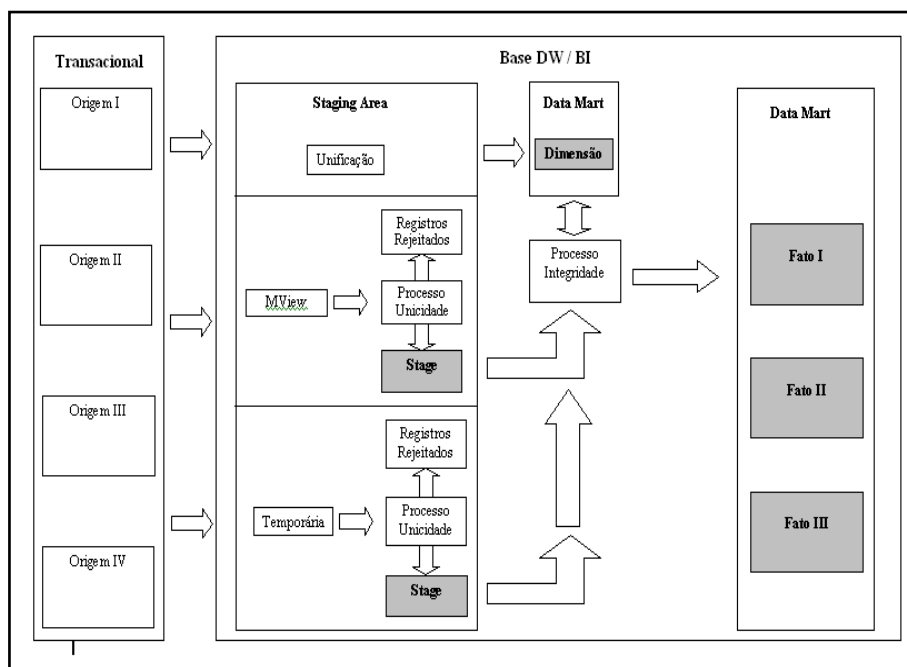


Figura 24: Arquitetura de Carga Fatos

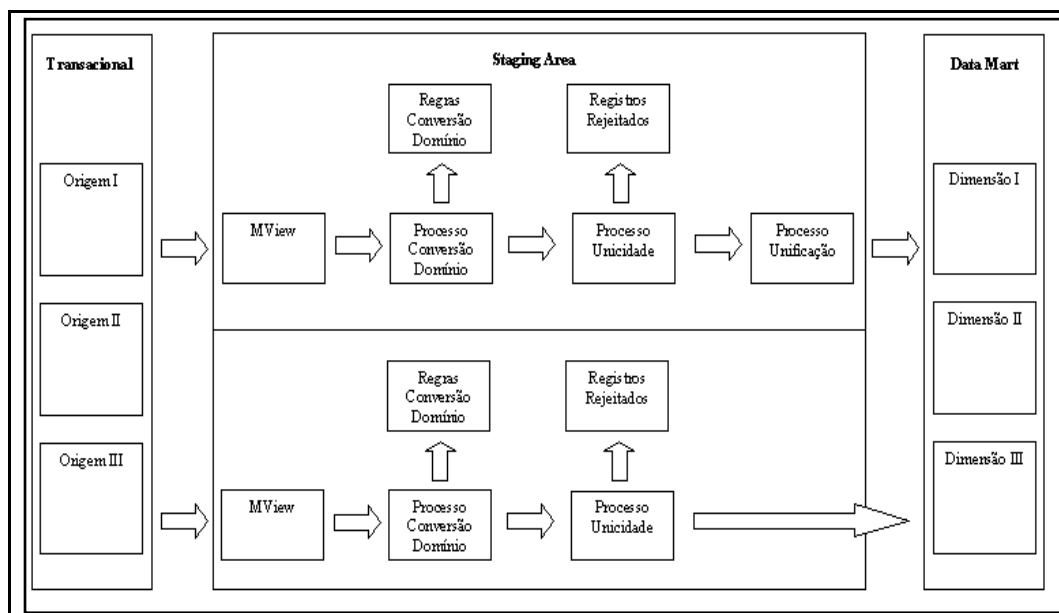


Figura 25: Arquitetura de Carga Dimensões

4.4. Painel de Indicadores – Compartilhar Resultados

O aplicativo adquirido pelo INCA, após avaliação de mais três outras soluções foi a MicroStrategy. Fundada em 1989, a MicroStrategy se especializou em aplicativos de BI. A MicroStrategy desenvolve software para relatórios integrados, análises e monitoramento, que auxiliam as organizações a tomarem melhores decisões de negócios. A escolha pela tecnologia MicroStrategy se dá por suas habilidades técnicas avançadas, capacidade de análise sofisticada e superior escalabilidade de dados e de usuários (www.microstrategy.com).

Toda a homologação dos dados (conteúdo) do DW é feita pelos usuários de relatórios que os próprios podem construir (Figura 26). Alguns relatórios mais complexos são construídos pelos analistas de BI para serem conferidos mais facilmente pelo usuário.

Paginado por: Ano: 2007							
UF	Escolaridade Métrica	NÃO INFORMADO	ANALFABETO	1º GRAU - INCOMPLETO	1º GRAU - COMPLETO	2º GRAU - COMPLETO	3º GRAU - COMPLETO
		nº	nº	nº	nº	nº	nº
	NÃO INFORMADO	8		2			
	RONDONIA	40.896	2.805	20.710	7.486	9.196	1.229
	ACRE	18.706	2.931	9.389	3.480	6.504	572
	AMAZONAS	55.550	4.395	34.727	13.108	28.341	2.892
	RORAIMA	34.201	31	24	14	12	7
	PARA	132.946	8.637	69.438	24.019	32.398	2.536
	AMAPA	1.348	1.387	7.077	2.984	6.720	545
	TOCANTINS	57.852	2.240	16.483	8.458	11.713	1.815
	MARANHAO	220.804	8.860	46.098	15.809	14.526	1.811
	PIAUÍ	169.993	14.456	66.026	23.675	31.657	5.095
	CEARA	534.812	7.348	46.444	12.684	14.756	2.298
	RIO GRANDE DO NORTE	161.559	5.474	39.835	12.877	18.755	1.989
	PARAIBA	241.053	3.793	28.909	8.734	10.547	1.505
	PERNAMBUCO	369.433	20.674	105.733	41.420	51.548	3.480
	ALAGOAS	11.411	1.703	4.629	1.077	1.097	84
	SERGIPE	23.464	1.326	6.080	1.194	1.208	117
	BAHIA	307.698	19.117	115.588	31.010	49.651	2.274
	MINAS GERAIS	1.004.398	18.905	155.467	46.018	58.639	9.036
	ESPIRITO SANTO	242.899	2.076	15.716	5.399	6.005	774
	RIO DE JANEIRO	618.775	11.990	38.278	16.137	14.869	1.144
	SÃO PAULO	1.735.119	36.350	358.467	163.400	223.221	27.988
	PARANA	523.200	9.399	86.300	31.798	39.931	6.088
	SANTA CATARINA	190.337	3.799	31.916	17.938	17.697	2.958
	RIO GRANDE DO SUL	405.182	5.743	79.363	20.459	22.586	2.052
	MATO GROSSO DO SUL	136.511	2.183	17.488	5.741	8.238	1.317
	MATO GROSSO	59.768	8.311	59.894	19.776	26.910	4.587
	GOIAS	251.686	1.441	18.402	5.481	6.461	730
	DISTRITO FEDERAL	73.658	998	10.535	5.313	8.247	877

Figura 26: Exemplo de relatório disponível para homologação do usuário.

5. Resultados

A partir de 1974, segundo TEIXEIRA e SÁ (1996), produziu-se no Brasil a saúde coletiva a saúde coletiva como campo de produção de saber e prática. Utilizando-se das informações coletadas em sistemas de informações fez-se o planejamento e administração em saúde que serviu de eixo aglutinador para objetos de investigação e propostas de intervenção social.

Com a criação do DW e uso de OLAP para expor o seu conteúdo o gestor consegue facilmente visualizar os indicadores comparando o seu conteúdo por município, por estado e nacional. Esse gestor poderá sem esforço escolher o período anual ou mensal das informações para criar uma série histórica.

O resultado final desse trabalho foi a criação na internet do “Painel de Indicadores do Câncer do Colo do Útero”. Esse painel abrange todos os indicadores especificados para auxiliar as ações nacionais referentes ao rastreamento do câncer do colo do útero.

Uma rápida visita ao site <http://www.inca.gov.br/painel> oferece ao leitor uma visualização do resultado deste trabalho (Figura 27).

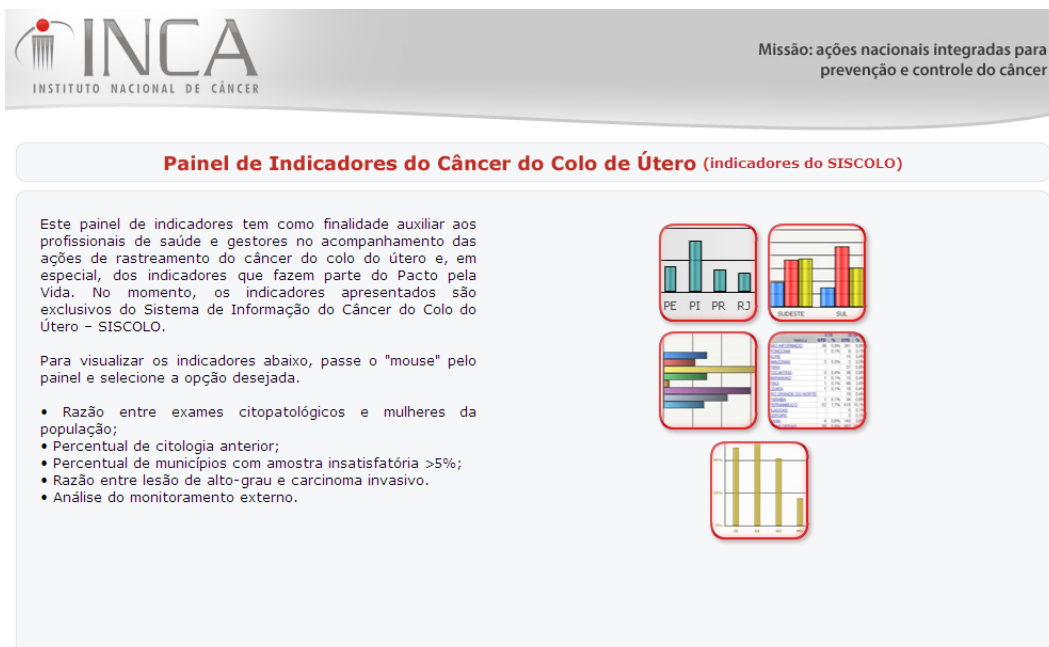


Figura 27: Painel de Indicadores - Site do INCA

Foram criados cinco painéis com os títulos abaixo:

1. Razão entre exames citopatológicos e mulheres da população;
2. Percentual de citologia anterior.
3. Percentual de municípios com amostra insatisfatória $\geq 5\%$.
4. Razão entre lesão de alto-grau e carcinoma invasivo.
5. Análise do monitoramento externo.

Informações que eram analisadas em relatórios estáticos ou pouco dinâmicos, após a implementação do BI, podem ser visualizadas em painéis que indicam, ao mesmo tempo, mais de um indicador, no período e localidade geográfica de escolha do usuário. Tornando a ferramenta um “self-service” de informações gerenciais de câncer do colo do útero.

Com o objetivo de ilustrar a riqueza de informação nos painéis, será explorado o painel de Razão entre exames citopatológicos e mulheres da população. Como se observa nesse painel (Figura 28) o gestor poderá selecionar ano e faixa etária para ser avaliada por estado comparando numericamente (em tabela do lado esquerdo) ou graficamente. Apresenta-se também no mesmo painel uma serie histórica de todos os anos consolidados no DW. O dinamismo do painel oferece ao gestor acurácia, clareza e fácil interpretação assim permitindo a correta análise de informações de saúde por uma faixa grande de usuários (BELL 2006). O maior desafio é fazer com que os painéis informem sem desinformar ou confundir o usuário, servindo então como ferramenta e não obstáculo de decisão. Existem varias evidências experimentais, (PINKER 1990), onde é sugerido que os gráficos oferecem aos humanos melhor compreensão e disponibilidade para razão dos fatos do que tabelas numéricas.

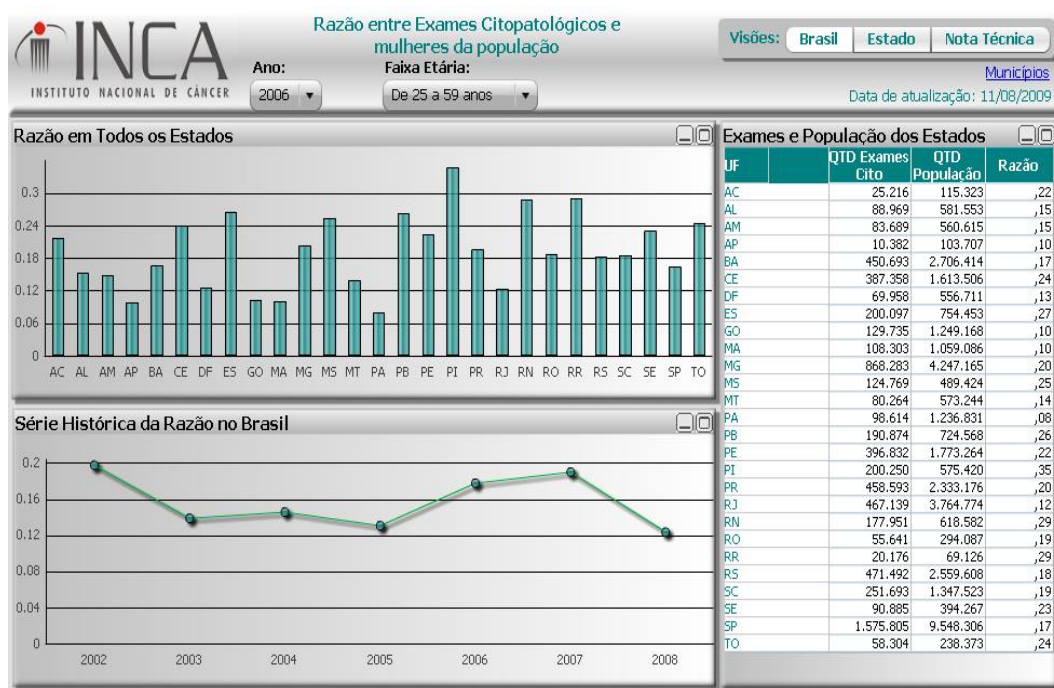


Figura 28: Exemplo Painel 01 - Brasil

Na visualização por estado (Figura 29) demonstra-se a mesma informação do painel anterior detalhado por região. Foi realizado, portanto o que se chama um *drill-down* ou detalhamento das informações de razão citopatológico e população.

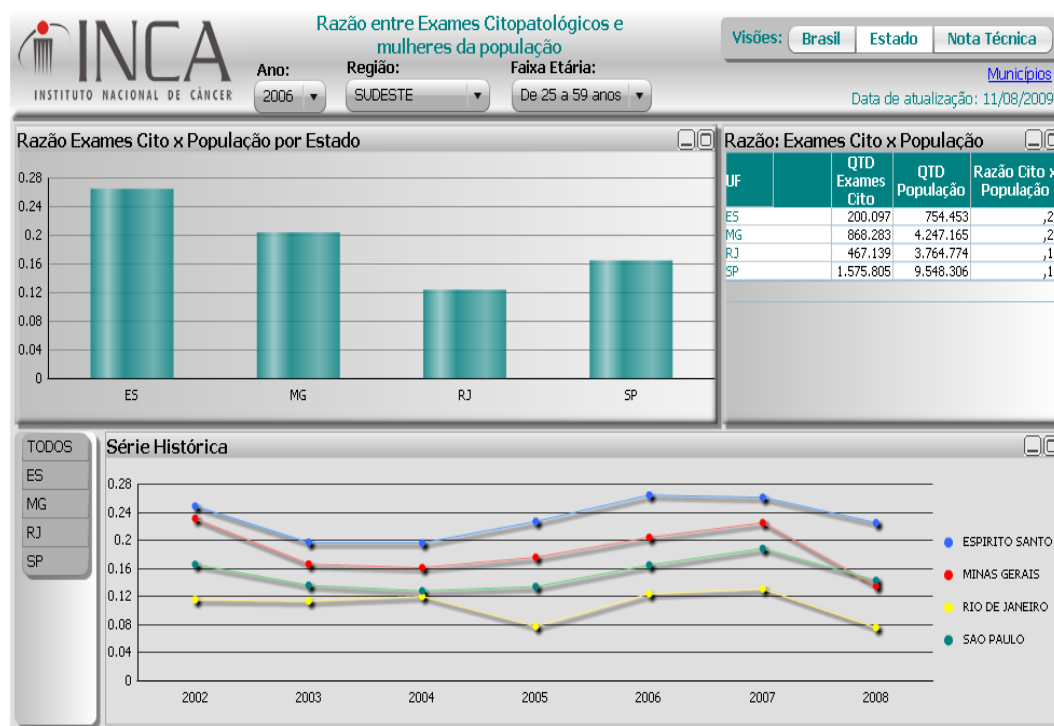


Figura 29: Exemplo Painel 01 - por Região

É importante ressaltar que anteriormente a esse trabalho as informações SISCOLO eram armazenadas em uma base relacional nacional, que causava baixo desempenho ao processar um relatório. Além disso, a visualização de diferentes faixas etária e/ou períodos num mesmo painel era inviável e se fosse possível demandaria experiência e tempo de um técnico de informática.

Os painéis de BI padronizam o resultado possibilitando cada nível hierárquico de informação ser comparada utilizando os mesmos cálculos e medidas.

Por conseqüências das dificuldades listadas acima, essas informações eram analisadas somente pela faixa etária preconizada (25 a 59 anos), com o novo aplicativo podemos visualizar as outras faixas etárias ou todas as idades (Figura 30) e assim utilizar toda a informação coletada para análise mais profunda da situação.

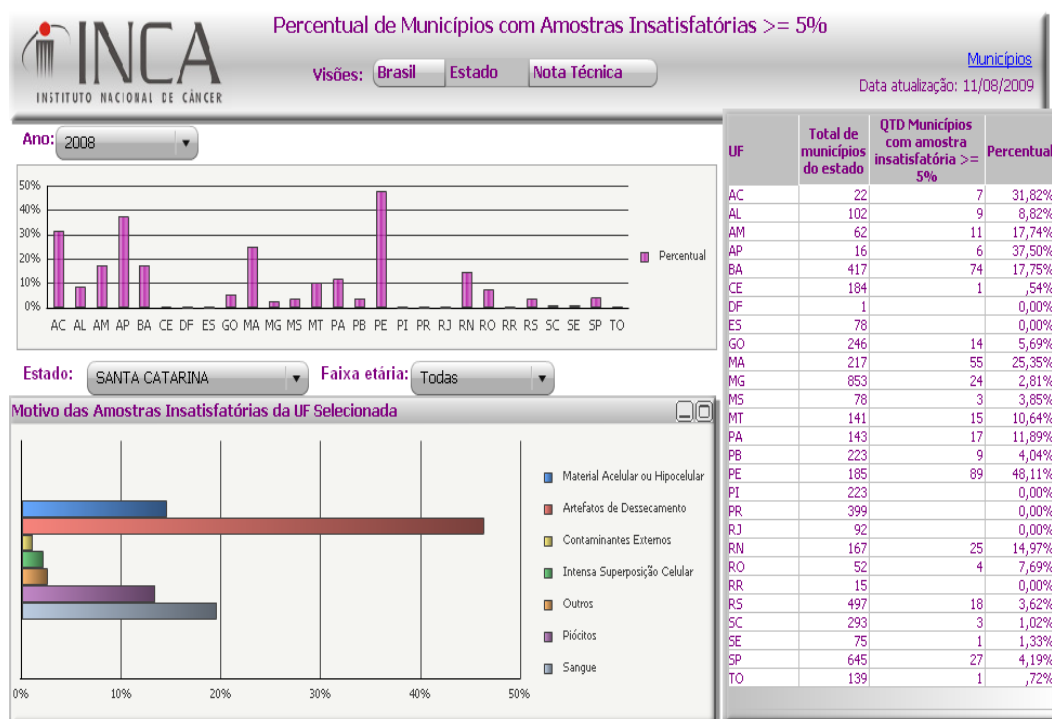


Figura 30: Exemplo Painel 03 - Todas as Faixas Etárias

O último gráfico representa as informações por estado por município em tabela (Figura 31). Podemos observar o total dos exames, a população (IBGE) e a razão entre os dois. Historicamente, a produção das informações em saúde em no Brasil deu-se de forma centralizada. Com o avanço do processo de municipalização dos serviços de saúde o estado, apesar dos níveis nacional e

estadual terem o papel de normatização dos sistemas e de fornecer apoio técnico e operacional, cabe aos municípios o desenvolvimento de suas atividades organizacionais. Para subsidiar esse trabalho o município era o que mais sofria com a falta de informação e acesso restrito e limitado. Esse painel possibilita uma transparência para o gestor municipal do seu próprio desenvolvimento e fazer uma análise comparativa a outros municípios, guardada as diferenças de população.

INCA INSTITUTO NACIONAL DE CÂNCER

Razão entre Exames Citopatológicos e mulheres da população

Visões: **Município** Nota Técnica

Ano: 2006 Faixa etária: De 25 a 59 anos Data de atualização: 11/08/2009

Razão Cito x População por Município

Município	Qtd Exames Cito	Qtd População	Razão Cito x População
CACAPAVA	9	18.864	0,00
PARAIBUNA	4	3.802	0,00
SALTINHO	3	1.490	0,00
UBATUBA	110	17.028	,01
GUARATINGUETA	372	26.784	,01
SANTA BRANCA	54	3.320	,02
SANTO ANTONIO DA ALEGRIA	26	1.302	,02
ITAPOLIS	364	9.353	,04
LEME	851	19.327	,04
MONGAGUA	440	9.758	,05
TORRE DE PEDRA	35	601	,06
SANTA ERNESTINA	67	1.120	,06
CAMPO LIMPO PAULISTA	1.079	17.249	,06
DOBRADA	89	1.413	,06
RINCAO	134	2.127	,06
LIMEIRA	4.102	64.983	,06
SÃO PEDRO	490	7.406	,07
PIRAPORA DO BOM JESUS	232	3.066	,08
JACI	80	1.055	,08
SÃO MANUEL	662	8.577	,08
SÃO VICENTE	6.120	76.950	,08
JABOTICABAL	1.374	16.976	,08
MONTE ARAZUEL	377	4.594	,08
CAIEIRAS	1.796	21.634	,08
IRACEMAPOLIS	365	4.177	,09
SANTA LUCIA	167	1.857	,09
ARIRANHA	185	1.932	,10
AMÉRICO BRASILIENSE	703	7.258	,10
MIGUELÓPOLIS	433	4.427	,10
PINDORAMA	301	3.067	,10
CEDRAL	165	1.670	,10
SÃO JOAQUIM DA BARRA	1.047	10.469	,10
PINDAMONHANGABA	3.210	31.773	,10

Figura 31: Exemplo Painel 01 - por município

O lançamento desse ambiente foi feito no final do ano de 2008 e contou com a presença de representantes de todas as secretarias estaduais do Brasil. Algumas discussões e sugestões foram feitas durante o workshop de apresentação, mas a aceitação do trabalho foi unânime. O acesso das secretarias estaduais e dos municípios tem sido constante.

Atualmente a direção do INCA assim como gestores municipais e estaduais tem utilizado e requisitado alterações para ampliar e aprofundar as análises já disponíveis.

6. Conclusões e Trabalhos Futuros

A criação de um ambiente BI com informações epidemiológicas do câncer do colo do útero é apenas o primeiro passo para se ter informações interativas e inteligentes de oncologia e saúde no Brasil.

Na medida em que outros *Data Marts* com informações de câncer forem criados o gestor da saúde poderá tomar decisões mais precisas sobre o recursos e foco de suas ações.

As informações poderão ser atualizadas periodicamente e compartilhados imediatamente permitindo uma avaliação mais precisa em menor tempo. As análises poderão ser comparadas nacionalmente, pois os critérios serão os mesmos. O usuário poderá criar seus próprios relatórios sem depender de profissionais da área de informática. Essas são alguns dos ganhos de se ter um ambiente BI no INCA.

Outros ambientes de consolidação de informação, como o Atlas de Mortalidade (Figura 32) surtiram efeitos positivos na disponibilidade de informações, mas, diferente do ambiente BI, a dificuldade de criar novos relatórios será sempre um impedimento para análises futuras.

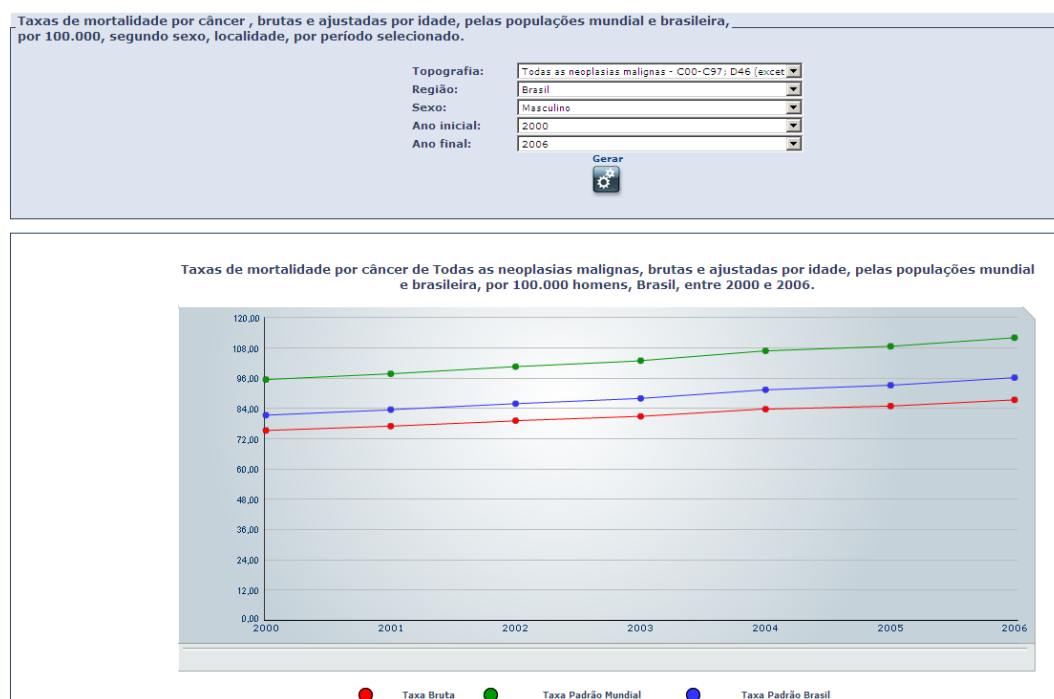


Figura 32: Atlas de Mortalidade por Câncer - BI sem DW

6.1. Outros Data Marts

As próximas fases da implantação do ambiente de BI no INCA serão a seguir:

Sistemas de autorização de procedimentos APAC / AIH que contém as informações sobre os procedimentos de alta complexidade e internação hospitalar e os dados para faturamento: dados do paciente, médico solicitante, unidade de saúde e localidade. Os principais cenários de exploração dos dados podem cruzar as informações de faturamento com os demais sistemas (Siscolo, Sismama e Registro de Câncer), visando fazer auditoria a quantidade de procedimentos faturados versus a quantidade de apresentada nos outros sistemas. Irá propiciar também maior controle sobre a qualidade das informações apresentadas.

Mortalidade que contempla as informações dos óbitos no país, como as causas, dados dos pacientes, região, estado e município. Permite mapear e controlar estatisticamente os principais motivos de óbito e demonstrar a proporção de óbitos por câncer versus demais causas no Brasil.

Um Registro de Câncer contém as notificações de câncer de um país e características do câncer (morfologia, localização do tumor, grau de evolução, dados do paciente e atendimento). Essa base permite mapear a incidência de câncer no país e sua correlação com região, morfologia e etc. O registro existente, e de gestão do INCA, que poderá compor esse *Data Mart* é o Registro de Base Populacional (RCBP). O cruzamento do RCBP e as estimativas de câncer poderão fornecer aos gestores informações passadas e futuras do mapeamento do câncer.

INCA atualmente faz a gestão de 4 hospitais que prestam assistência aos pacientes com câncer. Esses hospitais possuem um sistema de informação que contempla as informações sobre as internações hospitalares, os procedimentos médicos utilizados nos hospitais / ambulatorios do INCA e os dados do paciente. Possíveis análises podem acompanhar e controlar os procedimentos utilizados

em determinados tipos de internação comparativamente com outras, verificando possíveis inconsistências.

A área administrativa do INCA contempla os módulos de estoque e compra de materiais hospitalares do INCA. A criação de indicadores permitirá uma gestão mais efetiva e eficaz do estoque e do planejamento do setor de compras.

Após a implantação dos *Data Marts* acima citados, será possível avaliar a efetividade / eficácia do programa de prevenção na diminuição dos óbitos e nos registros nacionais de câncer; será possível ainda monitorar a qualidade do atendimento ambulatorial / hospitalar; melhorar a gestão de estoque e o planejamento de compras / suprimentos e conferir as informações de faturamento com os demais sistemas.

6.2. Data Mining

Apesar das informações clínicas estarem sendo acumuladas em grande quantidade poucas metodologias foram desenvolvidas e aplicadas para descobrir conhecimento oculto nestas bases. A técnica de *data mining* poderá ser aplicada a uma base de dados como a descrita acima adicionando algumas variáveis de tratamento e/ou custo de diagnóstico e tratamento de câncer. Tornando possível analisar padrões encontrados nos pacientes ou nas regiões registradas, ou traçar uma árvore de decisão com os tipos de tratamentos mais encontrados.

Como uma última nota é importante realçar que além de implantar um ambiente que permite de maneira inteligente tomar decisões gerenciais, a saúde no Brasil precisa de “implantar” a cultura de utilizar de meios computacionais para distribuição e análise de informações. Essa cultura irá modificar de modo significativo as políticas de saúde tirando o foco de ações localizadas e de escopo limitado para ações globais e visão abrangente.

Referências Bibliográficas³

- ALMEIDA, M. F. Descentralização de sistemas de informação e o uso das informações a nível municipal. Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo, IESUS, VII(3), jul/set 1998
- BELL, B.S.; HOSKINS, R.E.; PICKLE, L.W.; Wartenberg, D. Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public. International Journal of Health Geographics, 2006
- BRASIL. Ministério da Saúde. Norma Operacional Básica do Sistema Único de Saúde – NOB 1/96: portaria nº 2.203, 5 de novembro de 1996, DO 216, de 6/11/96.
- BRASIL. Ministério da Saúde / Instituto Nacional de Câncer (INCA). Nomenclatura Brasileira para Laudos Cervicais e Condutas Preconizadas (Recomendações para Profissionais de Saúde). 2.ed. Rio de Janeiro, 2006.
- BROWN, M. L.; RILEY, G. F.; POTOSKY, A. L.; ETZIONI, R. D. Obtaining long-term disease specific costs of care: application to Medicare enrollees diagnosed with colorectal cancer. Applied Research Program, National Cancer Institute, Bethesda, Maryland 20892-7344, USA.
- CAETANO, R.; MANSO C.; THULER, L. C.; GIRIANELLI, V. Custo-efetividade no diagnóstico precoce do câncer de colo uterino no Brasil. PHYSIS: Rev. Saúde Coletiva, Rio de Janeiro, 16(1):99-118, 2006.
- CODD, E. F.; CODD S. B.; SALLEY C. T. Providing OLAP to User-Analysts: An IT Mandate. 1993.
- DATE, C.J. An Introduction to Database Systems. 2003. 8th ed. Addison-Wesley
- INSTITUTO NACIONAL DE CÂNCER (INCA). Disponível em: <<http://www.inca.gov.br>>. Acesso em: jan. 2009.
- INSTITUTO NACIONAL DE CÂNCER (INCA). Estimativa 2008: incidência de câncer no Brasil. Rio de Janeiro. 2008.
- INTERNATIONAL BUSINESS MACHINES CORPORATION Centro de Data Warehouse Manual de Administração (DB2) versão 7 1996. Universal Database Disponível em: <<ftp://ftp.software.ibm.com/ps/products/db2/info/vr7/pdf/letter/nlv/db2ddb71.pdf>>. Acesso em: abr. 2009.
- LAURENTI, R. Análise da Informação em saúde: 1893-1993, cem anos da Classificação Internacional de Doenças. Revista de Saúde Pública, nº 6 São Paulo Dec1991 - SciELO Public Health.

³ Baseada na NBR-6023 de ago. de 2002, da Associação Brasileira de Normas Técnicas (ABNT). Abreviatura dos títulos dos periódicos em conformidade com o MEDLINE.

- O'CARROLL, P. W.; YASNOFF, W. A.; WARD, M. E.; RIPP, L. H.; MARTIN, E. L. Public health informatics and information systems Edition: illustrated Publicado por Springer, (ISBN 0387954740, 9780387954745) 2003
- PINKER, S. A Theory of Graph Comprehension. Artificial intelligence and the future of testing, Massachusetts Institute of Technology 1990
- RIBEIRO, J. I. HYGÉIA – Sistema de Acompanhamento da Saúde da Família e Solução de Apoio a Tomada de Decisão para os Gestores Públicos de Saúde Municipal Centro Universitário Feevale – Instituto de Ciências Exatas e Tecnológicas, Novo Hamburgo, set 2008.
- SANTOS, R.; GUTIERREZ, M. A. MINERSUS – Ambiente computacional para extração de informações para a gestão da saúde pública por meio da mineração dos dados do SUS. Revista Brasileira de Engenharia Biomédica, v. 24, n. 2, p. 77-90, agosto 2008© SBEB - Sociedade Brasileira de Engenharia Biomédica ISSN 1517-3151
- SANTOS, R. Ambiente para Extração de Informações através da Mineração das Bases de Dados do Sistema Único de Saúde. 2007. 254p. Tese (Programa de Pós-graduação em Informática em Saúde) - Universidade Federal de São Paulo. Escola Paulista de Medicina.. São Paulo, 2007.
- SATHER, J. The Assertive Cancer Patient: Living with cancer—and an attitude. Disponível em: < <http://www.assertivepatient.com/2007/02/index.html>>. Acesso em: 14 fev. 2007.
- SILVA, R. M.; LINARD, A.; DANTES E SILVA, F. A. Mulheres submetidas a tratamento para câncer de colo uterino - percepção de como enfrentam a realidade. Revista Brasileira de Cancerologia, 2002.
- TEIXEIRA, C; SÁ, M. C. Planejamento & gestão em saúde: situação atual e perspectivas para a pesquisa, o ensino e a cooperação técnica na área. Ciência e Saúde Coletiva 1(1): 82-103. 1996.
- WIKIPÉDIA Disponível em: <<http://pt.wikipedia.org>>. Acesso em: jan. 2009.
- ZEFERINO, L. C. O desafio de reduzir a mortalidade por câncer do colo do útero. Rev Bras Ginecol Obstet. 2008;30(5): 213-5 .