

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM SAÚDE COLETIVA**

Camila Soares Lima Corrêa

**Avaliação do rastreamento do câncer do colo do útero em Minas Gerais a partir
de dados do SISCOLO**

Juiz de Fora
2021

Camila Soares Lima Corrêa

**Avaliação do Rastreamento do Câncer do Colo do Útero em Minas Gerais a
partir de Dados do SISCOLO**

Tese apresentada ao Programa de Pós-Graduação em Saúde Coletiva da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutora em Saúde Coletiva.
Área de Concentração: Saúde Coletiva.

Orientadora: Prof^a Dra. Maria Teresa Bustamante-Teixeira

Coorientador: Prof. Dr. Mário Círio Nogueira

Juiz de Fora

2021

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Soares Lima Corrêa, Camila.

Avaliação do rastreamento do câncer do colo do útero em Minas Gerais a partir de dados do SISCOLO / Camila Soares Lima Corrêa. -- 2021.

244 p. : il.

Orientadora: Maria Teresa Bustamante-Teixeira

Coorientador: Mário Círio Nogueira

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Medicina. Programa de Pós-Graduação em Saúde Coletiva, 2021.

1. Neoplasias do Colo do Útero. 2. Programas de Rastreamento. 3. Sistemas de Informações em Saúde. I. Bustamante-Teixeira ; Maria Teresa, orient. II. Círio Nogueira , Mário, coorient. III. Título.

Camila Soares Lima Corrêa

**Avaliação do Rastreamento do Câncer do Colo do Útero em Minas Gerais a partir de
Dados do SISCOLO.**

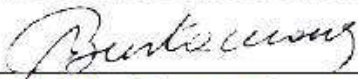
Tese apresentada ao Programa de Pós-
graduação em Saúde Coletiva da Universidade
Federal de Juiz de Fora como requisito parcial
à obtenção do título de Doutora em Saúde
Coletiva.
Área de Concentração: Saúde Coletiva.

Aprovada em 08 de junho de 2021

BANCA EXAMINADORA



Prof. Dra. Maria Teresa Bustamante Teixeira - Orientadora
Universidade Federal de Juiz de Fora



p/ Prof. Dr. Mário Círio Nogueira – Coorientador
Universidade Federal de Juiz de Fora



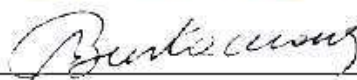
p/ Prof. Dra. Gulnar Azevedo e Silva
Universidade do Estado do Rio de Janeiro



p/ Prof. Dra. Jeane Glauca Tomazelli
Instituto Nacional do Câncer



p/ Prof. Dra. Joan Leigh Warren
National Cancer Institute - EUA



p/ Prof. Dra. Isabel Cristina Gonçalves Leite
Universidade Federal de Juiz de Fora

DEDICATÓRIA

Dedico este trabalho, com muito amor e gratidão, aos meus pais, Sérgio e Penha, por me mostrarem o valor dos estudos e por nunca terem medido esforços para o meu crescimento acadêmico e profissional.

AGRADECIMENTOS

A Deus, pela oportunidade de fazer o doutorado, por tantas bênçãos nesse período e por colocar tantas pessoas maravilhosas no meu caminho, que foram fundamentais em todo o processo!

Aos meus pais, Sérgio e Penha, e à minha irmã, Carina, pelas orações, incentivo e por serem o meu porto seguro. Ao meu amor, Pedro Augusto, que ao longo desse processo, foi namorado, noivo e, atualmente, esposo. Pelo companheirismo, paciência, por tantos ensinamentos e, principalmente, por estar sempre ao meu lado, me dando força, incentivando e transmitindo paz e tranquilidade.

Aos meus orientadores, Maria Teresa Bustamante e Mário Círio, por todo ensinamento e crescimento que me proporcionaram nesse período! Teíta, é uma orientadora com coração de mãe, que além de me ensinar muito sobre epidemiologia, me acolheu e me deu muito incentivo e força nos momentos mais desafiadores. Muito obrigada por todo apoio, minha querida! Mário, sem você, nada disso seria possível! Obrigada pela disponibilidade, por nos ensinar tanto sobre estatística, pelos *scripts*, paciência e por tanto empenho na pesquisa!

À Daniela de Almeida Pereira, minha dupla, meu maior presente do doutorado! Uma pessoa que me inspira e é um exemplo para mim! Obrigada pela amizade, por trazer leveza e alegria a esta caminhada, por tanto conhecimento compartilhado e por dividir essa jornada comigo! À Vívian Assis Fayer, pela parceria, por ter desbravado o caminho metodológico, ter conseguido os bancos e nos ensinado sobre o relacionamento entre as bases, que possibilitaram a realização dessa pesquisa. À Jeane Tomazelli e Caroline Madalena, pela grande disponibilidade em esclarecer nossas dúvidas sobre os sistemas de informação e metodologia adotada pelo grupo de pesquisa.

Às professoras componentes da banca, Gulnar Azevedo e Silva, Jeane Gláucia Tomazelli, Isabel Cristina Gonçalves Leite e Joan Leigh Warren, pela disponibilidade e grande contribuição com os questionamentos, reflexões e sugestões na qualificação. Foram fundamentais para a melhoria desse trabalho.

Ao Professor Alfredo Chaoubah, pela parceria e por nos possibilitar utilizar o servidor da Faculdade de Estatística da UFJF para rodar o relacionamento entre as bases de dados. A todos os funcionários do Nates, especialmente ao Anderson, pela disponibilidade e ajuda para rodar o relacionamento no servidor do Nates.

Educação não transforma o mundo.

Educação muda as pessoas.

Pessoas transformam o mundo.

Paulo Freire

RESUMO

O objetivo desta tese foi avaliar as ações de rastreamento do câncer do colo do útero (CCU) pelo Sistema Único de Saúde (SUS), em Minas Gerais e sua conformidade com as recomendações do Ministério da Saúde. Para isso, foram realizados três estudos: no primeiro, as ações de rastreamento foram avaliadas por meio de um estudo descritivo com base em dados agregados do SISCOLO, referentes ao período de 2006 a 2011, disponíveis no DATASUS. Foram analisados indicadores relacionados às características da oferta de exames citopatológicos, à qualidade do exame, à ocorrência de alterações celulares e ao seguimento informado dos casos de lesão de alto grau em Minas Gerais. Verificou-se que a razão de exames citopatológicos do colo do útero em mulheres de 25 a 59 anos manteve-se estável em Minas Gerais de 2006 a 2011. Aproximadamente 75% dos exames foram realizados em mulheres de 25 a 59 anos e houve progressiva redução na proporção de citopatológicos do colo do útero sem citologia anterior. Em média, 51,2% dos exames foram realizados no período de até um ano. Quanto à qualidade do exame, destaca-se o índice de positividade, categorizado como baixo durante todo o período analisado. O segundo estudo consistiu em um estudo descritivo, de coorte não-concorrente, baseado em dados identificados do SISCOLO com objetivo de estimar a cobertura do rastreamento do câncer do colo do útero em 2008 e a proporção de sobrerastreamento por periodicidade e faixa etária no estado. A cobertura foi calculada utilizando exames de mulheres de 25 a 59 anos e 25 a 64 anos após exclusão de duplicidades e separação dos exames de seguimento, ficando apenas um registro de exame por mulher nesse ano. O sobrerastreamento por faixa etária correspondeu ao percentual de mulheres com <25 e ≥ 65 anos. O sobrerastreamento por periodicidade foi avaliado a partir do seguimento de mulheres de 25 a 64 anos submetidas ao rastreamento em 2008 através do relacionamento probabilístico entre as bases de 2008 a 2012 (seguimento 2008-2012). O tempo até o próximo exame citopatológico foi definido como o intervalo entre a data de coleta do primeiro exame normal com amostra satisfatória realizado em 2008 e a data de coleta da primeira repetição deste exame. Para mulheres que iniciaram o rastreamento em 2008, o sobrerastreamento por periodicidade foi definido como a repetição em um período inferior a 9 meses e, para aquelas que iniciaram o rastreamento previamente, a repetição em um período inferior a 30 meses. Verificou-se que, em 2008, 1.211.014 mulheres foram submetidas ao rastreamento do CCU pelo SUS em Minas Gerais, sendo que 75,8% estavam na faixa etária de 25 a 59 anos e 79,8%, na faixa etária de 25 a 64 anos. A cobertura foi de 77,6% na faixa etária de 25 a 59 anos e de 76,2%, de 25 a 64 anos. Identificou-se o seguimento de 1.162.405 mulheres sendo o sobrerastreamento por periodicidade de 3,9% para aquelas que iniciaram o rastreamento em 2008 e de 62,6% para as que tinham exame prévio. O sobrerastreamento por faixa etária correspondeu a 20,2% do total de mulheres submetidas ao rastreamento em 2008, sendo que, em mulheres que estavam iniciando o rastreamento em 2008 foi de 49,6% e, naquelas com exame prévio, foi de

14,9%. No terceiro estudo, os indicadores obtidos no segundo foram utilizados para a realização de um estudo ecológico, cujas unidades de análise foram as 77 microrregiões de saúde de Minas Gerais. Tal estudo teve como objetivo avaliar a desigualdade regional da cobertura do rastreamento do câncer do colo do útero, em 2008, em Minas Gerais e sua associação com indicadores socioeconômicos e assistenciais, e descrever a distribuição espacial do sobrerastreamento nas microrregiões do estado. Os resultados indicaram que as microrregiões que apresentaram as menores coberturas estavam concentradas, em sua maioria, nas macrorregiões Nordeste, Jequitinhonha, Leste e Leste do Sul, que se caracterizam por piores indicadores socioeconômicos. O IDH apresentou associação positiva com a cobertura. Os sobrerastreamento por periodicidade e por faixa etária apresentaram padrão diferenciado de distribuição espacial entre si e entre mulheres que estão iniciando o rastreamento e aquelas que têm exame prévio. Pode-se concluir, então, que apesar de a cobertura no estado estar próxima da meta de 80%, havia desigualdade regional no acesso ao exame, com áreas de melhor nível socioeconômico apresentando maiores coberturas. Indicadores de qualidade do exame estavam abaixo da meta e o sobrerastreamento era elevado no estado. Quanto à periodicidade, o sobrerastreamento foi mais importante em mulheres que deveriam fazer o exame a cada três anos e quanto à faixa etária, em mulheres que estavam iniciando o rastreamento. Os resultados corroboram o caráter oportunístico do rastreamento do câncer do colo do útero e evidenciam a necessidade de aprimoramento do programa no estado de Minas Gerais. Ressalta-se que o presente trabalho possibilitou o desenvolvimento de uma metodologia que tornou possível a estimativa da cobertura e a realização do seguimento de mulheres submetidas ao rastreamento utilizando dados do SISCOLO. Tal metodologia pode ser utilizada em outros contextos com objetivo de monitorizar as ações de rastreamento e contribuir para o aperfeiçoamento do programa de rastreamento.

Palavras-chave: Neoplasias do Colo do Útero; Programas de Rastreamento; Sistemas de Informações em Saúde.

ABSTRACT

The objective of this thesis was to evaluate the cervical cancer screening actions by the Public Health System (SUS), in Minas Gerais and its compliance with the recommendations of the Ministry of Health. For this purpose, three studies were carried out: in the first, the actions of screening were evaluated through a descriptive study based on aggregated data from SISCOLO, referring to the period from 2006 to 2011, available at DATASUS. Indicators related to the characteristics of the offer of cytopathological exams, the quality of the exam, the occurrence of cellular changes and the informed follow-up of high-grade injury cases in Minas Gerais were analyzed. It was found that the ratio of cervical cytopathological examinations in women aged 25 to 59 years remained stable in Minas Gerais from 2006 to 2011. Approximately 75% of the examinations were performed in women aged 25 to 59 years and there was a progressive reduction in the proportion of cervical cytopathology without previous cytology. On average, 51.2% of the examinations were performed within a period of up to one year. Regarding the quality of the exam, the positivity index stands out, categorized as low throughout the analyzed period. The second study consisted of a descriptive, non-concurrent cohort study, based on identified data from SISCOLO in order to estimate the coverage of cervical cancer screening in 2008 and the proportion of over-screening by periodicity and age group in the state. Coverage was calculated using examinations of women aged 25 to 59 and 25 to 64 years after excluding duplicates and separating the follow-up exams, leaving only one exam record per woman in that year. Over-screening by age group corresponded to the percentage of women aged <25 and ≥ 65 years. Over-screening due to excessive periodicity was assessed based on the follow-up of women aged 25 to 64 years who were screened in 2008 through the probabilistic relationship between the bases from 2008 to 2012 (follow-up 2008-2012). The time until the next cytopathologic exam was defined as the interval between the date of the first normal exam with a satisfactory sample carried out in 2008 and the date of the first repetition of this exam. For women who started screening in 2008, over-screening by periodicity was defined as repetition in less than 9 months and, for those who started screening previously, repetition in less than 30 months. It was found that, in 2008, 1,211,014 women underwent CC screening by SUS in Minas Gerais, with 75.8% in the 25 to 59 age group and 79.8% in the 25 age group to 64 years. Coverage was 77.6% in the age group from 25 to 59 years old and 76.2%, from 25 to 64 years old. The follow-up of 1,162,405 women was identified, with over-screening at 3.9% intervals for those who started screening in 2008 and 62.6% for those who had a previous examination. Over-screening by age group corresponded to 20.2% of the total number of women submitted to screening in 2008, whereas in women who were starting screening in 2008 it was 49.6% and, in those with prior exam, it was 14.9%. In the third study, the indicators obtained in the second were used to carry out an ecological study, whose units of analysis were the 77 health micro-regions of Minas Gerais. This study aimed to assess the regional inequality of coverage of cervical cancer screening in 2008 in Minas Gerais and its association with

socioeconomic and assistance indicators, and to describe the spatial distribution of over-screening in the state's micro-regions. The results indicated that the microregions with the lowest coverage were mostly concentrated in the Northeast, Jequitinhonha, Leste and Leste do Sul macroregions, which are characterized by the worst socioeconomic indicators. The HDI showed a positive association with coverage. The over-screening by periodicity and by age group showed a different pattern of spatial distribution among themselves and among women who are starting the screening and those who have had a previous examination. It can be concluded, then, that although the coverage in the state is close to the 80% target, there was regional inequality in access to the exam, with areas of better socioeconomic level presenting greater coverage. Exam quality indicators were below target and over-screening was high in the state. Regarding the frequency, the over-screening was more important in women who should have the exam every three years and in terms of age, in women who were starting the screening. The results corroborate the opportunistic nature of cervical cancer screening and highlight the need to improve the program in the state of Minas Gerais. It is noteworthy that the present study enabled the development of a methodology that made it possible to estimate coverage and follow up women submitted to screening using data from SISCOLO. Such methodology can be used in other contexts with the objective of monitoring the actions and contributing to the improvement of the screening program.

Keywords: Uterine Cervical Neoplasms; Mass Screening; Health Information Systems;

LISTA DE ILUSTRAÇÕES

FIGURA 1	Curva <i>precision-recall</i> exemplificando um <i>linkage</i> perfeito	53
FIGURA 2	Limpezas da base 2008 do SISCOLO e formação do <i>Banco 2008</i>	62
FIGURA 3	Limpeza no <i>Banco 2008</i> e formação do <i>Banco Referência</i>	68
FIGURA 4	Construção do <i>Banco Seguimento (2008-2012)</i>	71
FIGURA 5	Microrregiões de saúde de Minas Gerais	78
FIGURA 6	IDH por macrorregião de Saúde em Minas Gerais, 2010	79
FIGURA 7	Variáveis explicativas do estudo separadas em dois blocos hierárquicos (distal e proximal)	80
ARTIGO 2		
FIGURA 1	Limpezas da base 2008 do SISCOLO e formação do <i>Banco 2008</i>	110
FIGURA 2	Construção do <i>Banco Seguimento (2008-2012)</i>	112
FIGURA 3	<i>Kaplan-Meier</i> do percentual cumulativo de mulheres sobrerastreadas estratificado por faixa etária no grupo de mulheres que (A) iniciou o rastreamento em 2008 e no grupo que (B) iniciou antes de 2008	118
ARTIGO 3		
FIGURA 1	IDH por macrorregião de Saúde em Minas Gerais, 2010	131
FIGURA 2	Variáveis explicativas da cobertura do rastreamento separadas em dois blocos hierárquicos (distal e proximal)	133
FIGURA 3	Distribuição espacial da cobertura do câncer do colo do útero nas microrregiões de saúde de Minas Gerais em mulheres de 25 a 59 anos e 25 a 64 anos e respectivo Moran Local, 2008	138
FIGURA 4	Distribuição espacial do sobrerastreamento nas microrregiões de saúde de Minas Gerais, para mulheres	142

que iniciaram o rastreamento em 2008 e para aquelas que iniciaram previamente

- | | | |
|-----------------|--|-----|
| FIGURA 5 | Moran Local do soberrastreamento nas microrregiões de saúde de Minas Gerais, para mulheres que iniciaram o rastreamento em 2008 e para aquelas que iniciaram previamente | 143 |
| FIGURA 6 | Distribuição espacial das variáveis explicativas do bloco distal (socioeconômicas) | 147 |
| FIGURA 7 | Distribuição espacial das variáveis explicativas do bloco proximal (financiamento e assistenciais) | 148 |

LISTA DE QUADROS

QUADRO 1	Dezesseis passos propostos inicialmente para serem empregados nos relacionamentos probabilísticos, com respectivas variáveis de blocagem e de pareamento	63
QUADRO 2	Critérios para classificação manual dos pares formados no relacionamento probabilístico	64
QUADRO 3	Descrição dos passos utilizados no relacionamento entre o <i>Banco 2008</i> e o banco de mulheres que tiveram exames alterados/rejeitados em 2007 e os respectivos parâmetros e pares verdadeiros identificados	66
QUADRO 4	Limpezas das bases de 2009 a 2012 e formação dos respectivos bancos anuais	70
QUADRO 5	Pares verdadeiros identificados em cada passo dos relacionamentos entre o banco referência e cada banco anual de 2009 a 2012	73
QUADRO 6	Descrição do numerador, denominador e cálculo da cobertura do rastreamento do câncer do colo do útero	74
QUADRO 7	Variáveis explicativas: indicadores sociodemográficos, econômicos e assistenciais, ano de referência e fonte dos dados	81
ARTIGO 1		
QUADRO 1	Descrição dos indicadores calculados e respectivas fontes dos dados	89
ARTIGO 2		
QUADRO 1	Descrição da formação dos <i>Bancos de comparação</i> referentes aos anos 2009 a 2012	114
ARTIGO 3		
QUADRO 1	Variáveis explicativas: indicadores sociodemográficos, econômicos e assistenciais, ano de referência e fonte dos dados	134

LISTA DE TABELAS

ARTIGO 1

TABELA 1	Razão de exames citopatológicos em mulheres de 25 a 59 anos nas macrorregiões de saúde de Minas Gerais e número de exames e população feminina na faixa etária alvo do rastreamento em Minas Gerais, 2006-2011	92
TABELA 2	Proporção de exames citopatológicos do colo do útero sem citologia anterior em mulheres de 25 a 59 anos, por ano, em Minas Gerais e suas macrorregiões de saúde. Minas Gerais, 2006-2011	93
TABELA 3	Índice de positividade de exames citopatológicos do colo do útero, segundo ano, em Minas Gerais e suas macrorregiões de saúde. Minas Gerais, 2006 a 2011	95
TABELA 4	Proporção de seguimento informado de mulheres com lesão intraepitelial de alto grau do colo do útero, segundo o ano de diagnóstico, em Minas Gerais e suas macrorregiões de saúde. Minas Gerais, 2006-2011	96

ARTIGO 2

TABELA 1	Cobertura do rastreamento do câncer do colo do útero, por faixa etária e macrorregiões de saúde. Minas Gerais, 2008	116
TABELA 2	Distribuição do tempo em meses entre a data de coleta do primeiro exame realizado em 2008 e a data da primeira repetição	117

ARTIGO 3

TABELA 1	Distribuição da cobertura do rastreamento do câncer do colo do útero em mulheres de 25 a 64 anos, por microrregião de saúde de Minas Gerais, para cada macrorregião e para o estado, 2008	139
TABELA 2	Distribuição do sobrerastreamento por faixa etária em mulheres que iniciaram o rastreio em 2008, por	144

microrregião de saúde de Minas Gerais, para cada macrorregião e para o estado

TABELA 3	Distribuição do sobrerastreamento por periodicidade em mulheres que iniciaram o rastreo antes de 2008, por microrregião de saúde de Minas Gerais, para cada macrorregião e para o estado.	145
TABELA 4	Autocorrelação espacial (Índice de Moran Global)	149
TABELA 5	Correlação de Pearson entre as variáveis explicativas e a cobertura do rastreamento	150
TABELA 6	Modelos de regressão simples entre a cobertura do rastreamento do câncer do colo do útero e as variáveis de exposição para as microrregiões de saúde de Minas Gerais, 2008	151
TABELA 7	Modelos múltiplos com as variáveis explicativas do bloco distal	152
TABELA 8	Modelos múltiplos com a inserção de variáveis do bloco proximal	153

LISTA DE ABREVIATURAS E SIGLAS

AAPC	Varição percentual anual média
ANS	Agência Nacional de Saúde Suplementar
ASC	Atipias de significado indeterminado em células escamosas
ASC-H	Atipias de significado indeterminado em células escamosas - não sendo possível excluir lesão intraepitelial de alto grau
ASC-US	Atipias de significado indeterminado em células escamosas - possivelmente não neoplásicas
CCU	Câncer do colo do útero
CNS	Número do cartão nacional de saúde
CPF	Número do registro no cadastro de pessoas físicas
DATASUS	Departamento de Informática do Sistema Único de Saúde
ESF	Estratégia Saúde da Família
EUA	Estados Unidos da América
GLOBOCAN	<i>Estimated Cancer Incidence, Mortality and Prevalence Worldwide</i>
GBD	<i>Global Burden of Disease</i>
GPTS	<i>Gasto Público Total em Saúde</i>
HSIL	Lesão intraepitelial escamosa de alto grau
HPV	Papilomavírus humano
IC95%	Intervalo de confiança de 95%
IDH	Índice de Desenvolvimento Humano
INCA	Instituto Nacional de Câncer José Alencar Gomes da Silva
IP	Índice de positividade
IVS	Índice de Vulnerabilidade na Saúde
LSIL	Lesão intraepitelial escamosa de baixo grau
MG	Minas Gerais
NIC	Neoplasia Intraepitelial Cervical
PIB	Produto Interno Bruto
RHC	Registros Hospitalares de Câncer
SIA	Sistema de Informações Ambulatoriais

SIS	Sistemas de Informação em Saúde
SISCAN	Sistema de Informação do Câncer
SISCOLO	Sistema de Informação do Câncer do Colo do Útero
SISMAMA	Sistema de Informação do Câncer de Mama
SISPACTO	Sistema de Informação do Pacto pela Saúde
SUS	Sistema Único de Saúde
OR	<i>Odds ratio</i>
QualiCito	Qualificação Nacional em Citopatologia
VPP	Valor preditivo positivo
ZT	Zona de transformação
WHO	<i>World Health Organization</i>

SUMÁRIO

1 INTRODUÇÃO	21
2 REFERENCIAL TEÓRICO	23
2.1 EPIDEMIOLOGIA DO CÂNCER DO COLO DO ÚTERO.....	23
2.2 HISTÓRIA NATURAL DA DOENÇA E FATORES DE RISCO DO CÂNCER DO COLO DO ÚTERO.....	25
2.3 PREVENÇÃO E CONTROLE DO CÂNCER DO COLO DO ÚTERO	27
2.3.1 Prevenção primária	27
2.3.2 Prevenção secundária	29
2.4 HISTÓRICO DAS PRINCIPAIS AÇÕES DE CONTROLE DO CCU NO BRASIL	32
2.5 AVALIAÇÃO DE INDICADORES RELACIONADOS AO RASTREAMENTO DO CÂNCER DO COLO DO ÚTERO NO BRASIL E FATORES ASSOCIADOS	35
2.6 RASTREAMENTO OPORTUNÍSTICO E SOBERRASTREAMENTO	42
2.7 SISTEMAS DE INFORMAÇÃO EM SAÚDE E METODOLOGIAS DE RELACIONAMENTO DE BASES DE DADOS.....	47
3 JUSTIFICATIVA	54
4 OBJETIVOS	56
4.1 GERAL.....	56
4.2 ESPECÍFICOS.....	56
5. MÉTODOS	57
5.1 MÉTODOS REFERENTES AO OBJETIVO ESPECÍFICO 1 (ARTIGO 1).....	57
5.2 MÉTODOS REFERENTES AO OBJETIVO ESPECÍFICO 2 (ARTIGO 2).....	57
5.2.1 Tipo de estudo e aspectos éticos	57
5.2.2 Fonte dos dados	57
5.2.3 População de estudo	58

5.2.4 Variáveis	58
5.2.5 Bases de dados utilizadas.....	58
5.2.6 Limpezas da base de 2008 do SISCOLO e formação do <i>Banco 2008</i>	59
5.2.7 Preparo e formação dos bancos de comparação.....	69
5.2.8 Construção do <i>Banco Seguimento (2008-2012)</i>	71
5.2.8 Análise dos dados.....	73
5.3 MÉTODOS REFERENTES AO OBJETIVO ESPECÍFICO 3 (ARTIGO 3)	77
5.3.1 Tipo de estudo e fonte dos dados	77
5.3.2 População de estudo	77
5.3.3 Variáveis de estudo.....	80
5.3.4 Análise dos dados.....	83
6 RESULTADOS E DISCUSSÃO	85
6.1 – ARTIGO 1: RASTREAMENTO DO CÂNCER DO COLO DO ÚTERO EM MINAS GERAIS: AVALIAÇÃO A PARTIR DE DADOS DO SISTEMA DE INFORMAÇÃO DO CÂNCER DO COLO DO ÚTERO (SISCOLO)	85
6.2 – ARTIGO 2: COBERTURA E SOBERRASTREAMENTO DO EXAME CITOPATOLÓGICO DO COLO DO ÚTERO EM MINAS GERAIS	105
6.3 – ARTIGO 3: DESIGUALDADES REGIONAIS E FATORES ASSOCIADOS NO RASTREIO DO CÂNCER DE COLO DE ÚTERO EM MINAS GERAIS	127
7 CONSIDERAÇÕES FINAIS	163
REFERÊNCIAS.....	165
APÊNDICES.....	178

1 INTRODUÇÃO

O câncer do colo do útero (CCU), também denominado de câncer cervical, é o quarto tipo de câncer mais comum em mulheres de todo o mundo, sendo que 85% dos casos e mortes pela doença ocorrem em países em desenvolvimento (FERLAY, 2015). Em 2020, a doença foi responsável por 342 mil mortes no mundo, representando 7,7% das mortes pelos 10 principais tipos de câncer feminino, ficando atrás apenas do câncer de mama (15,5%), pulmão (13,7%) e colorretal (9,5%) (SUNG *et al.*, 2021).

A mortalidade por câncer do colo do útero é inversamente relacionada ao índice de desenvolvimento humano (IDH), ao índice de desigualdade de gênero, a fatores socioeconômicos e ao gasto em saúde em uma determinada população (SINGH; AZUINE; SIAHPUSH, 2012). Nos países em desenvolvimento, o CCU representa o segundo tipo mais diagnosticado e a terceira principal causa de morte por câncer em mulheres (TORRE *et al.*, 2015).

No Brasil, estimativas do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA) apontam 16.590 casos novos da doença para cada ano do triênio 2020-2022. Excluindo o câncer de pele não melanoma, o CCU é o terceiro tipo de câncer mais frequente nas mulheres brasileiras (7,4% de todos os tipos de câncer), ficando atrás apenas do câncer de mama (29,7%) e de cólon (9,2%) - (INCA, 2019). Quanto à mortalidade pela doença, no período entre 1978 e 2017, houve declínio em todas as regiões do Brasil, exceto em mulheres do interior da região Norte, onde as taxas em 2017 chegaram a ser três vezes maiores em relação à região Sudeste (AZEVEDO E SILVA *et al.*, 2020). Desta forma, o controle de câncer do colo do útero constitui uma das prioridades da agenda de saúde do país e integra o *Plano de Ações Estratégicas para o Enfrentamento das Doenças Crônicas não Transmissíveis* (BRASIL, 2011).

O exame citopatológico do colo do útero, também conhecido como teste de *Papanicolaou*, historicamente foi considerado a principal estratégia para o rastreamento do câncer do colo do útero (HOLCOMB; RUNOWICZ, 2005). O objetivo fundamental é detectar precocemente as lesões precursoras e instituir rapidamente o tratamento antes da evolução para a forma invasiva da doença, tendo como objetivo final a redução da incidência e mortalidade por câncer cervical invasivo (PEIRSON *et al.*, 2013). Estima-se que, por meio de programas de rastreamento realizados com uma cobertura da população-alvo de 80 a 100% e acesso à confirmação diagnóstica

e ao tratamento adequado, seja possível reduzir a incidência do CCU em até 90% (WHO, 2002).

No Brasil, desde 2011, o Ministério da Saúde, através do INCA recomenda o rastreamento citopatológico do colo do útero em mulheres de 25 a 64 anos, com intervalo de três anos, após dois exames normais com intervalo anual (INCA, 2011a). Porém, o rastreamento ainda é realizado de forma oportunística no país, ou seja, a solicitação do exame depende da procura espontânea pelo serviço de saúde determinada por outros motivos que não o rastreamento do CCU (BRASIL, 2010). O rastreamento oportunístico está associado com o sobrerastreamento (*over-screening*) ou exames em excesso, que corresponde a exames realizados fora grupo etário alvo (sobrerastreamento por faixa etária) ou com periodicidade em intervalo inferior ao preconizado pelo Ministério da Saúde (sobrerastreamento por periodicidade) - (VALE *et al.*, 2010). O sobrerastreamento, além de sobrecarregar os serviços, pode comprometer o acesso das mulheres que realmente precisam ser rastreadas e encaminhadas para a investigação diagnóstica e tratamento das lesões precursoras (INCA, 2014a).

Diante deste contexto, torna-se relevante a análise de indicadores relacionados ao rastreamento do CCU, já que este constitui a principal forma de controle da doença no Brasil. Com a realização desse estudo foi possível fazer um diagnóstico do rastreamento em Minas Gerais e os resultados podem contribuir para o planejamento de ações mais resolutivas.

O presente estudo faz parte de um projeto maior intitulado *Desigualdades de Acesso às Ações de Controle do Cânceres de Mama e Colo do Útero no Brasil através dos Sistemas de Informação em Saúde (DAAC-SIS)*, desenvolvido por meio de parceria entre pesquisadores da Universidade Federal de Juiz de Fora, Universidade do Estado do Rio de Janeiro, Universidade de São Paulo, Universidade Federal da Bahia e *London School of Hygiene and Tropical Medicine* (LSHTM, do Reino Unido). Este projeto visa construir metodologia de análise que integre diversos sistemas de informação em saúde por meio do relacionamento das bases de dados, buscando identificar aspectos relacionados à detecção precoce do CCU e de mama e contribuir para o aprimoramento das políticas de controle dessas doenças.

A tese será apresentada em seis sessões: referencial teórico, justificativa, objetivos, métodos, resultados e discussão e considerações finais. A sessão referente aos resultados e discussão será apresentada sob a forma de três artigos.

2 REFERENCIAL TEÓRICO

2.1 EPIDEMIOLOGIA DO CÂNCER DO COLO DO ÚTERO

De acordo com dados do GLOBOCAN, o CCU é o quarto tipo de câncer mais frequentemente diagnosticado e a quarta causa de morte por câncer em mulheres no mundo, com uma estimativa de 604.000 novos casos e 342.000 mortes em 2020. As maiores taxas de incidência e mortalidade por 100 mil mulheres são vistas em países da África Oriental (40,1 e 28,6, respectivamente), África do Sul (36,4 e 20,6, respectivamente) e África Central (31,6 e 22,7, respectivamente). América do Sul (15,4 e 7,8, respectivamente), Europa Oriental (14,5 e 6,1, respectivamente) e América Central (13,8 e 6,8, respectivamente) apresentam taxas intermediárias e as menores taxas são observadas na América do Norte (6,2 e 2,1, respectivamente), Austrália e Nova Zelândia (5,6 e 1,6, respectivamente) e Ásia Ocidental (4,1 e 2,3, respectivamente) - (SUNG *et al.*, 2021).

Na América Latina, em 2012, os países com as maiores taxas padronizadas de incidência e mortalidade por 100 mil mulheres eram os seguintes: Bolívia (47,7; 21,0, respectivamente); Nicarágua (36,2; 18,3); Paraguai (34,2; 15,7); Venezuela (32,8; 12,3); Peru (32,7; 12,0) e República Dominicana (30,7; 12,3). E, os que apresentavam as menores taxas: Porto Rico (11,4; 2,8, respectivamente), Costa Rica (11,4; 4,4) e Chile (12,8; 6,0). Argentina, Uruguai, Panamá, Cuba e Brasil apresentaram taxas intermediárias (incidência variando de 16,3 no Brasil a 20,9 por 100 mil mulheres na Argentina e mortalidade variando de 6,7 em Cuba a 8,4 na Argentina) – (NEGRIN, 2015).

No Brasil, dentre as principais neoplasias, com exceção do câncer de pele não-melanoma, o CCU é o terceiro tipo mais frequente nas mulheres, com um risco estimado de 15,43 casos a cada 100 mil mulheres para o triênio 2020-2022. É o segundo tipo de câncer mais incidente nas Regiões Norte (21,20/100 mil), Nordeste (17,62/100 mil) e Centro-Oeste (15,92/100 mil). Já na Região Sul (17,48/100 mil), ocupa a quarta posição e, na Região Sudeste (12,01/100 mil), a quinta posição (INCA, 2019).

Estudo realizado com dados do GBD (*Global Burden of Disease*) verificou que a incidência padronizada por idade por 100.000 mulheres no país passou de 23,53 (IC

95% 22,79–24,26) no ano 2000 para 18,39 (IC 95% 17,63–19,17) em 2017. Em Minas Gerais, a taxa de incidência passou de 17,90 (IC 95% 16,26-19,62) em 2000 para 14,18 (IC95% 12,63-16,03) em 2017. As taxas nos estados do Norte e Nordeste foram as maiores do país, atingindo o dobro ou o triplo da taxa nacional. Por exemplo, no Amapá a taxa de incidência por 100.000 mulheres foi 37,47 (IC95% 32,94-42,40). Os dados mostram que, apesar da queda verificada no país ao comparar o ano 2000 com 2017, ainda há grande desigualdade regional e a taxa de incidência ainda é bem superior em relação a países desenvolvidos, como por exemplo, Inglaterra e Canadá (10,37 e 9,37, respectivamente) – (REIS *et al.*, 2020)

Quanto à mortalidade por CCU, estudo de série temporal de 30 anos (1980 a 2010), mostrou queda das taxas nas capitais e nos demais municípios das regiões Sudeste, Sul e Centro-Oeste. Mas, para as regiões Norte e Nordeste, houve queda nas capitais e aumento no interior do país (GIRIANELLI; GAMARRA; AZEVEDO e SILVA, 2014). No período entre 1978 e 2017, foi observado declínio da mortalidade em todas as regiões do Brasil, exceto entre mulheres do interior da região Norte, onde as taxas em 2017 chegam a ser três vezes maiores do que na região Sudeste (AZEVEDO E SILVA *et al.*, 2020). Outro estudo, que analisou a projeção da tendência da mortalidade pela doença até 2030, aponta que a discrepância regional será aprofundada, pois as regiões mais desenvolvidas deverão seguir reduzindo suas taxas, enquanto as mais pobres (Norte e Nordeste) deverão ter incremento (BARBOSA *et al.*, 2016).

Desta forma, observa-se que nas regiões Norte e Nordeste, especialmente a primeira, o perfil epidemiológico da doença assume um padrão diferenciado das demais e do país.

Para o ano de 2015, análise dos dados do GBD para a mortalidade por câncer no Brasil mostrou que, no sexo feminino, o câncer de mama foi a neoplasia com a maior taxa padronizada de mortalidade por 100 mil habitantes (16,3), seguido pelos cânceres de pulmão (12,8), cólon e reto (11,1), colo uterino (9,6) e estômago (7,0). Contudo, a análise das cinco principais localizações por estado apresentou padrão diferenciado. Em alguns estados da região Norte (Amazonas – 26,2; Pará –16,9; Amapá – 26,2 e Tocantins – 14,5) e Nordeste (Maranhão – 23,1 e Piauí – 13,6), o câncer do colo do útero foi o que apresentou a maior taxa de mortalidade. Ao comparar a taxa de mortalidade entre os anos 1990 e 2015, observou-se redução acentuada no país (-33,9%). Também foi observada redução em Rondônia (-38,5%), Roraima (-

34,7%), Alagoas (-33,3%), em todos os estados das regiões Sul e Sudeste e em quase todos os estados da região Centro-Oeste, sendo que a maior redução foi observada no Distrito Federal (-51%). No estado de Minas Gerais, a taxa de mortalidade padronizada para a doença, em 2015, foi 7,9 por 100 mil mulheres, com variação de -39,7% em relação ao ano de 1990 (GUERRA *et al.*, 2017).

Para o período de 2008 a 2012, a taxa de mortalidade bruta após a redistribuição dos óbitos por neoplasia de útero, porção não especificada, em Minas Gerais, foi de 5,7/100.000 mulheres. Microrregiões de saúde que apresentam maior Índice de Vulnerabilidade na Saúde, localizadas nas macrorregiões Norte, Noroeste, Jequitinhonha e Leste, apresentaram as maiores taxas de mortalidade por câncer de colo do útero no período analisado. Por outro lado, microrregiões com alto IDH e maior Grau de Urbanização localizadas, em sua maioria, nas macrorregiões do Triângulo Norte e Sul, Centro, Sul e Sudeste, apresentaram menor vulnerabilidade na saúde e taxas de baixa a intermediárias (DUARTE, 2016)

2.2 HISTÓRIA NATURAL DA DOENÇA E FATORES DE RISCO DO CÂNCER DO COLO DO ÚTERO

O CCU tem causa conhecida e, de todos os cânceres humanos, é o melhor compreendido (SANKARANARAYANAN *et al.*, 2008; CASTRO *et al.*, 2014). Possui história natural longa, começando com uma infecção persistente por um ou mais tipos oncogênicos de papilomavírus humano (HPV), seguindo alterações celulares que resultam em lesões intra-epiteliais cervicais ou adenocarcinomas *in situ* e finalmente, carcinoma ou adenocarcinoma invasor se as alterações celulares não forem tratadas (SANKARANARAYANAN *et al.*, 2008). Portanto, a infecção persistente por um ou mais tipos de HPV é considerada a causa primária do CCU (BOSCH *et al.*, 2002; WALBOOMERS *et al.*, 1999). Existem mais de 100 tipos de HPV e os mais carcinogênicos que acometem o colo do útero são: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 e 59 (COLPANI *et al.*, 2016). Estes tipos, considerados de alto risco oncogênico, são detectados em 99% dos cânceres de colo do útero, sendo os mais comuns os tipos 16 e 18, que em conjunto, são responsáveis por cerca de 70% dos casos (MUÑOZ *et al.*, 2006; WALBOOMERS *et al.*, 1999; WHO, 2014).

O genótipo mais carcinogênico do HPV é o HPV 16, que causa principalmente o carcinoma de células escamosas, enquanto o HPV 18 causa principalmente o adenocarcinoma, tipo menos frequente e mais agressivo. No entanto, a carcinogênese cervical pode surgir de infecções com muitos tipos de alto risco que agem sinergicamente (MCGRAW; FERRANTE, 2014).

A maioria das mulheres sexualmente ativas adquire a infecção pelo HPV ao longo de suas vidas, porém, a infecção tem curta duração e o corpo a elimina espontaneamente (SCHIFFMAN *et al.*, 2013). No Brasil, estudo de revisão sistemática que utilizou 14 artigos publicados entre 1989 e 2008 encontrou uma prevalência geral de infecção do colo do útero pelo HPV variando de 13,7% a 54,3%. Para mulheres com citologia normal, a prevalência variou entre 10,4% e 24,5% (AYRES; AZEVEDO e SILVA, 2010). Estudo transversal realizado com cerca de 2 mil mulheres de 20 a 59 anos residentes em área adscrita a duas unidades da Estratégia Saúde da Família (ESF), localizadas na periferia do município de Juiz de Fora (MG), encontrou uma prevalência global de infecção pelo HPV de 12,6% (AYRES *et al.*, 2017).

A maioria das infecções por HPV regridem espontaneamente dentro de seis meses a dois anos (RODRÍGUEZ *et al.*, 2008; SCHIFFMAN *et al.*, 2013). A minoria das infecções persiste e pode levar a transformações intraepiteliais progressivas, evoluindo para lesões intraepiteliais precursoras da doença, as quais, se não diagnosticadas e tratadas oportunamente, evoluem para o câncer do colo do útero (SANKARANARAYANAN *et al.*, 2008). O período de evolução de uma lesão cervical para a forma invasiva é cerca de 10 a 20 anos (ZEFERINO *et al.*, 1998; WHO, 2014). Lesões intraepiteliais cervicais de alto grau (HSIL) como NIC 3 têm o risco de progredir para a forma invasiva em aproximadamente 30% dos casos (NAUCLER *et al.*, 2007; MC CREDIE *et al.*, 2008).

A infecção pelo HPV, apesar de ser causa necessária para o desenvolvimento da doença, não é suficiente, havendo outros fatores envolvidos no processo de carcinogênese (KOLIOPOULOS *et al.*, 2017). Além de aspectos relacionados à própria infecção pelo HPV (subtipo e carga viral, infecção única ou múltipla), outros fatores relacionados ao desenvolvimento da doença incluem: uso de contraceptivos hormonais, tabagismo, início precoce da atividade sexual, múltiplos parceiros sexuais, multiparidade, presença de outras doenças sexualmente transmissíveis e imunossupressão (MUÑOZ *et al.*, 2006; EVERETT *et al.*, 2011).

O baixo nível socioeconômico é também considerado um fator de risco e está relacionado a dificuldades de acesso a serviços de saúde tanto para a detecção precoce da doença (PARIKH; BRENNAN; BOFFETTA, 2003), como para o recebimento de tratamento adequado e oportuno (RODRIGUES; BUSTAMANTE-TEIXEIRA, 2011).

2.3 PREVENÇÃO E CONTROLE DO CÂNCER DO COLO DO ÚTERO

Em virtude da longa história natural, composta por fases pré-clínicas, detectáveis e curáveis, o CCU é o câncer que apresenta um dos mais altos potenciais de prevenção e cura (SANKARANARAYANAN *et al.*, 2008; WHO, 2014). Com base no conhecimento do fator etiológico e na longa história natural da doença, existem duas abordagens complementares para a prevenção do câncer do colo do útero: a prevenção primária através da vacinação contra o HPV e a prevenção secundária, realizada por meio do rastreamento populacional, com o objetivo de detectar e tratar as lesões precursoras antes que a doença se torne invasiva (ELZEIN; RICHARDSON; FRANCO, 2016).

2.3.1 Prevenção primária

A descoberta da associação entre o CCU com a infecção persistente por um ou mais tipos oncogênicos de HPV no final da década de 1990 e início dos anos 2000 (WALBOOMERS *et al.*, 1999; BOSCH *et al.*, 2002), permitiu o desenvolvimento da prevenção primária por meio de vacinas contra o HPV. Desde o primeiro licenciamento dessas vacinas em 2006, elas provaram ser seguras, altamente imunogênicas e capazes de induzir forte proteção direta e indireta contra o vírus (GALLAGHER; LAMONTAGNE; WATSON-JONES, 2018).

A Escócia, por exemplo, possui um programa organizado de rastreamento e implementou no ano de 2008, um programa nacional de imunização contra o HPV direcionado a meninas de 12 e 13 anos. Em 2015, foi verificado que mulheres rastreadas aos 20 anos que receberam a vacina tiveram prevalência de NIC 3 89% menor em relação àquelas que não foram vacinadas (PALMER *et al.*, 2019).

Nos últimos 12 anos, mais de 80 países introduziram programas nacionais de vacinação contra o HPV, porém, a maioria desses países são de renda média ou

alta. O custo elevado da vacina a torna impraticável na saúde pública em muitos países em desenvolvimento. Desta forma, as barreiras à introdução da vacina contra o HPV continuam maiores nos países com maior carga de câncer do colo do útero e maior necessidade de vacinação (GALLAGHER; LAMONTAGNE; WATSON-JONES, 2018).

Uma revisão sistemática incluindo seis estudos de custo-efetividade da vacina contra o HPV para o Brasil concluiu que, apesar da vacinação aumentar o custo da prevenção do CCU, ela é altamente custo-efetiva e benéfica para o país (FONSECA; DE LIMA FERREIRA, 2014). Assumindo 70% de cobertura vacinal da população-alvo, a vacinação resultaria em uma redução de 42% no risco de CCU ao longo da vida da mulher. Uma abordagem combinada de vacinação de meninas e triagem em mulheres três vezes na vida após os 30 anos de idade, ambas com cobertura de 70%, proporcionaria benefício adicional com uma redução estimada de 53 a 70% no risco de câncer durante a vida. Os fatores mais importantes sobre o impacto previsto da vacina incluem o nível de cobertura e a capacidade de alcançar alta eficácia vacinando as meninas antes da atividade sexual (GOLDIE *et al.*, 2007).

No Brasil, o Ministério da Saúde introduziu a vacina contra o HPV no Calendário Nacional de Vacinação em 2014, tendo como objetivo principal a redução da incidência CCU. A implantação da vacina no Brasil foi gradativa e oferecida na rotina de vacinação das unidades de saúde do SUS e em escolas. No ano de 2014, o público alvo foi meninas de 11 a 13 anos de idade, a cobertura vacinal da primeira dose foi de 85%, superando a meta de 80%, porém, a cobertura da segunda dose foi apenas 60%. A partir de 2016, o esquema vacinal foi modificado para duas doses (0 e 6 meses). Em 2017, ampliou-se a vacinação para as meninas de nove a 14 anos e introduziu-se para os meninos de 11 a 14 anos de idade. Neste ano, a cobertura vacinal nas meninas foi de 82,6% para a primeira dose e de 52,8% para a segunda dose. Para os meninos, a cobertura vacinal com a primeira dose foi de 43,8%. Em 2018, a vacina continuou a ser oferecida por meio de duas doses (0 e 6 meses) para meninas a partir de 9 anos e meninos a partir de 11 anos com possibilidade de completarem o esquema até os 15 anos, 11 meses e 29 dias de idade (BRASIL, 2018a).

Estudo recente aponta que uma única dose de vacina quadrivalente é imunogênica e fornece proteção significativa e duradoura contra infecções por HPV 16 e 18, similar aos esquemas vacinais de três ou duas doses. Isso pode sustentar a

introdução de uma única dose da vacina em muitos países de baixa renda onde não há estratégias de prevenção e controle do CCU (SANKARANARAYANAN *et al.*, 2018)

Cabe destacar que a vacinação não substitui o rastreamento populacional, visto que não há proteção contra todos os HPV de alto risco e nem imunização de toda a população. Desta forma, a vacinação e o rastreamento são complementares (HARPER; DEMARS, 2017; GOLDIE *et al.*, 2007; COLPANI *et al.*, 2016).

2.3.2 Prevenção secundária

O exame de *Papanicolaou* (citologia convencional) é o método de rastreamento do CCU mais antigo e possibilita a detecção de lesões precursoras vários anos antes do surgimento da neoplasia do colo do útero. Desta forma, historicamente, foi considerada a principal estratégia de controle da doença no Brasil e no mundo (SANKARANARAYANAN *et al.*, 2008).

Este exame, desenvolvido na década de 1940 pelo Dr. George Papanicolaou, não foi objeto de nenhum grande ensaio clínico randomizado que confirmasse sua eficácia, não tendo sido, portanto, formalmente avaliado antes de ser implementado como teste de rastreamento. No entanto, estudos epidemiológicos globais demonstraram de forma convincente sua eficácia como uma estratégia de prevenção do câncer (LEES; ERICKSON; HUH, 2016). O objetivo final é diminuir a incidência e a subsequente mortalidade por CCU e esse impacto do rastreamento está bem estabelecido na literatura. Mesmo um único exame de triagem realizado durante a vida reduz a mortalidade e a incidência pela doença em comparação com nenhuma triagem (PEIRSON *et al.*, 2013).

Em vários países desenvolvidos, onde programas de rastreamento com o *Papanicolaou* foram bem estabelecidos, com alta cobertura da população alvo, houve impacto significativo na incidência e mortalidade por CCU. Nos EUA e no Reino Unido, por exemplo, em um período de 30 anos a partir das décadas de 50 e 70, respectivamente, houve diminuição de cerca de 70% na mortalidade por CCU (MCGRAW; FERRANTE, 2014; PESOLA; SASIENI, 2019). Os países nórdicos também foram muito bem-sucedidos na redução da carga da doença, sendo a maior alcançada pela Islândia, com redução de 80% na mortalidade (ALI; KUELKER; WASSIE, 2012).

No entanto, a história de sucesso na redução da carga da doença em países desenvolvidos não é replicada nos países em desenvolvimento, pois a implementação

de programas de rastreamento organizados e eficazes é difícil em ambientes com poucos recursos. O rastreamento do câncer do colo do útero requer infraestrutura, recursos humanos, garantia de qualidade, monitoramento e avaliação dos programas de triagem (ALI; KUELKER; WASSIE, 2012). Neste contexto, os componentes críticos para o sucesso de um programa de rastreamento são: alta cobertura da população alvo com testes de rastreamento de qualidade e acesso à investigação diagnóstica para mulheres com teste positivo no rastreamento e ao tratamento oportuno para mulheres com neoplasia cervical confirmada (SANKARANARAYANAN *et al.*, 2008).

O teste de Papanicolaou foi introduzido no Brasil no fim da década de 1950 e, a princípio, era realizado de forma incipiente, em nível individual e não em âmbito populacional (ALVES; GUERRA; BASTOS, 2009). A partir da década de 1980, quando as primeiras ações de rastreamento para a doença foram introduzidas pelo Ministério da Saúde, são observadas quedas na incidência e mortalidade pelo CCU (GONÇALVES *et al.*, 2011). No final da década de 1990, com a introdução do *Programa Nacional de Controle do Câncer do Colo do Útero*, houve grande expansão da prevenção e controle dessa neoplasia no Brasil (MEIRA *et al.*, 2012). Atualmente, o Ministério da Saúde recomenda o rastreamento citopatológico do colo do útero em mulheres de 25 a 64 anos, com intervalo de três anos, após dois exames normais com intervalo anual (INCA, 2016a).

Embora esse método subjetivo tenha levado a uma diminuição significativa na incidência e na mortalidade pelo CCU, principalmente em países com programas de rastreamento organizados e alta cobertura da população alvo, a citologia tem limitações como baixa sensibilidade e fraca reprodutibilidade (EL-ZEIN; RICHARDSON; FRANCO, 2016). A citologia tradicional demonstrou ter uma sensibilidade de apenas 51% (30-87%) e especificidade de 98% (86-100%). Além disso, devido à natureza subjetiva do teste de triagem, existe uma variabilidade interobservador significativa na interpretação da citologia, o que contribui ainda mais para suas taxas variáveis de sensibilidade e especificidade (LEES; ERICKSON; HUH, 2016).

Os avanços no conhecimento sobre a história natural da doença associado à implementação da vacinação contra o HPV têm mudado o foco do rastreamento em alguns países para detectar seu agente causal, com a adoção de testes de HPV (EL-ZEIN; RICHARDSON; FRANCO, 2016). O teste de DNA do HPV tem maior

sensibilidade (95%) quando comparado ao *Papanicolaou*, porém especificidade ligeiramente inferior (94%). Além disso, o teste de HPV é automatizado, tornando-o menos propenso a erros humanos e mais reprodutível em relação à citologia tradicional (TOTA *et al.*, 2017).

Estudos indicam a necessidade de mudança no paradigma do rastreamento do CCU na era pós-vacinação, já que a menor prevalência de lesões cervicais levará a uma queda no valor preditivo positivo (VPP) da citologia cervical. Além disso, em virtude da natureza subjetiva desta, a menor prevalência de anormalidades pode ter grandes consequências negativas na precisão, resultando em declínios ainda maiores no VPP e na eficiência dos programas de rastreamento (ELZEIN; RICHARDSON; FRANCO, 2016; TOTA *et al.*, 2017).

Nesta perspectiva, a citologia deixaria de ser um teste primário adequado para o rastreamento do CCU. Podendo ser realizado com precisão adequada naquelas mulheres que tiveram um teste de HPV positivo na triagem primária, já que a prevalência da doença será artificialmente alta. A estratégia de triagem primária com um teste de HPV e citologia posterior aproveitaria as propriedades desejáveis de ambos os testes (alta sensibilidade do primeiro e alta especificidade do segundo) - (MCGRAW; FERRANTE, 2014; TOTA *et al.*, 2017).

A principal vantagem do teste de HPV é que ele pode ser realizado com maiores intervalos de tempo (cinco anos) e com amostras autocoletadas. A possibilidade de realização de autocoleta permite aumentar a cobertura do rastreamento pela possibilidade de alcançar locais com menores recursos e de acesso restrito a serviços de saúde, como em muitas regiões do Brasil e pela maior aceitabilidade por mulheres que não fazem o exame por constrangimento. No entanto, o padrão ouro do teste do HPV ainda é a amostragem por um profissional. Já em relação às desvantagens, destacam-se o alto custo e a infra-estrutura necessária para sua implementação (LORENZI; SYRJÄNEN; LONGATTO-FILHO, 2015).

Porém, cabe destacar que a introdução do teste de HPV só é recomendada em cenários de rastreamento organizado, com cuidadosa monitorização da qualidade e dos custos (ARBYN *et al.*, 2010; ZEFERINO *et al.*, 2018). Como o rastreamento no Brasil ainda é oportunístico, não há controle das mulheres que realizam os exames e nem da periodicidade com que o fazem. Desta forma, não há ferramentas que garantam que o intervalo entre os exames será efetivamente ampliado a partir da adoção do teste de HPV, condição necessária para que se obtenha algum resultado

de custo-efetividade favorável. Essa deficiência representa importante obstáculo para o uso destes testes no país (INCA, 2011a).

2.4 HISTÓRICO DAS PRINCIPAIS AÇÕES DE CONTROLE DO CCU NO BRASIL

Na década de 1950, começaram a surgir os primeiros programas de rastreamento do CCU em países desenvolvidos e alguns médicos brasileiros começaram a estimular a utilização de campanhas para o rastreamento, como principal forma de controle da doença. Neste contexto, em meados dos anos 1960, surgiram algumas campanhas locais, sendo que, Porto Alegre, Bahia, e, principalmente, São Paulo foram precursores dessas iniciativas. A prevenção da doença começou, então, a deixar o âmbito dos consultórios particulares e hospitais especializados para se transformar numa atividade de saúde pública (TEIXEIRA; PORTO; SOUZA, 2012).

Diante da expansão das ações de rastreamento no país, verificou-se a necessidade de formar profissionais específicos visando garantir a qualidade do rastreamento. Houve, então, grande ampliação de centros formadores de citotécnicos no Brasil, principalmente a partir da década de 1970 (TEIXEIRA; PORTO; SOUZA, 2012).

No início da década de 1980, as primeiras ações de rastreamento foram instituídas pelo Ministério da Saúde no Brasil, com a implantação, em 1984, do *Programa de Atenção Integral à Saúde da Mulher* (PAISM). Este previa que os serviços básicos de saúde oferecessem às mulheres atividades de prevenção do câncer do colo do útero. A principal contribuição do programa foi introduzir e estimular a coleta de material para o exame citopatológico como procedimento de rotina nas consultas ginecológicas (INCA, 2011a).

No ano de 1997, foi instituído um projeto piloto de rastreamento em seis localidades brasileiras (Distrito Federal, Curitiba, Recife, Rio de Janeiro, Belém e no Estado de Sergipe), denominado programa *Viva Mulher*. Esse projeto desenvolvia ações de oferta de serviços para prevenção, detecção precoce e tratamento do CCU para mulheres de 35 a 49 anos (TEIXEIRA, 2015; AZEVEDO E SILVA *et al.*, 2015). Posteriormente, expandiu para todo o território nacional como parte do *Programa Nacional de Combate ao Câncer do Colo do Útero*, instituído pelo Ministério da Saúde,

em 1998. Este programa definiu estratégias para assegurar o acesso das mulheres, estabeleceu padrões de coleta do esfregaço cérvico-uterino e de terminologias a serem utilizadas nos laudos citopatológicos. Além disso, definiu a faixa etária alvo do rastreamento (25 a 59 anos) e a periodicidade recomendada para a realização dos exames (uma vez por ano e, após dois exames anuais consecutivos negativos, a cada três anos) - (BRASIL, 1998). Foi também criado o Sistema de Informação do Câncer do Colo do Útero (SISCOLO), o qual foi implantado oficialmente no país no ano seguinte com o objetivo de viabilizar o monitoramento e o gerenciamento das ações de rastreamento do CCU (BRASIL, 1999). Posteriormente, este sistema foi aperfeiçoado e, em 2006, foi implantada uma nova versão, atualizada de acordo com a *Nomenclatura Brasileira de Laudos Citopatológicos* (BRASIL, 2006a).

O SISCOLO é um subsistema do Sistema de Informação Ambulatorial do Sistema Único de Saúde (SIA/SUS), utilizado para cadastrar os exames citopatológicos e histopatológicos do colo do útero no âmbito do SUS. O sistema permite obter o Boletim de Produção Ambulatorial; registrar informações sobre condutas diagnósticas e terapêuticas relativas aos exames positivos/alterados (módulo seguimento da paciente); selecionar amostras para monitoramento externo da qualidade dos exames e coletar dados para construção de indicadores (INCA, 2011b).

Em 2005, foi lançada a *Política Nacional de Atenção Oncológica*, que estabeleceu o controle dos cânceres do colo do útero e de mama como componente fundamental a ser previsto nos planos estaduais e municipais de saúde (BRASIL, 2005). A importância da detecção precoce dessas neoplasias foi reafirmada no *Pacto pela Saúde* em 2006, por meio da inclusão de indicadores na pactuação de metas com estados e municípios (BRASIL, 2006b).

No ano de 2006, também foi publicada a *Nomenclatura Brasileira para Laudos Cervicais de Condutas Preconizadas*, padronizando os diagnósticos citopatológicos e orientando os profissionais envolvidos no cuidado às mulheres com alterações citológicas identificadas no rastreamento. Este documento reafirma a faixa etária alvo de 25 a 59 anos e a periodicidade de três anos para o rastreamento, conforme o *Programa Nacional de Combate ao Câncer do Colo do Útero* (INCA, 2006).

Em 2011, foram publicadas pelo INCA as *Diretrizes Brasileiras para o Rastreamento do Câncer do Colo do Útero* que apresentam as definições de população alvo, periodicidade e condutas clínicas para mulheres com resultados

alterados. A partir deste ano, a faixa etária recomendada para o rastreamento passou a ser de 25 a 64 anos (INCA, 2011a). No ano de 2011, o Ministério da Saúde também publicou o *Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis no Brasil, 2011 a 2022*. Uma das ações prioritárias é o aperfeiçoamento do rastreamento do CCU e “evoluir do modelo oportunístico para o modelo organizado, com garantia de confirmação diagnóstica, tratamento de lesões precursoras e referência dos casos confirmados de câncer para o nível terciário.” (BRASIL, 2011).

Em 2013, com o objetivo de substituir e integrar os sistemas de informação relacionados ao câncer feminino foi instituído o *Sistema de Informação de Câncer* (SISCAN). Trata-se de uma versão em plataforma *web* que integra o SISCOLO e o *Sistema de Informação do Câncer de Mama* (SISMAMA) e tem como unidade de observação a mulher, e não o exame (BRASIL, 2013a). Porém, este sistema ainda está em processo de implementação e alguns municípios ainda registram os exames no SISCOLO, não tendo havido uma transição definitiva para o SISCAN em todo o país.

Por meio da Portaria nº 3.388/2013 o Ministério da Saúde redefiniu a *Qualificação Nacional em Citopatologia* na prevenção do câncer do colo do útero (QualiCito). A QualiCito consiste na definição de padrões de qualidade e na avaliação da qualidade do exame citopatológico do colo do útero por meio do acompanhamento, pelos gestores do SUS, do desempenho dos laboratórios públicos e privados prestadores de serviços para o SUS (BRASIL, 2013b).

No ano de 2014, foram definidos os Serviços de Referência para Diagnóstico e Tratamento de Lesões Precursoras do Câncer do Colo do Útero, bem como os respectivos incentivos financeiros de custeio e de investimento para a sua implantação e os critérios para a habilitação das unidades, além do rol mínimo de exames necessários para o diagnóstico (BRASIL, 2014). Ainda em 2014, o Ministério da Saúde, por meio do Programa Nacional de Imunizações, implementou a vacinação de meninas adolescentes contra o HPV e, em 2017, incluiu também os meninos na população-alvo da vacinação (BRASIL, 2018a).

No ano de 2016, foi publicada a segunda edição revista, ampliada e atualizada das *Diretrizes Brasileiras para o Rastreamento do Câncer do Colo do Útero*. Este documento mantém as recomendações quanto à faixa etária e periodicidade publicadas em 2011: os dois primeiros exames devem ser realizados com intervalo

anual e, se ambos os resultados forem negativos, os próximos devem ser realizados a cada 3 anos; o início da coleta deve ser aos 25 anos de idade para as mulheres que já tiveram ou têm atividade sexual; o rastreamento antes dos 25 anos deve ser evitado; os exames periódicos devem seguir até os 64 anos de idade e devem ser interrompidos quando as mulheres tiverem pelo menos dois exames negativos consecutivos nos últimos cinco anos (INCA, 2016a).

Em consonância à Portaria nº 3.388/2013, foi publicado também em 2016, o *Manual de Gestão da Qualidade para Laboratório de Citopatologia*, atualizando a edição de 2012, tendo como objetivo contribuir para melhorar a qualidade e a confiabilidade dos exames citopatológicos (INCA, 2016b).

2.5 AVALIAÇÃO DE INDICADORES RELACIONADOS AO RASTREAMENTO DO CÂNCER DO COLO DO ÚTERO NO BRASIL E FATORES ASSOCIADOS

Pesquisa realizada com base em dados do VIGITEL (*Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico*), no ano de 2011, verificou que apenas dez capitais do país conseguiram atingir a meta de 80% para realização do *Papanicolaou* em mulheres da faixa etária alvo em algum momento da vida (Rio de Janeiro, Goiânia, Belo Horizonte, Campo Grande, Cuiabá, Palmas, Curitiba, Porto Alegre, Florianópolis e São Paulo) e sete capitais atingiram a meta para a realização do exame nos últimos três anos (Belo Horizonte, Cuiabá, Palmas, Florianópolis, Porto Alegre, Curitiba e São Paulo). Maceió, João Pessoa e Teresina foram as capitais com as menores coberturas, sendo que, em Maceió a proporção de mulheres que realizaram o exame pelo menos uma vez na vida foi inferior a 65%. Foi verificada forte correlação positiva entre o IDH e a cobertura de rastreamento do CCU. No modelo final ajustado por idade média, situação conjugal, tabagismo, Índice de Gini e cobertura pela ESF, a razão de prevalência para realização do exame nos últimos três anos foi de 1,07 (1,04-1,10) evidenciando que, a cada mudança em um desvio padrão do IDH, aumentava em 7% a proporção de realização do *Papanicolaou* (SADOVSKY *et al.*, 2015).

Resultados do VIGITEL para o ano de 2017 mostraram que no conjunto das 27 cidades avaliadas, a frequência de realização do exame de citopatológico nos últimos três anos, em mulheres entre 25 e 64 anos de idade, foi de 82,8%. As maiores

frequências de realização do exame nos últimos três anos foram observadas em Florianópolis (88,9%), São Paulo (88,3%) e Porto Alegre (87,8%) e as menores, em Maceió (66,6%), João Pessoa (68,7%) e Teresina (69,2%). Cinco capitais não atingiram a meta de 80% para realização do exame em algum momento da vida (Aracaju, Fortaleza, João Pessoa, Maceió e Teresina) e oito não atingiram a meta para realização do exame nos últimos três anos (Aracaju, Boa Vista, Cuiabá, Fortaleza, João Pessoa, Maceió, Natal, Teresina e Distrito Federal). Em Belo Horizonte, a proporção de mulheres da faixa etária alvo que relataram exame preventivo nos últimos três anos foi 85,5 (IC95% 82,4 - 88,5). Além disso, observou-se que a cobertura do exame aumentou com o aumento do nível de escolaridade (BRASIL, 2018b).

Dados da *Pesquisa Nacional de Saúde* (PNS) apontam que a cobertura do *Papanicolaou* em mulheres da faixa etária alvo do rastreamento no Brasil passou de 65,5%, em 2003, para 78,8%, em 2013, quase atingindo a meta preconizada de 80%. Em 2013, em todas as grandes regiões brasileiras, a realização do exame nos últimos três anos foi maior que 75% (THEME-FILHA *et al.*, 2016). Cabe ressaltar que essas coberturas, assim como aquelas avaliadas pelo VIGITEL, podem estar superestimadas, pois são baseadas em informações autorrelatadas (AZEVEDO e SILVA *et al.*, 2015). A PNS verificou ainda que mulheres que viviam nas regiões Sul e Sudeste, de pele branca, com maior nível de escolaridade, que viviam com um parceiro e, especialmente, aquelas que tinham plano de saúde privado eram mais propensas a submeter-se a um exame de *Papanicolaou* no intervalo recomendado (THEME-FILHA *et al.*, 2016).

Os resultados da PNS 2013 mostraram também que mulheres com idade entre 35 e 44 anos e 45 e 54 anos, apresentaram maiores prevalências de realização do *Papanicolaou* (83,2% e 81,6%, respectivamente). Houve aumento da cobertura conforme maior escolaridade, chegando a 88,8% (IC95% 86,9 - 90,4) entre as mulheres de nível superior completo. A prevalência de exame *Papanicolaou* foi maior entre as mulheres de raça/cor da pele branca (82,6%; IC95% 81,4 - 83,9) e nas residentes na área urbana (80,1%; IC95% 79,2 - 81,1). Quanto à região de residência, Sul (83%; IC95% 81,8 - 84,0) e Sudeste (81%; IC95% 79,5 - 82,6) foram as que apresentaram as maiores proporções; enquanto Norte (75,5%; IC95% 73,2 - 77,6) e Nordeste (75,1%; IC95% 73,2 - 77,6) foram as que tiveram as menores proporções (OLIVEIRA *et al.*, 2018a).

Cabe destacar que a proporção de mulheres das regiões Norte e Nordeste com plano de saúde privado foi 50% menor em relação às outras regiões do país. E, para mulheres com plano de saúde privado, o acesso ao teste de *Papanicolaou* aumentou em mais de 200% (OR ajustado=2,49). Portanto, os resultados da PNS 2013 indicam desigualdades ainda existentes no Brasil no acesso ao rastreamento (THEME-FILHA *et al.*, 2016).

Estudos realizados em diferentes regiões do país encontraram características semelhantes associadas a não realização do *Papanicolaou*: baixa escolaridade, ser solteira, não ter filho, não ter plano de saúde, não ter realizado consulta médica recente e não realizar outras práticas preventivas (ALBUQUERQUE *et al.*, 2009; MARTINS; VALENTE; THULER, 2009; OLIVEIRA *et al.*, 2006; ANDRADE *et al.*, 2014; BORGES *et al.*, 2012; GASPERIN; BOING; KUPEK, 2011; NAVARRO *et al.*, 2015; LAGE; PESSOA; MELÉNDEZ, 2013).

Corroborando estas características, estudo realizado no Estado de Roraima, em 2009, identificou que o perfil das pacientes com diagnóstico de CCU e HSIL assistidas em hospitais da rede pública foi: mulher parda ou indígena, desempregada, solteira, com baixa escolaridade, baixo nível socioeconômico, que reside em moradia sem saneamento básico, com início da atividade sexual precoce, que não faz teste de *Papanicolaou* regularmente ou nunca o fez (71% nunca haviam realizado um exame). As principais questões pessoais impeditivas à realização do exame preventivo foi o desconhecimento (47,4%). Das mulheres com câncer invasivo, 68,5% foram diagnosticadas com estadio avançado. O custo médio anual por paciente com CCU foi de R\$ 8.711,00. Diante dos resultados encontrados, os autores ressaltam a ineficiência do programa de rastreamento em alcançar e conscientizar as mulheres com maior risco para a doença, populações especiais como a indígena e aquelas com baixa escolaridade e perfil de exclusão social (FONSECA *et al.*, 2010).

Dentre os motivos mais referidos para a não realização do exame preventivo em um município da Bahia, destacaram-se a ausência de problemas ginecológicos (47,0%) e vergonha (23,5%) - (ANDRADE *et al.*, 2014). Geralmente, diante de sinais e sintomas ginecológicos, como prurido e leucorreia, as mulheres procuram assistência médica e recebem orientação para realizar o exame preventivo. Pois, para muitas mulheres, ainda há um entendimento de que o exame somente é necessário em condições de enfermidade ou sintomas, quando na verdade deveria ser realizado de forma periódica em mulheres assintomáticas. Assim, o desconhecimento da

doença e dos métodos de prevenção também são considerados fatores de risco para a não realização do rastreamento para importante parcela da população, principalmente a de baixa renda (NAVARRO *et al.*, 2015).

Estudo realizado com dados do SISCOLO para o ano de 2015, ao comparar as necessidades de exames de rastreamento para atingir cobertura de 100% da população-alvo que utiliza os serviços de saúde do SUS com a quantidade de exames efetivamente realizados, verificou um *déficit* de 46,9% no país. Sendo que os maiores *déficits* foram verificados nas regiões Norte (-66,2%), Nordeste (-59,5%) e Centro-Oeste (-56,4%) - (RIBEIRO; AZEVEDO e SILVA, 2018).

Nesta perspectiva, outro estudo também estimou a necessidade e avaliou a adequação da produção nacional dos procedimentos de rastreamento, investigação diagnóstica e tratamento de lesões precursoras do CCU realizados pelo SUS, em 2017. Para isso, foram utilizados dados do SISCOLO, dados de seguimento obtidos em prontuários médicos de uma unidade de referência em patologia cervical e dados dos Sistemas de Informações Ambulatoriais e Hospitalares. A estimativa de necessidade de exames citopatológicos de rastreamento correspondeu a 39,8% das mulheres na faixa etária alvo: considerando-se a periodicidade do exame (33,3% das mulheres), a realização de exames anuais para mulheres que fizeram o exame pela primeira vez (6% das mulheres) e a necessidade de repetição de exames insatisfatórios e rejeitados dentre os exames de rastreamento realizados (0,47%). Considerando as outras indicações de realização do exame (necessidade de repetição do exame entre 6 a 12 meses após exame com resultado de atipia de significado indeterminado em células escamosas - ASC-US e lesão escamosa de baixo grau - LSIL, de realização de exame para investigação de canal endocervical ou revisão de lâmina, seguimento de mulheres após biópsia com resultado normal/benigno e seguimento pós tratamento excisional), estimou-se a necessidade de 44,13% da população alvo. A análise do quantitativo de exames realizados no Brasil em 2017 comparada à estimativa de necessidade desses exames para rastreamento e seguimento de 100% das usuárias do SUS na faixa etária alvo revelou um *déficit* de 61%. Além disso, verificou-se *déficit* na produção de biópsias (-20,4%) e no tratamento de lesões precursoras, tanto em nível ambulatorial (-74,8%), quanto em nível hospitalar (-67,6%). Os autores destacam que a diferença dos resultados em relação ao estudo citado anteriormente é consequência das diferenças metodológicas para o cálculo das estimativas, sendo que neste, a utilização de fontes de informações

mais robustas e com identificação das mulheres permitiu melhor aproximação do cenário real de necessidades de exames (RIBEIRO *et al.*, 2019).

Análise dos dados do SISCOLO feita pelo INCA mostrou que, entre 2007 e 2013 no Brasil, a proporção de exames realizados na faixa etária alvo do rastreamento passou de 73,7% para 78,7%. Embora tenha ocorrido diminuição do percentual de exames realizados no intervalo de até um ano, de 55,9% em 2007 para 49,4% em 2013, ainda é muito frequente a realização de exames anuais em todas as regiões (INCA, 2015).

Em relação à periodicidade do rastreamento no Brasil, dados do SISCOLO referentes a 2006, revelam que 43% das mulheres que fizeram o exame de *Papanicolaou* naquele ano repetiram o exame em um período menor ou igual a um ano. Considerando que, neste ano, 8,8% das mulheres fizeram o exame pela primeira vez e que os percentuais de positividade e de amostras insatisfatórias foram, respectivamente, 2,9 e 1,1%, observa-se que o percentual de repetição em um ano é superior ao número de situações que justificariam novo exame neste prazo. Desta forma, verifica-se repetição desnecessária do exame, revelando um alcance provavelmente menor das ações de rastreamento, já que evidencia a existência de um mesmo grupo de mulheres fazendo exames além do recomendado (DIAS; TOMAZELLI; ASSIS, 2010).

O sucesso do rastreamento do câncer do colo do útero e de suas lesões precursoras está diretamente relacionado à qualidade do exame citopatológico (INCA, 2015), porém, a análise de indicadores de qualidade para o Brasil indica um cenário preocupante.

O exame citopatológico do colo do útero é classificado como insatisfatório quando, por algum problema, em geral relacionado à coleta do material, não é possível analisar a lâmina e emitir um resultado. Desta forma, a mulher terá que repetir o exame, o que implica em onerar mais o programa e na necessidade de solicitar que a mulher retorne à unidade de saúde para nova coleta. O percentual de amostras insatisfatórias acima de 5% aponta a necessidade de investimento na qualificação profissional. No período de 2007 a 2013, apesar da redução do percentual de 8,2% para 6,6% no Brasil, ele ainda é superior ao limite de 5%. Chamando atenção a região Nordeste, que durante todo o período analisado, apresentou valores acima de 14% para este indicador de qualidade do exame. Nas regiões Sul e Sudeste, o percentual foi inferior a 4% de 2007 a 2013 (INCA, 2015).

O índice de positividade (IP) é a proporção de exames citopatológicos com resultados alterados. Ele reflete tanto a prevalência das alterações na população examinada, quanto a capacidade dos laboratórios em identificá-las. A Portaria QualiCito determinou 3% como um valor mínimo aceitável para o IP, porém esse valor não foi alcançado no país. As Regiões Sudeste e Centro-Oeste foram as únicas que atingiram esta meta no período de 2007 a 2013 (INCA, 2015). A proporção de atípias de significado indeterminado em células escamosas (ASC), por ser um resultado inconclusivo, não deve representar 60% ou mais dos resultados alterados. No país, embora não tenha ultrapassado tal valor, os percentuais aumentaram no período analisado, ultrapassando os 50% a partir de 2008 (INCA, 2015).

Corroborando estes dados, um estudo que analisou dados agregados do SISCOLO, no período de 2006 a 2013, verificou que a porcentagem de exames rejeitados aumentou significativamente de 2010-2013, o que pode ser explicado pela falta de cuidado no manuseio, transporte e identificação das amostras. Em relação à presença de células epiteliais da zona de transformação, um indicador de qualidade da coleta, foi observada diminuição significativamente ao longo dos anos, o que pode ser consequência de problemas relacionados à coleta de amostras. O índice de positividade foi inferior ao recomendado, apresentando o menor valor em 2009 (2,48). As taxas de detecção de lesões intraepiteliais (LSIL e HSIL) diminuíram significativamente durante o período do estudo, ao passo em que foi verificado aumento nas porcentagens de “células escamosas atípicas de significado indeterminado possivelmente não neoplásicas” (ASC-US) e de “células escamosas atípicas não sendo possível excluir lesão intraepitelial de alto grau” (ASC-H), sugerindo problemas na fase de coleta e/ou a interpretação das amostras. De acordo com os autores, os indicadores de fase pré-analítica e analítica estudados sugerem um desempenho subótimo do rastreamento, que pode ser atribuído à baixa eficiência profissional na coleta, preparação e leitura das amostras. Assim, enfatizam a necessidade de melhor capacitação dos profissionais que coletam e analisam essas amostras (COSTA *et al.*, 2015).

Araújo e colaboradores (2014) analisaram 12.208 resultados de exames citopatológicos de mulheres atendidas nas Unidades de Atenção Básica a Saúde em Goiânia, no período de 2006 a 2008, utilizando os prontuários das Unidades de Atenção Básica e Média Complexidade. Os autores verificaram que do total de 900 mulheres com exames citopatológicos alterados, 61,22% tiveram seguimento

ignorado e, possivelmente, não foram orientadas e encaminhadas conforme as normas preconizadas pelo INCA. Das 297 mulheres com resultados citopatológicos classificados como ASC-H ou HSIL, 35,02% e 28,96% realizaram a colposcopia e a biópsia, respectivamente e 42,76% repetiram desnecessariamente o exame citopatológico em seis meses. Os resultados indicam que as ações desenvolvidas vão de encontro às diretrizes nacionais do programa de rastreamento e estão relacionadas com atrasos na confirmação diagnóstica e, conseqüentemente, do início do tratamento.

Estudo realizado em uma microrregião de saúde abrangendo sete municípios no estado do Mato Grosso do Sul, entrevistou profissionais da rede básica em relação às ações de rastreamento e diagnóstico do câncer do colo do útero, entre os meses de dezembro de 2013 e março de 2014. Também foi observado que as condutas da maioria dos profissionais da atenção básica não estão de acordo com as diretrizes. A maioria referiu encaminhar para outros níveis de atenção mulheres que apresentaram exame com resultados ASC-US (48%) e LSIL (79%) quando a recomendação é a repetição do exame em seis meses. Ao passo que, 42% dos profissionais afirmaram que diante de um resultado alterado, cuja recomendação do INCA seria o encaminhamento para colposcopia, repetem o exame citopatológico em seis meses (FARIAS; BARBIERI, 2016).

A proporção de casos da doença detectados em estadios avançados é um importante indicador relacionado ao desempenho dos programas de rastreamento. Quanto a esse indicador, estudo realizado com dados dos Registros Hospitalares de Câncer (RHC) com 37.638 casos de CCU no período de 2000 a 2009, concluiu que 70,6% das mulheres foram diagnosticadas em estadios avançados (\geq II B) da doença. Este desfecho foi associado à presença de carcinoma de células escamosas (OR=1,8; IC95% 1,7–2,0), idade \geq 50 anos (OR=1,5; IC95% 1,4–1,6), viver com companheiro (OR=1,3; IC95% 1,2–1,4), cor da pele preta (OR=1,2; IC95% 1,1–1,4) e baixo nível educacional (OR=1,2; IC95% 1,1–1,3) (THULER; AGUIAR; BERGMANN, 2014).

Outro estudo realizado com dados dos RHC verificou que o percentual de diagnósticos em estágio avançado aumentou no período de 2000 a 2012 (variação anual de 1,10%, IC95% 0,80;1,50). Mulheres com nível de escolaridade superior tiveram menores chances de estadiamento avançado, comparadas a analfabetas (OR=0,38; IC95% 0,31;0,47); indígenas (OR=2,38; IC95% 1,06;5,33) e negras (OR=1,16; IC95% 1,02;1,31) comparadas às brancas, e mulheres tratadas na região

Norte (OR=2,55; IC95% 2,26;2,89) em relação às do Sudeste, apresentaram maiores chances (RENNA JÚNIOR; AZEVEDO E SILVA, 2018).

Desta forma, observa-se que a despeito dos avanços na política de controle da doença no país, a maioria dos casos de CCU ainda são diagnosticados em estádios avançados e estão associados com as disparidades socioeconômicas presentes no país. Múltiplos fatores podem explicar o sucesso parcial alcançado no controle do CCU no Brasil. Dentre eles, destacam-se: o caráter oportunista do programa de rastreamento, no qual parte das mulheres faz exames em excesso, enquanto outras, permanecem sem acesso; grandes diferenças regionais na cobertura do rastreamento; a baixa capacidade da rede de laboratórios do SUS para identificação de lesões intraepiteliais; carência de recursos humanos especializados e falhas no seguimento de casos positivos (FONSECA, DE LIMA FERREIRA, 2014; AZEVEDO e SILVA *et al.*, 2015).

2.6 RASTREAMENTO OPORTUNÍSTICO E SOBERRASTREAMENTO

Um programa organizado para o rastreamento do CCU deve incluir os seguintes componentes: adoção de recomendações baseadas em evidências científicas, que inclui definição da população alvo e do intervalo entre as coletas, assim como elaboração de guias clínicos para o manejo dos casos suspeitos; monitoramento da população alvo e da periodicidade de realização dos exames, por meio de um sistema de informação de base populacional, que permita o recrutamento das mulheres na faixa etária alvo, de mulheres que nunca foram rastreadas ou que estão com exames atrasados, além de permitir o seguimento de mulheres com exames alterados; garantia da abordagem necessária para as mulheres com exames alterados; educação e comunicação e garantia de qualidade dos procedimentos realizados em todos os níveis do cuidado (BRASIL, 2010).

No período de 2006-2007, a Polônia implementou um programa de rastreamento organizado, que, combinado com a triagem oportunística ainda existente, alcançou uma cobertura de 70%. Tal implementação resultou em impactos significativos na carga do CCU no país, acelerando as tendências decrescentes na incidência e na mortalidade pela doença. A taxa de mulheres entre 20 e 69 anos que declararam ter sido submetidas a testes de *Papanicolaou* nos últimos três anos aumentou de 55%

em 2004 para 73% em 2009. Em 2004, 30% das mulheres polonesas declararam nunca ter realizado o exame, enquanto em 2014 essa taxa caiu para 10%. O declínio nas taxas de incidência da doença padronizadas por idade na faixa etária de rastreamento (25-59 anos) passou de -2,2% (IC95% -2,7-1,7%) entre 1993 e 2008 para -6,1% (IC95% -7,7 - -4,4%) anualmente após 2008, até 2014 (NOWAKOWSKI *et al.*, 2017).

No Brasil, o rastreamento do CCU falha em praticamente todos os componentes de um programa organizado e não há recrutamento ativo da população alvo (AZEVEDO e SILVA *et al.*, 2015). Existem apenas iniciativas pontuais como, por exemplo, do Instituto de Prevenção do Hospital de Câncer de Barretos, no estado de São Paulo, que desenvolveu e implementou um programa organizado. Neste programa, cartas-convite são enviadas para as mulheres na faixa etária de 25 a 64 anos, seguindo o protocolo de triagem e os exames são rotineiramente realizados em intervalos de três anos após dois exames negativos anuais. Um acompanhamento sistemático das mulheres que participam do programa de rastreamento é realizado usando um *software* desenvolvido pelo Hospital de Câncer de Barretos para esse fim. De 2012 a 2015, o percentual de exames coletados aumentou de 54,6% em 2012 para 71% em 2015. Das 31.939 mulheres que estavam programadas para repetir o exame em 2015, 78% compareceram. Desde que o sistema organizado foi implementado, 98% das mulheres compareceram à chamada para realização de colposcopia. As taxas de amostragem insatisfatória caíram de quase 0,83% usando esfregaços tradicionais para 0,14% do total de amostras coletadas com a introdução da citologia de base líquida. O índice de casos positivos aumentou de aproximadamente 1,0% (abaixo dos padrões internacionais) para quase 4,0% (dentro das recomendações internacionais). Entre 2011 e 2015, 89,4% de todos os carcinomas foram detectados no estágio clínico 0 ou I (carcinoma *in situ*), e apenas 5% nos estádios III e IV (POSSATI-RESENDE *et al.*, 2018)

Outro exemplo local de rastreamento organizado foi realizado por uma equipe de Saúde da Família do município do Rio de Janeiro, nos anos de 2014 e 2015. A equipe era responsável pelos cuidados de saúde primários de cerca de 3700 pessoas, sendo 1060 mulheres de 25 a 64 anos, com a maior parte de sua população vivendo em situação de pobreza ou extrema pobreza. A intervenção foi iniciada no mês de setembro de 2014 e sistematicamente mantida até dezembro de 2015, quando a

profissional que idealizou o projeto foi transferida para outra unidade de saúde (MAIA; SILVA; SANTOS, 2018).

A intervenção consistiu em oito etapas: 1) Mudança na organização da entrega de resultados de citopatológico na unidade de saúde; 2) Resgate e registro de exames arquivados na unidade; 3) Oferta racional de coleta facilitada para mulheres na unidade de saúde; 4) Busca ativa de mulheres que não frequentam a unidade, com agendamento imediato de coleta; 5) Solicitação de acesso a resultados de exames colhidos em outras instituições; 6) Avaliação e registro imediato em prontuário eletrônico de resultados de novos exames; 7) Aprazamento da próxima coleta, escrito no laudo do exame, que seria posteriormente entregue para a mulher, e também registrado em uma planilha eletrônica para controle de ações da própria equipe e 8) Ações imediatas para mulheres com resultados alterados (eram imediatamente convocadas para a comunicação do resultado e para a adoção das medidas pertinentes - aprazamento da repetição do exame ou agendamento de colposcopia) (MAIA; SILVA; SANTOS, 2018).

Em relação aos principais resultados desta intervenção destacam-se: o percentual de mulheres de 25 a 64 anos com pelo menos um resultado de exame citopatológico registrado em prontuário eletrônico passou de 10% para 44%; enquanto em outra equipe que não implementou tal intervenção, este percentual passou de 10 para 22%; nenhuma das 24 mulheres com exames alterados no período teve perda de seguimento e menos de cinco exames foram colhidos em mulheres fora da faixa etária alvo para o rastreamento, sendo estes, em mulheres de 65 anos ou mais sem história de exames prévios suficientes para a cessação do rastreamento, conforme recomendado pelas diretrizes nacionais. E, nenhum exame foi realizado em mulheres com menos de 25 anos (MAIA; SILVA; SANTOS, 2018).

Apesar da existência dessas experiências pontuais, o tipo predominante no Brasil ainda é o rastreamento oportunístico, que não possui uma sistematização e é caracterizado pela procura do serviço de saúde por algum outro motivo e o profissional de saúde aproveita o momento para realizar o rastreamento. A desvantagem deste tipo de rastreamento é que, além de ser menos efetivo no impacto sobre a morbidade e a mortalidade atribuídas à condição rastreada, também é mais oneroso para o sistema de saúde (BRASIL, 2010). O rastreamento oportunístico, por não realizar monitoramento e seguimento da população alvo, está relacionado com o sobrerastreamento, pois muitas mulheres não seguem as recomendações quanto à

faixa etária e periodicidade preconizadas pelo Ministério da Saúde e, desta forma, fazem exames em excesso, enquanto outras, permanecem sem acesso (AZEVEDO e SILVA *et al.*, 2015; VALE *et al.*, 2010).

Atualmente, não seria viável a implementação de um programa organizado a nível nacional, pois não há um sistema de informação com base populacional a partir do qual se identificariam as mulheres de risco, principalmente, aquelas que não estão frequentando os serviços de saúde ou não estão com seus controles atualizados (AZEVEDO e SILVA *et al.*, 2015). Uma possibilidade seria o cadastramento de mulheres e o controle de seu comparecimento para exame citopatológico por profissionais integrantes da ESF (INCA, 2011a).

Corroborando o caráter oportunístico do rastreamento do CCU no Brasil, estudo realizado em Campinas entre janeiro e dezembro de 2003, verificou que 63,4% dos exames foram considerados em excesso, ou seja, exames realizados fora das recomendações do Ministério da Saúde. Cerca de 25% das mulheres tinham menos de 25 anos e 6,5% tinham mais de 60 anos. Das mulheres submetidas ao rastreamento pela primeira vez, 76,2% tinham menos que 25 anos. A maioria dos exames foi realizada em um período de um ano após o último exame (44,5%). A porcentagem da população de 25 a 59 anos coberta pelo rastreamento foi 14,8%. Os autores estimaram que, se o excesso de exame fosse eliminado e os mesmos recursos fossem utilizados de acordo com as recomendações do Ministério da Saúde, a cobertura teria sido 65,4% (FREITAS *et al.*, 2008).

Com objetivo de avaliar o rastreamento do câncer do colo do útero na Estratégia Saúde da Família no Município de Amparo, no estado de São Paulo, foi realizado um estudo com mulheres submetidas a exame citológico na ESF de 2001 a 2007. Foram excluídos exames realizados com menos de um ano em relação ao último; aqueles em que as mulheres referiram exame anterior alterado e exames sem informação de periodicidade. O percentual de exames em excesso variou entre 61,2% em 2002, e 65,5% em 2005, o que correspondeu a quase dois terços do total de exames realizados. Mais de 50% das mulheres foram submetidas ao rastreamento com periodicidade anual e isso não se modificou ao longo dos anos. Estimou-se que, se os exames fossem realizados de acordo com as normas do Ministério da Saúde, todas as mulheres do grupo etário recomendado teriam realizado exames 2,2 vezes em sete anos, ou seja, seria atingida 100% de cobertura trienal no período. Desse modo o

rastreamento poderia se tornar mais efetivo, sem a necessidade de expandir a capacidade assistencial (VALE *et al.*, 2010).

Percebe-se que o percentual de exames em excesso no país é superior aos relatados em outros países. Estudo utilizando dados secundários de 8.018 mulheres nos Estados Unidos, em um período de seguimento de três anos, encontrou um percentual de 45% de uso excessivo da triagem. Idade mais jovem, mais visitas médicas, visitas médicas para o manejo de contraceptivos e visitas médicas ao ginecologista foram variáveis independentemente associadas ao sobrerastreamento. Semelhante ao observado aqui, nos EUA, a maioria dos profissionais de saúde continuam a solicitar triagem anual para suas pacientes e muitos desconhecem as diretrizes (ALMEIDA *et al.*, 2013)

Estudo de revisão sistemática realizado com 25 artigos publicados entre 1990 e 2016 verificou que o não seguimento das recomendações da Força-Tarefa de Serviços Preventivos dos EUA (*U.S. Preventive Services Task Force - USPSTF*) para o rastreamento do câncer do colo do útero variou de 38% a 61%. Os principais motivos para a não adesão dos profissionais de saúde às diretrizes foram: preocupações dos pacientes, discordância com recomendações, preocupação com o risco de ações judiciais por negligência e tempo limitado para discutir riscos e benefícios do rastreamento de baixo valor com os pacientes. Os resultados falso-positivos da triagem excessiva podem resultar em colposcopias e biópsias desnecessárias em pacientes que provavelmente não desenvolverão câncer invasivo, resultando em custos mais elevados, bem como dor e ansiedade (ALBER *et al.*, 2018).

Dentro deste contexto, cabe destacar o papel da prevenção quaternária voltada para evitar as consequências do intervencionismo médico excessivo, protegendo os indivíduos de intervenções inapropriadas ou desnecessárias (NORMAN; TESSER, 2009). Desta forma, tem havido uma crescente preocupação quanto ao excesso de cuidado em saúde. A sobremedicalização pode ser definida como a procura por cuidados em saúde em uma extensão na qual não confere benefícios em termos de saúde e bem-estar. Este termo se alinha aos fenômenos de sobrerastreamento, sobrediagnóstico e sobretratamento. Todos estes termos referem-se a questões clínicas do fenômeno de medicalização, desafios diariamente enfrentados pelos médicos de família e comunidade. Nesse sentido, a prevenção quaternária propõe uma série de ações para os médicos de família e comunidade “protegerem o paciente

dos excessos da medicina”, bem como para prevenirem a si próprios de sobremedicalizar (CARDOSO, 2015).

Reduzir a frequência dos exames em excesso significa otimizar os recursos disponíveis. O objetivo não deve ser a redução do número de coletas, mas sim desenvolver estratégias para que aquelas mulheres sob maior risco de desenvolver o CCU tenham acesso ao sistema, seja por demanda espontânea ou por busca ativa. Neste sentido, a ESF tem papel fundamental, pois a territorialização e o cadastramento das famílias favorecem que os agentes comunitários da saúde identifiquem a população de mulheres-alvo e convoquem aquelas que nunca fizeram o exame ou que estão em atraso (VALE *et al.*, 2010).

2.7 SISTEMAS DE INFORMAÇÃO EM SAÚDE E METODOLOGIAS DE RELACIONAMENTO DE BASES DE DADOS

Os sistemas de informação em saúde (SIS) são definidos como “*um conjunto de componentes interrelacionados que coletam, processam, armazenam e distribuem a informação para apoiar o processo de tomada de decisão e auxiliar no controle das organizações de saúde*”. Assim, os SIS são constituídos por vários subsistemas formados por conjunto de dados que fornecem informações e conhecimentos. Esses são utilizados para avaliação e planejamento das ações em saúde, subsidiando o processo de tomada de decisões (BRASIL, 2015).

No Brasil, há uma fragmentação das bases de informação do SUS, o que resulta em uma grande redundância na produção de informações em saúde, pois o mesmo indivíduo aparece em diversos sistemas, sem nenhuma integração das suas informações (BRASIL, 2015).

A integração ou relacionamento (*linkage*) das bases de dados é essencial, pois permite a identificação de um mesmo indivíduo em diversos SIS. Possibilita, assim, avaliar a linha de cuidado na rede do SUS e realizar o acompanhamento longitudinal dos indivíduos em um mesmo SIS (TOMAZELLI; GIRIANELLIII; AZEVEDO e SILVA, 2018). Por exemplo, por meio da integração entre o SISCOLO, o Sistema de Informações Ambulatoriais (SIA) e o Sistema de Informações Hospitalares (SIH) é possível fazer o seguimento de uma mulher que teve exame de *Papanicolaou* alterado, passando pela confirmação diagnóstica até o tratamento. E, relacionando

diferentes anos do SISCOLO, pode-se fazer o seguimento de mulheres ao longo do tempo para verificar a adequação do rastreamento quanto à periodicidade de repetição dos exames.

O relacionamento entre bases de dados pode ser feito de forma determinística ou probabilística. O relacionamento determinístico é realizado quando há um identificador único ou se as variáveis usadas para comparação são livres de erros e altamente discriminatórias, pois busca a concordância exata entre registros, classificando-os como pares ou não pares. Já o relacionamento probabilístico é baseado em probabilidade, levando em conta a incerteza que pode existir na comparação de variáveis usadas para comparação em ambos os arquivos (MACHADO; HILL, 2004; MOURA *et al.*, 2014).

O relacionamento entre bases de dados é simples quando há uma variável identificadora unívoca e comum entre os diferentes bancos de dados, como, por exemplo, o número do registro no cadastro de pessoas físicas (CPF) ou o número do cartão nacional de saúde (CNS) (CAMARGO JR; COELI, 2000; SPINETI *et al.*, 2016). Porém, no Brasil, em virtude da não obrigatoriedade de preenchimento destes dados em muitos SIS ou do mau preenchimento, o relacionamento entre os SIS não pode ser feito de forma direta ou determinística, sendo necessária a realização de relacionamento probabilístico entre as bases de dados (TOMAZELLI; GIRIANELLIII; AZEVEDO e SILVA, 2018).

No relacionamento probabilístico, são utilizadas várias variáveis identificadoras com diferentes graus de especificidade, tais como nome, nome da mãe, data de nascimento e endereço, que são comuns às bases a serem relacionadas. Esses campos são utilizados conjuntamente, por meio de algoritmos estatísticos, para a formação de pares de registros e cálculo de um escore, que traduz o grau de concordância entre os registros. O objetivo final, então, é verificar o quanto é provável que um par de registros se refira a um mesmo indivíduo (CAMARGO JR; COELI, 2000; BLAKELY; SALMOND, 2002; SOUSA *et al.*, 2008; CAMARGO JR., COELI, 2007).

No Brasil, um *software* muito utilizado para o relacionamento de bases de dados é o RecLink®, o qual foi desenvolvido para fazer o *linkage* de registros levando em consideração a fonética da língua portuguesa (CAMARGO JR.,; COELI, 2000; COELI; CAMARGO JR, 2002). Estudos que o utilizaram para o relacionamento entre bases de dados encontraram boa acurácia (SPINETI *et al.*, 2016; PERES *et al.*, 2014; COUTINHO; COELI, 2006).

O relacionamento probabilístico é composto por um conjunto de processos: padronização; blocagem; pareamento e classificação dos pares em verdadeiros ou falsos (COELI; CAMARGO JR, 2002).

Na padronização, os campos em comum que serão utilizados nas próximas etapas são homogeneizados para se tornarem comparáveis utilizando rotinas específicas visando minimizar a ocorrência de erros no processo de relacionamento (CAMARGO JR., COELI, 2007). São exemplos de rotinas utilizadas nesta etapa: retirada de preposições de campos nominais; transformação dos campos em maiúsculos; retirada de pontuações, símbolos e espaços duplos; transformação das variáveis a serem relacionadas em um mesmo formato e transformação do primeiro e último nome em código fonético (ABREU *et al.*, 2017; RIBEIRO, 2019; CAMARGO JR.,; COELI, 2000).

A transformação do primeiro e último nomes em código fonético visa minimizar o problema do erro de registro, sendo que o código fonético mais utilizado é o *soundex*. Este algoritmo converte a palavra em um código alfanumérico contendo uma letra e três números, com base na fonética das palavras (COELI; CAMARGO JR., 2002). Desta forma, pequenas diferenças na escrita ou pronúncia originam o mesmo código (OLIVEIRA *et al.*, 2018b).

O processo de blocagem consiste na criação de blocos lógicos de registros dentro dos arquivos a serem relacionados. Este processo permite que as bases de dados sejam logicamente divididas em blocos mutuamente exclusivos, sendo as comparações limitadas aos registros pertencentes a um mesmo bloco. Os blocos são constituídos de forma a aumentar a probabilidade de que os registros neles contidos representem pares verdadeiros. Se a chave de blocagem selecionada for o bairro de residência, por exemplo, só serão comparados os registros nas duas bases que apresentem o mesmo bairro de residência (CAMARGO JR.; COELI, 2000).

O principal objetivo desta etapa é otimizar a comparação entre os registros. O número de pares possíveis com a combinação de duas bases de dados seria equivalente ao produto entre o número de registros das duas bases. Com a blocagem, as comparações ficam limitadas a registros do mesmo bloco, ou seja, àqueles que possuem, em comum, valores de todas as variáveis contidas em cada chave de blocagem, o que diminui o total de pares formados a cada passo e aumenta a probabilidade de formação de pares verdadeiros (COELI; CAMARGO JR., 2002; MIGOWSKI *et al.*, 2011; CAMARGO JR., COELI, 2007).

Como registros de um mesmo indivíduo podem ser classificados em blocos diferentes em virtude de erros de registro, o processo de blocagem deve ser feito em múltiplos passos. Desta forma, os registros não pareados na primeira etapa podem ser blocados na próxima, ao ser utilizada outra variável para a blocagem (COELI; CAMARGO JR., 2002)

Após a blocagem, a etapa seguinte é o pareamento na qual calcula-se o escore atribuído ao relacionamento entre dois registros, que sumariza o grau de concordância global entre eles. Pares com pontuações altas têm maiores probabilidades de serem verdadeiros e pares com pontuações baixas têm menores probabilidades (CAPUANI *et al.*, 2014). Ressalta-se que, na rotina do pareamento não devem ser incluídas as mesmas variáveis utilizadas na blocagem (CAMARGO JR., COELI, 2007).

O escore final de cada par é construído a partir da soma dos escores ponderados de cada campo empregado nesta etapa, o que permite que cada um contribua de modo diferenciado para o escore total do par. Para cada campo (i), define-se a probabilidade (m_i) do campo identificar um par como verdadeiro quando ele realmente é verdadeiro (sensibilidade) e a probabilidade (u_i) de identificar um par como verdadeiro, quando na realidade ele é falso (1 - especificidade). Com base nestas probabilidades de concordância/discordância entre os pares são construídos os fatores de ponderação, um para a situação de concordância e o outro, para a de discordância. O escore total de determinado par é obtido a partir da soma dos fatores de ponderação atribuídos após a comparação de cada campo avaliado. Sendo que, o fator de concordância contribui positivamente para o escore final, enquanto o fator de discordância contribui negativamente (CAMARGO JR.; COELI, 2000).

Após o cálculo do escore dos pares formados, o grande desafio é definir um ponto de corte acima do qual os pares são classificados como verdadeiros (*match*) e abaixo dos quais os pares são categorizados como falsos (*non-match*) (BLAKELY, SALMOND, 2002). O procedimento mais utilizado e recomendado pela literatura para esta classificação é a inspeção manual de todos os pares. Porém, quando as bases de dados são muito grandes, tal procedimento torna-se inviável e uma opção pode ser a utilização de uma amostra dos pares para determinação do melhor ponto de corte e posterior aplicação deste para classificação automática de todos os pares (SILVEIRA, ARTMANN, 2009; PITA *et al.*, 2018).

Por ser probabilístico, este tipo de relacionamento é suscetível a erros no pareamento. Alguns pares referentes a indivíduos diferentes podem ser erroneamente

classificados como verdadeiros (falsos-positivos), enquanto outros, do mesmo indivíduo podem não ser identificados como pares verdadeiros (falsos-negativos) – (COUTINHO; COELI, 2006).

Uma forma de avaliar a qualidade dos métodos de relacionamento é por meio de estudos de acurácia. Para isso, é necessário comparar os resultados obtidos no processo de relacionamento com uma fonte de informação independente, ou seja, um padrão-ouro (SILVEIRA, ARTMANN, 2009). Esse pode ser obtido por meio de uma fonte de dados com identificadores completos ou de uma subamostra dos registros que foram revisados manualmente (HARRON *et al.*, 2017).

Os erros falsos-positivos ocorrem mais frequentemente quando são disponíveis poucos campos para a comparação, os campos de identificação apresentam baixa completude, a proporção de homônimos é elevada e as bases relacionadas apresentam grande volume de dados. Já os erros falsos-negativos ocorrem devido à informação obtida ser incorreta, a erros de digitação e à ocorrência de dados faltantes nas bases (COELI, 2015).

Os parâmetros utilizados para avaliar a qualidade do relacionamento probabilístico, em geral, são: sensibilidade, especificidade, acurácia e a curva ROC (*Receiver Operating Characteristics*) para ilustrar o desempenho. Enquanto para conjuntos de dados desequilibrados, nos quais o número de pares negativos supera o de positivos, os parâmetros recomendados são: *precision*, *recall* e medida F (SAITO; REHMSMEIER, 2015).

Precision, também conhecido como valor preditivo positivo, é definido como a proporção de pares formados que estão corretos. É calculado da seguinte forma (BOYD *et al.*, 2016):

$$Precision = \frac{\text{número de verdadeiros-positivos}}{\text{número de verdadeiros-positivos} + \text{número de falsos-positivos}}$$

Recall, também conhecido como sensibilidade, é a proporção de todos os pares verdadeiros que foram vinculados corretamente e pode ser calculado da seguinte forma (BOYD *et al.*, 2016):

$$Recall = \frac{\text{número de verdadeiros-positivos}}{\text{número de verdadeiros-positivos} + \text{número de falsos-negativos}}$$

Assim, relacionamentos com valores elevados de *precision* terão poucos falsos-positivos e, altos valores de *recall* estarão associados a poucos falsos-negativos (RANDALL *et al.*, 2013).

Outro parâmetro utilizado para avaliar a qualidade do *linkage* é a medida F, uma média harmônica dos parâmetros *precision* e *recall* (FERRANTE; BOYD, 2012):

$$\text{Medida F} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Altos valores de *precision* e de *recall* levam a uma medida F elevada, portanto, essa é uma maneira de encontrar a melhor combinação entre os dois parâmetros (FERRANTE; BOYD, 2012).

A curva *precision-recall* (PRC) mostra a relação entre valores de *precision* e de *recall* para cada *cut-off* possível e, semelhante à curva ROC, fornece uma avaliação ampla do modelo. O eixo y mostra os valores de *precision* e o eixo x, os valores de *recall* e cada ponto da curva representa um ponto de corte que corresponde a um determinado valor de *recall* e de *precision* (SAITO; REHMSMEIER, 2015). O *linkage* perfeito é caracterizado por apresentar uma curva *precision-recall* (PRC) que passa pelo canto superior direito (correspondendo a 100% de *precision* e 100% de *recall*) – Figura 1. Desta forma, quanto mais próxima a PRC estiver do canto superior direito, melhor será a qualidade do *linkage* (EKELUND, 2017).

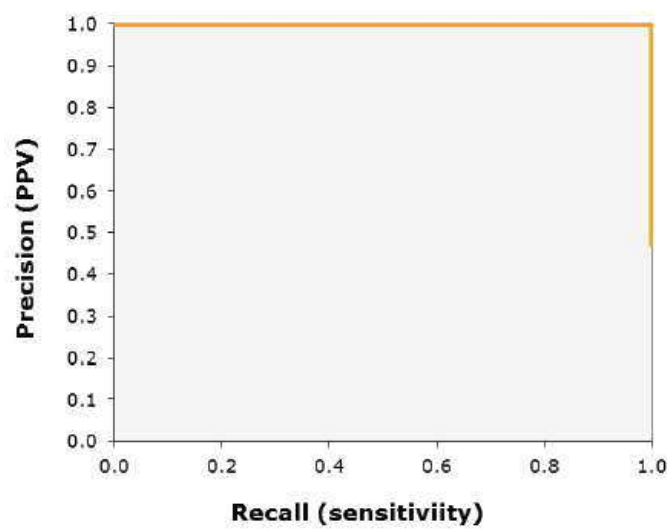


Figura 1 – Curva *precision-recall* exemplificando um *linkage* perfeito
Fonte: EKELUND, 2017

3 JUSTIFICATIVA

Apesar de ser uma doença evitável e da expansão das políticas de controle do CCU ao longo do tempo no Brasil, o câncer cervical ainda representa um desafio no país e percebem-se grandes diferenças regionais. Dados da Pesquisa Nacional de Saúde mostraram aumento da cobertura do *Papanicolaou* entre 2003 e 2013, porém, observa-se que apenas o acesso aos exames não é suficiente para gerar impacto no perfil epidemiológico do CCU, já que a maioria dos casos ainda é diagnosticada em fases avançadas. Além disso, o aumento da oferta não foi acompanhado por aprimoramento da qualidade do rastreamento e grande parte dos exames não segue as recomendações do Ministério da Saúde, principalmente, quanto à periodicidade de repetição. Desta forma, há um sobrerastreamento de um grupo de mulheres, enquanto muitas mulheres que apresentam maior risco para a doença permanecem sem acesso. O sobrerastreamento é típico de programas oportunistas e está associado com a má utilização dos recursos públicos e pouco impacto na carga da doença.

O SISCOLO, implantado em 1999, foi o sistema de informações do SUS criado para viabilizar o monitoramento das ações relacionadas ao rastreamento do CCU. Uma importante limitação deste sistema é o fato de o registro ser feito por exame ou procedimento e não por mulher, impossibilitando o acompanhamento longitudinal das mulheres no programa de rastreamento. Os dados do sistema permitem o cálculo de um *proxy* da cobertura, por meio da divisão entre o número de exames realizados na faixa etária alvo do rastreamento e a população feminina nesta faixa etária. Porém, como existe o problema do sobrerastreamento, muitas mulheres fazem mais de um exame por ano, o que pode superestimar este indicador. Por outro lado, este indicador não considera no denominador as mulheres cobertas por planos de saúde privados, subestimando-o. O sistema também possui uma variável que informa se é o primeiro exame da mulher e, caso não seja, qual o ano do exame anterior, o que permite estimar a periodicidade de realização do rastreamento. Porém, por ser uma informação autorrelatada, também está sujeita a viés. Além disso, não permite o conhecimento da periodicidade de exames anteriores.

Em 2013, foi instituído o SISCAN visando substituir e integrar os sistemas de informação relacionados ao câncer feminino. Tal sistema adota como unidade de observação a mulher, e não o exame. No entanto, este sistema ainda está em

processo de implementação e alguns municípios ainda registram os exames no SISCOLO, não tendo havido uma transição definitiva para o SISCAN em todo o país.

Este estudo se justifica pela necessidade de análise dos dados do SISCOLO tendo como unidade de análise a mulher e não o exame. Para tal, foi feito inicialmente, estudo descritivo com base em dados agregados do SISCOLO e, posteriormente, realizado o relacionamento probabilístico das bases do SISCOLO contendo dados identificados. Permitindo assim, a identificação das mulheres submetidas ao rastreamento do CCU, o cálculo da cobertura do programa de rastreamento e a realização de seguimento longitudinal das mulheres para verificar a adequação das ações em relação à periodicidade e faixa etária preconizadas pelo Ministério da Saúde.

Cabe destacar a importância de desenvolver metodologia adequada e efetiva para a avaliação do rastreamento visando orientar e aprimorar a política de controle do câncer do colo do útero no Brasil, alinhando-se aos objetivos da OMS de eliminar, até 2030, a doença em todos os países, baseado em metas de vacinação, rastreamento e tratamento (WHO, 2020).

4 OBJETIVOS

4.1 GERAL

Avaliar as ações de rastreamento do câncer do colo do útero pelo Sistema Único de Saúde, em Minas Gerais e sua conformidade com as recomendações do Ministério da Saúde, com base em dados do SISCOLO.

4.2 ESPECÍFICOS

1. Avaliar as ações de rastreamento com base em dados agregados de 2006 a 2011 do SISCOLO a partir da análise de indicadores relacionados às características da oferta de exames citopatológicos, à qualidade do exame, à ocorrência de alterações celulares nos exames e ao seguimento informado dos casos de lesão de alto grau (artigo 1);
2. Estimar a cobertura anual do rastreamento do câncer do colo do útero e a proporção de sobrerastreamento por periodicidade e faixa etária a partir da identificação, nas bases do SISCOLO, das mulheres rastreadas pelo SUS em Minas Gerais no ano de 2008 (artigo 2).
3. Avaliar a desigualdade regional da cobertura do rastreamento do câncer do colo do útero em Minas Gerais e sua associação com indicadores socioeconômicos e assistenciais e descrever a distribuição espacial do sobrerastreamento nas microrregiões do estado, no ano de 2008 (artigo 3)

5. MÉTODOS

5.1 MÉTODOS REFERENTES AO OBJETIVO ESPECÍFICO 1 (ARTIGO 1)

Os métodos referentes ao objetivo específico 1 estão descritos no artigo intitulado “Rastreamento do câncer do colo do útero em Minas Gerais: avaliação a partir de dados do Sistema de Informação do Câncer do Colo do Útero (SISCOLO)”, publicado na Cadernos Saúde Coletiva, em 2017 – página 82 (CORRÊA *et al.*, 2017-<https://doi.org/10.1590/1414-462x201700030201>).

5.2 MÉTODOS REFERENTES AO OBJETIVO ESPECÍFICO 2 (ARTIGO 2)

5.2.1 Tipo de estudo e aspectos éticos

Estudo descritivo, de coorte não-concorrente, baseado em dados identificados do SISCOLO.

O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora sob o parecer 1.237.274/2015. Durante todo o trabalho foram adotadas medidas preventivas para garantir a segurança dos dados pessoais disponibilizados. O arquivamento dos dados foi feito em unidade de armazenamento externo (HD externo) protegido por senha e sempre que os bancos de dados foram manipulados, a internet do notebook era desabilitada.

5.2.2 Fonte dos dados

Este estudo utilizou dados identificados do SISCOLO, referentes ao estado de Minas Gerais, de 2007 a 2012, cujas bases foram cedidas pela Secretaria de Saúde do Estado (SES-MG) e extraídas do módulo coordenação estadual do SISCOLO, por ano e mês de competência. Os arquivos foram disponibilizados no formato de competência mensal para o período de 01 de janeiro de 2007 a 31 de dezembro de 2012. Neste estudo foi utilizado apenas o módulo *CITCOLO*, referente aos registros de exames citopatológicos do colo do útero.

5.2.3 População de estudo

A população de estudo foi constituída por mulheres residentes no estado de Minas Gerais, submetidas ao rastreamento do câncer do colo do útero na rede pública de saúde, no ano de 2008.

5.2.4 Variáveis

Foram utilizadas as seguintes variáveis do SISCOLO: nome da mulher, nome da mãe, data de nascimento, endereço, município de residência, estado, CPF, cartão nacional de saúde (CNS), escolaridade, data da coleta do exame, exame citopatológico anterior, ano do exame anterior, amostra rejeitada, adequabilidade da amostra, epitélios representados na amostra, células atípicas de significado indeterminado, atipias em células escamosas e/ou glandulares, outras neoplasias malignas.

Foi criada uma variável identificadora (*ID*) que atribuiu um código único para cada exame da base inicial, possibilitando sua posterior identificação. A partir da variável *data de coleta do exame*, foram extraídas as variáveis *ano* e *mês* do exame. A variável estado de residência foi criada por meio da extração dos dois primeiros dígitos da variável código do município de residência. Além disso, foram incluídas as variáveis microrregião e macrorregião de residência da mulher nos bancos, identificadas a partir do município.

A variável idade foi criada a partir da subtração do ano de coleta do exame e o ano de nascimento e, posteriormente, categorizada nas seguintes faixas etárias: <25 anos; 25-29; 30-39; 40-49; 50-59; 60-64 e ≥65 anos.

5.2.5 Bases de dados utilizadas

Os arquivos das competências mensais do SISCOLO referentes aos anos de 2007 a 2012 foram reunidos em bases anuais e todas passaram pelo mesmo processo de limpeza, padronização das variáveis e formação dos respectivos bancos com objetivo de otimizar e permitir o posterior relacionamento entre eles.

Como o SISCOLO é utilizado para pagamento dos prestadores de serviços, em determinada base anual, podem existir exames que foram realizados em anos anteriores. Dessa forma, utilizando a variável *data de coleta do exame*, exames processados em outros anos foram transferidos para a base do respectivo ano de realização.

O banco de dados de referência, utilizado para montar o seguimento, foi baseado no *Banco 2008* após limpezas do mesmo. O *Banco 2007* foi utilizado para limpar o *Banco 2008*, visando evitar a superestimação do sobrerastreamento, permitindo a exclusão de mulheres que fizeram o exame em 2008 e tiveram exame alterado ou com amostra insatisfatória/rejeitada no ano anterior. Os demais anos (2009-2012) constituíram os *Bancos de comparação* que foram relacionados com o *Banco Referência* para montar o seguimento e obter as repetições de exames a partir do primeiro exame normal realizado em 2008.

5.2.6 Limpezas da base de 2008 do SISCOLO e formação do *Banco 2008*

Para compor o *Banco 2008* foram selecionadas mulheres residentes em Minas Gerais que realizaram exame citopatológico no ano de 2008. Foram excluídos registros com o campo nome em branco ou preenchido de forma que impossibilitasse a identificação da mulher e registros sem a data da realização do exame. Além disso, foram excluídos registros de mulheres com idade inferior a 12 anos, exames duplicados e exames de mulheres que tiveram resultado alterado ou com amostra insatisfatória/rejeitada em 2007. As repetições de exame pela mesma mulher em 2008 foram agrupadas em outro banco (*Seguimento 2008*).

Foram considerados como resultados alterados aqueles cujo laudo constava: atipias de significado indeterminado: escamosa - possivelmente não-neoplásica; escamosa - não se pode afastar lesão intraepitelial de alto grau; glandular - possivelmente não-neoplásica; glandular - não se pode afastar lesão intraepitelial de alto grau; origem indefinida - possivelmente não-neoplásicas e origem indefinida - não se pode afastar lesão intraepitelial de alto grau; lesões de baixo grau (lesão intraepitelial de baixo grau) e lesões de alto grau (lesão intraepitelial de alto grau, lesão intraepitelial de alto grau não podendo afastar micro-invasão); carcinoma escamoso (carcinoma epidermóide invasor); adenocarcinoma (adenocarcinoma *in situ* e adenocarcinoma invasor) e outras neoplasias.

5.2.6.1 Padronização

A padronização das variáveis foi realizada excluindo-se acentos, cedilhas, símbolos e números das variáveis nominais e especificando os formatos das variáveis numéricas e de data. Os campos nominais também foram padronizados com a

retirada de espaços duplos, pontuações e preposições e transformados em caixa alta. Os campos nome da mãe e data de nascimento que estavam em branco foram substituídos por um código (C08) e o mesmo procedimento foi feito nos demais bancos (2007: código C07) com objetivo de impedir, no momento da utilização da técnica de relacionamento probabilístico, a formação de pares entre registros que tivessem como características em comum apenas a falta de preenchimento. Os *scripts* utilizados no programa *R* (versão 4.0.2) para o preparo, limpezas e padronização da base 2008 estão descritos no apêndice 1. Nesta etapa, utilizou-se também o módulo *Padroniza* do programa *RecLink III*, para padronização das variáveis e transformação do primeiro e último nomes da mulher e da mãe em código *soundex* (apêndice 2). Além da base 2008, todas as outras utilizadas no estudo passaram por esse mesmo processo de padronização.

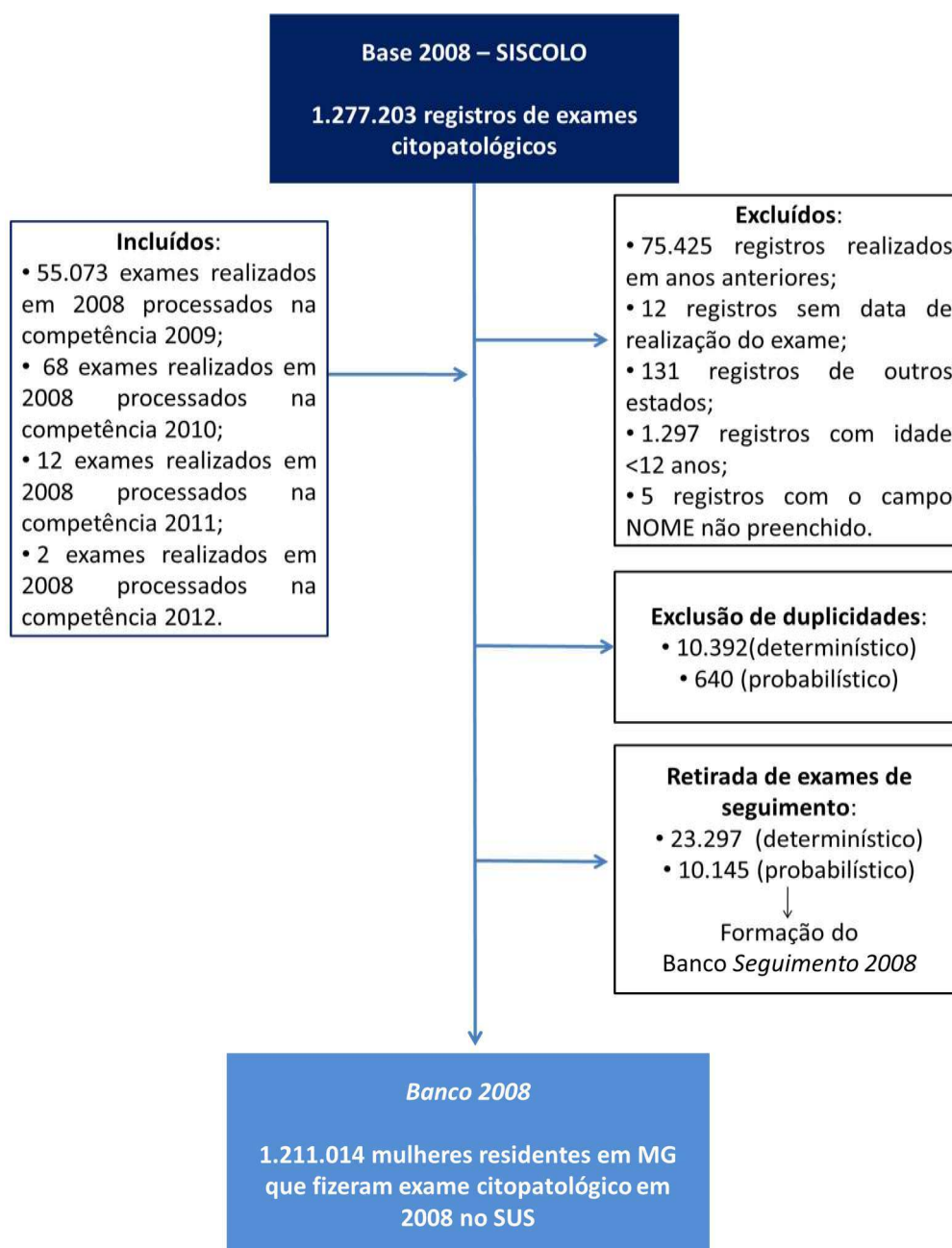
5.2.6.2 Retirada de duplicidades e exames de seguimento

Após a padronização da base de 2008, os registros duplicados (exames registrados mais de uma vez no sistema) foram excluídos, inicialmente, por meio de técnica de relacionamento determinístico utilizando o programa *R* (apêndice 3). Desta forma, foram excluídas 10.392 duplicidades, através da função *duplicated* utilizando as variáveis: nome, nome da mãe, data de nascimento; data de coleta do exame; CNES da unidade de saúde e código do município de residência.

Posteriormente, foi utilizado o módulo *Duplicidade* do *RecLink III* com técnica de relacionamento probabilístico em duas etapas (apêndice 3) e as duplicidades foram excluídas seguindo as orientações do manual do programa para este módulo (CAMARGO JR.; COELI, 2007). Na primeira etapa, foram excluídos 510 registros duplicados empregando como campos de bloqueio: códigos *soundex* do primeiro e último nome da mulher, data de coleta do exame, ano de nascimento e CNES da unidade de saúde e, como campos de comparação: nome da mulher, nome da mãe, data de nascimento e código do município de residência. Na segunda etapa, foram empregadas as mesmas variáveis de comparação da primeira, com os seguintes campos de bloqueio: códigos *soundex* do primeiro e último nome da mãe, data de coleta do exame, ano de nascimento, CNES da unidade de saúde e CNES do prestador (laboratório), sendo, então, excluídos 130 registros duplicados.

Os exames referentes ao seguimento, ou seja, mulheres que tiveram mais de um exame realizado em 2008 (repetições), foram retirados da base de 2008 (apêndice 4), visando manter apenas o primeiro exame realizado nesse ano. Para isso, essa base foi ordenada pela data de coleta do exame. Inicialmente, também foi utilizado relacionamento determinístico, empregando-se as variáveis: nome da mulher, nome da mãe, data de nascimento e código do município de residência. Nesta etapa, foram identificados 23.297 exames de seguimento em 2008. Posteriormente, foi conduzido relacionamento probabilístico em duas etapas utilizando o módulo *Duplicidade* do *RecLink III*. Identificaram-se 9.228 exames de seguimento na primeira etapa e 917 na segunda. Utilizaram-se como campos de bloqueio: código *soundex* do primeiro e do último nome da mulher e ano de nascimento na primeira etapa e código *soundex* do primeiro e do último nome da mãe e ano de nascimento na segunda; os campos de comparação utilizados em ambas as etapas foram: nome da mulher, nome da mãe, data de nascimento e código do município. Após a retirada desses exames, eles foram agrupados em outro banco (*Seguimento 2008*), contendo todos os exames de seguimento do ano 2008. Assim, após essas limpezas, formou-se o *Banco 2008*, cuja unidade de registro passou a ser a mulher, contendo apenas o seu primeiro exame realizado no ano de 2008 (1.211.014 mulheres). A figura 2 ilustra o processo de limpezas da base e formação do *Banco 2008*.

Figura 2 - Limpezas da base 2008 do SISCOLO e formação do *Banco 2008*



5.2.6.3 Exclusão de mulheres do Banco 2008 com resultados alterados ou amostra insatisfatória/rejeitada em 2007

Com objetivo de excluir mulheres com exames alterados ou que tiveram amostra insatisfatória ou rejeitada no ano de 2007, relacionou-se o *Banco 2008* (n=1.211.014 mulheres) com o banco contendo mulheres com exames alterados, rejeitados/amostra insatisfatória realizados em 2007 (n=36.228 mulheres).

Visando a formação do *Banco 2007* para este relacionamento, a base desse ano passou pelo mesmo processo de limpeza, padronização e retirada de duplicidades/seguimento ao qual a base de 2008 foi submetida. Sendo que, inicialmente, a base 2007 era composta por 1.303.932 exames registrados neste ano.

O relacionamento probabilístico entre os bancos 2007 e 2008 foi feito utilizando-se o programa *RecLink III* em quatro passos (APÊNDICE 5). A princípio, foram propostos 16 passos e as chaves de blocagem e pareamento foram adaptadas de estudo que também utilizou dados identificados do SISCOLO (RIBEIRO, 2019) – descritos no quadro 1. No entanto, seguindo a metodologia proposta por Duarte *et al.* (2019), verificou-se que, apenas quatro passos (2, 9, 12 e 16) eram suficientes e apresentavam boa acurácia na identificação dos pares verdadeiros. Os parâmetros para o relacionamento foram definidos de acordo com as orientações do manual do *RecLink* (CAMARGO JR.; COELI, 2007).

Quadro 1 – Dezesesseis passos propostos inicialmente para serem empregados nos relacionamentos probabilísticos, com respectivas variáveis de blocagem e de pareamento.

Passo	Variáveis de blocagem	Varáveis de pareamento
1	PN, UN, AN	NOME, MAE, DTNASC, MUN_COD
2	PN, UN, AN	NOME, DTNASC
3	CNS	NOME, MAE, DTNASC
4	PN, UN, PM, UM, AN	NOME, MAE, DTNASC
5	PN, UN, AN	NOME, MAE, DTNASC
6	PN, PM, UM, AN	NOME, MAE, DTNASC
7	PN, UM, AN	NOME, MAE, DTNASC
8	PN, PM, AN	NOME, MAE, DTNASC
9	PN, AN	NOME, MAE, DTNASC
10	PN, UN, PM, AN	NOME, MAE, DTNASC
11	PN, UN, PM	NOME, MAE, DTNASC
12	PN, UM	NOME, MAE, DTNASC
13	PN, PM, UM	NOME, MAE, DTNASC
14	PN, UM	NOME, MAE, DTNASC
15	UN, AN	NOME, MAE, DTNASC
16	AN	NOME, MAE, DTNASC

Legenda: PN: código *soundex* do primeiro nome da mulher; UN: código *soundex* do último nome da mulher; PM: código *soundex* do primeiro nome da mãe; UM: código *soundex* do último nome da mãe; AN: ano de nascimento; CNS: cartão nacional de saúde; NOME: nome padronizado da mulher; MAE: nome padronizado da mãe; DTNASC: data de nascimento; MUN_COD: código do município.

Nota: em negrito os quatro passos utilizados para o relacionamento do *Banco 2008* com o banco contendo mulheres com resultados alterados em 2007.

Em virtude do tamanho das bases, a inspeção manual para classificação dos pares em verdadeiro ou falso foi inviável. Desta forma, optou-se pela utilização de método automatizado de seleção dos pares verdadeiros com base na inspeção manual de uma amostra aleatória de 300 pares (DUARTE *et al.*, 2019).

Para os quatro passos utilizados no relacionamento foi selecionada aleatoriamente uma amostra de 300 pares com escore positivo. Estes foram, então, inspecionados manualmente e classificados como par ou não par utilizando o módulo *Combina – Inspecciona* do *RecLink III*. Os critérios utilizados para esta classificação estão descritos no quadro 2 (RIBEIRO, 2019). Os pares que atendiam a, pelo menos, um dos critérios foram considerados verdadeiros.

Quadro 2 – Critérios para classificação manual dos pares formados no relacionamento probabilístico

Critério	Descrição
A	Mesma data de nascimento + nome idêntico + nome da mãe idêntico ou abreviado ou faltando nomes do meio em um dos bancos de dados.
B	Nome idêntico + nome da mãe idêntico + data de nascimento até dois dígitos diferentes ou a troca do dia por mês.
C	Nome e/ou nome da mãe incomum (ex.: nome estrangeiro) mas semelhante entre bancos de dados e com a mesma data de nascimento ou mesmo endereço.
D	Nome ou nome da mãe ou data de nascimento diferente ou ausência de informações de nome da mãe ou data de nascimento em um dos bancos de dados; demais campos contendo informações idênticas ou muito semelhantes (como endereço), com correspondência em pelo menos três campos.

Fonte: Ribeiro, 2019

Após inspeção manual dos pares das amostras foram feitas as curvas *precision-recall* e calculados os valores de *precision*, *recall* e medida F para determinação do ponto de corte do escore de maior poder discriminatório na identificação dos pares verdadeiros (QUADRO 3) - (SAITO; REHMSMEIER, 2015; DUARTE *et al.*, 2019). O *script* utilizado no programa *R* para construção das curvas e

determinação do ponto de corte e dos respectivos valores de *precision*, *recall*, medida F e acurácia está descrito no apêndice 6; as curvas estão ilustradas no apêndice 7.

Quadro 3 – Descrição dos passos utilizados no relacionamento entre o Banco 2008 e o banco de mulheres que tiveram exames alterados/rejeitados em 2007 e os respectivos parâmetros e pares verdadeiros identificados

Passo	Blocagem	Pareamento	Número de pares formados	Score	Ponto de corte	Amostra de 300 pares			Total
						Precision; Recall	Medida F	Acurácia	
2	PN, UN, AN	NOME, DTNASC	104.989	-6,87 – 10,69	10,1722	0,93; 0,91	0,92	0,98	15.329
9	PN, AN	NOME, MAE, DTNASC	28.507	-10,50 – 17,21	6,7379	0,99; 0,98	0,98	0,98	16.412
12	PN, UN	NOME, MAE, DTNASC	37.034	-10,50 – 17,21	6,7207	0,91; 0,96	0,93	0,95	16.928
16	AN	NOME, MAE, DTNASC	62.416	-10,50 – 17,21	6,7379	1,00; 0,99	0,99	0,99	16.927

Legenda: PN: código *soundex* do primeiro nome da mulher; UN: código *soundex* do último nome da mulher; AN: ano de nascimento; NOME: nome padronizado da mulher; MAE: nome padronizado da mãe; DTNASC: data de nascimento.

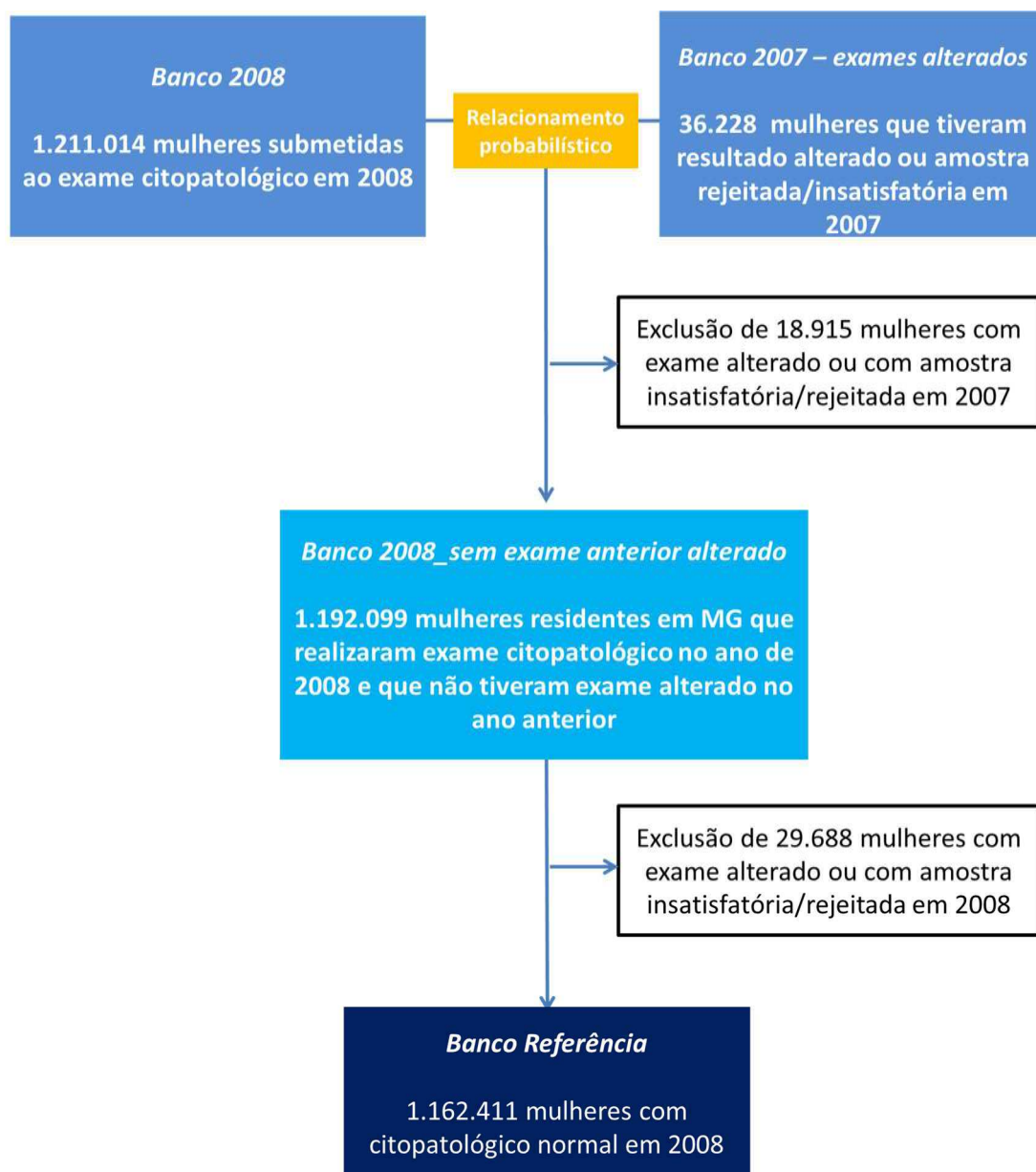
Posteriormente, aplicou-se o ponto de corte encontrado nas amostras ao total de pares dos passos 2, 9, 12 e 16 para a classificação automatizada dos pares, visando a seleção dos pares verdadeiros. Assim, foram excluídas 18.915 mulheres do Banco 2008 que tiveram exame alterado/rejeitado em 2007.

Após essas limpezas, o banco composto por mulheres que realizaram exame citopatológico do colo do útero em Minas Gerais neste ano e não tiveram exame alterado no ano anterior ficou com um total de 1.192.099 mulheres (Figura 3).

5.2.6.4 Criação do Banco Referência (coorte 2008)

O Banco Referência foi formado pelo Banco 2008_*sem exame anterior alterado* com a exclusão de 29.688 mulheres cujos exames coletados nesse ano tiveram resultado alterado (1,7% do Banco 2008_*sem exame anterior alterado*) ou amostra rejeitada (0,05% do Banco 2008_*sem exame anterior alterado*) ou insatisfatória (0,7% do Banco 2008_*sem exame anterior alterado*). Desta forma, o Banco Referência foi composto por 1.162.411 mulheres residentes em Minas Gerais, que tiveram exame citopatológico do colo do útero realizado em 2008 com resultado normal.

A figura 3 ilustra o processo de limpeza do Banco 2008 através da exclusão de mulheres que tiveram exame alterado em 2007 e em 2008 e a posterior formação do Banco Referência.

Figura 3- Limpeza no *Banco 2008* e formação do *Banco Referência*

5.2.7 Preparo e formação dos bancos de comparação

As bases anuais do SISCOLO referentes a exames realizados de 2009 a 2012 passaram pelo mesmo processo de limpeza, padronização e exclusão de duplicidades ao qual a base de 2008 foi submetida (descrito nos itens 5.2.6.1 e 5.2.6.2), exceto pela retirada de exames de seguimento. Após tais limpezas, estas bases formaram os respectivos bancos anuais (Banco 2009, Banco 2010, Banco 2011, Banco 2012) que, posteriormente, foram relacionados probabilisticamente com o *Banco Referência* para formar o seguimento 2008-2012.

O quadro 4 descreve as limpezas e formação dos bancos 2009, 2010, 2011 e 2012 e o tamanho final de cada base.

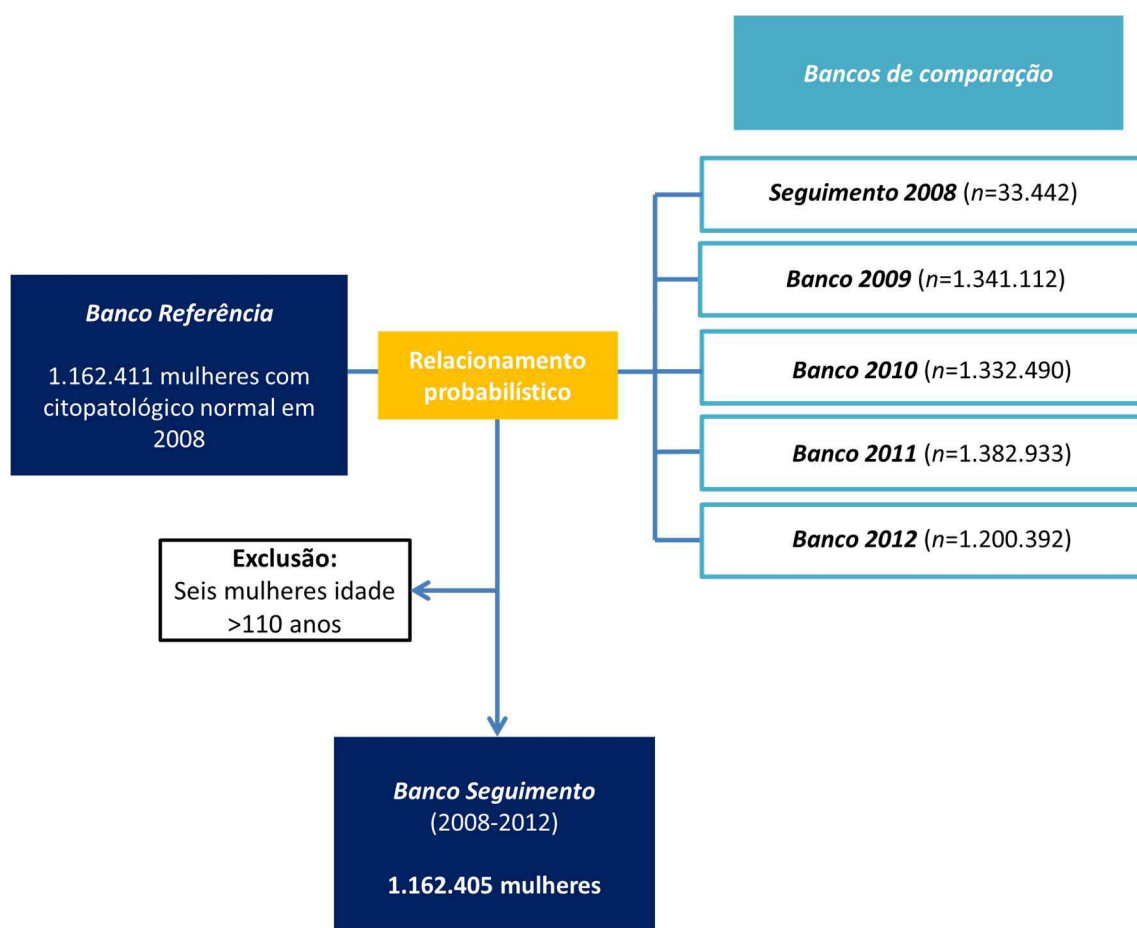
Quadro 4 – Limpezas das bases de 2009 a 2012 e formação dos respectivos bancos anuais

BANCO	BASE ORIGINAL	INCLUSÃO		EXCLUSÃO							Número final de registros
		Número de registros	Registrados em outras competências	Sem data de coleta do exame	Outros anos	Sem nome	Outros estados	Duplicidades		Probabilístico	
2009	1.348.128	72.255	3	55.836	10	153	5.484	1.369	1.341.112		
2010	1.337.089	70.225	3	72.501	7	216	3.770	606	1.332.490		
2011	1.388.702	72.945	1	70.489	6	135	4.687	941	1.382.933		
2012	1.211.348	----	14	73.325	1	122	10.310	523	1.200.392		

5.2.8 Construção do *Banco Seguimento* (2008-2012)

O seguimento das mulheres que tiveram resultado do citopatológico normal em 2008 foi construído através do relacionamento probabilístico entre o *Banco Referência* e os *Bancos de comparação* (*Seguimento 2008*; *Banco 2009*, *Banco 2010*, *Banco 2011* e *Banco 2012*), conforme ilustrado na figura 4.

Figura 4 – Construção do *Banco Seguimento* (2008-2012)



Para isso, foi utilizado o programa *RecLink III*, seguindo a metodologia descrita no item 5.2.6.3, ou seja, seguiu as recomendações do manual do programa e a metodologia proposta por Duarte *et al.* (2019). Cada relacionamento foi feito utilizando quatro passos; seleção aleatória de uma amostra de 300 pares para conferência manual dos pares formados em cada passo e posterior definição do ponto

de corte e seleção automatizada de pares verdadeiros. Para o relacionamento entre o *Banco Referência* e o *Seguimento 2008*, foram utilizados os passos 2, 9, 12 e 16, porém, para o relacionamento com cada banco anual de 2009 a 2012, em virtude do tamanho das bases, o passo 16 (chave de blocagem: ano de nascimento) tornou-se inviável de ser executado. Optou-se, então, pela substituição deste pelo passo 15 (chaves de blocagem: código *soundex* do último nome da mulher e ano de nascimento), pois foi verificado previamente que este identificava 95% dos pares identificados pelo passo 16 que os demais (2, 9 e 12) não identificavam. Desta forma, no relacionamento entre o *Banco Referência* e os bancos 2009, 2010, 2011 e 2012, foram utilizados os passos 2, 9, 12 e 15. Por problemas relacionados ao tamanho do banco gerado pelo relacionamento, foi necessário fracionar o passo 2 em quatro etapas, utilizando quatro partes diferentes do *Banco Referência*, de acordo com o trimestre de realização do exame: REF1 (n=282.205); REF2 (n=316.429); REF3 (n=301.247) e REF4 (n=262.530).

O apêndice 8 descreve os passos e os dados referentes ao relacionamento entre o *Banco Referência* e os *Bancos de comparação* para formação do seguimento 2008-2012.

Ao final do relacionamento, procedeu-se à união dos arquivos de pares dos respectivos passos (apêndice 9) e obteve-se um arquivo contendo 3.431.089 pares resultantes do relacionamento 2008-2012 (quadro 5). Este arquivo foi adicionado ao *Banco Referência* utilizando o programa *R* formando, assim, o *Banco Seguimento 2008-2012* que foi organizado para que cada linha representasse a mulher e as repetições dos exames foram ordenadas por cronologia nas colunas (apêndice 9). Após o relacionamento, identificaram-se seis mulheres com idade >110 anos no banco seguimento que foram excluídas. Assim, a coorte do estudo (*Banco Seguimento 2008-2012*) foi composta por 1.162.405 mulheres (figura 4).

Quadro 5 – Pares verdadeiros identificados em cada passo dos relacionamentos entre o banco referência e cada banco anual de 2009 a 2012

	Pares verdadeiros
Passo 2	1.477.950
Passo 9	1.670.224
Passo 12	1.638.346
Passo 15	1.526.687
P_total	3.431.089

O início do seguimento foi a data de coleta do primeiro exame normal, com amostra satisfatória realizado em 2008 e os exames de seguimento foram as repetições posteriores, organizadas em ordem cronológica até a quarta repetição. Para este estudo, será considerada apenas a primeira repetição.

5.2.8 Análise dos dados

5.2.8.1 Cobertura anual do rastreamento do câncer do colo do útero

A cobertura anual do rastreamento em Minas Gerais foi estimada para mulheres na faixa etária de 25 a 59 anos, população alvo recomendada pelo Ministério da Saúde no ano de 2008, e também para mulheres de 25 a 64 anos, conforme atualização das diretrizes do rastreamento, a partir de 2011. O cálculo foi feito pela razão entre o número de mulheres por faixa etária do *Banco 2008* (25-29; 30-39; 40-49; 50-59; 60-64 e total: 25 a 59 e 25 a 64 anos) e um terço da população feminina em 2008 que não tinha plano de saúde, na mesma faixa etária, por micro e macrorregiões de saúde (dados coletados em abril de 2020) - (INCA, 2014b). O quadro 6 descreve o numerador, o denominador, o cálculo do indicador e as respectivas fontes dos dados.

Quadro 6 – Descrição do numerador, denominador e cálculo da cobertura anual do rastreamento do câncer do colo do útero

	Descrição	Fonte dos dados
Numerador		
Mulheres na faixa etária alvo submetidas ao rastreamento em Minas Gerais no ano de 2008	Número de mulheres por faixa etária (25-29; 30-39; 40-49; 50-59; 60-64; 25-59 e 25-64) que realizaram o <i>papanicolaou</i> em Minas Gerais no ano de 2008, por micro e macrorregião de saúde	<i>Banco 2008</i> (figura 2, $n=1.211.014$)
Denominador		
População que depende exclusivamente do SUS (não possui plano privado de saúde)	População feminina por faixa etária (25-29; 30-39; 40-49; 50-59; 60-64; 25-59 e 25-64), em 2008 menos população feminina nesta faixa etária que possuía plano de saúde, em junho de 2008, por micro e macrorregião de saúde de Minas Gerais	IBGE ANS
Indicador		
Cobertura anual do rastreamento do colo do útero em Minas Gerais, por faixa etária, micro e macrorregião de saúde.	Mulheres submetidas ao rastreamento por faixa etária (25-29; 30-39; 40-49; 50-59; 60-64; 25-59 e 25-64) dividido por 1/3 da população feminina que depende exclusivamente do SUS, nesta faixa etária, no ano de 2008, por micro e macrorregião de saúde de Minas Gerais	<i>Banco 2008</i> IBGE ANS

Legenda: IBGE – Instituto Brasileiro de Geografia e Estatística; ANS – Agência Nacional de Saúde Suplementar

5.2.8.2 Sobrerrastreamento em relação à faixa etária

O sobrerrastreamento por faixa etária foi definido como o percentual de mulheres fora da faixa etária preconizada pelo Ministério da Saúde para o rastreamento (<25 e ≥65 anos) do total de mulheres submetidas ao rastreamento em 2008 (*Banco 2008*). E também foi estimado separadamente para mulheres que iniciaram o rastreamento em 2008 e para aquelas que iniciaram previamente.

5.2.8.3 Sobrerrastreamento em relação à periodicidade

Para avaliar o sobrerrastreamento foi utilizado o *Banco Seguimento 2008-2012*, considerando como início do seguimento a data de coleta do primeiro exame com amostra satisfatória e resultado normal realizado em 2008, e a falha (evento de interesse), a data de coleta da repetição do exame antes do período preconizado pelo Ministério da Saúde. Como a recomendação é diferente para mulheres que estão iniciando o rastreamento e mulheres que já têm exame prévio, o *Banco Seguimento 2008-2012* foi dividido, com base na variável *exame anterior*: 1) mulheres que relataram não ter sido submetidas ao citopatológico anteriormente e 2) mulheres que relataram exame anterior. Foram excluídas as mulheres que informaram não saber se já tinham iniciado o rastreamento ou o registro estava sem esta informação.

Para mulheres que iniciaram o rastreamento em 2008, o sobrerrastreamento por periodicidade foi definido como a repetição do exame em um período inferior a nove meses (274 dias). Nesse caso, a recomendação é a repetição do exame com intervalo de 12 meses, mas variações ≤ 3 meses do preconizado foram consideradas pouco relevantes. Aquelas mulheres que repetiram o exame em período maior ou igual a nove meses foram censuradas. Para esse grupo, o sobrerrastreamento foi calculado de acordo com a seguinte fórmula:

$$\text{Sobrerrastreamento} = \frac{\text{Nº mulheres que repetiram o exame} < 9 \text{ meses (274 dias)}}{\text{Total de mulheres que iniciaram o rastreamento em 2008}}$$

E, para as mulheres que iniciaram o rastreamento antes de 2008, a repetição em um período inferior a 30 meses (912 dias) foi considerada como sobrerrastreamento e, aquelas cuja repetição foi igual ou maior a 30 meses, foram censuradas. Para mulheres que já iniciaram previamente o rastreamento, a recomendação é a repetição do exame com intervalo de 36 meses, mas variações ≤ 6 meses foram consideradas pouco relevantes. O cálculo do sobrerrastreamento foi feito utilizando a fórmula a seguir para essas mulheres:

$$\text{Sobrerrastreamento} = \frac{\text{Nº mulheres que repetiram o exame} < 30 \text{ meses (912 dias)}}{\text{Total de mulheres que iniciaram o rastreamento antes de 2008}}$$

O método de *Kaplan Meier* foi utilizado para estimar a probabilidade cumulativa de sobrerrestreamento ao longo do período de seguimento e as curvas foram estratificadas por faixa etária. Para comparar os estratos, foi utilizado o teste de *log-rank*, com nível de significância de 5%. Todas as análises foram realizadas utilizando o programa *R* (versão 4.0.2 – pacote *survival*).

5.3 MÉTODOS REFERENTES AO OBJETIVO ESPECÍFICO 3 (ARTIGO 3)

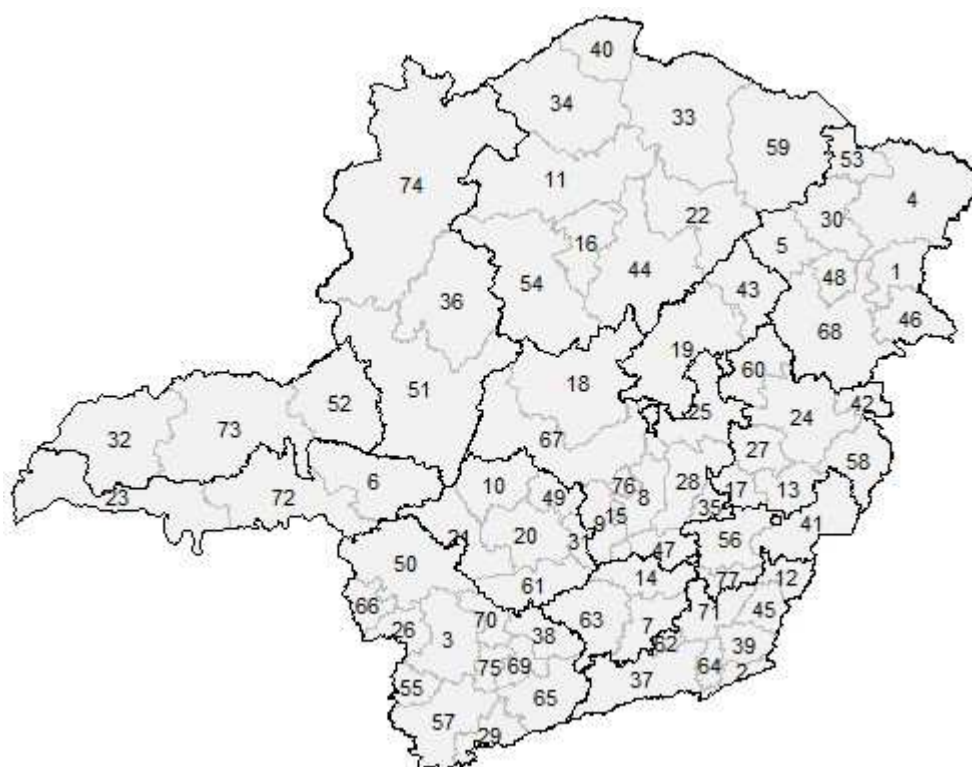
5.3.1 Tipo de estudo e fonte dos dados

Trata-se de um estudo ecológico exploratório que utiliza os indicadores referentes à cobertura do rastreamento do câncer do colo do útero e ao sobrerastreamento em Minas Gerais em 2008, descritos no item 5.2.

5.3.2 População de estudo

As 77 microrregiões de saúde do estado de Minas Gerais constituem a unidade de análise do estudo. Minas Gerais situa-se na região Sudeste do Brasil e é o quarto estado do Brasil em extensão territorial, ocupando uma superfície de 586.522,122 km² (IBGE, 2010). O estado possui 853 municípios, 77 microrregiões e 13 macrorregiões de saúde caracterizadas por grande disparidade socioeconômica, sendo considerado uma representação da estrutura regional brasileira, menos desenvolvida ao norte/nordeste e mais desenvolvida ao sul (DOMINGUES; MAGALHÃES; FARIA, 2009). A figura 5 ilustra as microrregiões de saúde de Minas Gerais e a figura 6, as 13 macrorregiões de saúde de acordo com o Índice de Desenvolvimento Humano (IDH), agrupados em quintis, referente ao ano de 2010.

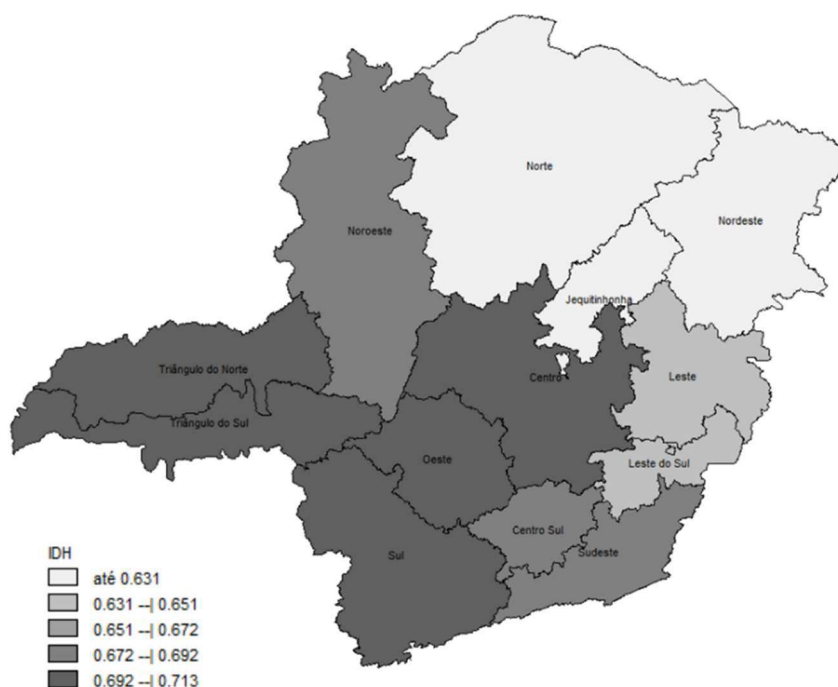
Figura 5 – Microrregiões de saúde de Minas Gerais



Legenda: 1- Águas Formosas; 2- Além Paraíba; 3- Alfenas/Machado; 4- Almenara; 5- Araçuaí; 6- Araxá; 7- Barbacena; 8- B.Horizonte/N.Lima/Caeté; 9-Betim; 10- Bom Despacho; 11-Brasília MG/S Francisco; 12-Carangola; 13- Caratinga; 14- Cons. Lafaiete/Congonhas; 15- Contagem; 16- Coração de Jesus; 17- Cel. Fabriciano/Timóteo; 18- Curvelo; 19- Diamantina; 20- Divinópolis/S Antôn. Monte; 21- Formiga; 22- Francisco Sá; 23- Frutal/Iturama; 24- Governador Valadares; 25- Guanhães; 26- Guaxupé; 27- Ipatinga; 28- Itabira; 29-Itajubá; 30-Itaobim; 31-Itaúna; 32-Ituiutaba; 33- Janaúba/Monte Azul; 34- Januária; 35- João Monlevade; 36- João Pinheiro; 37- J Fora/L Duarte/B Jardim; 38-Lavras; 39- Leopoldina/Cataguases; 40- Manga; 41-Manhuaçu; 42- Mantena; 43- M.Novas/Turmalin/Capelin; 44- Montes Claros/Bocaiúva; 45- Muriaé; 46.Nanuque;47- Ouro Preto; 48- Padre Paraíso; 49- Pará de Minas; 50- Passos/Piumhi; 51-Patos de Minas; 52- Patrocínio/Monte Carmelo; 53-Pedra Azul; 54- Pirapora; 55- Poços de Caldas; 56- Ponte Nova; 57- Pouso Alegre; 58- Resplendor; 59- Salinas/Taiobeiras; 60- S Maria Suaçuí/S Joao Eva; 61- S Antônio Amparo/C Belo; 62- Santos Dumont; 63- São João del Rei; 64- São João Nep./Bicas; 65- São Lourenço; 66- São Sebastião do Paraíso; 67- Sete Lagoas; 68- T Otoni/Malacacheta/Itambacuri; 69-Três Corações; 70- Três Pontas; 71- Ubá; 72. Uberaba; 73- Uberlândia/Araguari; 74- Unai; 75- Varginha; 76- Vespasiano; 77- Viçosa.

Fonte: <http://datasus.saude.gov.br/informacoes-de-saude/tabnet/demograficas-e-socioeconomicas>

Figura 6 – IDH por macrorregião de Saúde em Minas Gerais, 2010



Fonte: <http://datasus.saude.gov.br/informacoes-de-saude/tabnet/demograficas-e-socioeconomicas>

De acordo com dados do Censo, a população em 2010 era de 19.597.330 habitantes, sendo 50,8% (9.955.453) do sexo feminino e 5.189.853 mulheres (52,1%) encontravam-se na faixa etária de 25 a 64 anos (IBGE, 2010). O Índice de Desenvolvimento Humano (IDH) do estado em 2010 foi 0,731, ocupando a nona posição no país. A cobertura pela Estratégia Saúde da Família era 61,7% em 2008; 66,9% em 2010 e , em 2019, foi 79,7%, superior à do Brasil (48,4%; 51,3% e 63,5, respectivamente) (Fundação Oswaldo Cruz, 2010 PROADSS - <https://www.proadess.icict.fiocruz.br/index.php?pag=fic&cod=A26&tab=1>). Em 2008, 24,4% das mulheres de 25 a 64 anos em Minas Gerais tinham plano de saúde e, em 2020, a estimativa é de que 25,3% das mulheres nessa faixa etária sejam cobertas por plano privado de saúde (ANS, 2020 09/10/2020 - http://www.ans.gov.br/anstabnet/cgi-bin/tabnet?dados/tabnet_tx.def).

5.3.3 Variáveis de estudo

As variáveis de desfecho, obtidas no estudo descrito no item 5.2, são:

- cobertura do rastreamento do câncer do colo do útero no estado de Minas Gerais e suas microrregiões de saúde, no ano de 2008;
- sobrerastreamento do câncer do colo do útero quanto à faixa etária e quanto à periodicidade no estado e suas microrregiões de saúde, considerando mulheres submetidas ao rastreamento em 2008.

As variáveis explicativas foram compostas por indicadores sociodemográficos, econômicos e assistenciais baseados na microrregião de residência (coletados em abril e maio de 2020) e separadas em dois blocos hierárquicos para orientar a análise de regressão múltipla posteriormente. Tais variáveis foram selecionadas de acordo com a revisão de literature e a disponibilidade em bases de acesso público visando a um estudo exploratório sobre a associação entre a cobertura do rastreamento e características das microrregiões de saúde de Minas Gerais (figura 7).

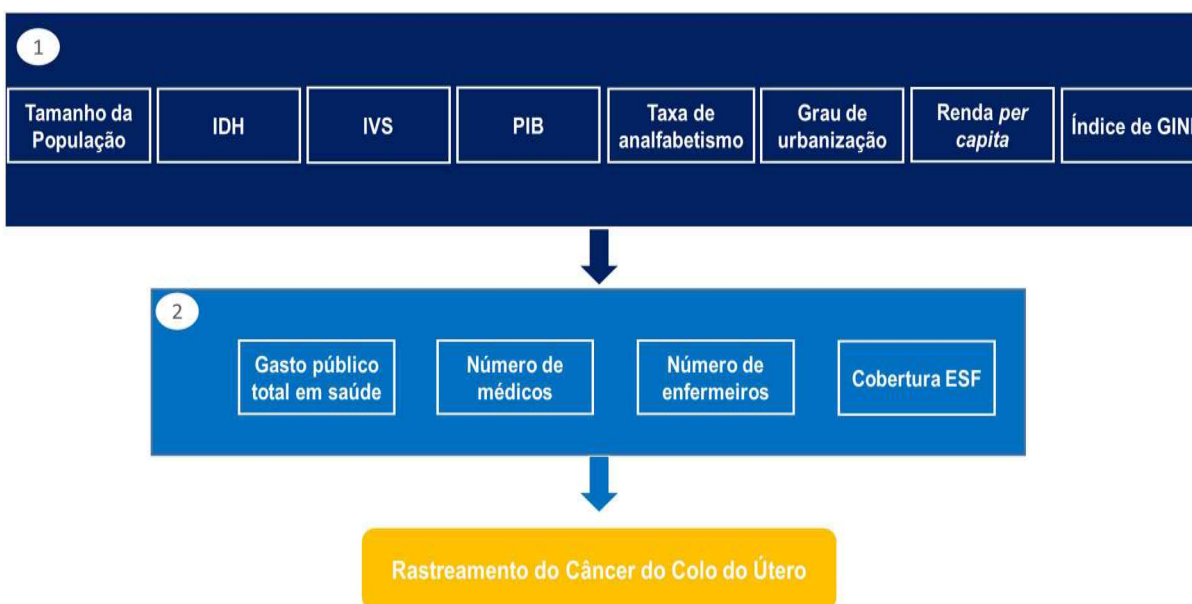


Figura 7: Variáveis explicativas do estudo separadas em dois blocos hierárquicos (distal e proximal)

Legenda: 1) bloco distal: variáveis sociodemográficas e econômicas; 2) bloco proximal: variáveis de financiamento e assistenciais

O quadro 7 descreve as variáveis explicativas e seus respectivos cálculos e fontes.

Quadro 7 – Variáveis explicativas: indicadores sociodemográficos, econômicos e assistenciais, ano de referência e fonte dos dados

Variáveis	Indicador, ano de referência	Fonte Descrição	Cálculo
População	Tamanho da população, 2010	População residente, IBGE - DATASUS http://tabnet.datasus.gov.br/cgi/defohtm.exe?ibge/cnv/popmg.def	Dado bruto, extraído por município
IDH	Índice de desenvolvimento humano, 2010	Proades https://www.proadess.ict.fiocruz.br/index.php?pag=fic_r&cod=L14&tab=1	Média geométrica das dimensões que compõem o IDH: longevidade, educação e renda. Para o cálculo do IDH das microrregiões, foram agregados os IDH municipais e realizada uma média simples.
RENDA	Renda média domiciliar per capita, 2010	Proades https://www.proadess.ict.fiocruz.br/index.php?pag=fic_r&cod=L09&tab=1	Numerador: soma dos rendimentos mensais dos moradores dos domicílios. Denominador: total de moradores residentes nos domicílios.
PIB	Produto interno bruto per capita, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/pibmunbmg	Indicador baixado por município e agregado por meio de média simples para o cálculo do PIB por microrregião.
IVS	Índice de Vulnerabilidade na Saúde, 2010	IBGE - Dados solicitados ao IBGE via portal de acesso à informação do governo federal (2015) Indicador composto pelas seguintes variáveis: <ul style="list-style-type: none"> • Dimensão Saneamento: (1) Percentual de domicílios particulares permanentes com abastecimento de água inadequado ou ausente; (2) Percentual de domicílios particulares permanentes com esgotamento sanitário inadequado ou ausente; (3) Percentual de domicílios particulares permanentes com destino do lixo de forma inadequada ou ausente; • Dimensão Socioeconômica: (1) Razão de moradores por domicílio; (2) Percentual de pessoas analfabetas; (3) Percentual de domicílios particulares com rendimento per capita até ½ SM; (4) Rendimento nominal mensal médio das pessoas responsáveis (invertido); (5) Percentual de pessoas de raça/cor preta, parda e indígena 	Realizado conforme: PITCHON, A.; <i>et al.</i> Minas Gerais. Índice de Vulnerabilidade Social (IVS) 2012. Belo Horizonte, 2013.15 p.
GINI	Índice de Gini, 2010	Atlas do desenvolvimento humano no Brasil http://www.atlasbrasil.org.br/2013/pt/consulta/ Mede o grau de desigualdade existente na distribuição de indivíduos segundo a renda domiciliar per capita. Seu valor é 0 quando não há desigualdade (a renda domiciliar per capita de todos os indivíduos têm o mesmo valor) e tende	Para o cálculo do índice de Gini das microrregiões, foram agregados os índices municipais e realizada média simples.

GRAU URBANIZAÇÃO	Grau de urbanização, 2010	a 1 à medida que a desigualdade aumenta. O universo de indivíduos é limitado àquelas que vivem em domicílios particulares permanentes. DATASUS http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/popmg.def % população urbana	(População urbana/População total) *100
ANALFABETISMO	Percentual da população com ≥ 15 anos analfabeta em 2010	Proades https://www.proadess.iciet.fiocruz.br/index.php?pag=fig_r&cod=L51&tab=1	Numerador: número de pessoas residentes de 15 anos ou mais de idade que não sabem ler e escrever um bilhete simples, no idioma que conhecem X 100. Denominador: população total residente de 15 anos ou mais de idade.
GPTS	Gasto Público Total em Saúde/habitante, 2008	Proades https://www.proadess.iciet.fiocruz.br/index.php?pag=fig_r&cod=K51&tab=1 Gasto em saúde, por habitante, advindo de todas as fontes, quer sejam impostos, transferências SUS de outras esferas governamentais, operações de créditos, e outras.	Numerador: gasto total em saúde (todas as fontes de financiamento). Denominador: população total residente.
MEDICOS	Nº Médicos/ 1000 hab, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def	(Número de médicos/habitantes)*1000
ENFERMEIROS	Nº Enfermeiros/ 1000 hab, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def	(Número de enfermeiros/habitantes)*1000
ESF	Percentual da população coberta pela Estratégia Saúde da Família, 2008	Proades https://www.proadess.iciet.fiocruz.br/index.php?pag=fig_r&cod=A26&tab=1	Numerador: número de pessoas que residem em domicílios cadastrados na Estratégia Saúde da Família X 100. Denominador: população total residente

5.3.4 Análise dos dados

Inicialmente, foram feitas análises espaciais descritivas por meio da construção de mapas temáticos para as variáveis explicativas e de desfecho, com tons mais claros para regiões com os menores valores e mais escuros para os maiores. Para isso, as categorias de cada variável foram baseadas em 0,5 ou 1,0 desvio-padrão abaixo e acima da média, a depender da distribuição e variabilidade de cada indicador.

Os mapas temáticos são úteis na identificação de áreas muito parecidas em relação ao indicador estudado e que possam ter dependência espacial, ou seja, o valor observado da variável em determinada região é dependente do valor dessa mesma variável nas regiões vizinhas. Porém, somente a visualização dos mapas temáticos não permite tal inferência, sendo necessário testar a autocorrelação visando investigar a aleatoriedade ou não da distribuição espacial verificada (LUZARDO; CASTAÑEDA FILHO; RUBIM, 2017).

Para testar a autocorrelação espacial geral foi usado o Índice de Moran Global (I), o qual, varia de -1 a 1 e a hipótese nula representa a independência espacial (ou seja, não há autocorrelação espacial, $I = 0$). Quanto mais próximo de 1, maior a semelhança entre vizinhos. O valor zero indica inexistência de autocorrelação e valores negativos indicam dessemelhança (BRASIL, 2007).

O indicador global de autocorrelação espacial fornece um valor único como medida de associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. Porém, quando se trabalha com um grande número de áreas em uma região, é muito provável que ocorram diferentes regimes de associação espacial. Assim, é desejável examinar os padrões mais detalhadamente. Para isso, a autocorrelação espacial local foi testada por meio do Índice de Moran Local ou LISA, que permite comparar o valor de cada região de saúde com seus vizinhos, objetivando identificar o agrupamento de regiões de saúde com dependência espacial significativa (LUZARDO; CASTAÑEDA FILHO; RUBIM, 2017).

Posteriormente, para verificar a associação entre a cobertura do rastreamento e os indicadores sociodemográficos, econômicos e assistenciais, foram utilizados modelos de regressão linear simples e múltipla. Nos modelos de regressão múltipla, foi seguida abordagem hierarquizada dos blocos de variáveis

explicativas, começando com o bloco distal e, posteriormente, inserindo as variáveis do bloco proximal – figura 7 - (VICTORA *et al.*, 1997). Para a seleção de variáveis do bloco distal, foram feitas combinações entre aquelas que não apresentaram correlação forte, avaliada por meio da correlação de *Pearson*. Em seguida, foram incluídas cada variável proximal nos modelos contendo as variáveis distais e foram testados os pressupostos de independência espacial (Teste de *Moran Global*), normalidade (Teste de *Jarque-Bera*) e homocedasticidade (Teste de *Breusch-Pagan*) dos resíduos. O modelo final foi escolhido com base nos seguintes critérios: significância estatística das variáveis explicativas; maior poder explicativo (maior R^2), menor número de variáveis para explicar o desfecho e atender a todos os pressupostos da regressão linear. O nível de significância adotado no estudo foi de 0,05.

As análises foram realizadas utilizando o programa R, versão 4.0.2 (<https://www.rproject.org/>).

6 RESULTADOS E DISCUSSÃO

Os resultados e discussão serão apresentados sob a forma de três artigos.

6.1 – ARTIGO 1: RASTREAMENTO DO CÂNCER DO COLO DO ÚTERO EM MINAS GERAIS: AVALIAÇÃO A PARTIR DE DADOS DO SISTEMA DE INFORMAÇÃO DO CÂNCER DO COLO DO ÚTERO (SISCOLO)

Artigo publicado na Revista Cadernos Saúde Coletiva (CORRÊA *et al.*, 2017 - <http://www.scielo.br/pdf/cadsc/v25n3/1414-462X-cadsc-25-3-315.pdf>)

Resumo

Objetivo: Avaliar os indicadores relacionados à oferta de exames citopatológicos do colo do útero, sua qualidade, frequência de alterações celulares e seguimento informado de lesões de alto grau, em Minas Gerais e suas macrorregiões de saúde.

Métodos: Estudo descritivo com base em dados do Sistema de Informação do Câncer do Colo do Útero (SISCOLO), de 2006 a 2011.

Resultados: A razão de exames citopatológicos do colo do útero em mulheres de 25 a 59 anos manteve-se estável, porém sem alcançar a meta estadual pactuada. Aproximadamente 75% dos exames foram realizados na população-alvo e houve progressiva redução na proporção de citopatológicos do colo do útero sem citologia anterior. Em média, 51,2% dos exames foram realizados no período de até um ano. Quanto à qualidade do exame, destaca-se o índice de positividade, categorizado como baixo durante todo o período analisado. Observou-se baixo percentual de seguimento informado no estado.

Conclusão: Os resultados evidenciam a necessidade de aprimoramento do programa de rastreamento em relação à oferta e qualidade do exame, bem como ao direcionamento correto quanto à faixa etária e à periodicidade, visando garantir o acesso da população-alvo e o encaminhamento para a investigação diagnóstica e tratamento das lesões precursoras, quando indicado.

Palavras-chave: Neoplasias do colo do útero; Programas de rastreamento; Saúde da Mulher; Sistemas de Informação.

Abstract

Objective: To evaluate the indicators related to cervical cytology, quality, frequency of cellular changes and informed follow-up of high-grade lesions in Minas Gerais and its macro-regions of health.

Methods: a descriptive study based on data from the Cervical Cancer Information System (SISCOLO) from 2006 to 2011.

Results: The ratio of cervical cytopathologic exams in women aged 25 to 59 years remained stable, but did not reach the goal set by the state. Approximately 75% of the exams were performed in the target population and there was a progressive reduction in the proportion of cervical cytopathologists without previous cytology. On average, 51.2% of the exams were performed within a year. Regarding the quality of the exam, the positivity index was categorized as low throughout the analyzed period. We observed a low percentage of follow-up reported in the state.

Conclusion: The results evidenced the need to improve the tracking program regarding the quality of the exam, as well as the correct orientation regarding the age group and the periodicity, aiming to guarantee the target population's access and referral for diagnostic investigation and treatment of precursor lesions, when indicated.

Keywords: Uterine cervical neoplasms; Mass Screening; Women's Health; Information Systems.

INTRODUÇÃO

O câncer do colo do útero (CCU) é o quarto tipo de câncer mais comum em mulheres no mundo (IARC, 2012). Em 2012, as estimativas apontaram 528 mil casos novos, dos quais cerca de 80% ocorreram em regiões menos desenvolvidas (WHO, 2014). Excluindo o câncer de pele não melanoma, é o terceiro tipo de câncer mais frequente nas mulheres brasileiras. Na região Norte, é o mais incidente, e, nas regiões Centro-Oeste e Nordeste, ele ocupa a segunda

posição (INCA, 2015). Quanto à mortalidade, apesar da queda observada no país nos últimos anos, ela não ocorre em todas as regiões, como no interior do Norte e Nordeste (GIRIANELLI; GAMARRA; AZEVEDO E SILVA, 2014).

O rastreamento realizado por meio do exame citopatológico do colo do útero é considerado a principal estratégia para a detecção precoce do CCU, sendo reconhecido mundialmente como seguro e eficiente. O objetivo fundamental é detectar e tratar precocemente as lesões precursoras antes da sua evolução para a doença invasiva (INCA, 2016a). Nesse sentido, com uma cobertura da população-alvo de, no mínimo, 80% e acesso à confirmação diagnóstica e ao tratamento adequado, é possível reduzir em até 90% a incidência do câncer cervical invasivo (WHO, 2002).

O Ministério da Saúde recomenda o rastreamento citopatológico em mulheres de 25 a 64 anos, a cada três anos após dois exames com resultados negativos realizados anualmente (INCA, 2016). Ao contrário de países desenvolvidos, onde há recrutamento ativo da população-alvo para o rastreamento, no Brasil o rastreamento ainda é oportunístico, com procura ocasional e espontânea dos serviços de saúde determinada por razões diversas que não o rastreamento do CCU (VALE *et al.*, 2010). O rastreamento oportunístico, além de ser menos efetivo no impacto sobre a mortalidade, é mais oneroso para o sistema de saúde (BRASIL, 2010).

O controle da doença é assumido como uma prioridade nacional (INCA, 2016). Entre as principais ações nesse sentido, destacam-se a instituição pelo Ministério da Saúde, em 1998, do Programa Nacional de Combate ao Câncer do Colo do Útero e a criação do Sistema de Informação do Câncer do Colo do Útero (SISCOLO). Esse sistema foi implantado oficialmente no país pela Portaria/SAS/MS nº 408, de 30 de julho de 1999, e visa viabilizar o gerenciamento das ações de rastreamento do CCU. Posteriormente, foi aperfeiçoado, e, em 2006, foi implantada uma nova versão (4.0) pela Portaria/SAS/MS nº 287, de 24 de abril de 2006. Em 2005, foi instituída a Política Nacional de Atenção Oncológica, e, em 2006, a importância da detecção precoce do CCU foi reafirmada no Pacto pela Saúde, por meio da inclusão de indicadores específicos na pactuação de metas com Estados e municípios (INCA, 2016).

Minas Gerais se caracteriza pela grande extensão e diversidade regional, sendo considerada uma representação da estrutura regional brasileira, menos

desenvolvida ao norte/nordeste e mais desenvolvida ao sul (DOMINGUES; MAGALHÃES; FARIA, 2009). Assim, torna-se relevante avaliar os indicadores relacionados ao rastreamento do CCU, com foco estadual, que permita monitorar indicadores pactuados e identificar diferenças regionais, podendo contribuir para o planejamento de ações mais resolutivas.

O presente estudo visa avaliar as ações de rastreamento com base em dados do SISCOLO em Minas Gerais e suas macrorregiões de saúde a partir da análise de indicadores relacionados às características da oferta de exames citopatológicos, à qualidade do exame, à ocorrência de alterações celulares nos exames e ao seguimento informado dos casos de lesão de alto grau.

MÉTODOS

Trata-se de um estudo descritivo com base em dados secundários do SISCOLO que analisa os indicadores, em Minas Gerais e suas macrorregiões de saúde, no período de 2006 a 2011.

Os dados foram extraídos do sítio eletrônico do Departamento de Informática do Sistema Único de Saúde (DATASUS) nos meses de abril e maio de 2016.

A oferta do exame foi avaliada pelo indicador razão de exames citopatológicos em mulheres de 25 a 59 anos, utilizado no Pacto pela Saúde (BRASIL, 2009) cujas metas estão disponíveis, a partir de 2007, no Sistema de Informação do Pacto pela Saúde (SISPACTO). Apesar de a recomendação do INCA quanto à faixa etária da população-alvo de rastreamento do CCU ter sido atualizada no ano de 2011 para 25 a 64 anos, optou-se por analisar a faixa etária de 25 a 59 anos, seguindo a recomendação vigente naquele período (INCA, 2006).

A oferta de exames foi caracterizada pela faixa etária e periodicidade da realização. Por meio desses indicadores, é possível avaliar se a realização dos exames quanto à idade e frequência de repetição está de acordo com as recomendações do Ministério da Saúde (BRASIL, 2014b).

A qualidade do exame foi analisada pelos indicadores: percentual de amostras insatisfatórias nos exames citopatológicos do colo do útero, percentual de exames com representatividade da Zona de Transformação (ZT) e o índice

de positividade. A análise da representatividade da ZT como indicador da qualidade da coleta deve-se ao fato de que é principalmente nessa área que surgem as lesões precursoras do CCU. A análise foi feita por faixa etária (<50 anos e ≥50 anos) a fim de diferenciar mulheres mais jovens daquelas na peri ou pós-menopausa, uma vez que estas últimas possuem maior probabilidade de apresentar mucosa vaginal atrófica, interferindo na coleta de uma amostra com representatividade da ZT (INCA, 2012a). O índice de positividade expressa a prevalência de alterações celulares nos exames e a sensibilidade do processo de rastreamento em detectar lesões na população examinada. Esse indicador pode ser categorizado em: muito baixo (<2%), baixo (2 a 2,9%), esperado (3 a 10%) e acima do esperado (>10%) – (INCA, 2012b).

As alterações celulares no exame citopatológico do colo do útero foram agrupadas nas categorias: atipias de significado indeterminado; lesões de baixo grau; lesões de alto grau; carcinoma escamoso; adenocarcinoma e outras neoplasias. A fim de verificar o percentual de mulheres com lesão intraepitelial de alto grau (HSIL) que tiveram seu seguimento informado no SISCOLO, utilizou-se o relatório “consolidado de dados” disponível no SISCOLO.

A descrição dos indicadores calculados e as respectivas fontes dos dados encontram-se no Quadro 1.

Quadro 1 – Descrição dos indicadores calculados e respectivas fontes dos dados

Indicador	Cálculo	Fonte dos dados
Razão de exames citopatológicos do colo do útero em mulheres de 25 a 59	Número total de exames citopatológico em mulheres de 25 a 59 anos/população feminina nesta faixa etária	SISCOLO ^a , IBGE ^b
Proporção de exames citopatológicos por faixa etária (<25 anos; 25-59 anos; ≥60 anos)	Número de exames citopatológicos em cada faixa etária (<25 anos; 25-59 anos; ≥60 anos) x 100/ (número total de exames em todas as faixas etárias)	SISCOLO ^{a,b}
Proporção de exames citopatológicos sem citologia anterior	Número de exames citopatológicos sem relato de exame anterior, em mulheres de 25 a 59 anos x100/(número total de exames nesta faixa etária – [número de exames nesta faixa etária nos quais a mulher não soube informar + número de exames nesta faixa etária sem esta informação])	SISCOLO ^{a,b}

Proporção de exames citopatológicos sem atraso no rastreamento	Número de exames citopatológicos do colo do útero, em mulheres de 25 a 59 anos, com relato de exame anterior há três anos (≤ 1 ano, 2 anos e 3 anos) $\times 100 /$ total de exames com relato de citopatologia anterior nesta faixa etária – número de exames nesta faixa etária com esta informação ignorada	SISCOLO ^{a,b}
Proporção de exames com citopatologia anterior há até um ano	Número de exames citopatológicos do colo do útero em mulheres de 25 a 59 anos que relataram citopatologia anterior em até um ano (≤ 1 ano) $\times 100 /$ total de exames com relato de citopatologia anterior nesta faixa etária – número de exames com esta informação ignorada	SISCOLO ^{a,b}
Proporção de amostras insatisfatórias em exames citopatológicos	Número de exames citopatológicos insatisfatórios $\times 100 /$ número total de exames	SISCOLO ^{a,b}
Proporção de exames com representatividade da Zona de Transformação entre exames citopatológicos com amostra satisfatória, por grupo etário (< 50 anos e ≥ 50 anos)	Número de exames citopatológicos com representatividade da ZT $\times 100 /$ número total de exames com amostra satisfatória	SISCOLO ^{a,b}
Índice de positividade de exames citopatológicos do colo do útero	Número de exames citopatológicos com resultados alterados $\times 100 /$ número de exames citopatológicos com amostra satisfatória.	SISCOLO ^{a,b}
Proporção dos tipos de alterações* por grupo etário (< 25 anos, 25 a 59 anos e ≥ 60 anos)	Número de exames com determinada alteração (atípicas, LSIL, HSIL, carcinoma escamoso, adenocarcinoma e outros tipos) $\times 100 /$ total de exames alterados	SISCOLO ^{a,b}

a) numerador b) denominador

Legenda: SISCOLO: Sistema de Informação do Controle do Câncer do Colo do Útero; IBGE: Fundação Instituto Brasileiro de Geografia e Estatística

Nota: *Atípicas de significado indeterminado: escamosa - possivelmente não-neoplásica; escamosa - não se pode afastar lesão intraepitelial de alto grau; glandular - possivelmente não-neoplásica; glandular - não se pode afastar lesão intraepitelial de alto grau; origem indefinida - possivelmente não-neoplásicas e origem indefinida - não se pode afastar lesão intraepitelial de alto grau); lesões de baixo grau (lesão intraepitelial de baixo grau – NIC I); lesões de alto grau (lesão intraepitelial de alto grau – NIC II e NIC III, lesão intraepitelial de alto grau não podendo afastar micro-invasão); carcinoma escamoso (carcinoma epidermóide invasor); adenocarcinoma (adenocarcinoma *in situ* adenocarcinoma invasor) e outras neoplasias.

Os dados brutos foram analisados de forma descritiva no programa Microsoft Excel 2007, por meio do cálculo dos indicadores expressos em razões e proporções.

Este artigo é parte de um projeto de pesquisa mais amplo, que objetivou analisar os resultados das ações previstas na Rede Viva Vida referentes ao rastreamento do CCU e de mama no período de 2006 a 2011. Foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora sob o Parecer nº 1.376.660/2015 e publicado na Revista Cadernos Saúde Coletiva em 2017.

RESULTADOS

Em 2011, a população de Minas Gerais era de 19.728.701 habitantes, dos quais 50,8% eram do sexo feminino, sendo 24,6% (4.846.569) mulheres de 25 a 59 anos. De 2006 a 2011, foram realizados 5.813.294 exames citopatológicos do colo do útero em mulheres nessa faixa etária no Estado.

No período analisado, observa-se que a razão de exames citopatológicos do colo do útero em mulheres de 25 a 59 anos manteve-se estável em Minas Gerais, porém sem alcançar a meta pactuada. Em 2008 e 2009, a meta foi de 0,30, igual ao parâmetro nacional, não tendo sido alcançada em nenhuma região. Norte, Nordeste e Jequitinhonha se destacaram superando a meta nos anos 2007, 2010 e 2011. No entanto, a Noroeste se destaca com o pior desempenho de 2007 a 2011 (Tabela 1).

Tabela 1. Razão de exames citopatológicos em mulheres de 25 a 59 anos nas macrorregiões de saúde de Minas Gerais e número de exames e população feminina na faixa etária alvo do rastreamento em Minas Gerais, 2006-2011

Macrorregiões	2006	2007	2008	2009	2010	2011
Centro	0,17	0,18	0,17	0,15	0,17	0,17
Centro-Sul	0,22	0,19	0,19	0,19	0,20	0,21
Jequitinhonha	0,25	0,26	0,23	0,24	0,28	0,28
Leste	0,19	0,21	0,20	0,21	0,23	0,22
Leste do Sul	0,22	0,21	0,21	0,21	0,22	0,26
Nordeste	0,25	0,25	0,22	0,24	0,28	0,26
Noroeste	0,23	0,11	0,10	0,10	0,10	0,11
Norte	0,26	0,28	0,26	0,27	0,25	0,28
Oeste	0,21	0,21	0,22	0,22	0,22	0,24
Sudeste	0,23	0,21	0,21	0,21	0,20	0,20
Sul	0,22	0,22	0,23	0,24	0,24	0,25
Triângulo do Norte	0,21	0,21	0,22	0,22	0,23	0,24
Triângulo do Sul	0,22	0,24	0,21	0,20	0,22	0,21
Minas Gerais						
Exames	868.279	970.519	951.677	966.032	1.009.463	1.047.324
População	4.247.165	4.798.458	4.886.044	4.981.484	5.002.386	5.035.579
Razão	0,20	0,21	0,20	0,20	0,21	0,22
Meta - SISPACTO	--	0,23	0,30	0,30	0,22	0,25

Fonte: Sistema de Informação do Câncer do Colo do Útero (SISCOLO) -

<http://datasus.saude.gov.br/>

A maior parte dos exames (75,2%) foi realizada em mulheres de 25 a 59 anos, população-alvo do programa de rastreamento do CCU até 2011, porém observou-se um número considerável de exames (16%) em mulheres com menos de 25 anos. De acordo com a Tabela 2, em Minas Gerais, houve progressiva redução na proporção de exames citopatológicos do colo do útero sem citologia anterior entre 2006 e 2011. Entre as macrorregiões, a Nordeste apresentou o maior percentual em 2006, e a Noroeste, de 2007 a 2011. Contudo, Centro-Sul e Triângulo do Norte tiveram os menores valores em 2006 e 2007, e a macrorregião Sul apresentou os menores nos de 2008 à 2011.

Tabela 2. Proporção de exames citopatológicos do colo do útero sem citologia anterior em mulheres de 25 a 59 anos, por ano, em Minas Gerais e suas macrorregiões de saúde. Minas Gerais, 2006-2011

Macrorregiões	2006	2007	2008	2009	2010	2011
Centro	10,0	10,0	6,6	5,1	4,0	3,3
Centro-Sul	8,0	7,6	6,0	5,4	5,0	4,3
Jequitinhonha	14,0	10,5	7,8	7,6	7,7	7,1
Leste	11,6	9,6	8,2	6,6	6,7	5,7
Leste do Sul	13,8	10,2	8,9	8,9	8,5	7,4
Nordeste	20,7	12,5	10,1	7,8	6,9	6,5
Noroeste	19,9	18,7	41,8	34,2	9,6	17,7
Norte	16,4	14,7	13,8	5,7	4,8	4,3
Oeste	12,2	6,3	5,0	4,7	4,1	3,6
Sudeste	12,6	6,1	8,4	7,0	4,9	4,5
Sul	10,3	5,2	4,3	2,7	3,0	3,3
Triângulo do Norte	8,0	4,3	5,0	5,5	5,5	4,5
Triângulo do Sul	9,7	6,4	6,0	7,0	7,1	5,1
MINAS GERAIS						
N	77222	77866	68090	52774	43553	42346
%	12,0	9,1	8,4	6,4	5,0	4,7

Fonte: Sistema de Informação do Câncer do Colo do Útero (SISCOLO) -

<http://datasus.saude.gov.br/>

A maioria dos exames foi realizada dentro da periodicidade recomendada (≤ 3 anos), com percentuais acima de 79% e pouca diferença entre as macrorregiões. Em Minas Gerais, no período de análise, cerca da metade dos exames anteriores em mulheres de 25 a 59 anos foi realizada no período de até um ano: 48,8%; 51,3%, 52,3%; 50,9%; 50,7% e 53%, em 2006, 2007, 2008, 2009, 2010 e 2011, respectivamente. A macrorregião Triângulo do Sul apresentou os maiores percentuais em 2006 (55,1%), 2008 (57,2%) e 2009 (55,6%); Sudeste, em 2007 (55,5%); Triângulo do Norte, em 2010 (57,8%); e Sul, em 2011 (60,3%).

O percentual de amostras insatisfatórias em Minas Gerais ficou praticamente estável no período, em torno de 0,7%. Os piores resultados foram observados na macrorregião Nordeste, que apresentou um considerável aumento no ano de 2009 (2,6%). No mesmo ano, Jequitinhonha também se destacou com 2,5% de amostras insatisfatórias.

No que se refere à representatividade da ZT em exames citopatológicos do colo do útero, em 2006, Minas Gerais obteve o pior resultado: 46,9% para a faixa etária inferior a 50 anos e 37,2% para mulheres com 50 anos ou mais. Em 2007 e 2008, apresentou os melhores desempenhos do período: 72,9% para a faixa etária inferior a 50 anos e 55,2% para mulheres com 50 anos ou mais e 71,1% para a faixa etária inferior a 50 anos e 52,8% para mulheres com 50 anos ou mais, respectivamente. Quanto às macrorregiões, a Norte foi a que apresentou o maior percentual de amostras com representatividade ZT nas duas faixas etárias analisadas, com percentuais que variaram de 56,2%, em 2006, para a faixa etária igual ou superior a 50 anos, a 93,7%, em 2007, para a faixa etária inferior a 50 anos. Triângulo do Norte e Leste do Sul apresentaram os piores percentuais em mulheres com menos de 50 anos (Triângulo do Norte no período 2006-2009: 28,1%; 44,6%; 43,5% e 42,8%, respectivamente; e Leste do Sul no período 2010-2011: 43,3% e 39,9%, respectivamente), e com 50 anos ou mais (Triângulo do Norte nos anos 2006 e 2011: 21,9% e 31,7%, respectivamente; e Leste do Sul no período 2007-2010: 33,6%; 33,6%; 29% e 28,5%, respectivamente). Em 2006, destacou-se a macrorregião Triângulo do Norte, que apresentou representatividade da ZT em apenas 28,1 e 21,9% dos exames realizados em mulheres com menos de 50 anos ou 50 anos e mais, respectivamente.

Na Tabela 3, observa-se que o índice de positividade no Estado durante todo o período de análise foi baixo (<3%). Triângulo do Sul foi a única macrorregião que apresentou categorização do índice de positividade dentro do esperado (3-10%), de 2006 a 2011. Jequitinhonha e Leste do Sul, no entanto, apresentaram percentuais abaixo do esperado em todos os anos. A macrorregião Centro destacou-se com uma gradativa melhora no período.

Tabela 3. Índice de positividade de exames citopatológicos do colo do útero, segundo ano, em Minas Gerais e suas macrorregiões de saúde. Minas Gerais, 2006 a 2011

Macrorregiões	2006	2007	2008	2009	2010	2011
Centro	1,9	2,7	2,2	2,3	2,2	3,1
Centro-Sul	1,5	2,2	1,7	1,5	1,3	1,4
Jequitinhonha	1,2	1,3	1,0	0,8	1,2	1,2
Leste	3,3	2,8	2,0	2,0	2,6	3,6
Leste do Sul	1,5	2,0	1,7	1,5	1,5	1,5
Nordeste	2,0	2,6	2,1	1,6	1,4	2,8
Noroeste	1,9	2,7	2,0	1,6	1,6	2,3
Norte	1,8	2,2	1,6	1,4	1,6	2,6
Oeste	1,1	1,7	1,3	1,3	1,7	2,2
Sudeste	2,2	4,8	2,8	2,5	2,8	3,5
Sul	3,0	3,4	2,5	2,4	2,8	4,0
Triângulo do Norte	1,7	3,0	2,0	1,5	1,6	2,5
Triângulo do Sul	3,9	5,8	4,4	5,2	6,0	7,6
MINAS GERAIS	2,2	2,2	2,2	2,1	2,3	2,3

Fonte: Sistema de Informação do Câncer do Colo do Útero (SISCOLO) -

<http://datasus.saude.gov.br/>

As atipias foram os principais tipos de alterações em Minas Gerais no período de 2006 a 2011, nas faixas etárias de 25 a 59 anos e ≥ 60 anos, representando 54,2% e 60,3% das alterações, respectivamente. Observou-se que o percentual de atipias cresceu com o aumento da faixa etária: em 2006, representou 38% das alterações em mulheres com menos de 25 anos, 50,5% das alterações em mulheres de 25 a 59 anos e 56,7% na faixa etária igual ou maior a 60 anos. Em 2011, representou 48,7%; 56,2% e 59,4%, respectivamente. Nas mulheres com menos de 25 anos, as lesões intraepiteliais de baixo grau LSIL corresponderam a 51,4% das alterações nessa faixa etária no período estudado, sendo as mais frequentes, exceto no ano de 2011, no qual as atipias foram o tipo de lesão mais frequente (atipias: 48,7%; LSIL: 48%).

As HSIL foram mais frequentes em mulheres com 60 anos ou mais, com uma média de 19,3% das alterações. Carcinoma escamoso, adenocarcinoma e outros tipos de neoplasias também foram mais frequentes nessa faixa etária, representando, respectivamente, 4,6, 1,6 e 0,4% das alterações. Além disso, verificou-se, nessa faixa etária, redução no percentual de carcinoma escamoso ao final do período analisado – 6,4% em 2006 e 2,9% em 2011.

Como pode ser observado na Tabela 4, em Minas Gerais, o percentual de seguimento informado de mulheres com lesões intraepiteliais de alto grau (HSIL) ficou em torno de 55%. A macrorregião Triângulo do Norte se destacou em 2006, com apenas 9,4% de seguimento informado, e, de 2007 a 2011, a Sudeste apresentou os piores resultados para esse indicador.

Tabela 4. Proporção de seguimento informado de mulheres com lesão intraepitelial de alto grau do colo do útero, segundo o ano de diagnóstico, em Minas Gerais e suas macrorregiões de saúde. Minas Gerais, 2006-2011

Macrorregiões	2006	2007	2008	2009	2010	2011
Centro	55,9	62,2	71,3	79,7	83,0	73,6
Centro-Sul	63,5	82,3	95,6	90,8	82,1	49,4
Jequitinhonha	66,7	84,6	62,5	69,2	57,9	38,9
Leste	47,6	36,8	57,3	83,8	63,8	49,3
Leste do Sul	70,6	80,0	94,7	85,7	88,1	91,7
Nordeste	43,4	44,2	59,1	60,0	60,6	38,1
Noroeste	40,0	53,5	46,0	47,5	46,8	48,1
Norte	26,2	30,4	53,2	44,7	40,5	31,8
Oeste	73,4	78,2	59,8	69,0	53,3	35,2
Sudeste	29,5	20,6	22,0	28,4	27,9	18,1
Sul	76,5	71,8	88,2	84,6	83,3	80,6
Triângulo do Norte	9,3	23,2	74,8	72,2	61,3	56,1
Triângulo do Sul	37,0	29,6	43,0	40,6	45,7	53,2
MINAS GERAIS	48,7	47,6	57,5	62,4	59,4	52,7

Fonte: Sistema de Informação do Câncer do Colo do Útero (SISCOLO) -

<http://datasus.saude.gov.br/>

DISCUSSÃO

Em Minas Gerais, no período de 2006 a 2011, observou-se estabilidade na razão de exames citopatológicos em mulheres de 25 a 59 anos, não tendo atingido a meta pactuada, o que sugere baixo acesso da população-alvo ao programa de rastreamento do CCU. A maior parte dos exames foi realizada na população-alvo, e houve progressiva redução na proporção de citopatológicos do colo do útero sem citologia anterior. A maioria dos exames ocorreu dentro da periodicidade recomendada, dos quais cerca da metade no período de até um

ano. Quanto à qualidade do exame, destaca-se que o índice de positividade foi categorizado como baixo ao longo de todo o período analisado. Além disso, observou-se baixo percentual de seguimento informado no Estado.

Minas Gerais se caracteriza pela grande disparidade socioeconômica entre suas macrorregiões de saúde, destacando-se dois blocos quanto ao valor do Produto Interno Bruto (PIB) *per capita* referente a 2011: por um lado, as macrorregiões Nordeste, Jequitinhonha, Norte e Leste do Sul com os menores valores e, por outro lado, Triângulo do Norte, Centro, Triângulo do Sul e Sul com os maiores valores.

Jequitinhonha, Nordeste e Norte, de menor nível socioeconômico, destacaram-se com os altos valores de razão de exames na população estudada, ultrapassando a meta estadual em 2007, 2010 e 2011. A razão de exames em mulheres de 25 a 59 anos é considerada um indicador de desempenho da Atenção Primária à Saúde (APS) – (INCA, 2012). Fontanive e colaboradores (2008) verificaram que municípios com maior cobertura da Estratégia Saúde da Família (ESF) no Rio Grande do Sul tiveram maior razão de citopatológicos na população-alvo. Dessa forma, representa um importante preditor de realização de exames citopatológicos, o que pode ser atribuído ao vínculo profissional-paciente fortalecido nesse nível devido à longitudinalidade da assistência. Além disso, a territorialização e o cadastramento das famílias pela ESF favorece a identificação de mulheres que nunca fizeram o exame citológico ou que estão em atraso (VALE *et al.*, 2010). Dentro desse contexto, Tomasi e colaboradores (2015) sugerem que a adoção da ESF como modelo de atenção básica parece contribuir para a melhoria da estrutura dos serviços de saúde e, conseqüentemente, para a prestação de serviços voltados à prevenção do câncer de colo uterino.

Estudo ecológico com base em dados secundários de 2008 a 2012 em Minas Gerais verificou maior cobertura de ESF em microrregiões de saúde do Estado com alta vulnerabilidade em saúde e baixo nível socioeconômico (Norte, Noroeste, Nordeste, Jequitinhonha e Leste) – (DUARTE, 2016). Assim, as altas razões de exames na população-alvo nas regiões Jequitinhonha, Nordeste e Norte observadas no presente estudo podem estar relacionadas à maior cobertura pela ESF. Contudo, a região Noroeste, apesar de possuir alta cobertura de ESF, apresentou resultados desfavoráveis, o que pode estar

relacionado à rotatividade de profissionais nessas equipes, interferindo negativamente no vínculo construído com as famílias, à dificuldade de acesso das usuárias ao SUS para a coleta do exame citopatológico ou às barreiras das próprias mulheres em relação ao exame de Papanicolaou (DUARTE, 2016).

Em relação a esse indicador, deve-se ressaltar que a razão de exames citopatológicos não mensura diretamente a cobertura, já que avalia a oferta de Papanicolau com base no número de exames, e não de mulheres examinadas. Além disso, pode estar subestimado em regiões com alta cobertura pela saúde suplementar, pois considera apenas os exames realizados pelo SUS (SANTOS; MELO; SANTOS, 2012).

Em outro estudo, também realizado utilizando dados do SISCOLO para Minas Gerais, referentes ao período 2000-2010, verificou-se melhora do percentual da meta alcançada do rastreamento do CCU em mulheres de 25 a 59 anos em 2010 quando comparada ao ano de 2000 (63,8% em 2000 e 89,3% em 2010), observando os melhores resultados nas macrorregiões Nordeste, Jequitinhonha e Sul (NASCIMENTO *et al.*, 2015).

A maioria dos exames em Minas Gerais foi realizada em mulheres na faixa etária preconizada pelo Ministério da Saúde, porém observou-se percentual considerável de exames em mulheres menores de 25 anos. Tal resultado confirma os achados de um estudo ecológico que também utilizou dados do SISCOLO para o Brasil e regiões no período de 2006 a 2009 (SANTOS; MELO; SANTOS, 2012). No presente estudo, de 2006 a 2011, observou-se uma média anual de 206.430 exames em mulheres com menos de 25 anos. Se esses exames tivessem sido realizados na faixa etária preconizada, a razão de exames seria: 0,25 em 2006, 2007 e 2011 e 0,24 em 2008, 2009 e 2010, o que permitiria o alcance da meta estadual nos anos 2007, 2010 e 2011.

Em relação à periodicidade de realização, assim como observado para o Brasil, a maioria dos exames foi repetida em até um ano, sugerindo o não cumprimento das recomendações nacionais (INCA, 2012). De acordo com o INCA (2016), apenas as mulheres que iniciaram o rastreamento, as que tiveram exames com amostra insatisfatória e as que apresentaram alterações com necessidade de controle em intervalo menor deveriam repetir o exame no período de um ano. Nesse contexto, o percentual de exames com periodicidade anual talvez seja superior às recomendações nesse prazo (DIAS; TOMAZELLI;

ASSIS, 2010). Dessa forma, a cobertura da população-alvo é provavelmente menor que o ideal, com as mesmas mulheres repetindo anualmente o citopatológico, enquanto outras continuam sem acesso (INCA, 2014).

Nessa perspectiva, considera-se que exames em excesso correspondem a todos aqueles realizados fora do grupo etário alvo ou àqueles com intervalo desde o último controle menor do que três anos. Estudo realizado no município de Amparo/SP encontrou percentuais de exames em excesso que variaram entre 61,2% em 2002 e 65,5% em 2005, o que correspondeu a quase dois terços do total dos exames realizados (VALE *et al.*, 2010). Quanto à faixa etária, cabe ressaltar que abaixo dos 25 anos predominam as LSIL, que regredem espontaneamente na maioria dos casos, e que a repetição anual do exame não aumenta significativamente o efeito protetor do rastreamento. A realização de exames fora da faixa etária alvo e da periodicidade recomendada, além de sobrecarregar os serviços, pode comprometer o acesso das mulheres que realmente precisam ser rastreadas e encaminhadas para a investigação diagnóstica e tratamento das lesões precursoras (INCA, 2014).

Observou-se diminuição progressiva de exames sem citologia anterior em Minas Gerais ao longo do período estudado. De acordo com o INCA (2014), em geral, há um decréscimo no percentual de exames em mulheres que nunca o realizaram, o que pode indicar que o programa aumentou a cobertura ao longo do tempo e, com isso, diminuiu a captação de mulheres sem citologia anterior (SANTOS; MELO; SANTOS, 2012). Porém, tal indicador deve ser visto com cautela, pois é esperado que, a cada ano, novas mulheres iniciem o rastreamento ao entrar na faixa etária alvo. Como no presente estudo observaram-se estabilidade da oferta de exames e alta proporção de exames com repetição anual, esses resultados podem refletir oferta inadequada de exames e menor capacidade para captação de mulheres que nunca realizaram o exame (DAMACENA; LUZ; MATTOS, 2017).

Em relação à qualidade do exame, o limite para o percentual de exames insatisfatórios é de 5%, mas a meta deve tender a zero para garantir a efetividade da ação do rastreamento (INCA, 2012). A análise desse indicador para Minas Gerais mostrou relativa estabilidade no período, com resultados satisfatórios, sempre abaixo de 1%. Nordeste e Jequitinhonha se destacaram com os piores percentuais, especialmente em 2009. Para o Brasil, estudo realizado com dados

referentes ao período de 2002 a 2006 verificou que o percentual de amostras insatisfatórias se manteve em torno de 1% (DIAS; TOMAZELLI; ASSIS, 2010).

Em Minas Gerais, o indicador representatividade da ZT apresentou o pior resultado em 2006, com melhora nos anos seguintes. Triângulo do Norte e Leste do Sul chamaram a atenção com resultados insatisfatórios, o que pode levar ao subdiagnóstico de alterações. Nesse sentido, a baixa qualidade observada pelos indicadores de qualidade do exame, especialmente nas macrorregiões Nordeste, Jequitinhonha, Triângulo do Norte e Leste do Sul, indica a necessidade de qualificação dos profissionais para a coleta do exame citopatológico do colo do útero (INCA, 2012).

O índice de positividade de exames de Papanicolaou no Estado foi baixo ao longo de todo o período analisado, semelhante ao encontrado para o Brasil (2,5%) para o período de análise de 2002 a 2011 (INCA, 2012). Jequitinhonha e Leste do Sul se destacaram com resultados muito baixos para esse indicador, o que pode sugerir que alterações suspeitas não estão sendo identificadas, resultando em exames falso-negativos. Nesse sentido, a avaliação da qualidade do exame citopatológico é fundamental, pois a coleta adequada possibilita diagnosticar correta e precocemente as lesões precursoras, reduzindo os resultados falso-negativos e a necessidade de repetições de exames. Assim, a qualidade do exame está diretamente relacionada ao sucesso do rastreamento do CCU (INCA, 2015).

Os resultados do estudo revelaram que os principais tipos de alterações no Estado foram as atipias, constatando-se aumento dos percentuais com a elevação da faixa etária. Dessa forma, destaca-se que, em mulheres com 60 anos ou mais, as atipias corresponderam a 60,3% do total de alterações no período analisado. De forma geral, na faixa etária inferior a 25 anos, houve predominância de LSIL. As HSIL foram mais frequentes na faixa etária igual ou maior a 60 anos. Resultados semelhantes foram verificados em estudo com base em dados do SISCOLO em Teresina, no Piauí, no período de 2006-2013 (DAMACENA; LUZ; MATTOS, 2017). Dentro desse contexto, destaca-se que o percentual de atipias indica indiretamente a qualidade da coleta e da análise laboratorial e aponta, quando predominante, para a necessidade de melhoria na qualidade dos exames (BORTOLON *et al.*, 2012).

No presente estudo, observou-se redução no percentual de carcinoma escamoso ao final do período analisado, assim como verificado por Dias e colaboradores para o Brasil, no período de 2002 a 2006, o que pode indicar aumento da resolutividade do Programa de Rastreamento de Câncer de Colo do Útero (DIAS; TOMAZELLI; ASSIS, 2010).

Verificou-se que o percentual de seguimento para mulheres diagnosticadas com HSIL em 2010 com seguimento em 2011 foi baixo no Estado (52,7%), mas bem superior ao verificado em âmbito nacional, que foi de 21,5%(INCA, 2012). Isso pode indicar que a informação de seguimento não está sendo registrada ou que mulheres com HSIL não estão sendo acompanhadas. O objetivo do rastreamento é identificar e tratar precocemente as lesões precursoras. Assim, a informação do seguimento das mulheres que tiveram esse diagnóstico é fundamental para avaliar as ações do programa do controle do CCU (INCA, 2014).

Para que haja uma efetiva redução da incidência e da mortalidade pelo CCU, é necessário que o programa de rastreamento seja ofertado àquelas mulheres pertencentes aos grupos de risco preconizados por meio de exames de qualidade e que o tratamento adequado seja propiciado às mulheres que apresentam resultados positivos (MARTÍNEZ-MESA *et al.*, 2013).

Os dados do SISCOLO permitiram a construção de indicadores fundamentais para avaliar as ações do rastreamento do CCU em Minas Gerais, podendo subsidiar o planejamento das ações de saúde. Com a implantação do SISCAN (Sistema de Informação do Câncer), que substitui e integra o SISCOLO e o SISMAMA (Sistema de Informação do Câncer de Mama), algumas limitações desses sistemas serão superadas e permitirão o acompanhamento longitudinal das usuárias, pois o registro terá como unidade de observação a mulher, e não o exame.

Diante dos achados, percebe-se que o programa de rastreamento do CCU em Minas Gerais apresenta limitações, podendo-se verificar as diferenças regionais em relação aos indicadores analisados. Os resultados evidenciam a necessidade de aprimoramento do programa de rastreamento em relação à oferta e qualidade do exame, bem como ao direcionamento correto quanto à faixa etária e à periodicidade, visando garantir o acesso da população-alvo e

encaminhamento para a investigação diagnóstica e tratamento das lesões precursoras, quando indicado.

REFERÊNCIAS

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **Rastreamento** [Internet]. Brasília: Ministério da Saúde; 2010 [citado em 2016 jul 5]. (Série A. Normas e Manuais Técnicos. Cadernos de Atenção Primária, 29). Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/caderno_atencao_primaria_29_rastreamento.pdf

BRASIL. Ministério da Saúde. **Departamento de Informática do Sistema Único de Saúde** [Internet]. 2016 [citado em 2016 ago 2]. Disponível em: <http://datasus.saude.gov.br/>

BRASIL. Ministério da Saúde. Departamento de Apoio à Gestão Descentralizada. **Instrutivo da Pactuação de Prioridades, Objetivos, Metas e Indicadores de Monitoramento do Pacto pela Vida e de Gestão para o Biênio 2010-2011**. Conforme Portaria nº 2669 GM/MS de 03 de novembro de 2009, Brasília: Ministério da Saúde; 2009.

BORTOLON, P.C.; SILVA, M.A.F.; CORRÊA, F.M.; DIAS, M.B.K.; KNUPP, V.M.A.O.; ASSIS, M.*et al.* Avaliação da qualidade dos laboratórios de citopatologia do colo do útero no Brasil. **Revista Brasileira de Cancerologia**, v. 58, n. 3, p. 435-44, 2012.

DAMACENA, A.M.; LUZ, L.L.; MATTOS, I.E. Rastreamento do câncer do colo do útero em Teresina, Piauí: estudo avaliativo dos dados do Sistema de Informação do Câncer do Colo do Útero, 2006-2013. **Epidemiologia e Serviços de Saúde**, v. 26, n. 1, p. 71-80, 2017

DIAS, M.B.K.; TOMAZELLI, J.G.; ASSIS, M. 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**, v. 19, n. 3, p. 293-306, 2010.

DOMINGUES, E.P.; MAGALHÃES, A.S.; FARIA, W.R. Infraestrutura, crescimento e desigualdade regional: uma projeção dos impactos dos investimentos do Programa de Aceleração do Crescimento (PAC) em Minas Gerais. **PPE**, v. 39, n. 1, p. 121-58, 2009.

DUARTE, D.A.P. **Iniquidade social e câncer em mulheres: análise da mortalidade por câncer de mama e colo do útero nas microrregiões de saúde de Minas Gerais no período de 2008-2012** [dissertação]. Juiz de Fora: Universidade Federal de Juiz de Fora; 2016.

INTERNATIONAL AGENCY FOR RESEARCH ON CANCER. **Globocan 2012** [Internet]. Lyon: IARC; 2012 [citado em 2016 set 23]. Disponível em: <http://globocan.iarc.fr/old/FactSheets/cancers/cervix-new.asp>

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação de Prevenção e Vigilância. **Estimativa 2016: incidência de câncer no Brasil** [Internet]. Rio de Janeiro: INCA; 2015 [citado em 2016 jun 7]. Disponível em: <http://www.inca.gov.br/estimativa/2016/estimativa-2016-v11.pdf>

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero** [Internet]. Rio de Janeiro: INCA; 2016 [citado em 2016 jun 1].

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama** [Internet]. Informativo Detecção Precoce. 2014 jan/abr;5(1):1-8. [citado em 2016 jun 20]. Disponível em: http://www1.inca.gov.br/inca/Arquivos/inform_detec_precoce_1_2014.pdf

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama** [Internet]. Informativo Detecção Precoce. 2015 jan/abr;6(1):1-8. [citado em 2016 ago 2]. Disponível em: http://www1.inca.gov.br/inca/Arquivos/comunicacao/deteccao_precoce_12015.pdf

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Nomenclatura brasileira para laudos cervicais e condutas preconizadas - recomendações para profissionais de saúde**. Revista Brasileira de Cancerologia, v. 52, n. 3, p. 213-236, 2006.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação Geral de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Ficha técnica de indicadores das ações de controle do câncer do colo do útero** [Internet]. Rio de Janeiro: INCA; 2014 [citado em 2016 jun 25]. Disponível em: <http://www2.inca.gov.br/wps/wcm/connect/84f26080469faa79859bed5120665fa8/FICHA+T%C3%89CNICA+Indicadores+Colo+14.pdf?MOD=AJPERES&CACHEID=84f26080469faa79859bed5120665fa8>

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama** [Internet]. Informativo Detecção Precoce. 2012 jan/mar;3(1):1-8. [citado em 2016 jun 19]. Disponível em: http://www1.inca.gov.br/inca/Arquivos/informativo_detecprecoce_1.2_2012.pdf

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama** [Internet]. Informativo Detecção Precoce. 2012 ago/dez;3(3):1-12. [citado

em 2016 jun 20]. Disponível em:
[http://bvsms.saude.gov.br/bvs/publicacoes/inca/
Informativo_Deteccao_Precece_agosto_dezembro_2012.pdf](http://bvsms.saude.gov.br/bvs/publicacoes/inca/Informativo_Deteccao_Precece_agosto_dezembro_2012.pdf)

FONTANIVE, P.V.; KOLLING, J.H.G.; CASTRO, F.O.E.D; HARZHEIM, E. Cobertura da Estratégia Saúde da Família e de citopatologia de colo uterino no Rio Grande do Sul. **Revista Brasileira de Medicina de Família e Comunidade**, v.4, n.14, p. 119-28, 2008.

GIRIANELLI, V.R.; GAMARRA, C.J.; AZEVEDO E SILVA, G. Os grandes contrastes na mortalidade por câncer do colo uterino e de mama no Brasil. **Revista de Saúde Pública**, v. 48, n. 3, 459-67, 2014.

MARTÍNEZ-MESA, J.; WERUTSKY, G.; CAMPANI, R.B.; WEHRMEISTER; F.C.; BARRIOS, C.H. Inequalities in Pap smear screening for cervical cancer in Brazil. **Preventive Medicine**. v. 57, n. 4, p. 366-71, 2013.

NASCIMENTO, G.W.C.; PEREIRA, C.C.A.; NASCIMENTO, D.I.C.; LOURENÇO, G.C.; MACHADO, C.J. Cobertura do exame citopatológico do colo do útero no Estado de Minas Gerais, Brasil, no período entre 2000-2010: um estudo a partir dos dados do Sistema de Informação do câncer do Colo do Útero (SISCOLO). **Cadernos Saúde Coletiva**, v. 23, n. 3, p. 253-60, 2015.

SALA, A.; MENDES, J.D.V. Perfil de Indicadores da Atenção Primária à Saúde no Estado de São Paulo: retrospectiva de 10 anos. **Saúde Sociedade**, v. 20, n. 4, p. 912- 26, 2011.

SANTOS, R.S.; MELO, E.C.P.; SANTOS, K.M. Análise espacial dos indicadores pactuados para o rastreamento do câncer do colo do útero no Brasil. **Texto Contexto Enfermagem**, v. 21, n. 4, p. 800-10, 2012.

TOMASI, E.; OLIVEIRA, T.F.; FERNANDES, P.A.A.; THUMÉ, E.; SILVEIRA, D.S.; SIQUEIRA, F.V.*et al.* Estrutura e processo de trabalho na prevenção do câncer de colo de útero na Atenção Básica à Saúde no Brasil: Programa de Melhoria do Acesso e da Qualidade – PMAQ. **Revista Brasileira de Saúde Materno Infantil**, v.15, n. 2, p. 171-80, 2015.

VALE, D.B.A.P.; MORAIS, S.S.; PIMENTA, A.L.; ZEFERINO, L.C. Avaliação do rastreamento do câncer do colo do útero na Estratégia Saúde da Família no Município de Amparo, São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 26, n. 2, p. 383-90, 2010.

WORLD HEALTH ORGANIZATION. **National cancer control programmes: policies and managerial guidelines**. 2 ed. Geneva: WHO; 2002.

WORLD HEALTH ORGANIZATION. **Comprehensive cervical cancer control: a guide to essential practice**. 2 ed. Geneva: WHO; 2014.

6.2 – ARTIGO 2: COBERTURA E SOBERRASTREAMENTO DO EXAME CITOPATOLÓGICO DO COLO DO ÚTERO EM MINAS GERAIS

Resumo

Objetivos: Estimar a cobertura do rastreamento do câncer do colo do útero e a proporção de soberrastreamento por periodicidade e faixa etária a partir da identificação, nas bases do SISCOLO, das mulheres rastreadas pelo SUS em Minas Gerais no ano de 2008.

Métodos: estudo descritivo, de coorte não-concorrente, baseado em dados identificados do SISCOLO. A cobertura foi calculada utilizando exames de mulheres de 25 a 59 anos e 25 a 64 anos após exclusão de duplicidades e separação dos exames de seguimento, ficando apenas um registro de exame por mulher nesse ano. O soberrastreamento por faixa etária correspondeu ao percentual de mulheres com <25 e ≥65 anos. O soberrastreamento por periodicidade foi avaliado a partir dos exames de seguimento de 2008 através do relacionamento probabilístico entre as bases de 2008 a 2012, para mulheres de 25 a 64 anos. O tempo até o próximo exame citopatológico foi definido como o intervalo entre a data de coleta do primeiro exame normal com amostra satisfatória realizado em 2008 e a data de coleta da primeira repetição deste exame. Para mulheres que iniciaram o rastreamento em 2008, o soberrastreamento por periodicidade foi definido como a repetição em um período inferior a 9 meses e, para aquelas que iniciaram o rastreamento previamente, a repetição em um período inferior a 30 meses.

Resultados: Em 2008, 1.211.014 mulheres foram submetidas ao rastreamento do CCU pelo SUS em Minas Gerais, sendo que 79,8% estavam na faixa etária de 25 a 64 anos. A cobertura foi de 77,6% na faixa etária de 25 a 59 anos e de 76,2%, de 25 a 64 anos. Identificou-se o seguimento de 1.162.405 mulheres sendo o soberrastreamento por periodicidade de 3,9% para aquelas que iniciaram o rastreamento em 2008 e de 62,6% para as que tinham exame prévio. O soberrastreamento por faixa etária foi de 49,6% em mulheres que estavam iniciando o rastreamento em 2008 e de 14,9% naquelas com exame prévio.

Conclusão: A cobertura do estado, em 2008, estava próxima da meta de 80%. Os resultados apontam condutas diferentes das estabelecidas nas diretrizes. Quanto à periodicidade, o soberrastreamento é mais importante em mulheres

que deveriam fazer o exame a cada três anos e quanto à faixa etária, em mulheres que estão iniciando o rastreamento.

Palavras-chave: Neoplasias do Colo do Útero; Programas de Rastreamento; Sistema de informação em Saúde; Relacionamento de bases de dados; Excesso de exames.

Abstract

Objectives: To estimate the coverage of cervical cancer screening and the proportion of over-screening by periodicity and age group based on the identification, on the basis of SISCOLO, of the women screened by SUS in Minas Gerais in 2008.

Methods: descriptive study, non-concurrent cohort, based on identified data from SISCOLO. Coverage was calculated using examinations of women aged 25 to 59 and 25 to 64 years after excluding duplicates and separating the follow-up exams, leaving only one exam record per woman in that year. Over-screening by age group was 49.6% in women who were starting screening in 2008 and 14.9% in those without prior examination. Over-screening due to excessive periodicity was assessed from the 2008 follow-up exams through the probabilistic relationship between the bases from 2008 to 2012, for women aged 25 to 64 years. The time until the next cytopathological examination was defined as the interval between the date of the first normal examination with a satisfactory sample carried out in 2008 and the date of the first repetition of this examination. For women who started screening in 2008, over-screening by periodicity was defined as repetition in less than 9 months and, for those who started screening previously, repetition in less than 30 months. Over-screening by age group corresponded to the percentage of women aged <25 and ≥65 years.

Results: In 2008, 1,211,014 women were submitted to CCU screening by SUS in Minas Gerais, 79.8% of whom were aged between 25 and 64 years. Coverage was 77.6% in the age group from 25 to 59 years old and 76.2%, from 25 to 64 years old. The follow-up of 1,162,405 women was identified, with over-screening by periodicity at 3.9% for those who started screening in 2008 and 62.6% for those who had a previous examination.

Conclusion: State coverage, in 2008, was close to the 80% target. The results show different behaviors than those established in the guidelines. In terms of

frequency, over-screening is more important in women who should be tested every three years, and in terms of age, in women who are starting screening.

INTRODUÇÃO

O câncer do colo do útero (CCU) é o quarto tipo de câncer mais comum em mulheres no mundo, com 85% dos casos e mortes pela doença ocorrendo em países em desenvolvimento (FERLAY et al., 2015). No Brasil, excluindo o câncer de pele não melanoma, o CCU é o terceiro tipo de câncer mais frequente nas mulheres (7,4% de todos os tipos de câncer), ficando atrás apenas do câncer de mama e de cólon (INCA, 2019a). Quanto à mortalidade pela doença, no período entre 1978 e 2017, houve declínio em todas as regiões do Brasil, exceto no interior da região Norte, onde as taxas em 2017 chegaram a ser três vezes maiores em relação à região Sudeste (AZEVEDO E SILVA et al., 2020).

O exame citopatológico do colo do útero é considerado a principal estratégia para o rastreamento do CCU no Brasil (SANKARANARAYANAN *et al.*, 2008), por detectar precocemente as lesões precursoras e possibilitar o tratamento antes da evolução para a forma invasiva da doença, tendo como objetivo final a diminuição da incidência e mortalidade por CCU (EL-ZEIN; RICHARDSON; FRANCO, 2016).

O *Programa Nacional de Combate ao Câncer do Colo do Útero*, instituído no Brasil em 1998, estabeleceu como faixa etária alvo do rastreamento mulheres com 25 a 59 anos e periodicidade a cada três anos após dois exames anuais consecutivos negativos. A partir de 2011, as diretrizes do rastreamento ampliaram a faixa etária para 25 a 64 anos, mantendo a periodicidade (INCA, 2016). Porém, por ser predominantemente oportunístico, grande parte dos exames no país não segue essas recomendações. Desta forma, o rastreamento está associado ao sobrerastreamento, ou seja, exames em excesso realizados fora da faixa etária alvo ou com periodicidade em intervalo inferior ao preconizado pelo Ministério da Saúde. O sobrerastreamento, além de sobrecarregar os serviços, pode comprometer o acesso das mulheres que realmente precisam ser rastreadas (INCA, 2014a).

O *Sistema de Informação do Câncer do Colo do Útero (SISCOLO)*, implantado em 1999, é componente fundamental para monitoramento e organização das ações de rastreamento no país. Uma importante limitação deste sistema é o fato de ter como unidade de registro o procedimento realizado e não a mulher, impossibilitando o cálculo direto da taxa de cobertura e o acompanhamento longitudinal das mulheres no programa de rastreamento (BRASIL, 2011).

Diante desse contexto, é fundamental analisar os dados do SISCOLO tendo como unidade de registro a mulher e não o exame. Para tal, foi realizado relacionamento probabilístico das bases do SISCOLO contendo dados identificados, visando estimar a cobertura do rastreamento pelo SUS em Minas Gerais e suas macrorregiões de saúde no ano de 2008 e avaliar a proporção de sobrerastreamento quanto à faixa etária e quanto à periodicidade no estado, a partir do seguimento de 2008 a 2012.

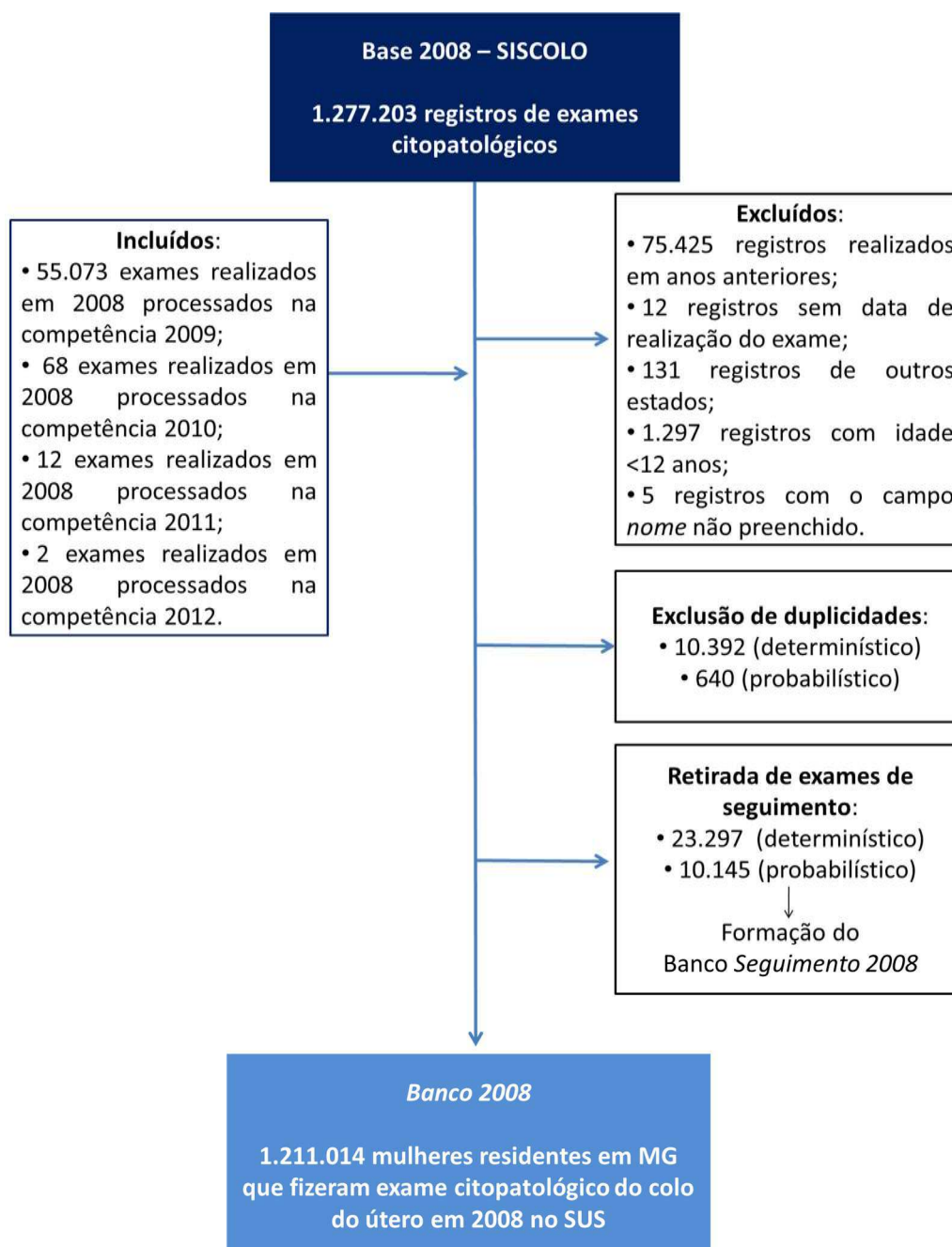
METODOLOGIA

Trata-se de estudo descritivo, de coorte não-concorrente, baseado em dados identificados do SISCOLO. A população de estudo foi constituída por mulheres residentes no estado de Minas Gerais (MG), submetidas ao rastreamento do câncer do colo do útero e com exames registrados no SISCOLO, no ano de 2008.

Foram utilizadas as bases do SISCOLO, referentes aos anos de 2007 a 2012, cedidas pela Secretaria de Saúde de MG e todas passaram pelo mesmo processo de padronização visando o posterior relacionamento probabilístico. O registro nessas bases é por exame e não por mulher, assim, visando transformar a unidade de análise em mulheres, foram excluídas as duplicidades (mesmo exame registrado mais de uma vez) da base de 2008. Foram ainda excluídos os registros com o campo nome em branco ou preenchido de forma que impossibilitasse a identificação da mulher e aqueles sem a data da realização do exame. As repetições de exame pela mesma mulher em 2008 foram agrupadas em outro banco (*Seguimento 2008*). Para isso, utilizaram-se técnicas de relacionamento determinístico e probabilístico, usando os programas *R* e *RecLink 3*, respectivamente, conforme metodologia proposta por estudo com

dados identificados do SISMAMA (DUARTE *et al.*, 2019). Após essas limpezas, formou-se, então, o *Banco 2008*, que passou a ter como unidade de registro a mulher, contendo apenas o seu primeiro exame realizado nesse ano. Tal banco foi composto por 1.211.014 mulheres submetidas ao rastreamento em 2008 em Minas Gerais. A figura 1 ilustra o processo de limpeza da base de 2008 do SISCOLO e formação do *Banco 2008*.

Figura 1 - Limpezas da base 2008 do SISCOLO e formação do *Banco 2008*



A cobertura do rastreamento em MG foi estimada para mulheres na faixa etária de 25 a 59 anos, que era a população alvo vigente no ano de 2008, e também para mulheres de 25 a 64 anos, conforme as diretrizes atuais (INCA, 2016). A cobertura foi calculada pela razão entre o número de mulheres por faixa

etária do *Banco 2008* e um terço da população feminina que não tinha plano de saúde, na mesma faixa etária (INCA, 2014b).

O sobrerastreamento por faixa etária foi definido como o percentual de mulheres fora da faixa etária preconizada pelo Ministério da Saúde (<25 e ≥65 anos) do total de mulheres submetidas ao rastreamento em 2008.

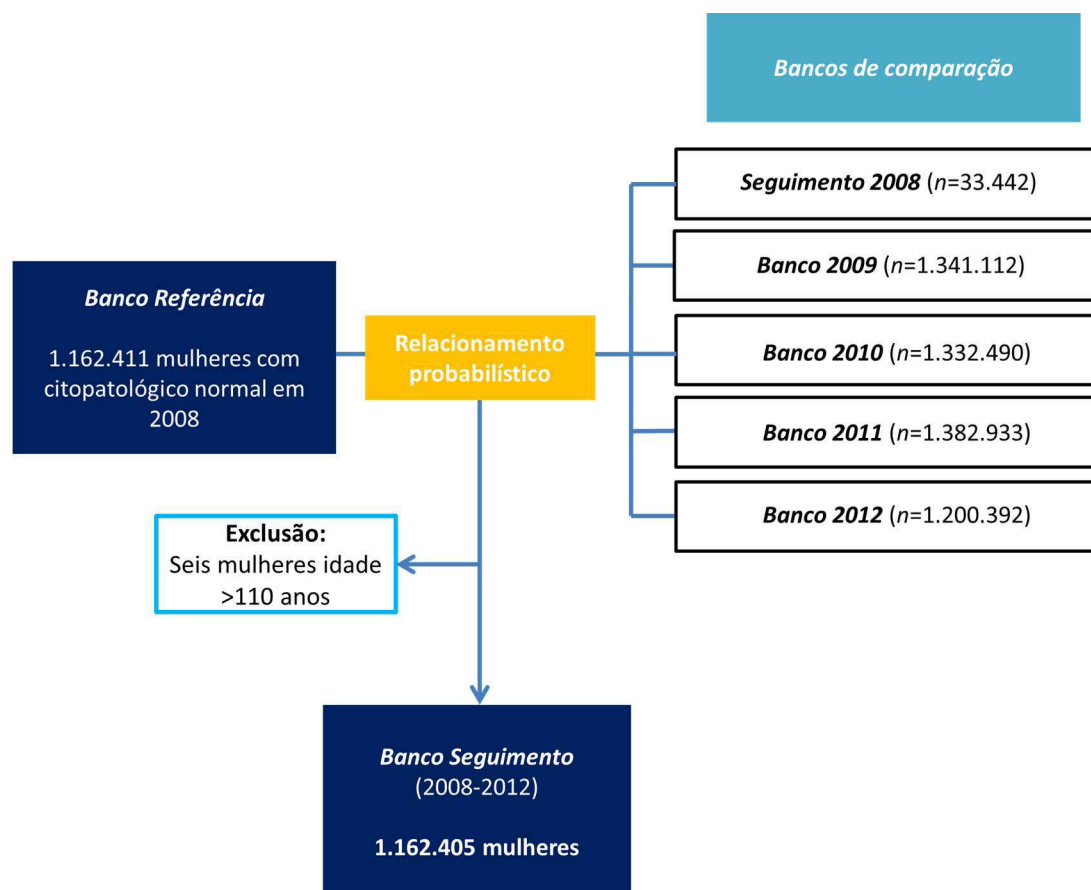
Para avaliar o sobrerastreamento por periodicidade, foi montado o seguimento das mulheres submetidas ao rastreamento em 2008 que tiveram resultado normal através do relacionamento probabilístico entre o *Banco Referência* (*Banco 2008* após retirada de exames alterados) e os exames de seguimento do mesmo ano (*Seguimento 2008*) e cada banco anual de 2009 a 2012. Então, as mulheres que tiveram exame alterado ou com amostra insatisfatória/rejeitada no ano anterior (2007) ou em 2008 foram excluídas visando evitar a superestimação do sobrerastreamento, uma vez que a recomendação de repetição do exame é diferente para essas situações (INCA, 2016).

Para excluir mulheres que tinham exame alterado anteriormente, foi realizado relacionamento probabilístico entre o *Banco 2008* e o banco contendo apenas mulheres com exame alterado em 2007 ($n= 36.228$ mulheres) por meio do programa *RecLink III*, utilizando os parâmetros de acordo com as orientações do manual do programa (CAMARGO JR.; COELI, 2007). O relacionamento entre os bancos foi feito em quatro passos de um total de 16 testados previamente e descritos por Duarte e colaboradores (2019), os quais apresentaram boa acurácia para identificação de pares verdadeiros. Para a seleção dos pares verdadeiros foi utilizado método automatizado por meio da inspeção manual de uma amostra aleatória de 300 pares (DUARTE *et al.*, 2019). Dessa forma, foram excluídas 18.915 mulheres do *Banco 2008* que tiveram exame alterado em 2007. Após essa limpeza, o banco ficou com um total de 1.192.099 mulheres. Posteriormente, foi feita a exclusão de 29.688 mulheres cujos exames coletados em 2008 tiveram resultado alterado ou amostra rejeitada ou insatisfatória. Assim, o banco utilizado para avaliar o sobrerastreio (*Banco referência*) foi composto por 1.162.411 mulheres que tiveram exame citopatológico do colo do útero realizado em 2008 com resultado normal.

O seguimento das 1.162.411 mulheres foi construído através do relacionamento probabilístico dessas mulheres (*Banco Referência*) com o *Banco*

Seguimento 2008 e cada banco anual de 2009 a 2012, conforme ilustrado na figura 2.

Figura 2 – Construção do *Banco Seguimento (2008-2012)*



Para isso, foi adotada a mesma metodologia utilizada no relacionamento entre as bases 2008 e 2007 (DUARTE *et al.*, 2019). Após o relacionamento identificaram-se seis mulheres com idade >110 anos que foram excluídas. Assim, a coorte do estudo (*Banco Seguimento 2008-2012*) foi composta por 1.162.405 mulheres.

As bases anuais contendo exames de 2009 a 2012 também passaram pelo mesmo processo de limpeza e retirada de duplicidades, exceto pela retirada de exames de seguimento. Verificou-se que alguns exames estavam contidos em anos diferentes do ano de coleta do mesmo. Então, esses exames foram incluídos no respectivo banco referente ao ano da coleta. O quadro 1 apresenta a descrição da formação destes bancos após limpezas das bases, com inclusão

de exames processados em outras competências e exclusão de exames sem data de realização, com data de outros anos, sem nome, de outros estados e exames duplicados.

Ao final do relacionamento, procedeu-se à união dos arquivos de pares dos respectivos passos. Com isso, obteve-se um arquivo contendo 3.431.089 pares resultantes do relacionamento 2008 a 2012, que foram unidos ao *Banco Referência* utilizando o programa *R*. Posteriormente, o *Banco Seguimento 2008-2012* foi organizado para que cada linha representasse a mulher e as repetições dos exames foram ordenadas por cronologia nas colunas. Para este estudo, foi considerada apenas a primeira repetição.

Calculou-se, então, o tempo entre a data da coleta do primeiro exame com amostra satisfatória e resultado normal realizado em 2008 (início do seguimento) e a data da primeira repetição para mulheres de 25 a 64 anos. Para estimar o sobrerastreamento por periodicidade, a falha (evento de interesse) foi considerada como a repetição da coleta do exame antes do período preconizado pelo Ministério da Saúde. Como a recomendação é diferente para mulheres que estão iniciando o rastreamento e mulheres que já têm exame prévio, o *Banco Seguimento 2008-2012* foi dividido, com base na variável *exame anterior*: 1) mulheres que relataram não ter sido submetidas ao citopatológico anteriormente e 2) mulheres que relataram exame anterior. Foram excluídas as mulheres que informaram não saber se já tinham iniciado o rastreamento ou o registro estava sem esta informação. A distribuição do tempo foi apresentada para os dois grupos de mulheres, em valor mínimo, primeiro quartil, mediana, média, terceiro quartil e valor máximo.

Quadro 1 – Descrição da formação dos Bancos de comparação referentes aos anos 2009 a 2012

BANCO	BASE ORIGINAL	INCLUSÃO	EXCLUSÃO							Número final de registros	
			Número de registros	Registrados em outras competências	Sem data de coleta do exame	Outros anos	Sem nome	Outros estados	Duplicidades		
									Determinístico		Probabilístico
2009	1.348.128	72.255	3	55.836	10	153	5.484	1.369	1.341.112		
2010	1.337.089	70.225	3	72.501	7	216	3.770	606	1.332.490		
2011	1.388.702	72.945	1	70.489	6	135	4.687	941	1.382.933		
2012	1.211.348	----	14	73.325	1	122	10.310	523	1.200.392		

Para mulheres que iniciaram o rastreamento em 2008, o sobrerastreamento por periodicidade foi definido como a repetição do exame em um período inferior a nove meses (274 dias). Nesse caso, a recomendação é a repetição do exame com intervalo de 12 meses, mas variações ≤ 3 meses do preconizado foram consideradas pouco relevantes. Aquelas mulheres que repetiram o exame em período maior ou igual a nove meses foram censuradas. Para esse grupo, o sobrerastreamento foi calculado de acordo com a seguinte fórmula:

$$\text{Sobrerastreamento} = \frac{N^{\circ} \text{ mulheres que repetiram o exame} < 9 \text{ meses (274 dias)}}{\text{Total de mulheres que iniciaram o rastreamento em 2008}}$$

Para as mulheres que iniciaram o rastreamento antes de 2008, a repetição em um período inferior a 30 meses (912 dias) foi considerada como sobrerastreamento e, aquelas cuja repetição foi igual ou maior a 30 meses, foram censuradas. Para mulheres

que já iniciaram previamente o rastreamento, a recomendação é a repetição do exame com intervalo de 36 meses, mas variações ≤ 6 meses foram consideradas pouco relevantes. O cálculo do sobrerastreamento foi feito utilizando a fórmula a seguir para essas mulheres:

$$\text{Sobrerastreamento} = \frac{\text{N}^{\circ} \text{mulheres que repetiram o exame} < 30 \text{ meses (912 dias)}}{\text{Total de mulheres que iniciaram o rastreamento antes de 2008}}$$

O método de *Kaplan Meier* foi utilizado para estimar a probabilidade cumulativa de sobrerastreamento por periodicidade ao longo do período de seguimento em cada grupo e as curvas foram estratificadas por faixa etária. Para comparar os estratos, foi utilizado o teste de *log-rank*, com nível de significância de 5%. Todas as análises foram realizadas utilizando o programa *R* versão 4.0.2, pacote *survival*.

RESULTADOS

Em 2008, 1.211.014 mulheres foram submetidas ao rastreamento pelo SUS em Minas Gerais, sendo que, 917.436 (75,8%) estavam na faixa etária de 25 a 59 anos e 966.522 (79,8%), na faixa de 25 a 64 anos.

A coorte do estudo foi composta por 1.162.405 mulheres (cinco mulheres foram excluídas após o relacionamento, pois tinham idade >110 anos), destas, 11,7% (135.635 mulheres) iniciaram o rastreamento em 2008 e 73,4% ($n=852.777$) o iniciaram previamente. As restantes foram excluídas, pois relataram não saber se já tinham iniciado o rastreamento (6,2% - $n=71.750$) ou o registro estava sem esta informação (8,8% - $n=102.243$). Das 135.635 mulheres que iniciaram o rastreamento em 2008, 68.311 (50,4%) tinham de 25 a 64 anos. E das 852.777 que tinham exame prévio, 724.772 (85,0%) estavam nessa faixa etária.

A cobertura do rastreamento do câncer do útero pelo SUS em MG no ano de 2008 foi de 77,6% considerando a faixa etária de 25 a 59 anos e de 76,2%, de 25 a 64 anos. A tabela 1 mostra a cobertura por macrorregião de saúde e faixa etária no estado. Para a faixa etária de 25 a 64 anos, as macrorregiões

Nordeste, Leste do Sul e Jequitinhonha foram as que apresentaram as menores coberturas (62,9; 66,4 e 67,7%, respectivamente) e Triângulo do Norte, Sul e Norte, as maiores (85,0; 82,5 e 80,0%, respectivamente). Em 2008, a cobertura por plano privado de saúde em mulheres de 25 a 64 anos em MG foi de 24,4%, sendo que, as macrorregiões Centro, Triângulo do Norte e Sul apresentaram as maiores coberturas por plano de saúde no estado (36,2; 27,0 e 23,2, respectivamente), enquanto, Jequitinhonha, Nordeste e Norte, apresentaram as menores (4,8%; 5,4% e 7,6%).

Tabela 1 – Cobertura do rastreamento do câncer do colo do útero, por faixa etária e macrorregiões de saúde. Minas Gerais, 2008.

Macrorregiões	25-29 anos	30-39 anos	40-49 anos	50-59 anos	60-64 anos	25-59 anos	25-64 anos
Centro	74,7	77,8	81,8	77,3	61,2	78,3	77,2
Centro Sul	61,9	74,7	72,1	66,4	50,4	69,9	68,5
Jequitinhonha	60,6	80,8	70,6	61,2	44,0	69,6	67,7
Leste	60,5	75,0	73,6	68,9	48,6	70,5	69,0
Leste do Sul	62,5	69,9	69,1	66,0	51,8	67,4	66,4
Nordeste	57,7	73,3	66,5	58,5	40,9	64,9	62,9
Noroeste	69,1	82,3	78,8	70,0	51,7	76,4	74,8
Norte	73,5	90,2	85,6	75,2	51,5	82,3	80,1
Oeste	71,3	82,3	83,5	77,5	60,6	79,7	78,4
Sudeste	69,3	77,0	77,5	75,3	61,7	75,4	74,4
Sul	79,6	86,8	85,4	80,7	64,7	83,8	82,5
Triângulo do Norte	78,8	84,4	90,8	86,5	75,2	85,7	85,0
Triângulo do Sul	73,9	77,0	77,5	73,2	55,5	75,7	74,3
TOTAL - MG	71,6	80,2	80,4	75,0	57,9	77,6	76,2

Nota: a faixa etária de 60 a 64 anos passou a ser considerada como alvo do rastreamento a partir de 2011.

Observa-se que as faixas etárias de 30-39 e 40-49 anos foram as que apresentaram as maiores coberturas, com cerca de 80%. Enquanto, as faixas de 60-64 e 25-29 anos foram as que apresentaram os menores valores.

O sobrerastreamento por faixa etária, correspondeu a 20,2% das mulheres submetidas ao rastreamento em 2008, sendo que, 15,2% tinha menos de 25 anos. Para mulheres que iniciaram o rastreamento em 2008, o

sobrerastreamento por faixa etária foi de 49,6%, sendo que 45,5% tinha menos de 25 anos. E para aquelas que tinham exame prévio, foi de 15,0%.

A tabela 2 apresenta a distribuição do tempo entre a data de coleta do primeiro exame realizado em 2008 (exame de referência) e a data da primeira repetição no período de seguimento para mulheres de 25 a 64 anos que estavam iniciando o rastreamento em 2008 (n=68.311) e para aquelas, na mesma faixa etária, que iniciaram o rastreamento antes de 2008 (n=724.774).

Tabela 2 – Distribuição do tempo em meses entre a data de coleta do primeiro exame realizado em 2008 e a data da primeira repetição.

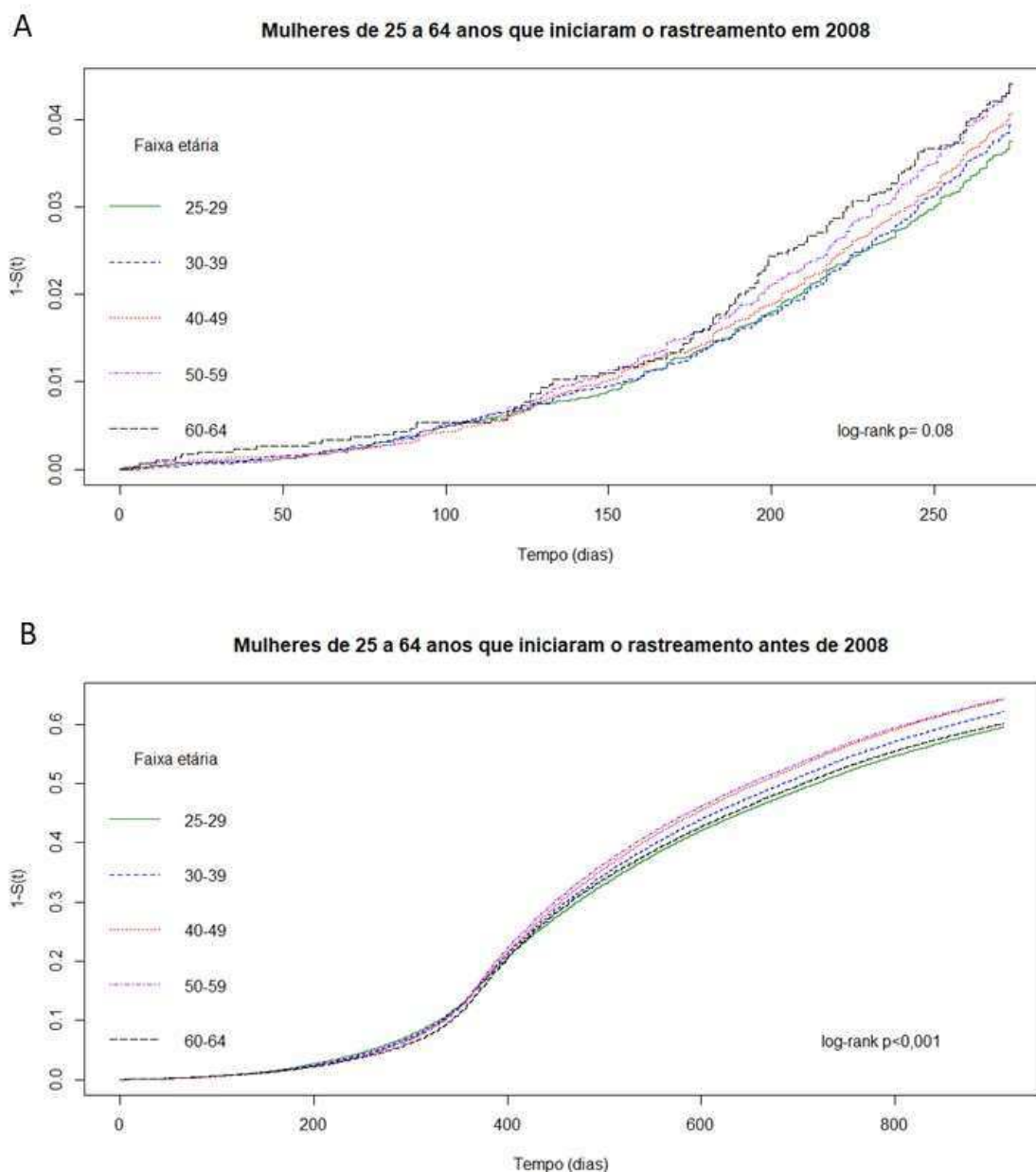
	Mín	Q1	Mediana	Média	Q3	Máx
Mulheres que iniciaram o rastreamento em 2008	1	16	31	33	52	60
Mulheres que iniciaram o rastreamento antes de 2008	1	14	22	28	45	60

Legenda: Mín: valor mínimo; Q1: primeiro quartil; Q3: terceiro quartil; Máx: valor máximo

Observa-se que a média e a mediana do tempo entre o primeiro exame realizado em 2008 e a repetição posterior foi maior em mulheres que estavam iniciando o rastreamento em 2008. Em mulheres que iniciaram o rastreamento em 2008, 30% repetiram o exame dentro de 18 meses (12,3%: ≤ 12 meses; 17,7%: >12 e ≤ 18 meses) e, no outro grupo, esse percentual foi de 41,2% (17,8%: ≤ 12 meses; 23,4%: >12 e ≤ 18 meses).

Quanto ao sobrerastreamento por periodicidade, 3,9% (n=2.691) das mulheres de 25 a 64 anos que iniciaram o rastreamento em 2008, repetiram o exame em um período inferior a nove meses e não houve diferença entre as faixas etárias - log-rank: $p < 0,08$ (figura 3A). Nas mulheres que iniciaram o rastreamento antes de 2008, o sobrerastreamento por periodicidade foi 62,6% (n=453.558). Destas, 28,6% repetiram o exame em período ≤ 12 meses; 37,3%, > 12 e ≤ 18 meses; 21,1%, > 18 e ≤ 24 meses e 12,9%, >24 e <30 meses. Ao final do período de seguimento, o sobrerastreamento foi maior nas faixas de 40-49 e 50-59 anos, seguido pela faixa 30-39 anos e depois pelas faixas 60-64 e 25-29 anos - log-rank: $p < 0,001$ (figura 3B).

Figura 3 - Gráfico de *Kaplan-Meier* do percentual cumulativo de mulheres sobrerastreadas estratificado por faixa etária no grupo de mulheres que (A) iniciou o rastreamento em 2008 ($n=68.311$) e no grupo que (B) iniciou antes de 2008 ($n=724.772$).



DISCUSSÃO

A cobertura do rastreamento do CCU em mulheres que utilizam exclusivamente o SUS no estado de Minas Gerais foi de 77,6% e 76,2% nas faixas etárias recomendadas pelo Ministério da Saúde antes e após 2011 (25 a

59 anos e 25 a 64 anos, respectivamente). Assim, verifica-se que, o estado quase atingiu a meta de 80% de cobertura em 2008 que é preconizada pela OMS (WHO, 2002). Não houve diferença da cobertura utilizando as duas faixas etárias de população alvo. Observou-se que mulheres de 30 a 49 anos foram as que apresentaram as maiores coberturas no estado.

O estado de Minas Gerais é caracterizado por grandes disparidades socioeconômicas entre suas macrorregiões e é considerado uma representação da estrutura regional brasileira, menos desenvolvida ao Norte/Nordeste e mais desenvolvida ao Sul, incluindo o Triângulo e o Centro (DOMINGUES; MAGALHÃES; FARIA, 2009). Observou-se que, de forma geral, macrorregiões de melhor desenvolvimento socioeconômico, apresentaram maiores coberturas, exceto a região Norte que se caracteriza por menor desenvolvimento, mas conseguiu alcançar a meta de rastreamento.

Segundo a Pesquisa Nacional de Saúde (PNS), no Brasil, em 2013, a cobertura do rastreamento do CCU em mulheres de 25 a 64 anos foi de 79,4%. Na Região Sudeste, a cobertura foi de 81,1% e, em Belo Horizonte, 82,1%. Cabe destacar que a PNS é um inquérito domiciliar de abrangência nacional e que essa cobertura inclui, portanto, as mulheres que têm acesso a planos privados de saúde. Além disso, foi obtida a partir do autorrelato das mulheres sobre a realização do citopatológico pelo menos uma vez nos últimos três anos (OLIVEIRA et al., 2018). Semelhante aos resultados do presente estudo, mulheres entre 35 e 44 anos e, entre 45 e 54 anos, apresentaram as maiores coberturas: 83,2% e 81,6%, respectivamente. Na Espanha, o Inquérito Nacional de Saúde de 2011 também verificou maiores coberturas nas faixas etárias mais jovens (25-39 anos e 40-52 anos, 72,5 e 73,7%, respectivamente) em relação à faixa etária de mulheres mais velhas (53-65 anos: 61,5%) - (RICARDO-RODRIGUES et al. 2015).

O indicador *razão de exames citopatológicos em mulheres na faixa etária alvo do rastreamento* é considerado um *proxy* da cobertura. Pressupõe-se a razão de 0,30 como a oferta mínima para o alcance de 80% da população-alvo (INCA, 2010a). Ao comparar os resultados desse indicador publicado em estudo anterior (CORREA et al., 2017) para a faixa etária de 25 a 59 anos com a cobertura estimada no presente estudo para essa mesma faixa etária, observa-se diferenças quanto ao desempenho das macrorregiões. Norte, Jequitinhonha

e Sul tiveram o melhor desempenho (0,26; 0,23 e 0,23, respectivamente), enquanto Centro Sul, Centro e Noroeste (0,19; 0,17 e 0,10, respectivamente), o pior desempenho. Já os presentes resultados mostraram melhor desempenho de cobertura para as macrorregiões Triângulo do Norte (85,7); Sul (83,8) e Norte (82,3) e pior desempenho para Jequitinhonha (69,5); Leste do Sul (67,4) e Nordeste (64,9). Dessa forma, fica evidente a limitação do indicador, pois avalia a oferta de *Papanicolaou* com base no número de exames e não de mulheres examinadas. Além disso, pode estar subestimado em regiões com alta cobertura pela saúde suplementar, pois considera apenas os exames realizados pelo SUS, enquanto, no denominador, considera a população feminina total.

Cabe considerar, por outro lado, que o cálculo da cobertura anual também possui limitações e trata-se de uma estimativa aproximada. Ao utilizar 1/3 da população alvo do rastreamento no denominador, pode-se superestimar o indicador, pois mulheres que estão iniciando o rastreamento deveriam repetir o exame em um período de 12 meses e aquelas que possuem exames rejeitados ou insatisfatórios deveriam repeti-lo de forma imediata. No entanto, assim como o estimado pelo INCA (6% e 0,5%, respectivamente), o percentual de exames correspondentes a estas situações foi baixo no presente estudo (7% e 0,6%, respectivamente) (INCA, 2019b).

Os resultados mostraram que o sobrerastreamento por periodicidade é mais considerável para aquelas mulheres que deveriam fazer o exame a cada três anos. Nesse grupo, observou-se a repetição do exame antes do preconizado em 62,6% das mulheres. Dessas, a maioria (66%) repetiu o exame com 18 meses ou menos. Para as mulheres que iniciaram o rastreamento em 2008, o sobrerastreamento por faixa etária é mais importante, pois 49,6% não estava na faixa etária preconizada. Fica evidente, portanto, o não seguimento das diretrizes do INCA pelos profissionais de saúde do SUS que atuam no rastreamento do CCU.

Os achados quanto ao sobrerastreamento por periodicidade estão de acordo com outros estudos. Estudo norte-americano realizado com 2.087 profissionais de saúde verificou que de 67,1% a 93,8% deles solicitou o exame antes do recomendado pelas diretrizes (BERKOWITZ *et al.*, 2013). Estudo de coorte também utilizando dados secundários de 2.299.177 mulheres entre 30 e 65 anos que se submeteram ao rastreamento do CCU de 2013 a 2014 nos EUA,

verificou que 65,8% repetiu o exame antes de três anos. A faixa etária de 60-64 anos foi a que apresentou o menor percentual de sobrerastreamento (WRIGHT *et al.*, 2021). De acordo com dados do INCA, baseado no autorrelato de exame anterior, a periodicidade anual foi de 52,6% no Brasil em 2008 e 45,6% na região Sudeste (INCA, 2012). Em um município do interior de São Paulo foi verificado que cerca de metade das mulheres submetidas ao rastreamento entre 2001 e 2007 na atenção primária repetiu o exame com periodicidade anual e cerca de 25% a cada dois anos. Os autores estimam que, se os exames fossem realizados de acordo com as recomendações, poderia ser atingida uma cobertura trienal de 100% no período. É ressaltado, ainda, que, dessa forma, o rastreamento poderia se tornar mais efetivo, com otimização dos recursos e direcionamento adequado da oferta, sem a necessidade de expandir a capacidade assistencial (VALE *et al.*, 2010).

Quanto à faixa etária, 244.492 mulheres submetidas ao rastreamento em Minas Gerais em 2008 não estavam na faixa etária preconizada, ou seja, o sobrerastreamento foi de 20%. Esse resultado está de acordo com dados do INCA para o Brasil, segundo os quais, em 2011, 22% dos exames foram realizados em mulheres fora da faixa etária preconizada. E, semelhante aos nossos achados, o problema é maior para a faixa etária com menos de 25 anos, que no Brasil chegou a 17% do total de exames (INCA, 2012). Isso pode ser devido à crença de muitos profissionais de que o rastreamento deve começar quando a mulher inicia a vida sexual, já que em 2001, a recomendação da Federação Brasileira das Sociedades de Ginecologia e Obstetrícia era a realização do exame em todas as mulheres sexualmente ativas independentemente da idade (FEBRASGO, 2001). Nesse contexto, cabe ressaltar que nossos resultados mostraram que quase metade das mulheres que iniciaram o rastreamento em 2008 tinha menos de 25 anos. Sendo que, nessa faixa etária predominam lesões de baixo grau que têm grande probabilidade de regressão espontânea e podem resultar em um número significativo de colposcopias e procedimentos diagnósticos e terapêuticos desnecessários (BRASIL, 2013a).

De acordo com Fernandes e colaboradores (2009), o médico é a principal fonte de informação sobre o exame citopatológico do colo do útero e desta forma, interfere diretamente na prática do rastreamento, bem como no

sobrerastreamento. Os motivos mais citados para a não adesão às recomendações por profissionais de saúde nos Estados Unidos foram o não conhecimento das diretrizes e a demanda da paciente (TEOH *et al.*, 2015). Neste sentido, em programas de rastreamento oportunistas, mesmo que as mulheres conheçam as recomendações, geralmente preferem continuar o rastreamento anual do câncer do colo do útero. E os médicos, em geral, temem que exames menos frequentes resultem na perda de acompanhamento e se preocupam com a responsabilidade médica (WRIGHT *et al.*, 2021). Portanto, é fundamental identificar as barreiras para a não adesão às diretrizes visando desenvolver estratégias adequadas para a disseminação e implementação das mesmas na prática clínica (FISCHER *et al.*, 2016).

Destaca-se como limitação do presente estudo a qualidade dos dados secundários utilizados. Em virtude do alto índice de não preenchimento de informações pessoais no SISCOLO, não foi possível investigar a associação entre a cobertura e o sobrerastreamento com variáveis socioeconômicas. Além disso, há uma subnotificação no SISCOLO em relação ao SIA, pois o segundo está vinculado ao pagamento do exame. Em 2009, a diferença de exames citopatológicos registrados no SIA em relação ao SISCOLO foi de 7% (INCA, 2010b). Outra limitação que deve ser considerada é o fato da estimativa da cobertura não contemplar mulheres assistidas pela saúde suplementar. Para o cálculo da cobertura populacional são necessários inquéritos que abranjam todo o conjunto da população feminina (INCA, 2014b).

Em 2013, com o objetivo de substituir e integrar os sistemas de informação relacionados ao câncer feminino (SISCOLO e Sistema de Informação do Câncer de Mama - SISMAMA) foi instituído o Sistema de Informação de Câncer (SISCAN), que tem como unidade de observação a mulher (BRASIL, 2013b). Porém, este sistema ainda está em processo de implementação e alguns municípios ainda registram os exames no SISCOLO, não tendo havido uma transição definitiva para o SISCAN em todo o país. Dessa forma, o presente estudo possibilitou o desenvolvimento de uma metodologia para estimar a cobertura e a realização do seguimento de mulheres submetidas ao rastreamento utilizando dados do SISCOLO. Em virtude da falta de um identificador único no SISCOLO foi necessário utilizar técnica de relacionamento probabilístico para retirar as duplicidades e repetições do banco referência e

permitir a montagem do seguimento das mulheres. Com isso, a limitação do SISCOLO por não ter a mulher como unidade de registro foi superada. Tal metodologia pode ser utilizada em outros contextos com objetivo de monitorizar as ações de rastreamento e contribuir para o planejamento de ações estratégicas visando o aperfeiçoamento do programa de rastreamento.

Constata-se que a recomendação de periodicidade não tem sido adotada, embora seja a mesma desde 1998. Dados do SISCAN mostram que dos 1.050.616 exames registrados em 2018, cerca de 50% das mulheres referiram exame anterior em um período menor ou igual a um ano; cerca de 30%, no período de dois anos e apenas 12% no período de três anos. Além disso, em mulheres sem exame prévio, 53,2% tinha menos de 25 anos (DATASUS, 2021). Reforçando que o não seguimento das diretrizes pelos profissionais de saúde, tanto em relação à periodicidade quanto à faixa etária do rastreamento persiste e os percentuais são semelhantes aos observados no presente estudo.

Considera-se que o objetivo final do rastreamento do câncer de colo do útero é a redução da incidência e mortalidade pela doença. O impacto do rastreamento, contudo, depende da cobertura da população-alvo, com exame de qualidade, periodicidade adequada e acesso oportuno ao tratamento adequado das lesões identificadas. Nesse sentido, o sobrerastreamento compromete a efetividade do rastreamento, além de ser mais oneroso para o sistema de saúde (BRASIL, 2010).

Os resultados desse estudo permitiram a avaliação do programa de rastreamento no estado e a identificação de macrorregiões que necessitam expandir o acesso ao exame visando alcançar a meta de 80% de cobertura. Além disso, apontam a necessidade de adequações nas ações de rastreamento e a importância de esforços voltados para o planejamento de intervenções que aumentem a adesão dos profissionais de saúde às diretrizes, visando otimizar os recursos, direcionar adequadamente a oferta e atingir a eficiência adequada do programa de rastreamento com impacto na mortalidade.

REFERÊNCIAS:

AZEVEDO E SILVA, G.; JARDIM, B. C.; FERREIRA, V. M., et al. Cancer mortality in the Capitals and in the interior of Brazil: a four-decade analysis. **Revista Saúde Pública**, v. 54, p. 126, 2020.

BERKOWITZ, Z.; SARAIYA, M.; SAWAYA, G.F. Cervical Cancer Screening Intervals, 2006 to 2009: Moving Beyond Annual Testing. **JAMA internal medicine**, v. 8, 2013.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **Rastreamento**. Brasília, 2010.

BRASIL. Ministério da Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação Geral de Ações Estratégicas. Divisão de Apoio à Rede de Atenção Oncológica. **Sistema de informação do controle do câncer de mama (SISMAMA) e do câncer do colo do útero (SISCOLO): manual gerencial** – Rio de Janeiro: INCA, 2011.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **Controle dos cânceres do colo do útero e da mama** 2. ed. – Brasília: Editora do Ministério da Saúde, 2013a.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Instituto Nacional de Câncer José Alencar Gomes da Silva. **Sistema de informação do câncer: manual preliminar para apoio à implantação**. Rio de Janeiro: Instituto Nacional de Câncer José Alencar Gomes da Silva; 2013b.

CAMARGO JR, K.R.; COELI, C.M. **RecLink**. Rio de Janeiro; 2007.

CORRÊA, CSL; LIMA, AS; LEITE, ICG et al. Rastreamento do câncer do colo do útero em Minas Gerais: avaliação a partir de dados do Sistema de Informação do Câncer do Colo do Útero (SISCOLO). **Cadernos Saúde Coletiva**, v. 25, n. 3, p. 315-323, 2017.

DATASUS. **SISCAN**. Disponível em: <https://datasus.saude.gov.br/aceso-a-informacao/sistema-de-informacao-do-cancer-siscan-colo-do-utero-e-mama/>. Acesso em 10 mai 2021

DOMINGUES, E.P.; MAGALHÃES, A.S.; FARIA, W.R. Infraestrutura, crescimento e desigualdade regional: uma projeção dos impactos dos investimentos do Programa de Aceleração do Crescimento (PAC) em Minas Gerais. **PPE**, v. 39, n. 1, p. 121-58, 2009.

DUARTE, D.A.P.; CORRÊA, C.SL.; FAYER, V.A; NOGUEIRA, M.C.; BUSTAMANTE-TEIXEIRA, M.T. Avaliação de método para classificação automatizada de pares em relacionamentos probabilísticos de bancos de dados. **Cadernos de Saúde Pública (Online)**; v. 35; n. 11; e00066419, 2019.

EL-ZEIN, M.; RICHARDSON, L.; FRANCO, E.L. Cervical câncer screening of HPV vaccinated populations: Cytology, molecular testing, both or none. **Journal of Clinical Virology**, 2016.

FEDERAÇÃO BRASILEIRA DAS SOCIEDADES DE GINECOLOGIA E OBSTETRÍCIA (Febrasgo), Sociedade Brasileira de Cancerologia. **Projeto Diretrizes: Rastreamento, diagnóstico e tratamento do carcinoma do colo do útero**, 2001.

FERLAY, J.; SOERJOMATARAM, I.; DIKSHIT, R., et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. **International Journal of Cancer**, v. 136, n.5, p. E359-386, 2015.

FERNANDES, J.V.; RODRIGUES, S.H.L.; COSTA, Y.G.A.S et al. Conhecimentos, atitudes e prática do exame de Papanicolaou por mulheres, Nordeste do Brasil. **Revista Saúde Pública**, v. 43, n. 5, p. 851-8, 2009.

FISCHER, F.; LANGE, K.; KLOSE, K. et al. Barriers and Strategies in Guideline Implementation-A Scoping Review. **Healthcare**, v. 4, n. 3, 2016.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim n. 1, jan/mar 2010. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-1-2010.pdf>. Acesso: fev, 2021. INCA, 2010a.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim n. 2, abr/jun 2010. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-2-2010.pdf>. Acesso: fev., 2021. INCA, 2010b.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim ano 3, n. 3 agosto/dezembro 2012. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-3-2012.pdf>. Acesso: fev. 2021. INCA, 2012.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim ano 5, n. 1 janeiro/ abril 2014. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-1-2014.pdf>. Acesso em: fev, 2021. INCA, 2014a.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação Geral de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Ficha técnica de indicadores das ações de controle do câncer do colo do útero** [Internet]. Rio de Janeiro: INCA; 2014b. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//fichatecnicaindicadorescolo14.pdf>. Acesso: fev. 2021. INCA; 2014b

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio a Organização de Rede. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero – 2. ed. rev. atual.** – Rio de Janeiro: INCA, 2016.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Estimativa 2020: incidência de câncer no Brasil** / Instituto Nacional de Câncer José Alencar Gomes da Silva. – Rio de Janeiro: INCA, 2019a.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Parâmetros técnicos para o rastreamento do câncer do colo do útero** / Instituto Nacional de Câncer José Alencar Gomes da Silva; Maria Beatriz Kneipp Dias; Caroline Madalena Ribeiro (organizadores). - Rio de Janeiro: INCA, 2019b.

RICARDO-RODRIGUES, I.; JIMÉNEZ-GARCÍA,R.; HERNÁNDEZ-BARRERA,V.; CARRASCO-GARRIDO, P.; JIMÉNEZ-TRUJILLO, I.; LÓPEZ DE ANDRÉS, A. Social disparities in access to breast and cervical cancer screening by women living in Spain. **Public Health**, v. 129, p. 881-888, 2015.

SANKARANARAYANAN, R.; SOMANATHAN, T.; SANTOS, C.; MUWONGE, R.; TWALIB, N. O desafio de controlar o cancer do colo do útero no mundo em desenvolvimento. In: COELHO, F.R.G.; SOARES, F.A.; FOCCHI, J.; FREGNANI, J.H.T.G; ZEFERINO, L.C.; VILLA, L.L.; FEDERICO, M.H.; NOVAES, P.E.R.S; COSTA, R.L.R. **Câncer do colo do útero**. Tecmeddd, São Paulo – SP, 2008.

TEOH DGK, MARRIOTT AE, ISAKSSON VOGEL R, et al. Adherence to the 2012 national cervical cancer screening guidelines: a pilot study. **American Journal of Obstetrics & Gynecology**, v. 212; n. 62; e1-9, 2015.

OLIVEIRA, M.M.; ANDRADE, S.S.C.A.; OLIVEIRA, P.P.V., AZEVEDO E SILVA, G.; SILVA, M.M.A.; MALTA, D.C. Cobertura de exame Papanicolaou em mulheres de 25 a 64 anos, segundo a Pesquisa Nacional de Saúde e o Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico, 2013. **Revista Brasileira de Epidemiologia**, v. 21, e180014, 2018.

VALE, D.B.A.P.; MORAIS, S.S.; PIMENTA, A.L.; ZEFERINO, L.C. Avaliação do rastreamento do câncer do colo do útero na Estratégia Saúde da Família no Município de Amparo, São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 26, n. 2, p. 383-90, 2010.

WORLD HEALTH ORGANIZATION. **National cancer control programmes: policies and managerial guidelines**. 2.ed. Geneva: WHO, 2002.

WRIGHT, J.D.; CHEN, L.; TERGAS, A.I. *et al.* Overuse of Cervical Cancer Screening Tests Among Women With Average Risk in the United States From 2013 to 2014. **JAMA Network Open**, v. 4, n. 4, e218373, 2021.

6.3 – ARTIGO 3: DESIGUALDADES REGIONAIS E FATORES ASSOCIADOS NO RASTREIO DO CÂNCER DE COLO DE ÚTERO EM MINAS GERAIS

Resumo:

Objetivos: avaliar a desigualdade regional da cobertura do rastreamento do câncer do colo do útero, em 2008, em Minas Gerais e sua associação com indicadores socioeconômicos e assistenciais, e descrever a distribuição espacial do sobrerastreamento nas microrregiões do estado.

Métodos: estudo ecológico cujas unidades de análises foram as 77 microrregiões de saúde de Minas Gerais que utilizou dados de cobertura e sobrerastreamento baseados em exames registrados no SISCOLO. O cálculo da cobertura foi feito para mulheres de 25 a 64 anos, no ano de 2008. A distribuição espacial do sobrerastreamento foi analisada de forma separada para mulheres que iniciaram o rastreamento em 2008 (deveriam repetir o exame em um ano) e para aquelas que tinham exame anterior a 2008 (deveriam repetir o exame em três anos). O sobrerastreamento por faixa etária correspondeu ao percentual de mulheres <25 e ≥65 anos em cada grupo. O sobrerastreamento por periodicidade foi obtido por meio da razão entre mulheres de 25 a 64 anos que repetiram o exame antes de 9 ou 30 meses pelo total de mulheres na mesma faixa etária que estavam iniciando o rastreamento em 2008 ou que tinham exame prévio, respectivamente.

Resultados: Em 2008, a cobertura média estimada do rastreamento do câncer do colo do útero nas microrregiões de Minas Gerais foi de 74,5%. As microrregiões que apresentaram as menores coberturas estavam concentradas, em sua maioria, nas macrorregiões Nordeste, Jequitinhonha, Leste e Leste do Sul, que se caracterizam por piores indicadores socioeconômicos. O IDH apresentou associação positiva com a cobertura. O sobrerastreamento médio por faixa etária nas microrregiões de Minas Gerais foi de 54,5% em mulheres que iniciaram o rastreio em 2008 e 14,6% em mulheres que iniciaram anteriormente; e, por periodicidade foi de 3,7% e 64,2%, respectivamente. Os dois tipos de rastreamento apresentaram padrão diferenciado de distribuição espacial entre si e entre mulheres que estão iniciando o rastreamento e aquelas que têm exame prévio.

Conclusão: Os resultados indicam desigualdade no acesso ao rastreamento no estado de Minas Gerais, com microrregiões de maior IDH apresentando maior

cobertura. O sobrerastreamento nas microrregiões do estado é elevado indicando o não seguimento das diretrizes do programa nacional de rastreamento.

Palavras-chave: Neoplasias do Colo do Útero; Programas de Rastreamento; Excesso de exames; Análise espacial.

Abstract

Objectives: to assess the regional inequality of coverage of cervical cancer screening in 2008 in Minas Gerais and its association with socioeconomic and care indicators, and to describe the spatial distribution of over-screening in the state's micro-regions.

Methods: ecological study form analysis units were like 77 health micro-regions of Minas Gerais that use coverage and over-data based on exams registered with SISCOLO. Coverage calculation was made for women aged 25 to 64 years, in the year of 2008. The spatial distribution of the over-screening was analyzed separately for women who started the screening in 2008 (research repeating the exam in one year) and for those who took the exam before 2008 (research repeating the exam in three years). Over-screening by age group corresponds to the percentage of women <25 and ≥65 years old in each group. Over-screening due to excessive periodicity was established by the ratio between women aged 25 to 64 years who repeated the exam before 9 or 30 months by the total number of women in the same age group who were starting the screening in 2008 or who had a previous exam, respectively.

Results: In 2008, the estimated average coverage of cervical cancer screening in the microregions of Minas Gerais was 74.5%. The microregions with the lowest coverage were mostly concentrated in the Northeast, Jequitinhonha, East and East South regions, which are characterized by the worst socioeconomic indicators. The HDI showed a positive association with coverage. The average of over-screening by age group in the micro-regions of Minas Gerais was 54.5% in women who started screening in 2008 and 14.6% in women who started previously; and, due to excessive periodicity, it was 3.7% and 64.2%, respectively. The two types of screening showed a different pattern of spatial distribution among themselves and among women who are starting the screening and those who have had a previous examination.

Conclusion: The results indicate inequality in access to screening in the state of Minas Gerais, with microregions of higher HDI presenting greater coverage. Over-screening

in the state's micro-regions is high, indicating non-compliance with the guidelines of the national screening program.

Key words: Uterine Cervical Neoplasms; Mass Screening; Over-screening; Spatial Analysis.

INTRODUÇÃO

O câncer do colo do útero ainda constitui um importante problema de saúde pública no Brasil e no mundo, apesar de ser uma neoplasia evitável, com medidas de prevenção primária e secundária eficazes e bem estabelecidas (TSUCHIYA *et al.*, 2017). Por isso, o objetivo da OMS é eliminar a doença em todos os países, ou seja, atingir uma taxa de incidência inferior a quatro por 100.000 mulheres até 2030, baseado em metas de vacinação, rastreamento e tratamento (WHO, 2020).

A doença é o quarto tipo de câncer mais frequentemente diagnosticado e a quarta principal causa de morte por câncer em mulheres no mundo, com uma estimativa de 570.000 casos e 311.000 mortes em 2018 (BRAY *et al.*, 2018). No Brasil, dentre as principais neoplasias, com exceção do câncer de pele não-melanoma, o CCU é o terceiro tipo mais frequente nas mulheres, com um risco estimado de 15,43 casos a cada 100 mil mulheres para o triênio 2020-2022 (INCA, 2019). Ao longo dos últimos anos, houve redução nas taxas de incidência padronizadas por idade de 23,53 (IC 95% 22,79-24,26) por 100.000 mulheres em 2000 para 18,39 (IC 95% 17,63-19,17) por 100.000 mulheres em 2017. Os estados das regiões Sul e Sudeste apresentaram as menores taxas de incidência, enquanto os estados do Norte e Nordeste as maiores, atingindo o dobro ou o triplo da taxa nacional (REIS *et al.*, 2020).

Quanto à mortalidade pela doença, no período entre 1978 e 2017, observou-se declínio em todas as regiões do Brasil, exceto entre mulheres do interior da região Norte, onde as taxas em 2017 chegam a ser três vezes mais frequentes do que na região Sudeste (AZEVEDO E SILVA *et al.*, 2020). Desta forma, observa-se importante desigualdade regional na carga da doença no Brasil.

O rastreamento da doença por meio do exame de *Papanicolaou* (citopatológico do colo do útero) possibilita a detecção de lesões precursoras vários anos antes do surgimento da neoplasia do colo do útero (PEIRSON *et al.*, 2013) e, historicamente, foi considerada a principal estratégia de controle da doença no Brasil e no mundo (SANKARANARAYANAN *et al.*, 2008; LEES; ERICKSON; HUH, 2016). Estima-se

que, por meio de programas de rastreamento realizados com uma cobertura da população-alvo de 80 a 100% e acesso à confirmação diagnóstica e ao tratamento adequado, seja possível reduzir a incidência do CCU em até 90% (WHO, 2002).

No Brasil, o Ministério da Saúde, através do Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), recomenda a realização do *Papanicolaou* com intervalo trienal para mulheres de 25 a 64 anos após dois exames normais com intervalo anual (INCA, 2016). A meta do país é alcançar uma cobertura de 85% entre mulheres na faixa etária preconizada, até 2022 (BRASIL, 2011).

Dados da Pesquisa Nacional de Saúde (PNS) apontam que a cobertura do *Papanicolaou* em mulheres da faixa etária alvo do rastreamento no Brasil passou de 65,5%, em 2003, para 78,8%, em 2013. Os resultados da PNS mostraram associação positiva entre a cobertura e condições socioeconômicas (THEME-FILHA *et al.*, 2016) e diferenças regionais, com menores coberturas nas regiões Norte e Nordeste, apontando para as desigualdades no acesso e na realização do exame no país. (OLIVEIRA *et al.*, 2018a).

Ao contrário de países desenvolvidos, o rastreamento no Brasil ainda é realizado de forma oportunística, não havendo monitoramento e seguimento da população alvo e da periodicidade de realização dos exames. Este tipo de rastreamento é menos efetivo em relação ao controle da doença e é mais oneroso para o sistema de saúde (BRASIL, 2010), pois está relacionado com o sobrerastreamento. Nesse contexto, muitas mulheres não seguem as recomendações do Ministério da Saúde e, desta forma, fazem exames em excesso, enquanto outras permanecem sem acesso (AZEVEDO e SILVA *et al.*, 2015; VALE *et al.*, 2010).

Diante desse contexto, os objetivos do presente estudo são avaliar a desigualdade regional da cobertura do rastreamento do câncer do colo do útero em 2008 nas microrregiões de saúde de Minas Gerais e sua associação com indicadores socioeconômicos e assistenciais, e, ainda, descrever a distribuição espacial do sobrerastreamento por faixa etária e periodicidade nas microrregiões do estado.

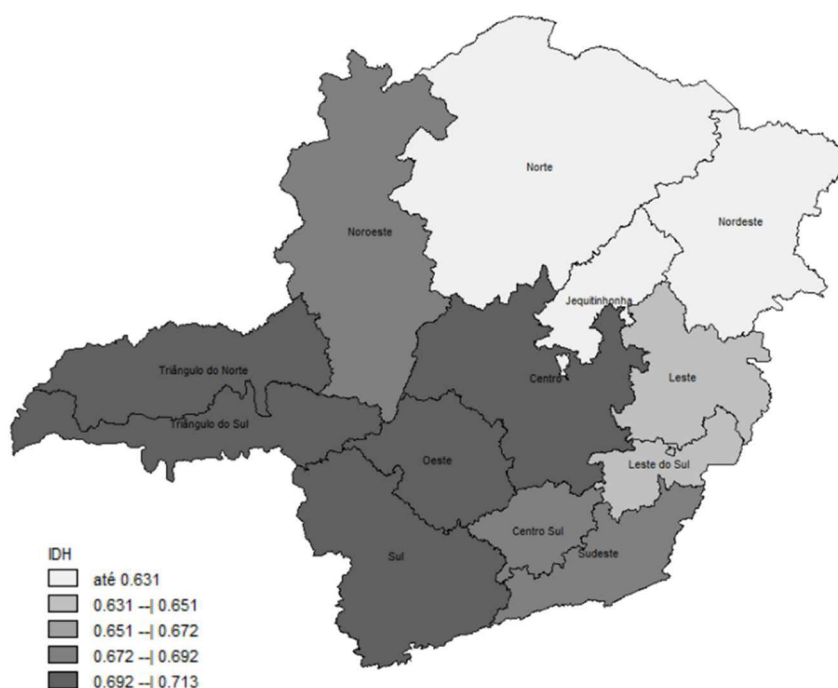
METODOLOGIA

Tipo de estudo e população

Trata-se de um estudo ecológico exploratório cujas unidades de análises foram as 77 microrregiões de saúde de Minas Gerais. Minas Gerais situa-se na região

Sudeste do Brasil e se caracteriza por grande disparidade, com regiões de menor nível socioeconômico no Nordeste, Norte e Jequitinhonha e maior, no Triângulo, Sul, Oeste e Centro. A figura 1 ilustra as 13 macrorregiões de saúde de acordo com o Índice de Desenvolvimento Humano (IDH), agrupados em quintis, referente ao ano de 2010.

Figura 1 – IDH por macrorregião de Saúde em Minas Gerais, 2010



Fonte: <http://datasus.saude.gov.br/informacoes-de-saude/tabnet/demograficas-e-socioeconomicas>

De acordo com dados do Censo, a população em 2010 era de 19.597.330 habitantes, sendo 50,8% (9.955.453) do sexo feminino e 5.189.853 mulheres (52,1%) encontravam-se na faixa etária de 25 a 64 anos (IBGE, 2010 - <https://datasus.saude.gov.br/populacao-residente>). O Índice de Desenvolvimento Humano (IDH) do estado em 2010 foi 0,731, ocupando a nona posição no país. A cobertura pela Estratégia Saúde da Família era 61,7% em 2008; 66,9% em 2010 e, em 2019, foi 79,7%, superior à do Brasil (48,4%; 51,3% e 63,5, respectivamente) (Fundação Oswaldo Cruz, 2010 PROADSS - <https://www.proadess.icict.fiocruz.br/index.php?pag=fic&cod=A26&tab=1>). Em 2008, 24,4% das mulheres de 25 a 64 anos em Minas Gerais tinham plano de saúde e, em 2020, a estimativa era de que 25,3% das mulheres nessa faixa etária estavam

cobertas por plano privado de saúde (ANS, 2020 09/10/2020 - http://www.ans.gov.br/anstabnet/cgi-bin/tabnet?dados/tabnet_tx.def).

Variáveis de estudo

Variáveis de desfecho:

- cobertura do rastreamento do câncer do colo do útero no estado de Minas Gerais e suas microrregiões de saúde, no ano de 2008;
- sobrerastreamento do câncer do colo do útero quanto à faixa etária e quanto à periodicidade no estado e suas microrregiões de saúde, considerando mulheres submetidas ao rastreamento em 2008.

O cálculo dessas variáveis foi feito utilizando bases identificadas do SISCOLO, que foram cedidas pela Secretaria Estadual de Saúde (SES-MG).

Para o cálculo da cobertura, utilizou-se a base de 2008 do SISCOLO e foram excluídos registros duplicados e de seguimento, ou seja, repetição do exame dentro do mesmo ano pela mesma mulher, visando manter apenas um registro de exame (o primeiro) por mulher, no ano de 2008. Então, dividiu-se o número de mulheres submetidas ao rastreamento na faixa etária alvo por 1/3 da população feminina que utilizava exclusivamente o SUS, na mesma faixa etária, no ano de 2008, por microrregião de saúde de Minas Gerais. Considerou-se a faixa etária de 25 a 59 anos, pois era a recomendação do Ministério da Saúde em 2008 (INCA, 2006) e foi calculado também para a faixa de 25 a 64, pois é a atual recomendação (INCA, 2016).

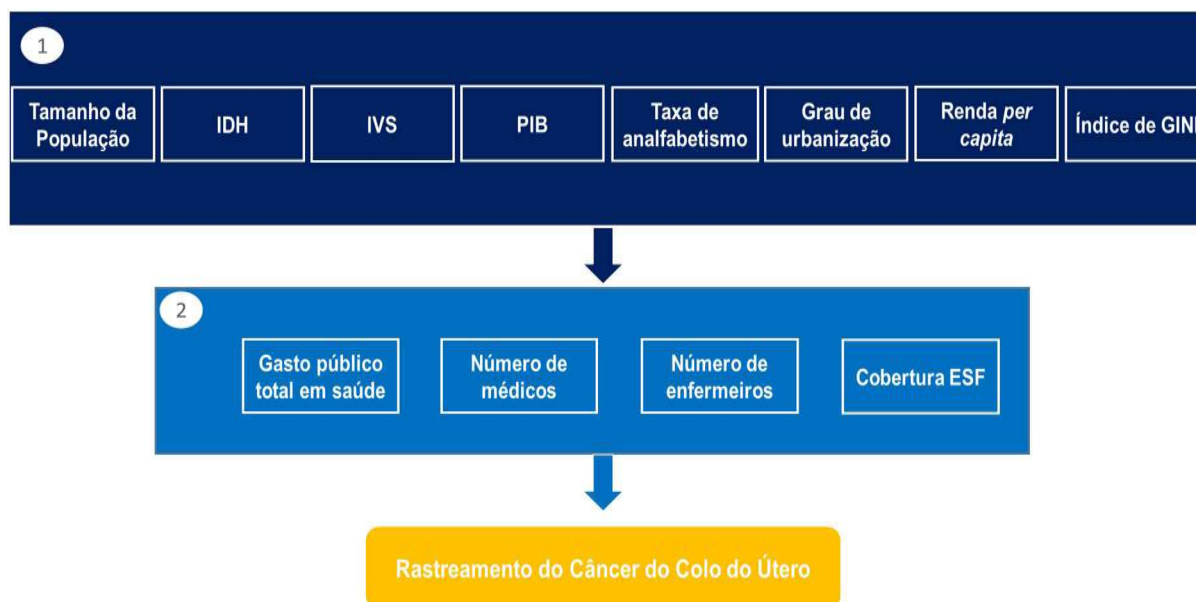
Para avaliar o sobrerastreamento por periodicidade, foi montado o seguimento das mulheres de 25 a 64 anos submetidas ao rastreamento em 2008 através do relacionamento probabilístico entre o banco contendo esses exames e cada banco anual de 2009 a 2012. Visando evitar a superestimação do indicador, foram excluídas mulheres que tinham exame alterado ou insatisfatório no ano anterior (2007) e em 2008. Para mulheres que estavam iniciando o rastreamento em 2008, o sobrerastreamento por periodicidade foi considerado como a repetição do exame com menos de 9 meses. E, para aquelas que relataram exame anterior, a repetição em um período inferior a 30 meses. Para cada situação, a proporção de sobrerastreamento por periodicidade foi obtida dividindo o número de mulheres que repetiram o exame antes do período preconizado pelo total de mulheres submetidas ao rastreamento, por microrregião de residência.

O sobrerastreamento por faixa etária, em cada microrregião de saúde, correspondeu ao percentual de mulheres com menos de 25 anos ou mais de 64 no grupo de mulheres que iniciaram o rastreamento em 2008 e naquelas que iniciaram antes de 2008.

Variáveis explicativas:

Foram compostas por indicadores sociodemográficos, econômicos e assistenciais baseados na microrregião de residência (coletados em abril e maio de 2020) e separadas em dois blocos hierárquicos para orientar a análise de regressão múltipla posteriormente (figura 2). Tais variáveis foram selecionadas de acordo com a revisão de literature e a disponibilidade em bases de acesso público visando a um estudo exploratório sobre a associação entre a cobertura do rastreamento e características das microrregiões de saúde de Minas Gerais.

Figura 2: Variáveis explicativas da cobertura do rastreamento separadas em dois blocos hierárquicos (distal e proximal)



Legenda: 1) bloco distal: variáveis sociodemográficas e econômicas; 2) bloco proximal: variáveis de financiamento e assistenciais

O quadro 1 descreve as variáveis explicativas e seus respectivos cálculos e fontes.

Quadro 1 – Variáveis explicativas: indicadores sociodemográficos, econômicos e assistenciais, ano de referência e fonte dos dados

Variáveis	Indicador, ano de referência	Fonte Descrição	Cálculo
População	Tamanho da população, 2010	População residente, IBGE - DATASUS http://tabnet.datasus.gov.br/cgi/defothm.exe?lbge/cnv/popmg.def	Dado bruto, extraído por município
IDH	Índice de desenvolvimento humano, 2010	Proades https://www.proadess.icict.fiocruz.br/index.php?pag=fic_r&cod=L14&tab=1	Média geométrica das dimensões que compõem o IDH: longevidade, educação e renda. Para o cálculo do IDH das microrregiões, foram agregados os IDH municipais e realizada uma média simples.
RENDA	Renda média domiciliar per capita, 2010	Proades https://www.proadess.icict.fiocruz.br/index.php?pag=fic_r&cod=L09&tab=1	Numerador: soma dos rendimentos mensais dos moradores dos domicílios. Denominador: total de moradores residentes nos domicílios.
PIB	Produto interno bruto per capita, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/tabcgi.exe?lbge/cnv/pibmumbmg	Indicador baixado por município e agregado por meio de média simples para o cálculo do PIB por microrregião.
IVS	Índice de Vulnerabilidade na Saúde, 2010	IBGE - Dados solicitados ao IBGE via portal de acesso à informação do governo federal (2015) Indicador composto pelas seguintes variáveis: <ul style="list-style-type: none"> • Dimensão Saneamento: (1) Percentual de domicílios particulares permanentes com abastecimento de água inadequado ou ausente; (2) Percentual de domicílios particulares permanentes com esgotamento sanitário inadequado ou ausente; (3) Percentual de domicílios particulares permanentes com destino do lixo de forma inadequada ou ausente; • Dimensão Socioeconômica: (1) Razão de moradores por domicílio; (2) Percentual de pessoas analfabetas; (3) Percentual de domicílios particulares com rendimento per capita até ½ SM; (4) Rendimento nominal mensal médio das pessoas responsáveis (invertido); (5) Percentual de pessoas de raça/cor preta, parda e indígena 	Realizado conforme: PITCHON, A.; <i>et al.</i> Minas Gerais. Índice de Vulnerabilidade Social (IVS) 2012. Belo Horizonte, 2013.15 p.
GINI	Índice de Gini, 2010	Atlas do desenvolvimento humano no Brasil http://www.atlasbrasil.org.br/2013/pt/consulta/ Mede o grau de desigualdade existente na distribuição de indivíduos segundo a renda domiciliar per capita. Seu valor é 0 quando não há	Para o cálculo do índice de Gini das microrregiões, foram agregados os índices municipais e realizada média simples.

		desigualdade (a renda domiciliar per capita de todos os indivíduos têm o mesmo valor) e tende a 1 à medida que a desigualdade aumenta. O universo de indivíduos é limitado àquelas que vivem em domicílios particulares permanentes.		
GRAU URBANIZAÇÃO	Grau de urbanização, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/tabcgi.exe?ibge/cnv/popmg.def % população urbana		(População urbana/População total) *100
ANALFABETISMO	Percentual da população com ≥ 15 anos analfabeta em 2010	Proades https://www.proadess.iciict.fiocruz.br/index.php?pag=fic_r&cod=L51&tab=1		Numerador: número de pessoas residentes de 15 anos ou mais de idade que não sabem ler e escrever um bilhete simples, no idioma que conhecem X 100. Denominador: população total residente de 15 anos ou mais de idade.
GPTS	Gasto Público Total em Saúde/habitante, 2008	Proades https://www.proadess.iciict.fiocruz.br/index.php?pag=fic_r&cod=K51&tab=1 Gasto em saúde, por habitante, advindo de todas as fontes, quer sejam impostos, transferências SUS de outras esferas governamentais, operações de créditos, e outras.		Numerador: gasto total em saúde (todas as fontes de financiamento). Denominador: população total residente.
MEDICOS	Nº Médicos/ 1000 hab, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def		(Número de médicos/habitantes)*1000
ENFERMEIROS	Nº Enfermeiros/ 1000 hab, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def		(Número de enfermeiros/habitantes)*1000
ESF	Percentual da população coberta pela Estratégia Saúde da Família, 2008	Proades https://www.proadess.iciict.fiocruz.br/index.php?pag=fic_r&cod=A26&tab=1		Numerador: número de pessoas que residem em domicílios cadastrados na Estratégia Saúde da Família X 100. Denominador: população total residente

Análise dos dados

Foram realizadas análises espaciais descritivas por meio de mapas temáticos, com tons mais claros para regiões com os menores valores e mais escuros para os maiores. Para isso, foram criadas categorias para cada variável utilizando 0,5 e 1,0 desvio-padrão abaixo e acima da média, a depender da distribuição e variabilidade de cada indicador.

Além disso, foram feitas tabelas descritivas dos resultados e calculadas medidas de autocorrelação espacial global (Índice de Moran Global) para investigar a dependência espacial. A autocorrelação espacial local da cobertura e do sobrerastreamento foi avaliada por meio do Índice de Moran Local ou LISA, que permite comparar o valor de cada microrregião com seus vizinhos, objetivando identificar o agrupamento de regiões com dependência espacial significativa.

Os coeficientes de correlação de Pearson (r) foram estimados para analisar a correlação entre as variáveis explicativas da cobertura, com o objetivo de definir aquelas que seriam incluídas nos modelos múltiplos de regressão. A força da correlação foi classificada da seguinte forma: (1) correlação fraca: r entre 0,10 e 0,39; (2) correlação moderada: r entre 0,40 e 0,69; (3) correlação forte: r entre 0,70 e 1,00 (DANCEY; REIDY, 2006).

Posteriormente, para verificar a associação entre a cobertura do rastreamento do câncer do colo do útero com os indicadores sociodemográficos, econômicos e assistenciais, foram utilizados modelos de regressão linear simples e múltipla. Nos modelos de regressão múltipla, foi seguida abordagem hierarquizada dos blocos de variáveis explicativas, começando com o bloco distal e, posteriormente, inserindo as variáveis do bloco proximal – figura 2 - (VICTORA *et al.*, 1997). Para a seleção de variáveis do bloco distal, foram feitas combinações entre aquelas que não apresentaram correlação forte entre si avaliada por meio da correlação de *Pearson*, para evitar a multicolinearidade. Em seguida, foram incluídas cada variável proximal nos modelos contendo as variáveis distais que apresentaram significância estatística. Para os modelos cujas variáveis explicativas foram estatisticamente significativas, foram testados os pressupostos de independência espacial (Teste de *Moran Global*), normalidade (Teste de *Jarque-Bera*) e homocedasticidade (Teste de *Breusch-Pagan*) dos resíduos. O modelo final foi escolhido com base nos seguintes

critérios: significância estatística das variáveis explicativas; maior poder explicativo (maior R^2), menor número de variáveis para explicar o desfecho e atender a todos os pressupostos da regressão linear. O nível de significância adotado no estudo foi de 0,05.

As análises foram realizadas utilizando o programa R, versão 4.0.2.

RESULTADOS

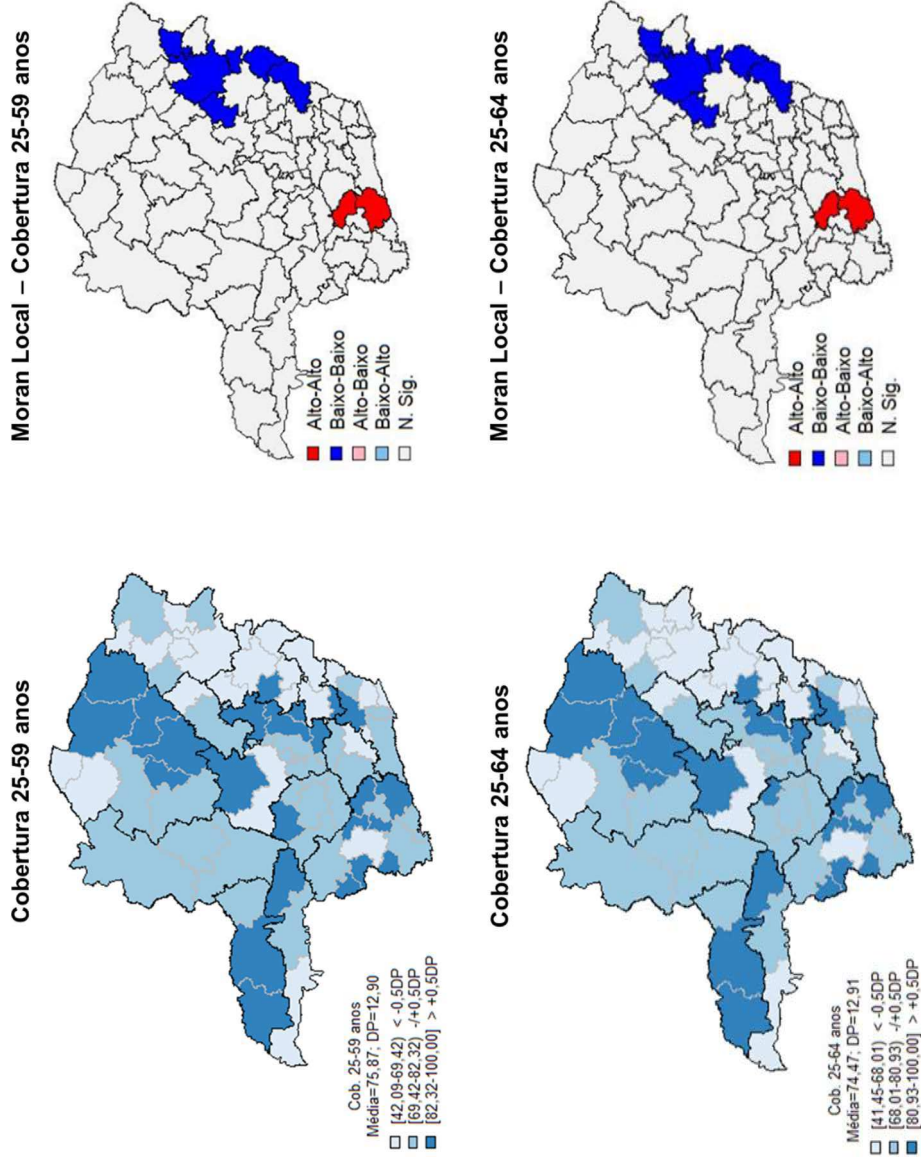
A cobertura média do rastreamento do CCU pelo SUS nas microrregiões de Saúde de MG em 2008 foi de 75,87% em mulheres de 25 a 59 anos e de 74,47% em mulheres de 25 a 64 anos. A distribuição espacial da cobertura por microrregião de saúde para estas faixas etárias e o respectivo Moran Local são apresentados na figura 3.

Ao comparar a distribuição da cobertura entre as duas faixas etárias observou-se grande semelhança, desta forma, optou-se por descrever os resultados para a última, já que é a recomendação mais recente do Ministério da Saúde.

Como pode ser observado na figura 3, há grande variabilidade da cobertura do rastreamento no estado, com coberturas variando de 41,5 a 100%. As menores coberturas estão concentradas, em sua maioria, nas macrorregiões Nordeste, Jequitinhonha, Leste e Leste do Sul e, as maiores, no Triângulo do Norte, Norte, Sul e Centro. Nas microrregiões Araxá (Triângulo do Sul), Itabira (Centro), Lavras (Sul) e Poços de Caldas (Sul) a cobertura foi máxima (100%). A maioria das microrregiões de saúde teve cobertura inferior a 80%; 33,8% das microrregiões teve cobertura inferior a 70% e apenas 36,3% alcançou cobertura igual ou superior a 80% na faixa etária preconizada para o rastreamento.

Quanto ao Índice de Moran Local (figura 3), observa-se que microrregiões do Nordeste (duas), Leste (três) e Leste do Sul (uma) apresentam baixas coberturas e são rodeadas por microrregiões que também têm baixa cobertura e duas microrregiões da macro Sul apresentam cobertura alta e são rodeadas por microrregiões que também têm alta cobertura. O restante das regiões não apresentou autocorrelação espacial local.

Figura 3 – Distribuição espacial da cobertura do rastreamento do câncer do colo do útero nas microrregiões de saúde de Minas Gerais em mulheres de 25 a 59 anos e 25 a 64 anos e respectivo Moran Local, 2008.



A tabela 1 mostra a distribuição da cobertura em mulheres de 25 a 64 anos, por microrregião de saúde nas macrorregiões do estado. Observa-se grande variabilidade no Triângulo do Sul (DP=29,3), o qual possui uma microrregião com a menor cobertura do estado (41,5) e outra, com a cobertura máxima (100). Por outro lado, o Noroeste apresenta a menor variabilidade, com cobertura mínima de 73,8 e máxima de 77,5.

Tabela 1 - Distribuição da cobertura do rastreamento do câncer do colo do útero em mulheres de 25 a 64 anos, por microrregião de saúde de Minas Gerais, para cada macrorregião e para o estado, 2008.

Macrorregião	Nº micros	Média	DP	Min	Q1	Mediana	Q3	Max
Centro	10	81,6	9,4	67,6	76,4	79,2	86,2	100,0
Centro Sul	3	68,2	8,4	58,5	65,7	72,9	73,1	73,3
Jequitinhonha	2	66,8	5,9	62,6	64,7	66,8	68,9	71,0
Leste	7	64,6	13,2	47,6	59,8	63,5	64,9	91,3
Leste do Sul	3	69,7	12,5	62,4	62,5	62,5	73,3	84,2
Nordeste	8	63,2	9,8	45,6	59,4	63,2	68,5	78,1
Noroeste	3	75,4	1,9	73,8	74,4	75,0	76,2	77,5
Norte	9	76,7	12,6	49,5	72,4	82,9	83,6	90,5
Oeste	6	78,0	4,8	68,6	78,1	80,2	80,6	81,2
Sudeste	8	69,5	11,4	54,4	62,2	69,2	74,3	90,6
Sul	12	84,2	11,2	64,6	76,0	84,1	94,1	100,0
Triângulo do Norte	3	83,4	3,0	80,5	81,8	83,1	84,8	86,5
Triângulo do Sul	3	70,9	29,3	41,5	56,4	71,3	85,6	100,0
Total – MG	77	74,5	12,9	41,5	64,6	75,8	83,1	100,0

Legenda: DP=desvio padrão; min: valor mínimo; Q1: primeiro quartil; Q3: terceiro quartil; max: valor máximo.

O sobrerastreamento médio por periodicidade nas microrregiões de saúde de MG em mulheres que iniciaram o rastreamento em 2008 foi de 3,7%, e naquelas que já tinham feito o exame antes, foi de 64,2%. No primeiro grupo, 68,8% das microrregiões tiveram esse percentual abaixo de 4%, e, no segundo, em 83,1% das microrregiões o sobrerastreamento foi igual ou superior a 60%.

O sobrerastreamento médio por faixa etária nas microrregiões de saúde de MG em mulheres que iniciaram o rastreamento em 2008 foi de 54,5% e para mulheres que iniciaram o rastreamento antes de 2008, foi de 14,6%.

A distribuição espacial do sobrerastreamento nas microrregiões de saúde de Minas Gerais, para mulheres que iniciaram o rastreamento em 2008 e para aquelas que iniciaram previamente é apresentada na figura 4. Para as primeiras, quanto ao sobrerastreamento por periodicidade, observa-se que os maiores valores estão concentrados em sua maioria nas macrorregiões do Triângulo e Sul e em relação ao sobrerastreamento por faixa etária, observa-se maior concentração de maiores valores na região Sul e menores valores na macrorregião Noroeste. Para mulheres que iniciaram o rastreamento antes de 2008, quanto ao sobrerastreamento por periodicidade, observa-se que os maiores valores estão concentrados em sua maioria no Norte, Nordeste e Oeste e os menores, na região Leste. Para o sobrerastreamento por faixa etária, os maiores valores estão concentrados no Triângulo.

A figura 5 ilustra o Moran Local para o sobrerastreamento. Para mulheres que iniciaram o rastreamento em 2008, em relação ao sobrerastreamento por periodicidade, observa-se que microrregiões do Triângulo do Sul (três), Triângulo do Norte (uma) e Sul (uma) apresentam altas proporções e são rodeadas por microrregiões que também apresentam altos valores. E, microrregiões do Norte (uma) e do Leste (uma) apresentam baixos valores e são rodeadas por microrregiões que também possuem baixos valores. Quanto ao sobrerastreamento por faixa etária, a autocorrelação espacial local foi significativa alto-alto em três microrregiões no Sul e baixo-baixo em três microrregiões do Noroeste e duas no Norte. Para mulheres que iniciaram o rastreamento antes de 2008, quanto ao sobrerastreamento por periodicidade, microrregiões do Sul (uma), Oeste (uma), Norte (três) e Nordeste (três) apresentaram autocorrelação espacial local alto-alto e microrregiões do Leste (duas) e Centro (duas), baixo-baixo. E para o sobrerastreamento por faixa

etária, microrregiões do Triângulo do Sul (três), Triângulo do Norte (duas) e Sudeste (uma) apresentaram autocorrelação espacial local alto-alto e microrregiões do Norte (duas) e Leste (duas), baixo-baixo.

Figura 4 – Distribuição espacial do sobrerastreamento nas microrregiões de saúde de Minas Gerais, para mulheres que iniciaram o rastreamento em 2008 e para aquelas que iniciaram previamente.

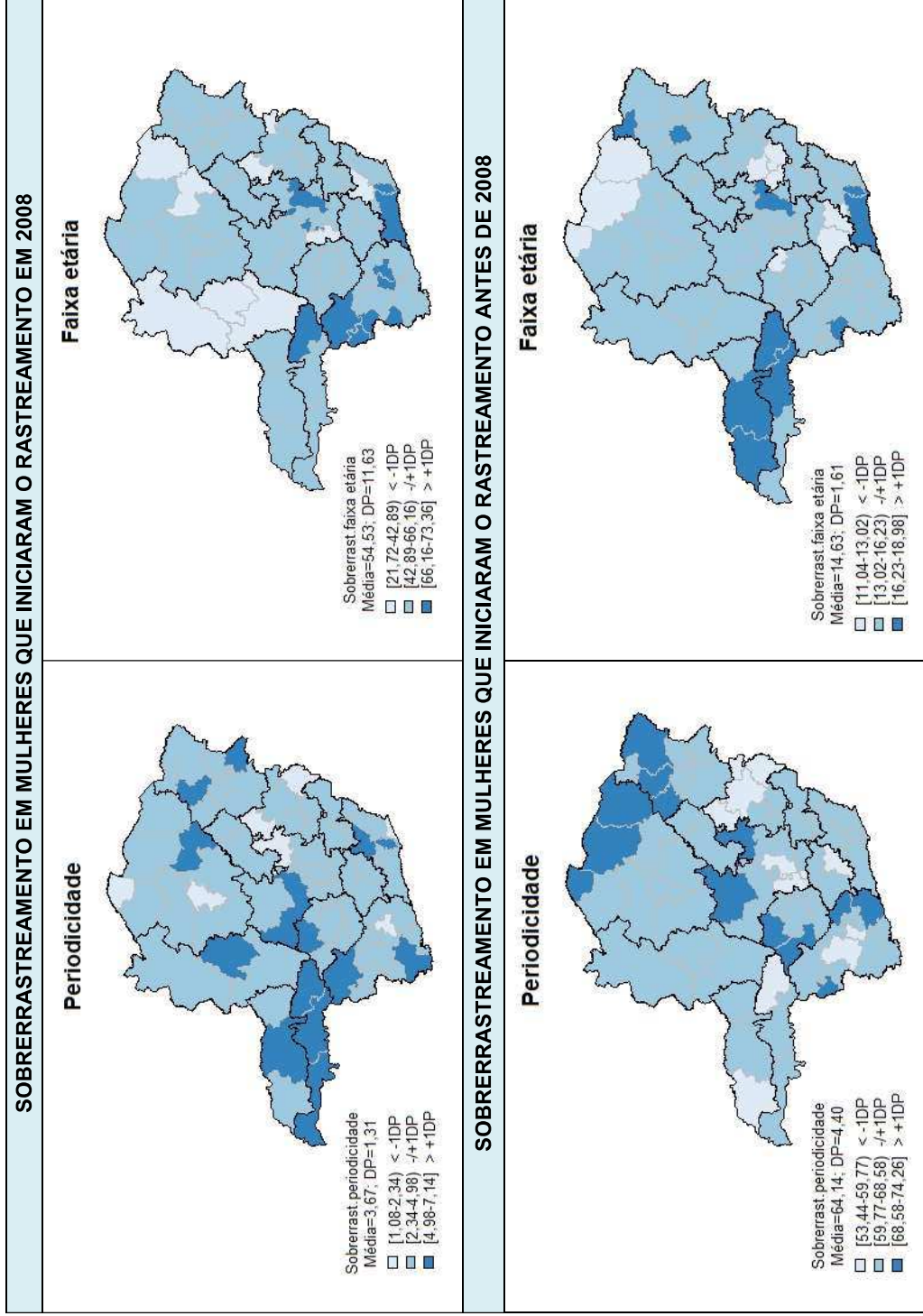
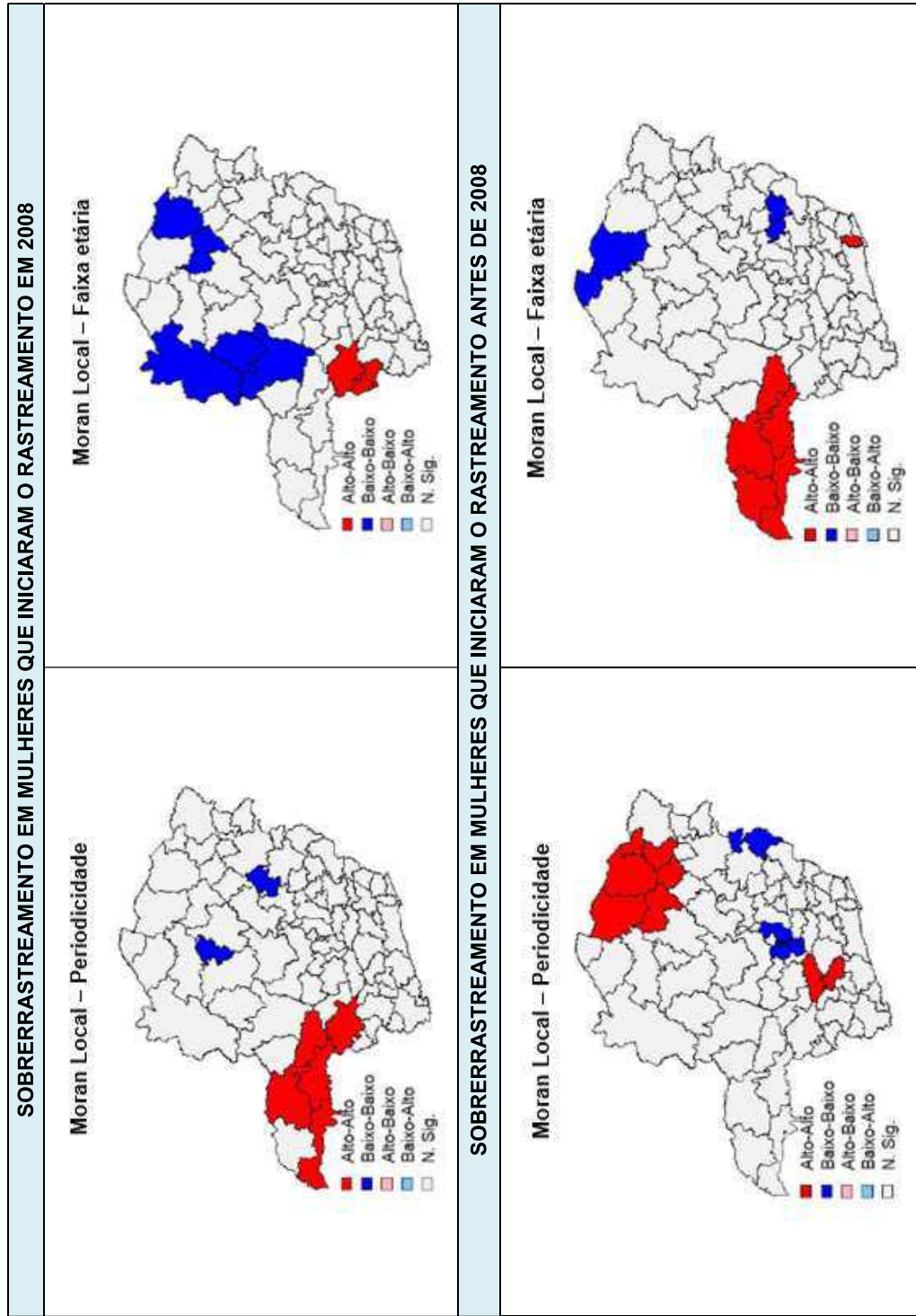


Figura 5 – Moran Local do sobrerastreamento nas microrregiões de saúde de Minas Gerais, para mulheres que iniciaram o rastreamento em 2008 e para aquelas que iniciaram previamente.



Como o sobrerastreamento por faixa etária foi mais importante nas mulheres que estavam iniciando o rastreio em 2008 e o sobrerastreamento por periodicidade foi maior em mulheres sem exame prévio, optou-se por apresentar a distribuição apenas destas situações nas macrorregiões do estado (tabelas 2 e 3, respectivamente).

Observa-se na tabela 2, que a macrorregião Noroeste apresentou a menor média de sobrerastreamento por faixa etária (25,4%), enquanto a macro Sul, a maior (64,5%). A microrregião com menor sobrerastreamento do estado localiza-se na macro Norte (21,7%) e aquela com o maior percentual, pertence à macrorregião Sudeste (73,4%).

Tabela 2 – Distribuição do sobrerastreamento por faixa etária em mulheres que iniciaram o rastreio em 2008, por microrregião de saúde de Minas Gerais, para cada macrorregião e para o estado.

Macrorregião	Média	DP	Min	Q1	Mediana	Q3	Max
Centro	56,8	10,8	31,0	53,6	56,9	65,2	67,7
Centro Sul	52,9	3,4	49,6	51,2	52,7	54,5	56,3
Jequitinhonha	54,2	0,1	54,1	54,2	54,2	54,3	54,3
Leste	49,6	5,7	42,1	44,8	51,4	54,1	55,9
Leste do Sul	51,9	5,1	48,6	49,0	49,4	53,6	57,7
Nordeste	49,7	5,2	45,3	46,1	47,4	52,1	58,5
Noroeste	25,4	1,3	24,3	24,6	25,0	25,9	26,8
Norte	47,6	14,1	21,7	48,7	53,9	54,6	61,0
Oeste	59,3	3,0	56,5	57,0	58,3	61,0	64,0
Sudeste	58,6	14,2	28,5	55,7	61,3	64,6	73,4
Sul	64,5	5,7	55,0	60,2	65,9	68,3	73,1
Triângulo do Norte	60,1	6,6	52,8	57,3	61,8	63,8	65,7
Triângulo do Sul	60,0	10,5	49,2	54,9	60,5	65,4	70,2
Total -MG	54,5	11,6	21,7	49,6	56,0	61,8	73,4

Legenda: DP=desvio padrão; min: valor mínimo; Q1: primeiro quartil; Q3: terceiro quartil; max: valor máximo.

Na tabela 3, verifica-se que o sobrerastreamento por periodicidade em mulheres que iniciaram o rastreio antes de 2008 apresenta pequena variabilidade no estado. A macrorregião Triângulo do Sul apresentou a menor média nas microrregiões de sobrerastreamento por periodicidade (60,3%),

enquanto a macro Oeste, a maior (67,5%). A microrregião com menor sobrerastreamento do estado localiza-se na macro Centro (53,4%) e aquela com o maior percentual, pertence à macrorregião Norte (74,3%).

Tabela 3 – Distribuição do sobrerastreamento por periodicidade em mulheres que iniciaram o rastreio antes de 2008, por microrregião de saúde de Minas Gerais, para cada macrorregião e para o estado.

MACRO	Media	DP	Min	Q1	Mediana	Q3	Max
Centro	63,2	6,0	53,4	59,6	61,7	67,0	72,2
Centro Sul	61,9	4,3	59,1	59,5	59,9	63,4	66,8
Jequitinhonha	65,6	1,0	64,9	65,3	65,6	66,0	66,3
Leste	61,3	4,5	56,4	58,4	59,1	64,8	67,4
Leste do Sul	62,3	0,6	61,7	62,0	62,3	62,6	62,9
Nordeste	66,8	3,2	62,4	64,7	67,2	69,3	71,0
Noroeste	62,5	1,4	60,8	62,0	63,1	63,3	63,4
Norte	66,9	4,1	61,3	63,4	67,5	69,0	74,3
Oeste	67,5	2,5	63,7	66,0	67,8	69,6	70,2
Sudeste	62,3	2,2	60,1	61,1	61,5	63,3	66,7
Sul	64,8	4,8	55,5	61,8	65,6	68,2	72,5
Triângulo do Norte	63,6	4,9	58,5	61,2	63,9	66,1	68,3
Triângulo do Sul	60,3	4,5	55,3	58,3	61,2	62,7	64,2
Total - MG	64,2	4,4	53,4	61,3	63,7	67,4	74,3

Legenda: DP=desvio padrão; min: valor mínimo; Q1: primeiro quartil; Q3: terceiro quartil; max: valor máximo.

Em relação às variáveis socioeconômicas (figura 6), observa-se padrão de distribuição espacial semelhante entre IDH, renda, grau de urbanização, IVS e analfabetismo, sendo a distribuição dessas duas últimas, o inverso das primeiras. Há grande variabilidade socioeconômica entre as microrregiões de Minas Gerais, sendo que, de forma geral, aquelas localizadas no Nordeste, Norte e Jequitinhonha apresentam os piores indicadores do estado. E microrregiões do Triângulo, Sul, Centro, Oeste e Sudeste, os melhores. No Triângulo do Sul e do Norte, o grau de urbanização médio das microrregiões é em torno de 90%, enquanto no Jequitinhonha, Norte e Nordeste, é cerca de 63%. A média de analfabetismo nas microrregiões do Nordeste é de 22,6%, do Norte 17,7% e do

Jequitinhonha, 17,2%, enquanto naquelas no Centro-Sul, Oeste e Triângulo do Sul é em torno de 6%.

Quanto às variáveis de financiamento e assistenciais (figura 7), em média, as microrregiões do Jequitinhonha; Leste do Sul e Norte apresentam os maiores valores de cobertura pela ESF; e Triângulo do Norte; Sul e Triângulo do Sul, as menores coberturas. No estado, as coberturas mínimas (cerca de 30%) estão localizadas na macrorregião Sul; e as máximas, acima de 96%, nas macrorregiões Norte e Nordeste. O número de médicos por 1.000 habitantes apresenta menores valores ao norte e nordeste e maiores valores ao sul, sudeste e triângulo. Na macro Norte, com exceção da região de Montes Claros/Bocaiuva, que tem 1,53 médicos/1.000 habitantes, todas as outras microrregiões têm entre 0,31 e 0,62 médicos/1.000 habitantes.

Figura 6 – Distribuição espacial das variáveis explicativas do bloco distal (socioeconômicas)

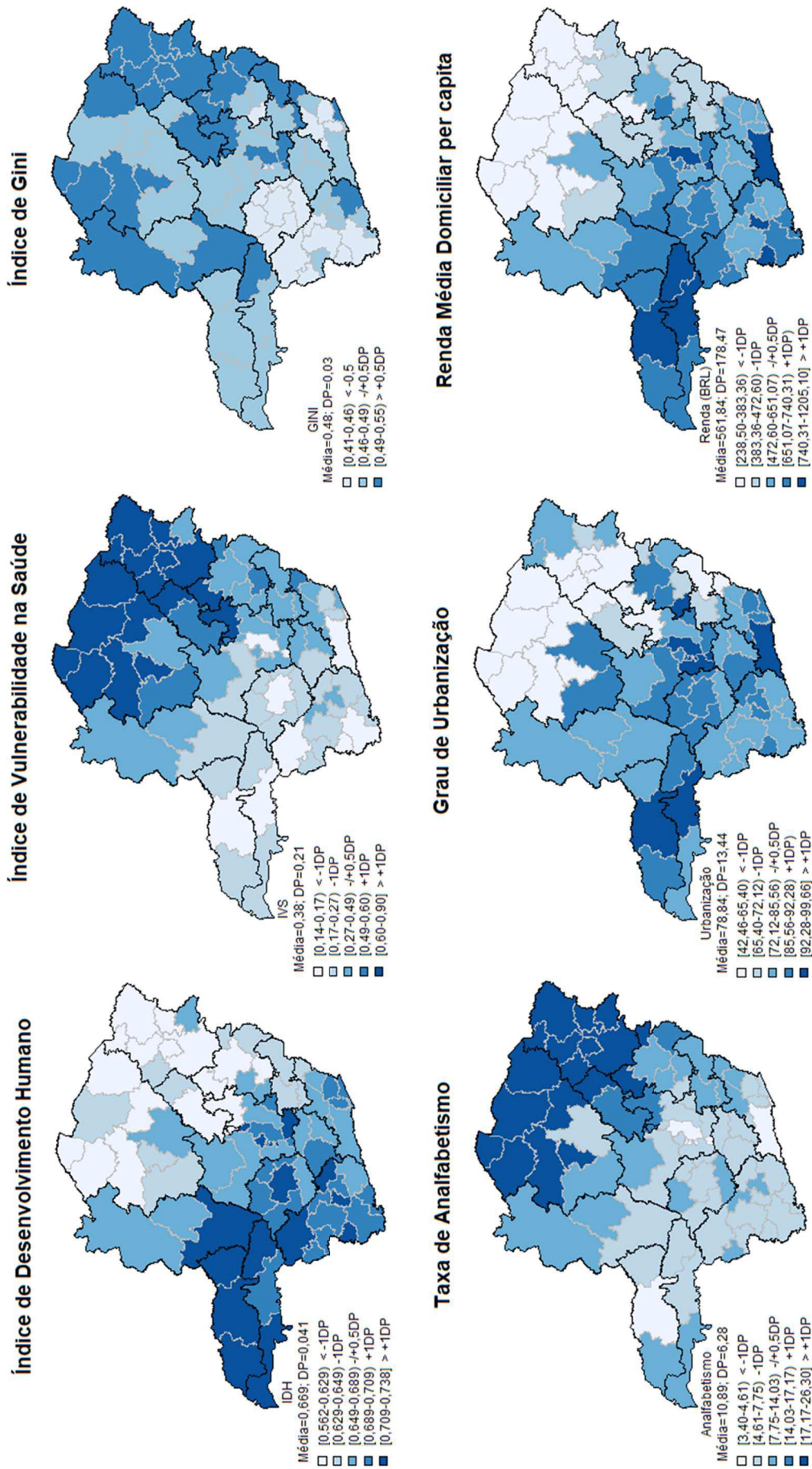
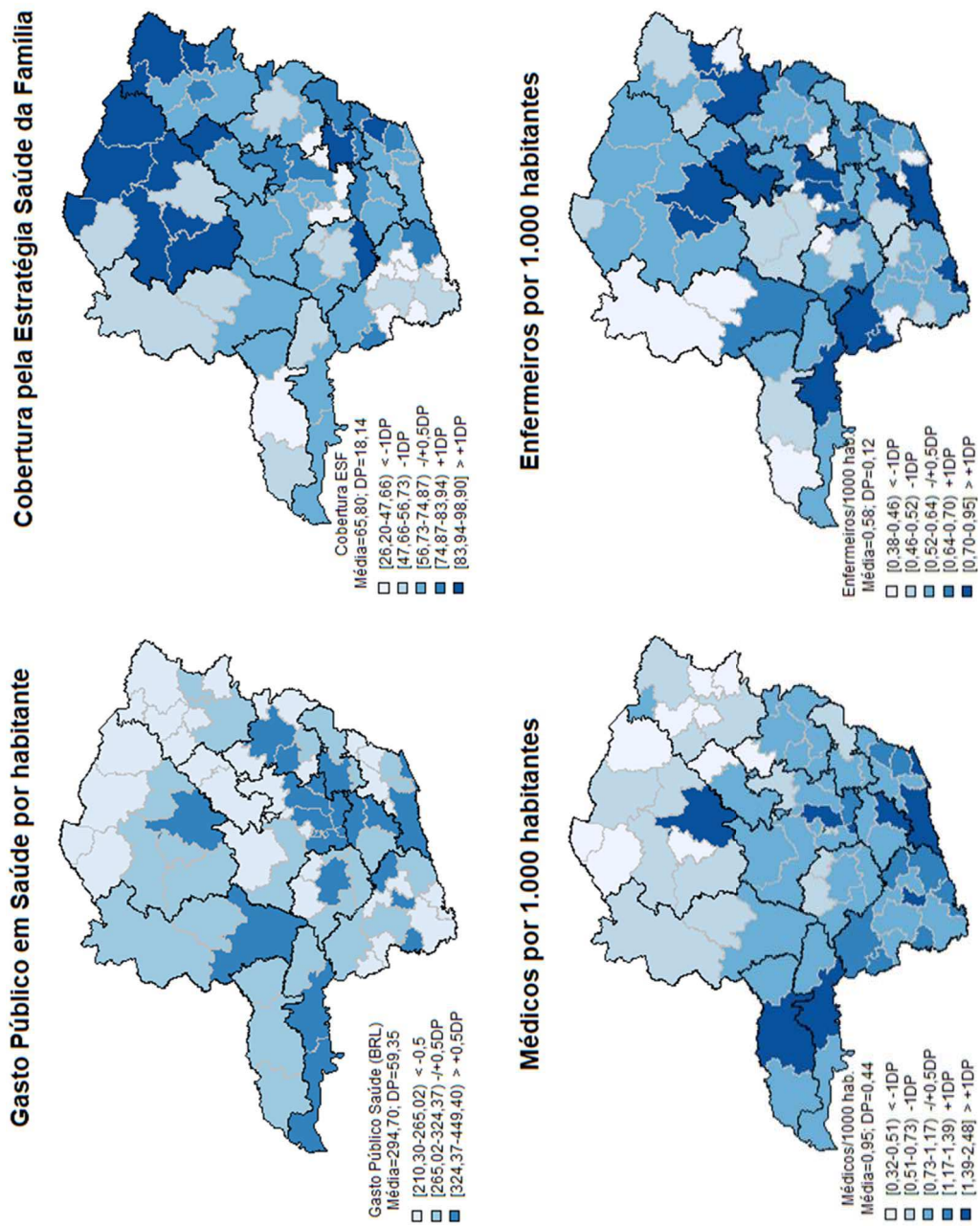


Figura 7 – Distribuição espacial das variáveis explicativas do bloco proximal (financiamento e assistenciais)



Todas as variáveis estudadas apresentaram autocorrelação espacial positiva significativa, exceto o número de enfermeiros por 1.000 habitantes, indicando a dependência espacial entre as microrregiões de saúde. Analfabetismo, IVS e IDH foram as variáveis que apresentaram os maiores valores de Índice de Moran (tabela 4).

Tabela 4 – Autocorrelação espacial (Índice de Moran Global)

Variável	Moran Global	p-valor
Cobertura – faixa etária 25 a 59 anos	0,15	0,011
Cobertura – faixa etária 25 a 64 anos	0,16	0,007
Sobrerrastreamento - início do rastreamento em 2008		
Faixa etária	0,26	<0,001
Periodicidade	0,13	0,020
Sobrerrastreamento – início do rastreamento antes de 2008		
Faixa etária	0,24	<0,001
Periodicidade	0,24	<0,001
População	0,13	<0,001
IDH	0,72	<0,001
IVS	0,78	<0,001
PIB <i>per capita</i>	0,41	<0,001
Analfabetismo	0,80	<0,001
Grau de urbanização	0,59	<0,001
Renda	0,62	<0,001
Índice de Gini	0,54	<0,001
Gasto público total em saúde	0,21	<0,001
Cobertura Estratégia Saúde da Família	0,26	<0,001
Médicos por 1.000 habitantes	0,32	<0,001
Enfermeiros por 1.000 habitantes	-0,12	0,936

A correlação de Pearson entre as variáveis explicativas e a cobertura do rastreamento está apresentada na tabela 5. Os coeficientes de Person (r) estatisticamente significativos foram destacados em negrito.

Tabela 5 – Correlação de Pearson entre as variáveis explicativas e a cobertura do rastreamento

	POP	IDH	IVS	PIB	ANALF	URBAN	RENDA	GINI	GPTS	ESF	MED	ENF	Cob 25-59	Cob 25-64
POP	1,00	0,23	-0,3	0,15	-0,31	0,33	0,57	-0,08	0,44	-0,16	0,49	0,31	0,06	0,07
IDH	0,23	1,00	-0,91	0,65	-0,89	0,83	0,83	-0,64	0,38	-0,57	0,50	-0,09	0,37	0,40
IVS	-0,30	-0,91	1,00	-0,52	0,94	-0,88	-0,90	0,71	-0,42	0,54	-0,64	-0,04	-0,32	-0,36
PIB	0,15	0,65	-0,52	1,00	-0,56	0,54	0,56	-0,25	0,45	-0,40	0,37	0,03	0,32	0,34
ANALF	-0,31	-0,89	0,94	-0,56	1,00	-0,84	-0,87	0,69	-0,48	0,58	-0,61	0,00	-0,36	-0,40
URBAN	0,33	0,83	-0,88	0,54	-0,84	1,00	0,82	-0,59	0,49	-0,48	0,56	-0,03	0,28	0,31
RENDA	0,57	0,83	-0,90	0,56	-0,87	0,82	1,00	-0,54	0,53	-0,54	0,76	0,19	0,33	0,36
GINI	-0,08	-0,64	0,71	-0,25	0,69	-0,59	-0,54	1,00	-0,20	0,32	-0,30	0,07	-0,31	-0,33
GPTS	0,44	0,38	-0,42	0,45	-0,48	0,49	0,53	-0,20	1,00	-0,24	0,52	0,27	0,22	0,24
ESF	-0,16	-0,57	0,54	-0,40	0,58	-0,48	-0,54	0,32	-0,24	1,00	-0,30	0,23	-0,14	-0,16
MED	0,49	0,50	-0,64	0,37	-0,61	0,56	0,76	-0,30	0,52	-0,30	1,00	0,53	0,17	0,19
ENF	0,31	-0,09	-0,04	0,03	0,00	-0,03	0,19	0,07	0,27	0,23	0,53	1,00	-0,03	-0,03
Cob25-59	0,06	0,37	-0,32	0,32	-0,36	0,28	0,33	-0,31	0,22	-0,14	0,17	-0,03	1,00	1,00
Cob25-64	0,07	0,40	-0,36	0,34	-0,40	0,31	0,36	-0,33	0,24	-0,16	0,19	-0,03	1,00	1,00

Nos modelos de regressão simples, todas as variáveis do bloco distal tiveram associação com a cobertura e, do bloco proximal, apenas gasto público em saúde apresentou associação (tabela 6).

Tabela 6 – Modelos de regressão simples entre a cobertura do rastreamento do câncer do colo do útero e as variáveis de exposição para as microrregiões de saúde de Minas Gerais, 2008.

Variáveis distais	Coefficiente	R² (%)	p-valor
População (log)	4,096	5,65	0,037
IDH (*100)	1,266	16,02	<0,001
IVS (*100)	-0,214	12,73	0,001
PIB (/1000)	0,511	11,71	0,002
Analfabetismo	-0,813	15,65	<0,001
Grau de urbanização	0,302	9,92	0,005
Renda (/100)	2,629	13,21	0,001
Gini (*100)	-1,595	10,87	0,003
Variáveis proximais			
Gasto público em saúde (/100)	5,299	5,93	0,032
Cobertura ESF	-0,111	2,45	0,173
Médicos/1000 hab.	5,641	3,78	0,089
Enfermeiros/1000 hab.	-3,148	0,08	0,805

Para a construção dos modelos de regressão múltipla, foram selecionadas inicialmente, as variáveis do bloco distal que não tiveram alta correlação entre si. Na tabela 7, são apresentados os modelos cujas variáveis foram estatisticamente significativas. Observa-se que nenhum modelo apresentou maior poder explicativo em relação ao modelo simples com o IDH. Desta forma, foi feita a inserção das variáveis proximais nesse modelo simples e verificou-se que apenas o IDH manteve-se estatisticamente significativo (tabela 8).

Tabela 7 - Modelos múltiplos com as variáveis explicativas do bloco distal

Variáveis	Variáveis com		Coeficiente das		Pressupostos (p-valor)			
	significância estatística ou p<0,10	variáveis com signif. ou p<0,10	R ²	R ² ajust	Moran Global I	Jarque-Bera	Breusch-Pagan	
IDH, POP (log)	IDH p<0,05	1,14	16,80	14,55	0,06 (0,21)	0,59 (0,74)	1,01 (0,60)	
IDH (x100), GINI (x100)	IDH p<0,05	1,01	16,94	14,69	0,05 (0,26)	0,77 (0,67)	0,24 (0,88)	
IDH (x100), GINI (x100), POP (log)	IDH p<0,05	0,93	17,53	14,14	0,05 (0,25)	0,37 (0,82)	1,57 (0,66)	
IVS, POP (log)	IVS p<0,05	-0,19	13,11	10,76	0,09 (0,10)	0,11 (0,94)	1,24 (0,53)	
GINI (x100), POP (log)	GINI p<0,05	-1,36	12,67	10,31	0,05(0,29)	0,06 (0,96)	0,91 (0,63)	
GINI, renda	RENDA p<0,05	0,01	15,70	13,42	0,07 (0,15)	0,17 (0,91)	0,25 (0,88)	
GINI, analfabetismo	ANALFAB p<0,05	-0,65	16,27	14,01	0,05 (0,22)	0,14 (0,93)	0,01 (0,99)	
Renda; POP (log)	RENDA p<0,05	0,02	13,22	10,88	0,09 (0,08)	0,34 (0,84)	0,71 (0,69)	
Analfabetismo; POP	ANALFAB p<0,05	-0,76	15,81	13,54	0,06 (0,19)	0,09 (0,95)	0,80 (0,66)	
PIB, POP	PIB p<0,05	0,44	14,07	11,75	0,10 (0,07)	0,55 (0,75)	1,22 (0,54)	
PIB, IDH	IDH p<0,05	0,97	17,19	14,95	0,06 (0,17)	2,14 (0,34)	0,33 (0,84)	
PIB, GINI	PIB p<0,05; GINI p<0,05	0,41; -1,25	18,04	15,83	0,05 (0,25)	1,03 (0,59)	0,26 (0,87)	
PIB, GINI, LOGPOP	PIB p<0,05; GINI p<0,05	0,38; -1,13	18,75	15,41	0,05 (0,22)	0,55 (0,75)	1,57 (0,66)	
PIB, IVS	PIB p=0,08; IVS p=0,05	0,32; -0,14	16,10	13,84	0,09 (0,08)	1,11 (0,57)	0,53 (0,76)	
PIB, analfabetismo	ANALFAB p<0,05	-0,61	17,78	15,56	0,07 (0,15)	0,77 (0,67)	0,59 (0,74)	

Tabela 8 - Modelos múltiplos com a inserção de variáveis do bloco proximal

Variáveis	Variáveis com significância estatística ou $p < 0,10$	Coeficiente das variáveis com signif. ou $p < 0,10$	R ²	R ² ajust	Pressupostos (p-valor)		
					Moran Global I	Jarque-Bera	Breusch-Pagan
IDH; GPTS	IDH $p < 0,05$	1,13	16,99	14,74	0,07 (0,17)	2,77 (0,25)	3,05 (0,21)
IDH; ESF	IDH $p < 0,05$	1,45	16,78	14,53	0,05 (0,23)	1,48 (0,47)	0,73 (0,69)
IDH, MED	IDH $p < 0,05$	1,27	16,03	13,76	0,06 (0,22)	1,25 (0,53)	0,04 (0,97)
IDH; ENF	IDH $p < 0,05$	1,26	16,03	13,76	0,06 (0,24)	1,20 (0,54)	0,04 (0,97)

Além disso, a adição de variáveis distais aos modelos não os tornou mais explicativos. Desta forma, foi selecionado o modelo simples com a variável IDH, pois com apenas uma variável, este modelo explica 16% da variabilidade da cobertura do rastreamento do câncer do colo do útero nas microrregiões de saúde de Minas Gerais e atende a todos os pressupostos da regressão linear.

DISCUSSÃO

A cobertura média estimada do rastreamento do câncer do colo do útero nas microrregiões de Minas Gerais, em 2008, foi de 74,5% (variando de 41,5 a 100%) e a maioria das microrregiões (65%) não alcançou a meta de 80% de cobertura estabelecida no Pacto pela Saúde (BRASIL, 2006). As menores coberturas estão concentradas, em sua maioria, nas macrorregiões Nordeste, Jequitinhonha, Leste e Leste do Sul, as quais, de forma geral, se caracterizam por apresentarem piores indicadores socioeconômicos (baixo IDH, alto IVS, baixo PIB per capita, alto analfabetismo e alto índice de Gini).

Os resultados do presente estudo indicam desigualdade no acesso ao rastreamento do câncer do colo do útero pelo SUS no estado de Minas Gerais, pois regiões que possuem maior IDH apresentam maior cobertura. O IDH explicou 16% da variação da cobertura e a cada aumento de 1% nesse indicador, em média, houve um aumento de 1,26% na cobertura. Destaca-se que a educação é uma das dimensões do IDH e, no presente estudo, a taxa de analfabetismo explicou 15,6% da variação da cobertura, sugerindo importante contribuição dessa dimensão. Esse resultado está de acordo com estudo ecológico realizado com dados de 2011 do VIGITEL para as 26 capitais brasileiras e o Distrito Federal, o qual também verificou que o IDH apresentou forte associação com a proporção de exames de *Papanicolaou* realizados (SADOVSKY *et al.*, 2015). Dados do VIGITEL de 2017 também mostraram maiores coberturas do rastreamento com o aumento do nível de escolaridade (BRASIL, 2018).

Cabe ressaltar que as regiões brasileiras com maiores taxas de desigualdade social (Índice de Gini) e menores níveis de desenvolvimento humano (IDH) apresentam as maiores taxas padronizadas de mortalidade por câncer de colo do útero, independente da distribuição dos serviços de saúde oferecidos à população (OLIVEIRA *et al.*, 2020).

Ao analisar dados do SISCOLO de 2006 a 2013, foi verificada que a tendência da taxa de produtividade (razão de *Papanicolaou*) no Brasil permaneceu estacionária ao longo dos anos, exceto no Nordeste, onde foi observada queda significativa no período de 2006 a 2013, e no Centro-Oeste, onde houve aumento significativo de 2006 a 2008 e significativo decréscimo de 2008 a 2013. Além disso, as tendências não mostram melhora nos indicadores de qualidade no período estudado, ressaltando

que o número de exames insatisfatórios no Norte e Nordeste foi quatro vezes maior em relação à região Sul (COSTA *et al.*, 2018).

Uma pesquisa realizada com mulheres das regiões Nordeste e Sul do país, encontrou uma cobertura de *papanicolaou* nos últimos três anos de 70,7 % (IC95%: 69,3-72,1) para o total da amostra, sendo 69,4% (IC95%: 67,3-71,6) em mulheres do Sul e 71,8 (IC95%: 69,9-73,7), do Nordeste. E, também foi observado que, mulheres com maior escolaridade (82,8% em mulheres com 9 anos ou mais de estudo; 62,2% em mulheres que estudaram de 0 a 5 anos) e de maior nível socioeconômico apresentaram maior cobertura (81,6% em mulheres dos estratos B e C da ABEP e 60,9%, estrato E) - (CORRÊA *et al.*, 2012), reforçando as desigualdades no acesso ao rastreamento do câncer do colo do útero no país.

O maior acesso ao rastreamento por mulheres com melhor nível socioeconômico pode estar relacionado ao maior acesso a informações sobre práticas preventivas e maior contato com profissionais de saúde, e, portanto, maior probabilidade de serem submetidas ao rastreamento (RICARDO-RODRIGUES *et al.*, 2015). Além disso, provavelmente, estas mulheres enfrentam barreiras menores ao rastreamento, incluindo responsabilidades domésticas e barreiras culturais (SONEJI; FUKUI, 2013). Portanto, programas oportunistas de rastreamento podem ser mais facilmente acessados por mulheres com maior renda e maior escolaridade (RICARDO-RODRIGUES *et al.*, 2015).

No presente estudo, considerando apenas as mulheres que iniciaram o rastreamento em 2008, a média de sobrerastreamento por periodicidade nas microrregiões foi de 3,7%, variando de 1,3 a 7,1% e por faixa etária, de 54,5%, variando de 21,7 a 73,4%. Para mulheres que iniciaram previamente, a média do sobrerastreamento por periodicidade foi de 64,2%, variando de 53,4 a 74,3% e por faixa etária, de 14,6%, variando de 11,0 a 19,0. Desta forma, observa-se que o sobrerastreamento por faixa etária é mais importante em mulheres que estão iniciando o rastreio, enquanto o sobrerastreamento por periodicidade é mais importante em mulheres que já se submeteram previamente ao exame e que deveriam repeti-lo com periodicidade trienal. Além disso, chama atenção o elevado percentual deste último em todo o estado, com pequena variabilidade entre as microrregiões e o padrão diferenciado de distribuição espacial dos dois tipos de sobrerastreamento nos dois grupos de mulheres. Tal fato pode indicar que o sobrerastreamento está mais relacionado ao conhecimento dos profissionais de saúde da atenção básica e o não

seguimento das diretrizes pelos mesmos do que com características regionais socioeconômicas e assistenciais.

Em consonância com esses resultados, análise dos dados do SISCOLO nos anos 2012 e 2013, verificou que, com base em informação de autorrelato do período do exame anterior, 54% dos exames foram realizados em intervalo de até um ano no Brasil. Em Minas Gerais esse percentual foi de 51% no período; o percentual de repetição bianual foi em torno de 30% no estado e, apenas 10% dos exames foram realizados no intervalo de três anos. Sendo importante ressaltar que o percentual esperado de exames realizados a cada três anos deve ser próximo a 100%, pois o percentual de mulheres com exames alterados em que a conduta é a repetição semestral ou anual do citopatológico é baixo (INCA, 2014a).

Em mulheres submetidas ao rastreamento em um município no interior de São Paulo, verificou-se que 51% repetiu em um intervalo de 1 ano e 24%, de 2 anos. O estudo considerou como exames em excesso aqueles realizados em mulheres fora da faixa etária alvo e antes do período preconizado e encontrou um percentual de 62,7%, em 2007. Esses resultados, reforçam o caráter oportunístico do programa e o não seguimento das diretrizes de rastreamento do Ministério da Saúde pelos profissionais de saúde (VALE *et al.*, 2010).

No que se refere ao conhecimento das mulheres em relação à periodicidade do exame, nenhuma das 240 mulheres de um município do Nordeste conhecia a recomendação de que o exame deve ser realizado trienalmente, após dois exames negativos, com intervalo anual (MALTA *et al.*, 2017). Outro estudo, realizado com 171 usuárias, de 25 a 64 anos, de uma unidade básica de saúde (UBS) no município de Juiz de Fora, verificou que todas as mulheres não tinham conhecimento adequado sobre a periodicidade do exame e mais de 95% não sabia a faixa etária recomendada para realização do mesmo. Refletindo esse desconhecimento, mais de 80% das usuárias tiveram prática inadequada de rastreamento, embora mais da metade dessas mulheres frequentassem a mesma UBS há mais de dez anos (MASCARENHAS *et al.*, 2020). As mulheres realizam o exame quando solicitado pelos profissionais e estes, porém, não promovem o conhecimento adequado por meio de educação em saúde (MELO *et al.*, 2019). Desta forma, torna-se essencial que os profissionais envolvidos com o rastreamento tenham conhecimento e prática adequada da periodicidade recomendada, de acordo com as diretrizes do Ministério da Saúde.

No âmbito da gestão, um estudo investigou o conhecimento dos gestores do SUS quanto às diretrizes de rastreamento do câncer do colo do útero e verificou que apenas 70% tinham conhecimento adequado quanto à faixa etária e 56% em relação à periodicidade. Quanto às barreiras para o seguimento das diretrizes, 25,0% indicaram a pouca tradição organizacional no uso de diretrizes, 21,4%, a baixa adesão dos profissionais e 21,4%, a desorganização dos serviços de saúde. Gestores de algumas regiões também relataram como barreira a alta rotatividade de profissionais da atenção básica, dificultando ações de capacitação (SANTOS; RAMOS; MIGOWSKI, 2019).

Estudo realizado com profissionais de 1.600 Unidades de Atenção Primária Saúde no Brasil, mostrou que a maioria dos médicos e enfermeiros (68,7% e 76,9%, respectivamente) declarou que as recomendações nacionais para rastreamento do câncer do colo do útero são muito influentes na sua unidade de saúde; no entanto, 92,6% relataram fazer o rastreamento das mulheres anualmente e 75,1% relataram começar a fazer o rastreamento das mulheres antes dos 25 anos, sendo os enfermeiros mais propensos a relatar essas duas práticas (STORMO; DE MOURA; SARAIYA, 2014). No município de Juiz de Fora, um estudo também realizado com 170 médicos e enfermeiros da ESF verificou que apenas 40% tinha conhecimento adequado em relação à faixa etária e periodicidade recomendados pelo Ministério da Saúde (FERREIRA, 2020). Dessa forma, constata-se que, apenas a publicação de diretrizes não resulta, por si só, em sua aplicação na prática clínica, sendo fundamental, nesse contexto, identificar as barreiras e desenvolver estratégias adequadas para a disseminação e implementação das diretrizes (FISCHER *et al.*, 2016).

Se os exames fossem realizados de acordo com as diretrizes do INCA, seria possível aumentar a cobertura, sem a necessidade de expandir a capacidade assistencial, apenas com o direcionamento adequado da oferta de exames e, desse modo, o rastreamento poderia se tornar mais efetivo (VALE *et al.*, 2010).

Nesse sentido, destaca-se a importância de instituir programas de educação permanente na atenção primária à saúde direcionados aos profissionais envolvidos no rastreamento do câncer do colo do útero. Estudo de intervenção realizado em Unidades Básicas de Saúde do município de Goiânia (GO) com médicos, enfermeiros e agentes comunitários da saúde, verificou impacto positivo de um programa de capacitação no seguimento das recomendações quanto à periodicidade, população

alvo e na qualidade da coleta. A realização do exame em intervalo menor que um ano passou de 5,3% para 2,8%, e, no intervalo de um ano, passou de 40,1% para 31,4%. Houve redução da frequência de realização do exame citopatológico em mulheres com menos de 25 anos de idade, de 22,0% para 17,9% e a proporção de amostras satisfatórias, passou de 70,4% para 80,2% após a capacitação (AMARAL *et al.*, 2014).

Apesar de nossos resultados não mostrarem associação entre a cobertura do rastreamento com a cobertura pela Estratégia Saúde da Família, ressalta-se o grande potencial da mesma para a organização do rastreamento do câncer do colo do útero no Brasil. Por meio da adscrição da população, é possível identificar as mulheres que não realizam exames regularmente e então incluí-las no rastreamento por meio de busca ativa realizada por agentes comunitários da saúde. Para isso, é imprescindível a adequada formação desses profissionais (VALE *et al.*, 2010; FERREIRA *et al.*, 2013)

As principais limitações do presente estudo estão relacionadas à qualidade dos dados de bases secundárias e ao fato de não considerar exames realizados por serviços privados ou saúde suplementar. No que se refere à qualidade dos dados do SISCOLO, destacam-se a subnotificação em relação ao SIA, a possibilidade de erros de digitação e a não existência de sistemas de validação da informação. Em 2009, de acordo com dados do INCA, no Brasil, foram registrados 8% a mais de exames do SIA em relação ao SISCOLO e, em Minas Gerais, essa diferença foi de 7% (INCA, 2010). Destaca-se que a cobertura estimada no presente trabalho se refere a mulheres que foram submetidas ao rastreamento no SUS, cujos exames foram registrados no SISCOLO. A cobertura populacional só pode ser aferida por meio de inquéritos que abranjam todo o conjunto da população feminina e, assim, também incluam mulheres que não dependam exclusivamente do SUS (INCA 2014b). A limitação do SISCOLO em relação à unidade de registro ser o exame e não a mulher foi superada nesse estudo, utilizando técnicas de relacionamento probabilístico para limpar a base de 2008, deixá-la apenas com um registro por mulher e montar o seguimento para verificar a periodicidade da repetição do exame.

Os achados do presente estudo sugerem a necessidade de aprimoramento do programa de rastreamento em Minas Gerais, com objetivo de aumentar o acesso, principalmente, em regiões mais vulneráveis, com menor nível socioeconômico. E, ainda, a necessidade de instituir programas de educação permanente direcionados aos profissionais da atenção primária, visando seguir as diretrizes do INCA quanto à periodicidade e faixa etária para o rastreamento do câncer do colo do útero. Por fim,

cabe ressaltar que o CCU é uma doença evitável e curável e para o seu controle, com impacto na incidência e na mortalidade, é fundamental garantir elevada cobertura do rastreamento na população pertencente à faixa etária alvo com exames de qualidade, periodicidade adequada e acesso à confirmação diagnóstica e ao tratamento oportuno.

REFERÊNCIAS

AMARAL, A.F.; ARAÚJO, E.S; MAGALHÃES, J.C, *et al.* Impacto da capacitação dos profissionais de saúde sobre o rastreamento do câncer do colo do útero em unidades básicas de saúde. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 36, n. 4, p. 182-187, 2014.

AZEVEDO E SILVA, G.; JARDIM, B. C.; FERREIRA, V. M., *et al.* Cancer mortality in the Capitals and in the interior of Brazil: a four-decade analysis. **Revista Saúde Pública**. v. 54, p. 126, 2020.

AZEVEDO e SILVA, G.; ZEFERINO, L.C.; THULER, L.C.S.; BUSTAMANTE-TEIXEIRA, M.T.; GUERRA, M.R. A situação dos cânceres do colo do útero e da mama no Brasil. In: Teixeira L. **Câncer de mama, câncer de colo de útero: conhecimentos, políticas e práticas**. Rio de Janeiro: Outras Letras; 2015. p. 41-73.

BRASIL. Ministério da Saúde. **Portaria nº 399/GM, de 22 fevereiro de 2006**. Divulga o Pacto pela Saúde 2006 – Consolidação do SUS e aprova as Diretrizes Operacionais do Referido Pacto. Diário Oficial da União, Brasília, 22 fev 2006. Seção 1, p.43-51. 2006.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **Rastreamento**. Brasília, 2010.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022**. Brasília, 2011.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. **Vigitel Brasil 2017: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2017**. Brasília: Ministério da Saúde, 2018.

BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; SIEGEL, R.L.; TORRE, L.A.; JEMAL, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality

worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinician**, v. 68, n. 6, p.394-424, 2018.

BRITTANY F. LEES, MD; BRITT K. ERICKSON, MD; WARNER K. HUH, MD. Cervical cancer screening: evidence behind the guidelines. **American Journal of Obstetrics & Gynecology**, 2016.

CORREA, M.S; SILVEIRA, D.S.; SIQUEIRA, F.V. et al. Cobertura e adequação do exame citopatológico de colo uterino em estados das regiões Sul e Nordeste do Brasil. **Cadernos de Saúde Pública**, v. 28, n. 12, p. 2257-2266, 2012.

COSTA, R.F.A.; LONGATTO-FILHO, A.; VAZQUEZ, F.L. Trend analysis of the quality indicators for the Brazilian cervical cancer screening programme by region and state from 2006 to 2013. **BMC Cancer**, v. 18; n.126, 2018.

DANCEY, C.; REIDY, J. Estatística Sem Matemática para Psicologia: Usando SPSS para Windows. Porto Alegre, Artmed, 2006.

FERREIRA, M.C.M. **Conhecimentos, atitudes e práticas de médicos e enfermeiros da estratégia saúde da família de Juiz de Fora sobre o controle dos cânceres da mama e do colo do útero** [dissertação]. Juiz de Fora: Universidade Federal de Juiz de Fora; 2020.

FERREIRA, T.X.A.M; TAVERES, S.B.N.; REZENDE, I.R. *et al.* Capacitação do agente comunitário de saúde visando reorganização do rastreamento do câncer do colo do útero. **Revista APS**, v. 16, n. 1, p. 75-82, 2013.

FISCHER, F.; LANGE, K.; KLOSE, K. et al. Barriers and Strategies in Guideline Implementation-A Scoping Review. **Healthcare**, v. 4, n. 3, 2016.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). Coordenação de Prevenção e Vigilância. **Nomenclatura brasileira para laudos cervicais e condutas preconizadas: recomendações para profissionais de saúde**. INCA, 2006.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim n. 2, abr/jun 2010. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-2-2010.pdf>. Acesso: fev., 2021. INCA, 2010b

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim ano 5, n. 1 janeiro/ abril 2014. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-1-2014.pdf>. Acesso em: fev, 2021. INCA, 2014a

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação Geral de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Ficha técnica de indicadores das ações de controle do**

câncer do colo do útero [Internet]. Rio de Janeiro: INCA; 2014b. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//fichatecnicaindicadorescolo14.pdf>. Acesso: fev. 2021. INCA, 2014b

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio a Organização de Rede. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero – 2. ed. rev. atual.** – Rio de Janeiro: INCA, 2016.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. **Estimativa 2020: incidência de câncer no Brasil** / Instituto Nacional de Câncer José Alencar Gomes da Silva. – Rio de Janeiro: INCA, 2019.

LEES, B.F.; ERICKSON, B.K.; HUH, W.K. Cervical cancer screening: evidence behind the guidelines. **American Journal of Obstetrics & Gynecology**, 2016.

MALTA, E.F.G.D.; GUBERT, F.A.; VASCONCELOS, C.T.M. *et al.* Prática inadequada de mulheres acerca do papanicolaou. **Texto & Contexto Enfermagem**, v. 26, n. 1, e5050015, 2017.

MASCARENHAS, M.S.; FARIA, L.V.; MORAIS, L.P.; LAURINDO, D.C.; NOGUEIRA, M.C. Conhecimentos e Práticas de Usuárias da Atenção Primária à Saúde sobre o Controle do Câncer do Colo do Útero. **Revista Brasileira de Cancerologia**, v. 66, n. 3, e-011030, 2020.

MELO, E.M.F.; LINHARES, F.M.P.; SILVA, T.M. *et al.* Câncer cérvico-uterino: conhecimento, atitude e prática sobre o exame de prevenção. **Revista Brasileira de Enfermagem**, v. 72, p. 30-6, 2019.

OLIVEIRA, M.M.; ANDRADE, S.S.C.A.; OLIVEIRA, P.P.V., AZEVEDO E SILVA, G.; SILVA, M.M.A.; MALTA, D.C. Cobertura de exame Papanicolaou em mulheres de 25 a 64 anos, segundo a Pesquisa Nacional de Saúde e o Sistema de Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico, 2013. **Revista Brasileira de Epidemiologia**, v. 21, e180014, 2018.

PEIRSON, L.; FITZPATRICK-LEWIS, D.; CILISKA, D.; WARREN, R. Screening for cervical cancer: a systematic review and meta-analysis. **Systematic Reviews**, v. 24, 2013.

REIS, N.V.S; ANDRADE, B.B.; GUERRA, M.R.; et al. The Global Burden of Disease Study Estimates of Brazil's Cervical Cancer Burden. **Annals of Global Health**, v.86, n.1, p.1-12, 2020.

RICARDO-RODRIGUES, I.; JIMÉNEZ-GARCÍA, R.; HERNÁNDEZ-BARRERA, V.; CARRASCO-GARRIDO, P.; JIMÉNEZ-TRUJILLO, I.; LÓPEZ DE ANDRÉS, A. Social disparities in access to breast and cervical cancer screening by women living in Spain. **Public Health**, v. 129, p. 881-888, 2015.

SADOVSKY, A.D.I.; POTON, W.L.; REIS-SANTOS, B.; BARCELOS, M.R.B.; SILVA, I.C.M. Índice de Desenvolvimento Humano e prevenção secundária de câncer de

mama e colo do útero: um estudo ecológico. **Cadernos de Saúde Pública**, v. 31, n. 7, p. 1539-1550, 2015.

SANKARANARAYANAN, R.; SOMANATHAN, T.; SANTOS, C.; MUWONGE, R.; TWALIB, N. O desafio de controlar o cancer do colo do útero no mundo em desenvolvimento. In: COELHO, F.R.G.; SOARES, F.A.; FOCCHI, J.; FREGNANI, J.H.T.G; ZEFERINO, L.C.; VILLA, L.L.; FEDERICO, M.H.; NOVAES, P.E.R.S; COSTA, R.L.R. **Câncer do colo do útero**. Tecmeddd, São Paulo – SP, 2008.

SANTOS, RENATA OLIVEIRA MACIEL DOS; RAMOS, DANIELLE NOGUEIRA and MIGOWSKI, ARN. **Barreiras na implementação das diretrizes de detecção precoce dos cânceres de mama e colo do útero no Brasil**. *Physis* [online]. 2019, vol.29, n.4, e290402.

SONEJI, s.; FUKUI, N. Socioeconomic determinants of cervical cancer screening in Latin America. **Revista Panamericana de Salud Pública**, v. 33, n.3, p. 174–82, 2013.

STORMO, A.R.; DEMOURA, L.; SARAIYA, M. Cervical Cancer-Related Knowledge, Attitudes, and Practices of Health Professionals Working in Brazil's Network of Primary Care Units. **The Oncologist**, v. 19, p. 375–382, 2014.

THEME FILHA, M.M.T.; LEAL, M.C.; OLIVEIRA, E.F.V.; ESTEVES-PEREIRA, A.P.; GAMA, S.G.N. Regional and social inequalities in the performance of Pap test and screening mammography and their correlation with lifestyle: Brazilian national health survey, 2013. **International Journal for Equity in Health**, v. 15, n. 1, p.136, 2016.

TSUCHIYA, C.T.; LAWRENCE, T.; KLEN, M.S.; FERNANDES, R.A.; ALVES, M.R. O câncer de colo do útero no Brasil: uma retrospectiva sobre as políticas públicas voltadas à saúde da mulher. **Jornal Brasileiro de Economia da Saúde**, v. 9; n. 1; p. 137-47, 2017.

VALE, D.B.A.P.; MORAIS, S.S.; PIMENTA, A.L.; ZEFERINO, L.C. Avaliação do rastreamento do câncer do colo do útero na Estratégia Saúde da Família no Município de Amparo, São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 26, n. 2, p. 383-90, 2010.

VICTORA, C.G.; HUTTLY, S.R.; FUCHS, S.C.; OLINTO, M.T. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. **International Journal of Epidemiology**, v. 26, p. 224-7, 1997.

WORLD HEALTH ORGANIZATION. **National cancer control programmes: policies and managerial guidelines**. 2 ed. Geneva: WHO; 2002.

WORLD HEALTH ORGANIZATION. **Global strategy to accelerate the elimination of cervical cancer as a public health problem**. 2020. Disponível em: <https://www.who.int/publications/i/item/9789240014107>. Acesso: fev, 2021.

7. CONSIDERAÇÕES FINAIS

Os resultados do primeiro artigo mostraram que, no período de 2006 a 2011, a razão de exames citopatológicos do colo do útero em mulheres de 25 a 59 anos não alcançou a meta pactuada. Cerca de um quarto dos exames citopatológicos do colo do útero realizados no período estava fora da faixa etária de 25 a 59 anos e o estado poderia ter alcançado a meta se esses exames tivessem sido realizados na faixa etária preconizada. Além do excesso de exames fora da faixa etária alvo, cerca de metade das mulheres relataram exame prévio no período de até um ano. O índice de positividade é um importante indicador de qualidade, pois caracteriza a sensibilidade do rastreamento em detectar lesões na população examinada. Durante o período analisado, esse indicador foi classificado como baixo no estado, podendo indicar que alterações suspeitas não estavam sendo identificadas.

O estudo de coorte não concorrente permitiu transformar a unidade de análise do SISCOLO de exames para mulheres. Sendo possível estimar cobertura e realizar o seguimento de mulheres submetidas ao citopatológico do colo do útero em 2008 e, com isso, avaliar o sobrerastreamento em relação à repetição do exame antes do preconizado. Os resultados mostraram uma cobertura próxima à meta de 80% para mulheres de 25 a 64 anos. Por outro lado, revelaram um cenário preocupante em relação ao excesso de exames. A maioria das mulheres que deveriam repetir o exame com periodicidade trienal o fizeram com menos de 18 meses. Em mulheres que estavam iniciando o rastreamento, o problema maior foi o sobrerastreamento em relação à faixa etária, pois quase metade das mulheres tinha menos de 25 anos.

O estudo ecológico evidenciou desigualdes regionais na cobertura do rastreamento de Minas Gerais, com microrregiões de melhor IDH apresentando maior cobertura. Esse resultado reforça as desigualdes sociais no acesso ao exame citopatológico existentes no país. A distribuição espacial do sobrerastreamento no estado mostrou padrão diferenciado entre o sobrerastreamento por periodicidade e faixa etária e em mulheres que estavam iniciando o rastreamento e mulheres com exame prévio. O sobrerastreamento por periodicidade foi elevado no estado com pequena variabilidade entre as microrregiões. Dessa forma, esses resultados podem indicar que o sobrerastreamento estava mais relacionado ao conhecimento dos profissionais de saúde da atenção básica e o não seguimento das diretrizes pelos mesmos do que com características regionais socioeconômicas e assistenciais.

Os dados do SISCOLO, apesar de suas limitações, permitiram a construção de indicadores fundamentais para avaliar as ações do rastreamento em Minas Gerais, bem como, a identificação de regiões mais vulneráveis, que devem ser foco de políticas regionais de controle do CCU.

Os resultados do presente estudo corroboram o caráter oportunístico do rastreamento do CCU e evidenciam a necessidade de aprimoramento do programa no estado de Minas Gerais. O sucesso do rastreamento, com impacto na incidência e na mortalidade pela doença depende da elevada cobertura da população-alvo, com exames de qualidade, periodicidade adequada e acesso oportuno ao tratamento das lesões identificadas. Assim, é imprescindível aumentar o acesso ao exame, especialmente em regiões mais vulneráveis, de menor nível socioeconômico; melhorar a qualidade do rastreamento através do investimento na capacitação de todos os profissionais envolvidos, tanto na fase de coleta, quanto na fase analítica; identificar as barreiras para o não seguimento das diretrizes quanto à faixa etária e periodicidade do exame pelos profissionais de saúde e desenvolver estratégias para a adesão dos mesmos. Nesse sentido, destaca-se a importância da implementação de programas de educação permanente na atenção primária à saúde direcionados aos profissionais envolvidos no rastreamento do CCU.

Além do impacto financeiro que o sobrerastreamento gera para o sistema de saúde, ele compromete o acesso de mulheres com maior risco para o desenvolvimento da doença e pode resultar em intervenções desnecessárias em mulheres com menos de 25 anos, cujas lesões são de caráter transitório.

Ressalta-se, então, a necessidade de implementação do modelo organizado de rastreamento, com monitoramento da população alvo e da periodicidade de realização dos exames, através de um sistema de informação de base populacional, que permita o recrutamento das mulheres na faixa etária alvo e o seguimento daquelas com exames alterados. Enquanto isso não for possível, uma possibilidade com grande potencial de sucesso é a Estratégia Saúde da Família, conforme relatos de experiência publicados no país. A adscrição da população possibilita identificar as mulheres na faixa etária alvo e, através de uma planilha eletrônica, fazer o seguimento, registrar a periodicidade e os resultados dos exames e fazer busca ativa através de agentes comunitários da saúde capacitados.

Por fim, cabe enfatizar que a OMS tem como objetivo eliminar o CCU no mundo até o ano de 2030 com base em metas de vacinação, rastreamento e tratamento

oportuno. Assim, os resultados do presente estudo apontam a necessidade de aprimorar o rastreamento do CCU, tendo em vista sua contribuição fundamental para que este objetivo de eliminação da doença seja alcançado até o final dessa década.

REFERÊNCIAS

ABREU, R.G.; SOUSA, A.I.A.; OLIVEIRA, M.R.F.; SANCHEZ, M.N. Tuberculose e diabetes: relacionamento probabilístico de bases de dados para o estudo da associação entre ambas doenças. **Epidemiologia e Serviços de Saúde**, v. 26, n. 2, p. 359-368, 2017.

ALBER, J.M.; BREWER, N.T.; MELVIN, C.; YACKLE, A.; SMITH, J.S.; KO L.K.; CRAWFORD, A.; GLANZ, K. Reducing overuse of cervical cancer screening: A systematic review. **Preventive Medicine**, v. 116, p. 51-59, 2018.

ALBUQUERQUE, K.M.; FRIAS, P.G.; ANDRADE, C.L.T.; AQUINO, E.M.L.; MENEZES, G.; SZWARCOWALD, C.L. Cobertura do teste de Papanicolaou e fatores associados a não-realização: um olhar sobre o Programa de Prevenção do Câncer do Colo do Útero em Pernambuco, Brasil. **Cadernos de Saúde Pública**, v. 25, Sup 2:S301-S309, 2009.

ALI, F; KUELKER, R; WASSIE, B. Understanding cervical cancer in the context of developing countries. **Annals of Tropical Medicine and Public Health**; v. 5; p. 3-15, 2012.

ALMEIDA, C.M.; RODRIGUEZ, M.A.; SKOOTSKY, S.; PREGLER, J.; STEERS, N.; WENGER, N.S. Cervical Cancer Screening Overuse and Underuse: Patient and Physician Factors. **American Journal of Managed Care**, v. 19, n. 6, p. 482-489, 2013.

ALVES, C.M.M.; GUERRA, M.R.; BASTOS, R. R. Tendência de mortalidade por câncer de colo de útero para o Estado de Minas Gerais, Brasil, 1980-2005. **Cadernos de Saúde Pública**, v.25, n.8, p.1693-1700, 2009.

ANDRADE, M.S; ALMEIDA, M.M.G.; ARAÚJO, T.M.; SANTOS, K.O.B. Fatores associados a não adesão ao Papanicolaou entre mulheres atendidas pela Estratégia Saúde da Família em Feira de Santana, Bahia, 2010. **Epidemiologia e Serviços de Saúde**, v. 23, n.1, p.111-120, 2014.

ARAÚJO, E.S.; BARBOSA, F.M.; ÁZARA, C.Z.S.; FERREIRA, T.X.A.M.; TAVARES, S.B.N.; AMARAL, R.G. Avaliação do Seguimento de Mulheres com Exames Citopatológicos Alterados de acordo com as Condutas Preconizadas pelo Ministério da Saúde do Brasil em Goiânia, Goiás. **Revista Brasileira de Cancerologia**, v. 60, n. 1, p.7-13, 2014.

ARBYN, M.; ANTTILA, A.; JORDAN, J.; RONCO, G.; SCHENCK, U.; SEGNAN, N.; WIENER, H.; HERBERT, A.; VON KARSA, L. European Guidelines for Quality Assurance in Cervical Cancer Screening. Second edition--summary document. **Annals of Oncology**, v. 21, n. 3, p. 448-58, 2010.

AYRES, A.R.G.; AZEVEDO e SILVA, G. Prevalência de infecção do colo do útero pelo HPV no Brasil: revisão sistemática. **Revista de Saúde Pública**, v. 44, n. 5, p. 963-974, 2010.

AYRES, A.R.G.; AZEVEDO E SILVA, G.; BUSTAMANTE-TEIXEIRA, M.T. *et al.* HPV in women assisted by the Family Health Strategy. **Revista de Saúde Pública**, v. 51, n. 92, 2017.

AZEVEDO e SILVA, G.; ZEFERINO, L.C.; THULER, L.C.S.; BUSTAMANTE-TEIXEIRA, M.T.; GUERRA, M.R. A situação dos cânceres do colo do útero e da mama no Brasil. In: Teixeira L. **Câncer de mama, câncer de colo de útero: conhecimentos, políticas e práticas**. Rio de Janeiro: Outras Letras; 2015. p. 41-73.

AZEVEDO E SILVA, G.; JARDIM, B. C.; FERREIRA, V. M., et al. Cancer mortality in the Capitals and in the interior of Brazil: a four-decade analysis. **Revista Saúde Pública**. v. 54, p. 126, 2020.

BARBOSA, I.R.; SOUZA, D.L.B.; BERNAL, M.M.; COSTA, I.C.C. Desigualdades regionais na mortalidade por câncer de colo de útero no Brasil: tendências e projeções até o ano 2030. **Ciência & Saúde Coletiva**, 21(1), 253-262, 2016.

BLAKELY, T.; SALMOND, C. Probabilistic record linkage and a method to calculate the positive predictive value. **International Journal of Epidemiology**, v. 31, n. 6, p. 1246-1252, 2002.

BOSCH, F.X.; LORINCZ, A.; MUÑOZ, N.; MEIJER, C.J.L.M.; SHAH, K.V. The causal relation between human papillomavirus and cervical cancer. **Journal of Clinical Pathology**, v.55, n.4, p. 244-65, 2002.

BOYD, J.H.; GUIVER, T.; RANDALL, S.M.; FERRANTE AM, SEMMENS JB, ANDERSON P, DICKINSON T. A Simple Sampling Method for Estimating the Accuracy of Large Scale Record Linkage Projects. **Methods of Information in Medicine**, v. 55, n. 3, p. 276-83, 2016.

BRASIL. Ministério da Saúde. **Portaria nº 3.040, de 21 de junho de 1998**. Institui o Programa Nacional de Combate ao Câncer de Colo Uterino. Diário Oficial da União, Brasília (DF), 23/06/1998, Seção 1, pág. 102.

BRASIL. Ministério da Saúde. **Portaria nº 408, de 30 de agosto de 1999**. Diário Oficial da União. Brasília, 2 agosto de 1999, Seção 1, p. 14.

BRASIL. Ministério da Saúde. **Portaria nº 2.439/GM de 8 de dezembro de 2005**. Institui a Política Nacional de Atenção Oncológica. Diário Oficial da União, Brasília, 9 dez 2005. Seção 1, p.80-8.

BRASIL. Ministério da Saúde. **Portaria nº 287, de 24 de abril de 2006**. Diário Oficial da União. Brasília, 05/06/2006, Seção 1, pág. 31. 2006a.

BRASIL. Ministério da Saúde. **Portaria nº 399/GM, de 22 fevereiro de 2006**. Divulga o Pacto pela Saúde 2006 – Consolidação do SUS e aprova as Diretrizes Operacionais do Referido Pacto. Diário Oficial da União, Brasília, 22 fev 2006. Seção 1, p.43-51. 2006b.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Fundação Oswaldo Cruz. **Introdução à Estatística Espacial para a Saúde Pública**. Vols. 1, 2, 3. Brasília: Ministério da Saúde, 2007.

BRASIL. Ministério da Saúde. Departamento de Apoio à Gestão Descentralizada. **Instrutivo da Pactuação de Prioridades, Objetivos, Metas e Indicadores de Monitoramento do Pacto pela Vida e de Gestão para o Biênio 2010-2011**. Conforme Portaria nº 2669 GM/MS de 03 de novembro de 2009, Brasília: Ministério da Saúde; 2009.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. **Rastreamento**. Brasília, 2010.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Plano de ações estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil 2011-2022**. Brasília, 2011.

BRASIL. Ministério da Saúde. **Portaria nº 3.394, de 30 de dezembro de 2013**. Institui o Sistema de Informação de Câncer (SICAN) no âmbito do Sistema Único de Saúde (SUS). Brasília: Ministério da Saúde, 2013a.

BRASIL. Ministério da Saúde. **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. Diário Oficial da União, Brasília, 30 dez 2013. 2013b.

BRASIL. Ministério da Saúde. **Portaria nº 189, de 31 de janeiro de 2014**. Institui o Serviço de Referência para Diagnóstico e Tratamento de Lesões Precursoras do Câncer do Colo de Útero (SRC), o Serviço de Referência para Diagnóstico de Câncer de Mama (SDM) e os respectivos incentivos financeiros de custeio e de investimento para a sua implantação. Brasília, 2014.

BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Regulação, Avaliação e Controle. **Sistemas de Informação da Atenção à Saúde: Contextos Históricos, Avanços e Perspectivas no SUS**. Brasília, 2015.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde Departamento de Vigilância das Doenças Transmissíveis Coordenação-Geral do Programa Nacional de Imunizações. **Informe técnico da ampliação da oferta das vacinas papilomavírus humano 6, 11, 16 e 18 (recombinante) – vacina HPV quadrivalente e meningocócica C (conjugada)**. Brasília, 2018a.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças e Agravos não Transmissíveis e Promoção da Saúde. **Vigitel Brasil 2017: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico: estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2017**. Brasília: Ministério da Saúde, 2018b.

BORGES, M.F.S.O.; DOTTO, L.M.G; KOIFMAN, R.J.; CUNHA, M.A.; MUNIZ, P.T. Prevalência do exame preventivo de câncer do colo do útero em Rio Branco, Acre, Brasil, e fatores associados à não-realização do exame. **Cadernos de Saúde Pública**, v. 28, n. 6, p. 1156-1166, 2012.

BRAY, F.; FERLAY, J.; SOERJOMATARAM, I.; *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA: A Cancer Journal for Clinician**, v. 68, n. 6, p.394-424, 2018.

CAMARGO JR., K.R.; COELI, C.M. Reclink: aplicativo para o relacionamento de bases de dados, implementando o método probabilisticrecordlinkag. **Cadernos de Saúde Pública**, v. 16, n.2, p. 439-447, 2000.

CAMARGO JR, K.R.; COELI, C.M. **RecLink**. Rio de Janeiro; 2007.

CAPUANI, L.; BIERRENBACH, A.L.; ABREU, F.; TAKECIAN, P.L.; FERREIRA, J.E.; SABINO, E.C. Accuracy of a probabilistic record-linkage methodology used to track blood donors in the Mortality Information System database. **Cadernos de Saúde Pública**, v. 30, n. 8, p.1623-1632, 2014.

CARDOSO, R.V. Prevenção quaternária: um olhar sobre a medicalização na prática dos médicos de família. **Revista Brasileira de Medicina de Família e Comunidade**, v.10, n. 35, p. 1-10, 2015.

CASTRO, B.; RIBEIRO, D.P.; OLIVEIRA, J.; PEREIRA, M.B; SOUSA, J.C.; YAPHE, J. Rastreo do câncer do colo do útero: limites etários, periodicidade e exame ideal: revisão da evidência recente e comparação com o indicador de desempenho avaliado em Portugal. **Ciência & Saúde Coletiva**, v. 19, n. 4, p. 1113-1122, 2014.

COELI, C.M.; CAMARGO JR., K.R. Avaliação de diferentes estratégias de blocagem no relacionamento probabilístico de registros. **Revista Brasileira de Epidemiologia**, v. 5, n. 2, p. 185-196, 2002.

COELI, C.M. A qualidade do linkagede dados precisa de mais atenção. **Cadernos de Saúde Pública**, v. 31, n. 7, p. 1349-1350, 2015.

COLPANI, V.; BIDINOTTO, A.B.; FALAVIGNA, M.; GIOZZA, S.P; BENZAKEN, A.S.; PIMENTA, C. *et al.* Prevalence of papilloma virus in Brazil: a systematic review protocol. **BMJ Open**, v. 6, e011884, 2016.

CORRÊA, C.S.L.; LIMA, A.S.; LEITE, I.C.G. et al. Rastreamento do câncer do colo do útero em Minas Gerais: avaliação a partir de dados do Sistema de Informação do Câncer do Colo do Útero (SISCOLO). **Cadernos Saúde Coletiva**, v. 25, n. 3, p. 315-323, 2017.

COSTA, R.F.; LONGATTO-FILHO, A.; PINHEIRO, C.; ZEFERINO, L.C.; FREGNANI, J.H. Historical Analysis of the Brazilian Cervical Cancer Screening Program from 2006 to 2013: A Time for Reflection. **PLoSOne**, v. 10, n. 9, e0138945, 2015.

COUTINHO, E.S.F.; COELI, C.M. Acurácia da metodologia de relacionamento probabilístico de registros para identificação de óbitos em estudos de sobrevivência. **Cadernos de Saúde Pública**, v. 22, n. 10, p. 2249-2252, 2006.

DA COSTA, A.M.; HASHIM, D.; FREGNANI, J.H.T.G.; WEIDERPASS, E. Overall survival and time trends in breast and cervical cancer incidence and mortality in the Regional Health District (RHD) of Barretos, São Paulo, Brazil. **BMC Cancer**, v. 18, n. 1, p. 1079, 2018.

DIAS, M.B.K.; TOMAZELLI, J.G.; ASSIS, M. 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**, v. 19, n. 3, p. 293-306, 2010.

DUARTE, D.A.P. **Iniquidade social e câncer em mulheres: análise da mortalidade por câncer de mama e colo do útero nas microrregiões de saúde de Minas Gerais no período de 2008-2012** [dissertação]. Juiz de Fora: Universidade Federal de Juiz de Fora; 2016.

DUARTE, D.A.P.; CORRÊA, C.S.L.; FAYER, V.A.; NOGUEIRA, M.C.; BUSTAMANTE-TEIXEIRA, M.T. Avaliação de método para classificação automatizada de pares em relacionamentos probabilísticos de bancos de dados. **Cadernos de Saúde Pública**, v. 35, n.11, e00066419, 2019.

DOMINGUES, E.P.; MAGALHÃES, A.S.; FARIA, W.R. Infraestrutura, crescimento e desigualdade regional: uma projeção dos impactos dos investimentos do Programa de Aceleração do Crescimento (PAC) em Minas Gerais. **PPE**, v. 39, n. 1, p. 121-58, 2009.

EKELUND, S. **Precision-recall curves – what are they and how are they used?** Disponível em: <https://acutecaretesting.org/en/articles/precision-recall-curves-what-are-they-and-how-are-they-used>. Acesso em 03 de junho de 2019.

EL-ZEIN, M.; RICHARDSON, L.; FRANCO, E.L. Cervical cancer screening of HPV vaccinated populations: Cytology, molecular testing, both or none. **Journal of Clinical Virology**, 2016.

EVERETT, T.; BRYANT, A.; GRIFFIN, M.F.; MARTIN-HIRSCH, P.P.; FORBES, C.A.; JEPSON, R.G. Interventions targeted at women to encourage the uptake of cervical screening. **Cochrane Database of Systematic Reviews**, CD002834, 2011.

FARIAS, A.C.B.; BARBIERI, A.R. Seguimento do câncer de colo de útero: Estudo da continuidade da assistência à paciente em uma região de saúde. **Escola Anna Nery**, v. 20, n. 4, 2016.

FERRANTE, A.; BOYD, J. A transparent and transportable methodology for evaluating Data Linkage software. **Journal of Biomedical Informatics**, v. 45; n. 1; p. 165-172, 2012.

FERLAY, J.; SOERJOMATARAM, I.; DIKSHIT, R. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. **International Journal of Cancer**, v.136, E359–E386, 2015.

FONSECA, A.J.; FERREIRA, L.P.; DALLA-BENETTA, A.C.; ROLDAN, C.N.; FERREIRA, M.L.S. Epidemiologia e impacto econômico do câncer de colo de útero no Estado de Roraima: a perspectiva do SUS. **Revista Brasileira de Ginecologia e Obstetria**, v. 32, n. 8, p. 386-392, 2010.

FONSECA, A.J.; DE LIMA FERREIRA, L.C. Systematic review of the cost-effectiveness of the vaccination against HPV in Brazil. **Human Vaccines & Immunotherapeutics**, v.10, n. 12, p. 3484-3490, 2014.

FREITAS, R.A.P.; CARVASAN, G.A.F.; MORAIS, S.S.; ZEFERINO, L.C. Excessive pap smears due to opportunistic cervical cancer screening. **European Journal of Gynaecological Oncology**, n. 5, p. 479-82, 2008.

GALLAGHER, K.E; LAMONTAGNE, D.S; WATSON-JONES, D. Status of HPV vaccine introduction and barriers to country uptake. **Vaccine**, v. 36, p. 4761–4767, 2018.

GASPERIN, S.I.; BOING, A.F.; KUPEK, E. Cobertura e fatores associados à realização do exame de detecção do câncer de colo de útero em área urbana no Sul do Brasil: estudo de base populacional. **Cadernos de Saúde Pública**, v.27, n.7, p.1312-1322, 2011.

GIRIANELLI, V.R.; GAMARRA, C.J.; AZEVEDO e SILVA, G. Os grandes contrastes na mortalidade por câncer do colo uterino e de mama no Brasil. **Revista de Saúde Pública**, v. 48, n. 3, p. 459-67, 2014.

GOLDIE, S.J.; KIM, J.J.; KOBUS, K.; GOLDBERGER-FIEBERT, J.D.; SALOMON, J.; O'SHEA, M.K.; XAVIER BOSCH, F.; DE SANJOSÉ, S.; FRANCO, E.L. Cost-effectiveness of HPV 16, 18 vaccination in Brazil. **Vaccine**, v. 25, p. 6257–6270, 2007.

GONÇALVES, C.V.; SASSI, R.M.; NETTO, I.O.; CASTRO, N.B.; BORTOLOMEDI, A.P. Cobertura do citopatológico do colo uterino em Unidades Básicas de Saúde da Família. **Revista Brasileira de Ginecologia e Obstetria**, v.33, n.9, p.258-63, 2011.

GUERRA, M.R.; BUSTAMANTE-TEIXEIRA, M.T.; CORRÊA, C.S.L.; ABREU, D.M.X.; CURADO, M.P.; MOONEY, M.; *et al.* Magnitude e variação da carga da

mortalidade por câncer no Brasil e Unidades da Federação, 1990 e 2015. **Revista Brasileira de Epidemiologia**, v. 20 (Supl. 1), p. 102-115, 2017.

HARPER, D.M.; DEMARS, L.R. HPV vaccines – A review of the first decade. **Gynecologic Oncology**, v. 146, p. 196–204 , 2017.

HARRON, K.L.; DOIDGE, J.C.; KNIGHT, H.E.; GILBERT, R.E.; GOLDSTEIN, H.; CROMWELL, D.A.*et al.* A guide to evaluating linkage quality for the analysis of linked data. **International Journal of Epidemiology**, v. 46, n. 5, p. 1699–710, 2017.

HOLCOMB, K.; RUNOWICZ, C.D. Cervical Cancer Screening. **Surgical Oncology Clinics of North America**, v. 14, p. 777–797, 2005.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação de Prevenção e Vigilância. **Nomenclatura brasileira para laudos cervicais e condutas preconizadas: recomendações para profissionais de saúde**. - Rio de Janeiro: INCA, 2006.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação Geral de Ações Estratégicas. Divisão de Apoio à Rede de Atenção Oncológica. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero**. Rio de Janeiro: INCA, 2011A.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação Geral de Ações Estratégicas. Divisão de Apoio à Rede de Atenção Oncológica. **Sistema de informação do controle do câncer de mama (SISMAMA) e do câncer do colo do útero (SISCOLO): manual gerencial** – Rio de Janeiro: INCA, 2011B.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). **Monitoramento das ações de controle dos cânceres do colo do útero e de mama**. Informativo Detecção Precoce, Boletim ano 5, n. 1 janeiro/ abril 2014. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//informativo-deteccao-precoce-1-2014.pdf>. Acesso em: 06 fev. 2019. INCA, 2014a.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA. Coordenação Geral de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Ficha técnica de indicadores das ações de controle do câncer do colo do útero** [Internet]. Rio de Janeiro: INCA; 2014b. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//fichatecnicaindicadorescolo14.pdf>. Acesso: fev 2020.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA (INCA). **Monitoramento das ações de controle dos cânceres do colo do útero e de mama** [Internet]. Informativo Detecção Precoce, Boletim ano 6, n.1 janeiro/abril de 2015. Disponível em: https://www.inca.gov.br/bvscontrolecancer/publicacoes/Informativo_deteccao_precoc_e_2015.pdf. Acesso em: 06 fev. 2019. INCA, 2015.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA.
Coordenação de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Diretrizes brasileiras para o rastreamento do câncer do colo do útero – 2. ed. rev. atual.** – Rio de Janeiro: INCA, 2016a.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA.
Coordenação de Prevenção e Vigilância. Divisão de Detecção Precoce e Apoio à Organização de Rede. **Manual de gestão da qualidade para laboratório de citopatologia.** – 2. ed. rev. ampl. – Rio de Janeiro: INCA, 2016b.

INSTITUTO NACIONAL DE CÂNCER JOSÉ ALENCAR GOMES DA SILVA.
Estimativa 2020: incidência de câncer no Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva. Rio de Janeiro: INCA, 2019.

KOLIOPOULOS, G.; NYAGA, V.N.; SANTESSO, N.; BRYANT, A.; MARTIN-HIRSCH, P.P.; MUSTAFA, R.A. *et al.* Cytology versus HPV testing for cervical cancer screening in the general population. **Cochrane Data base of Systematic Reviews**, CD008587, 2017.

LAGE, A.C.; PESSOA, M.C.; MELÉNDEZ, J.G.V. Fatores associados à não realização do Teste de Papanicolaou em Belo Horizonte, Minas Gerais, 2008. **Revista Mineira de enfermagem**, 2013.

LEES, B.F.; ERICKSON, B.K.; HUH, W.K. Cervical cancer screening: evidence behind the guidelines. **American Journal of Obstetrics & Gynecology**, 2016.

LORENZI, A.T.; SYRJÄNEN, K.J.; LONGATTO-FILHO, A. Human papillomavirus (HPV) screening and cervical cancer burden. A Brazilian perspective. **Virology Journal**, 2015.

LUZARDO, A.J.R.; CASTAÑEDA FILHO, R.M.; RUBIM, I.B. Análise espacial exploratória com o emprego do Índice de Moran. **GEOgraphia**, v. 19, n. 40, 2017.

MACHADO, C.J.; HILL, K. Ligação de registro probabilístico e um procedimento automatizado para minimizar o problema do par de correspondência indecisa. **Cadernos de Saúde Pública**, v. 20, n. 4, p. 915-925, 2004.

MAIA, M.N.; SILVA, R.P.O.; SANTOS, L.P.R. A organização do rastreamento do câncer do colo uterino por uma equipe de Saúde da Família no Rio de Janeiro, Brasil. **Revista Brasileira de Medicina de Família e Comunidade**, v 13, n.40, p. 1-10, 2018.

MARTINS, L.F.L.; VALENTE, J.G.; THULER, L.C.S. Factors related to inadequate cervical cancer screening in two Brazilian state capitals. **Revista de Saúde Pública**, v. 43, n. 2, p. 318-325, 2009.

MCCREDIE, M.R.; SHARPLES, K.J.; PAUL, C.; BARANYAI, J.; MEDLEY, G.; JONES, R.W.; SKEGG, D.C. Natural history of cervical neoplasia and risk of

invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. **Lancet Oncology**, p. 425–34, 2008.

MCGRAW, S.L.; FERRANTE, J.M. Update on prevention and screening of cervical cancer. **World Journal of Clinical Oncology**, v. 5, n. 4, p. 744-752, 2014.

MIGOWSKI, A.; CHAVES, R.B.M; COELI, C.M.; RIBEIRO, A.L.P; TURA, B.R; KUSCHNIR, M.C.C. *et al.* Acurácia do relacionamento probabilístico na avaliação da alta complexidade em cardiologia. **Revista de Saúde Pública**, v. 45, n.2, p. 269-75, 2011.

MEIRA, K.C.; FERREIRA, A.A.; SILVA, C.M.F.P.; VALENTE, J.G.; SANTOS J. Mortalidade por câncer do colo do útero no estado de Minas Gerais, análise do efeito da idade-período-coorte de nascimento. **Cadernos Saúde Coletiva**, v. 20, n.3, p. 381-8, 2012.

MOURA, L.; PRESTES, I.V.; DUNCAN, B.B.; SCHMIDT, M.I. Construção de base de dados nacional de pacientes em tratamento dialítico no Sistema Único de Saúde, 2000 – 2012. **Epidemiologia e Serviços de Saúde**, v. 23, n. 2, p. 227-238, 2014.

MUÑOZ, N.; CASTELLSAGUÉ, X.; GONZÁLEZ, A.B.; GISSMANN, L. HPV in the etiology of human câncer. **Vaccine**, v. 24, sup. 3, p. S1-S10, 2006.

NAUCLER, P.; RYD, W.; TÖRNBERG, S.; STRAND, A.; WADELL, G., ELFGREN, K. *et al.* Human papillomavirus and Papanicolaou tests to screen for cervical cancer. **New England Journal of Medicine**, v. 357, n. 16, p. 1589-97, 2007.

NAVARRO, C.; FONSECA, A.J.; SIBAJEV, A.; SOUZA, C.I.A.; ARAÚJO, D.S.; TELES, D.A.F. *et al.* Cobertura do rastreamento do câncer de colo de útero em região de alta incidência. **Revista de Saúde Pública**, v. 49, n. 17, 2015.

NEGRIN, L.G.C. Epidemiology of cervical cancer in Latin America. **Ecancer**, v. 9, p.577, 2015.

NORMAN, A.H.; TESSER, C.D. Prevenção quaternária na atenção primária à saúde: uma necessidade do Sistema Único de Saúde. **Cadernos de Saúde Pública**, v. 25, n. 9, p. 2012-2020, 2009.

NOWAKOWSKI, A.; WOJCIECHOWSKA, U.; WIESZCZY, P.; CYBULSKI, M.; KAMIŃSKI, M.F.; DIDKOWSKA, J. Trends in cervical cancer incidence and mortality in Poland: is there an impact of the introduction of the organised screening? **European Journal of Epidemiology**, v. 32, n. 6, p. 529-32, 2017.

OLIVEIRA, M.M.H.N.; SILVA, A.A.M.; BRITO, L.M.O; COIMBRA, L.C. Cobertura e fatores associados à não realização do exame preventivo de Papanicolaou em São Luís, Maranhão. **Revista Brasileira de Epidemiologia**, v. 9, n. 3, 2006.

OLIVEIRA, M.M.; ANDRADE, S.S.C.A.; OLIVEIRA, P.P.V., AZEVEDO E SILVA, G.; SILVA, M.M.A.; MALTA, D.C. Cobertura de exame Papanicolaou em mulheres de 25 a 64 anos, segundo a Pesquisa Nacional de Saúde e o Sistema de Vigilância de

Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico, 2013. **Revista Brasileira de Epidemiologia**, v. 21, e180014, 2018a.

OLIVEIRA, E.C.A; PIMENTEL, T.J.F.; ARAUJO, J.P.M; OLIVEIRA, L.C.S; FERNANDO, V.C.N. *et al.* Investigação sobre os casos e óbitos por esquistossomose na cidade do Recife, Pernambuco, Brasil, 2005-2013. **Epidemiologia e Serviços de Saúde**, v. 27, n. 4, 2018b.

PALMER, T.; WALLACE, L.; POLLOCK, K.G.; CUSCHIERI, K.; ROBERTSON, C.; KAVANAGH, K.; CRUICKSHANK, M. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12-13 in Scotland: retrospective population study. **BMJ**, 2019.

PARIKH, S.; BRENNAN, P.; BOFFETTA, P. Meta-analysis of social inequality and the risk of cervical cancer. **International Journal of Cancer**, v. 105, n.5, p.687-91, 2003.

PEIRSON, L.; FITZPATRICK-LEWIS, D.; CILISKA, D.; WARREN, R. Screening for cervical cancer: a systematic review and meta-analysis. **Systematic Reviews**, v. 24, 2013.

PERES, S.V.; LATORRE, M.R.D.O.; MICHELS, F.A.S.; TANAKA, L.F.; COELI, C.M; ALMEIDA, M.F. Determinação de um ponto de corte para a identificação de pares verdadeiros pelo método probabilístico de linkage de base de dados. **Cadernos Saúde Coletiva**, v. 22, n. 4, p. 428-36, 2014.

PESOLA, F.; SASIENI, P. Impact of screening on cervical cancer incidence in England: a time trend analysis. **BMJ Open**, 2019.

PITA, R.; PINTO, C; SENA, S.; FIACCONE, R.; AMORIM, L; REIS, S., *et al.* On the Accuracy and Scalability of Probabilistic Data Linkage Over the Brazilian 114 Million Cohort. **IEEE Journal of Biomedical and Health Informatics**, v. 22, n. 2, 2018.

POSSATI-RESENDE, J.C.; VAZQUEZ, F.L.; BIOT, S.T.; MAUAD, E.C.; TALARICO, T.; FREGNANI, J.H.T.G.; LONGATTO-FILHO, A. Organized Cervical Cancer Screening Program in Barretos, Brazil: Experience in 18 Municipalities of São Paulo State. **Acta Cytologica**, v.62, p.19-27, 2018.

RANDALL, S.M.; FERRANTE, A.M.; BOYD, J.H. *et al.* The effect of data cleaning on record linkage quality. **BMC Medical Informatics and Decision Making**, v.13, n. 64, 2013.

REIS, N.V.S; ANDRADE, B.B.; GUERRA, M.R.; *et al.* The Global Burden of Disease Study Estimates of Brazil's Cervical Cancer Burden. **Annals of Global Health**, v.86, n.1, p.1-12, 2020.

RENNA JUNIOR, N.L.; AZEVEDO e SILVA, G. Tendências temporais e fatores associados ao diagnóstico em estágio avançado de câncer do colo uterino: análise dos dados dos registros hospitalares de câncer no Brasil, 2000-2012. **Epidemiologia e Serviços de Saúde**, v. 27, n.2, e2017285, 2018.

RIBEIRO, C.M. **Rastreamento do câncer do colo do útero: uma análise da continuidade do cuidado com base nos sistemas de informação do SUS. 2019. 160 p.** Tese (Doutorado em Saúde Coletiva) - Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2019.

RIBEIRO, C.M.; AZEVEDO e SILVA, G. Avaliação da produção de procedimentos da linha de cuidado do câncer do colo do útero no Sistema Único de Saúde do Brasil em 2015. **Epidemiologia e Serviços de Saúde**, v .27, n.1, e20172124, 2018.

RIBEIRO, C.M.; DIAS, M.B.K.; PLA, M.A.S; CORREA, F.M.; RUSSOMANO, F.B.; TOMAZELLI, J.G. Parâmetros para a programação de procedimentos da linha de cuidado do câncer do colo do útero no Brasil. **Cadernos de Saúde Pública**, v. 35, n. 6, 2019.

RODRÍGUEZ, A.C.; SCHIFFMAN, M.; HERRERO, R.; WACHOLDER, S.; HILDESHEIM, A.; CASTLE, P.E.; *et al.* Rapid Clearance of Human Papillomavirus and Implications for Clinical Focus on Persistent Infections. **Journal of the National Cancer Institute**, v. 100, n. 7, p. 513–517, 2008.

RODRIGUES, A.D.; BUSTAMANTE-TEIXEIRA, M.T. Mortalidade por câncer de mama e câncer de colo do útero em município de porte médio da Região Sudeste do Brasil, 1980-2006. **Cadernos de Saúde Pública**, v. 27, n. 2, p. 241-248, 2011.

SADOVSKY, A.D.I.; POTON, W.L.; REIS-SANTOS, B.; BARCELOS, M.R.B.; SILVA, I.C.M. Índice de Desenvolvimento Humano e prevenção secundária de câncer de mama e colo do útero: um estudo ecológico. **Cadernos de Saúde Pública**, v. 31, n. 7, p. 1539-1550, 2015.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. **PLoS One**, v. 10, n. 3:e0118432, 2015.

SANKARANARAYANAN, R.; SOMANATHAN, T.; SANTOS, C.; MUWONGE, R.; TWALIB, N. O desafio de controlar o cancer do colo do útero no mundo em desenvolvimento. In: COELHO, F.R.G.; SOARES, F.A.; FOCCHI, J.; FREGNANI, J.H.T.G; ZEFERINO, L.C.; VILLA, L.L.; FEDERICO, M.H.; NOVAES, P.E.R.S; COSTA, R.L.R. **Câncer do colo do útero**. Tecmedd, São Paulo – SP, 2008.

SANKARANARAYANAN HARPER, R.; JOSHI, S.; MUWONGE, R.; ESMY, P.O.; BASU, P.; PRABHU, P. Can a single dose of human papillomavirus (HPV) vaccine prevent cervical cancer? Early findings from an Indian study. **Vaccine**, v. 36, p. 4783–4791, 2018.

SCHIFFMAN, M.; WENTZENSEN, N. Human papillomavirus infection and the multistage carcinogenesis of cervical cancer. **Cancer Epidemiology, Biomarkers & Prevention**, v. 22, n. 4, p. 553-60, 2013.

SILVEIRA, D.P.; ARTMANN, E. Acurácia em métodos de relacionamento probabilístico de bases de dados em saúde: revisão sistemática. **Revista de Saúde Pública**, v. 43, n. 5, p. 875-82, 2009.

SINGH, G.K., AZUINE, R.E., SIAHPUSH, M. Global inequalities in cervical cancer incidence and mortality are linked to deprivation, low socioeconomic status, and human development. **International Journal of MCH and AIDS**, v. 1, p.17–30, 2012.

SPINETI, P.P.M; SOUZA, A.S.; FEIJO, L.A.; GARCIA, M.I.; XAVIER, S.S. Acurácia do relacionamento probabilístico de registros na identificação de óbitos em uma coorte de pacientes com insuficiência cardíaca descompensada. **Cadernos de Saúde Pública**, v. 32, n. 1, e00097415, 2016.

SOUSA, M.H; CECATTI, J.G.; HARDY, E.; SERRUYA, S.J. Relacionamento probabilístico de registros: uma aplicação na área de morbidade materna grave (near miss) e mortalidade materna. **Cadernos de Saúde Pública**, v. 24, n. 3, p. 653-662, 2008.

SUNG, H.; FERLAY, J.; SIEGEL, R.L. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 71, p. 209-249, 2021.

TEIXEIRA, L.A.; PORTO, M.A.T; SOUZA, L.P.A. A expansão do rastreamento do câncer do colo do útero e a formação de citotécnicos no Brasil. **Physis Revista de Saúde Coletiva**, v. 22, n. 2, p. 713-731, 2012.

TEIXEIRA, L.A. Dos gabinetes de ginecologia às campanhas de rastreamento: a trajetória da prevenção ao câncer de colo do útero no Brasil. **História, Ciências, Saúde**, v.22, n.1, p.221-240, 2015.

THEME-FILHA, M.M.T.; LEAL, M.C.; OLIVEIRA, E.F.V.; ESTEVES-PEREIRA, A.P.; GAMA, S.G.N. Regional and social inequalities in the performance of Pap test and screening mammography and their correlation with lifestyle: Brazilian national health survey, 2013. **International Journal for Equity in Health**, v. 15, n. 1, p.136, 2016.

THULER, L.C.S.; AGUIAR, S.S.; BERGMANN, A. Determinantes do diagnóstico em estadios avançados do câncer do colo do útero no Brasil. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 36, n. 6, 237-243, 2014.

TOMAZELLI, J.G.; GIRIANELLI, V.R.; AZEVEDO e SILVA, G. Estratégias usadas no relacionamento entre Sistemas de Informações em Saúde para seguimento das mulheres com mamografias suspeitas no Sistema Único de Saúde. **Revista Brasileira de Epidemiologia**, v. 21, e180015, 2018.

TOTA, J.E.; BENTLEY, J.; BLAKE, J.; COUPLÉE, F.; DUGGAN, M.A.; FERENCZY, A. *et al.* Introduction of molecular HPV testing as the primary technology in cervical cancer screening: Acting on evidence to change the current paradigm. **Preventive Medicine**, v. 98, p. 5-14, 2017.

TORRE, L.A.; BRAY, F.; SIEGEL, R.L.; FERLAY, J.; LORTET-TIEULENT, J.; JEMAL, A. Global Cancer Statistics, 2012. **CA: A Cancer Journal for Clinicians**, v. 65, n. 2, p.87-108, 2015.

VALE, D.B.A.P.; MORAIS, S.S.; PIMENTA, A.L.; ZEFERINO, L.C. Avaliação do rastreamento do câncer do colo do útero na Estratégia Saúde da Família no Município de Amparo, São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 26, n. 2, p. 383-90, 2010.

VICTORA, C.G. et al. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. **International Journal of Epidemiology**, v. 26, n.1, p.224-227, 1997.

ZEFERINO, L.C.; BEDONE, A.J.; FAÚNDES, A.; OYAKAWA, N. Duração da Neoplasia Intra-Epitelial e do Carcinoma Invasor do Colo Uterino: Estudo Epidemiológico. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 20, n. 10, p. 565-569, 1998.

ZEFERINO, L.C.; BASTOS, J.B.; VALE, D.B.A.P.; ZANINE, R.M.; MELO, Y.L.M.F.; PRIMO, W.Q.S.P.; *et al.* Guidelines for HPV-DNA Testing for Cervical Cancer Screening in Brazil. **Revista Brasileira de Ginecologia e Obstetrícia**, v. 40, n. 6, p. 360-368, 2018.

WALBOOMERS, J.M.; JACOBS, M.V.; MANOS, M.M. *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. **The Journal of Pathology**, v. 189, n. 1, p.12-9, 1999.

WORLD HEALTH ORGANIZATION. **National cancer control programmes: policies and managerial guidelines**. 2 ed. Geneva: WHO; 2002.

WORLD HEALTH ORGANIZATION. **Comprehensive cervical cancer control: a guide to essential practice**. 2 ed. Geneva: WHO; 2014.

APÊNDICES

APÊNDICE 1 – PREPARO, LIMPEZAS E PADRONIZAÇÃO DA BASE 2008 NO PROGRAMA R

```
# Ler os bancos de dados no formato dbf
```

```
C200801 <- read.dbf("C200801.dbf")
C200802 <- read.dbf("C200802.dbf")
C200803 <- read.dbf("C200803.dbf")
C200804 <- read.dbf("C200804.dbf")
C200805 <- read.dbf("C200805.dbf")
C200806 <- read.dbf("C200806.dbf")
C200807 <- read.dbf("C200807.dbf")
C200808 <- read.dbf("C200808.dbf")
C200809 <- read.dbf("C200809.dbf")
C200810 <- read.dbf("C200810.dbf")
C200811 <- read.dbf("C200811.dbf")
C200812 <- read.dbf("C200812.dbf")
```

```
# Juntar os bancos mensais de 2008
```

```
banco2008<-
```

```
rbind(C200801,C200802,C200803,C200804,C200805,C200806,C200807,C200808,C200809,C200810,C200811,C200812)
```

banco2008	1277203 obs. of 102 variables
C200801	93836 obs. of 102 variables
C200802	88868 obs. of 102 variables
C200803	106578 obs. of 102 variables
C200804	119196 obs. of 102 variables
C200805	104995 obs. of 102 variables
C200806	109119 obs. of 102 variables
C200807	118046 obs. of 102 variables
C200808	103102 obs. of 102 variables
C200809	113435 obs. of 102 variables
C200810	111929 obs. of 102 variables
C200811	106291 obs. of 102 variables
C200812	101808 obs. of 102 variables

```
# Transformando as variáveis com data no formato data (mm/dd/aaaa) -> criação das variáveis D_COL_US2 (data de coleta 2, no formato data) e DIDDTNASC2 (Dt de nascimento 2, no formato data)
```

```
library(zoo)
```

```
banco2008$D_COL_US2 <- as.Date(banco2008$D_COL_US, "%d/%m/%Y")
```

```
banco2008$DIDDTNASC2 <- as.Date(banco2008$DIDDTNASC, "%d/%m/%Y")
```

```
# Extraindo o ano da data de coleta do exame
```

```
library(lubridate)
```

```
banco2008$ano <- year(banco2008$D_COL_US2)
```

```
# Ver variáveis do banco
```

```
names(banco2008)
```

```
[1] "C_CNES" "C_COMP" "C_EXAME" "C_ID_NOME" "CIDNOMEFON"
 [6] "CIDNOMEMAE" "IDNOMMAEFO" "C_ID_APEL" "DIDDTNASC" "C_ID_IDAD"
[11] "C_ID_IDENT" "C_ID_EMIT" "C_ID_UFIDE" "C_ID_CIC" "C_ID_ESCO"
[16] "IDENDERECO" "CIDNUMERO" "CIDCOMPLEM" "CIDBAIRRO" "C_ID_UF"
[21] "C_IBGE" "C_ID_CEP" "C_ID_REFE" "C_ID_FONE" "C_ID_SUS"
[26] "C_US_UF" "C_US_IBGE" "C_US_CNES" "C_US_NOME" "C_PRON"
[31] "C_ANM_PREV" "C_ANM_PANO" "C_ANM_DIU" "C_ANM_GRAV" "C_ANM_CONC"
```

```
[36] "C_ANM_HORM" "C_ANM_RADI" "C_ANM_DMES" "C_ANM_RSEX" "C_ANM_MENO"
[41] "C_CLI_INSP" "C_CLI_SINA" "C_RES_NORM" "C_BEM_INFL" "CBEMMETAS"
[46] "CBEMREPAR" "CBEMATROF" "CBEMRADIA" "C_BEM_OUT" "C_BEM_OUTR"
[51] "C_MIC_LACT" "C_MIC_COCO" "C_MIC_BACI" "C_MIC_CHLA" "C_MIC_ACTI"
[56] "C_MIC_CAND" "C_MIC_TRIC" "C_MIC_HERP" "C_MIC_GARD" "C_MIC_OUT"
[61] "C_MIC_OUTR" "C_NEO_MALI" "CNEOMOUT" "C_CEL_ESCA" "CRESTIPAN"
[66] "C_REJ_AUS" "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT"
[71] "C_REJ_OESP" "C_EPI_ESC" "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT"
[76] "C_ADQ_ACEL" "C_ADQ_SANG" "C_ADQ_PIO" "C_ADQ_ART" "C_ADQ_CONT"
[81] "C_ADQ_INT" "C_ADQ_OUT" "C_ADQ_OUTR" "C_ATI_ESC" "C_ATI_GLA"
[86] "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "C_OBS_GER" "D_COL_US"
[91] "D_RECEBE" "D_LIBERA" "C_PAT_CNS" "C_PAT_CIC" "C_FXET"
[96] "C_ID_FIEXT" "C_FCOM" "C_ID_COMP" "C_FLG_EXT" "C_FLG_POSI"
[101] "NSEQUENCIA" "H_EVENTO" "D_COL_US2" "DIDDTNASC2" "ano"
```

```
# Ver variável ano no banco, usando NA para células vazias
table(banco2008$ano,useNA = "ifany")
200 207 208 1973 1991 2000 2001 2002 2003
 34  4 10  1  1 134 19 24 40
2004 2005 2006 2007 2008 <NA>
 26 207 292 74633 1201766 12
```

```
# Selecionando apenas exames realizados em 2008
banco2008date <- banco2008[banco2008$ano=="2008",]
```

```
# Ver variável ano no banco, usando NA para células vazias
table(banco2008date$ano,useNA = "ifany")
2008 <NA>
1201766 12
```

```
# Excluir registros com dado faltante para a variável ano, criando o Banco2008_1
Banco2008_1 <- banco2008date[complete.cases(banco2008date[,c("ano")]),]
```

```
# Ver variável ano no banco, usando NA para células vazias
table(Banco2008_1$ano,useNA = "ifany")
2008
1201766
```

```
# Selecionando exames realizados em 2007 no banco de 2008
Banco2008_2007 <- banco2008[banco2008$ano=="2007",]
```

```
# Ver variável ano no banco, usando NA para células vazias
table(Banco2008_2007$ano,useNA = "ifany")
```

```
2007 <NA>
74633 12
```

```
# Excluir registros com dado faltante para a variável ano, criando o Banco 2008_2007_02
Banco2008_2007_2 <- Banco2008_2007[complete.cases(Banco2008_2007[,c("ano")]),]
```

```
# Ver variável ano no banco, usando NA para células vazias
table(Banco2008_2007_2$ano,useNA = "ifany")
```

```
2007
74633
```

Data	
banco2008	1277203 obs. of 105 variables
Banco2008_1	1201766 obs. of 105 variables
Banco2008_2007	74645 obs. of 105 variables
Banco2008_2007_2	74633 obs. of 105 variables
banco2008date	1201778 obs. of 105 variables

EXAMES REALIZADOS EM 2008 PROCESSADOS EM OUTROS ANOS

```
Banco2008_1 <- read.dbf("Banco2008_1.dbf")
Banco2009_2008_1 <- read.dbf("Banco2009_2008_1.dbf")
Banco2010_2008_1 <- read.dbf("Banco2010_2008_1.dbf")
Banco2011_2008_1 <- read.dbf("Banco2011_2008_1.dbf")
Banco2012_2008_1 <- read.dbf("Banco2012_2008_1.dbf")
```

Juntar os bancos contendo exames realizados em 2008

```
Banco2008_total<-
rbind(Banco2008_1,Banco2009_2008_1,Banco2010_2008_1,Banco2011_2008_1,Banco2012_2008_1)
```

Banco2008_1	1201766 obs. of 105 variables
Banco2009_2008_1	55073 obs. of 105 variables
Banco2010_2008_1	68 obs. of 105 variables
Banco2011_2008_1	12 obs. of 105 variables
Banco2012_2008_1	2 obs. of 105 variables
Banco2008_total1	1256921 obs. of 105 variables

Limpezas Banco2008

Renomear o banco

```
Banco2008total1 <- Banco2008total
```

Banco2008total	1256921 obs. of 105 variables
Banco2008total1	1256921 obs. of 105 variables

```
names(Banco2008total1)
```

```
[1] "C_CNES" "C_COMP" "C_EXAME" "C_ID_NOME" "CIDNOMEFON" "CIDNOMEMAE"
[7] "IDNOMMAEFO" "C_ID_APEL" "DIDDTNASC" "C_ID_IDAD" "C_ID_IDENT" "C_ID_EMIS"
[13] "C_ID_UFIDE" "C_ID_CIC" "C_ID_ESCO" "IDENDERECO" "CIDNUMERO" "CIDCOMPLEM"
[19] "CIDBAIRRO" "C_ID_UF" "C_IBGE" "C_ID_CEP" "C_ID_REFE" "C_ID_FONE"
[25] "C_ID_SUS" "C_US_UF" "C_US_IBGE" "C_US_CNES" "C_US_NOME" "C_PRON"
[31] "C_ANM_PREV" "C_ANM_PANO" "C_ANM_DIU" "C_ANM_GRAV" "C_ANM_CONC" "C_ANM_HORM"
[37] "C_ANM_RADI" "C_ANM_DMES" "C_ANM_RSEX" "C_ANM_MENO" "C_CLI_INSP" "C_CLI_SINA"
[43] "C_RES_NORM" "C_BEM_INFL" "CBEMMETAS" "CBEMREPAR" "CBEMATROF" "CBEMRADIA"
[49] "C_BEM_OUT" "C_BEM_OUTR" "C_MIC_LACT" "C_MIC_COCO" "C_MIC_BACI" "C_MIC_CHLA"
[55] "C_MIC_ACTI" "C_MIC_CAND" "C_MIC_TRIC" "C_MIC_HERP" "C_MIC_GARD" "C_MIC_OUT"
[61] "C_MIC_OUTR" "C_NEO_MALI" "CNEOMOUT" "C_CEL_ESCA" "CRESTIPAN" "C_REJ_AUS"
[67] "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT" "C_REJ_OESP" "C_EPI_ESC"
[73] "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT" "C_ADQ_ACEL" "C_ADQ_SANG" "C_ADQ_PIO"
[79] "C_ADQ_ART" "C_ADQ_CONT" "C_ADQ_INT" "C_ADQ_OUT" "C_ADQ_OUTR" "C_ATI_ESC"
[85] "C_ATI_GLA" "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "C_OBS_GER" "D_COL_US"
[91] "D_RECEBE" "D_LIBERA" "C_PAT_CNS" "C_PAT_CIC" "C_FXET" "C_ID_FIEXT"
[97] "C_FCOM" "C_ID_COMP" "C_FLG_EXT" "C_FLG_POSI" "NSEQUENCIA" "H_EVENTO"
[103] "D_COL_US2" "DIDDTNASC2" "ano"
```

```
# Extrair o mês da data de coleta do exame
library(lubridate)
Banco2008total1$mes <- month(Banco2008total1$D_COL_US2)

> names(Banco2008total1)
[1] "C_CNES" "C_COMP" "C_EXAME" "C_ID_NOME" "CIDNOMEFON" "CIDNOMEMAE"
[7] "IDNOMMAEFO" "C_ID_APEL" "DIDDTNASC" "C_ID_IDAD" "C_ID_IDENT" "C_ID_EMIS"
[13] "C_ID_UFIDE" "C_ID_CIC" "C_ID_ESCO" "IDENDERECO" "CIDNUMERO" "CIDCOMPLEM"
[19] "CIDBAIRRO" "C_ID_UF" "C_IBGE" "C_ID_CEP" "C_ID_REFE" "C_ID_FONE"
[25] "C_ID_SUS" "C_US_UF" "C_US_IBGE" "C_US_CNES" "C_US_NOME" "C_PRON"
[31] "C_ANM_PREV" "C_ANM_PANO" "C_ANM_DIU" "C_ANM_GRAV" "C_ANM_CONC" "C_ANM_HORM"
[37] "C_ANM_RADI" "C_ANM_DMES" "C_ANM_RSEX" "C_ANM_MENO" "C_CLI_INSP" "C_CLI_SINA"
[43] "C_RES_NORM" "C_BEM_INFL" "CBEMMETAS" "CBEMREPAR" "CBEMATROF" "CBEMRADIA"
[49] "C_BEM_OUT" "C_BEM_OUTR" "C_MIC_LACT" "C_MIC_COCO" "C_MIC_BACI" "C_MIC_CHLA"
[55] "C_MIC_ACTI" "C_MIC_CAND" "C_MIC_TRIC" "C_MIC_HERP" "C_MIC_GARD" "C_MIC_OUT"
[61] "C_MIC_OUTR" "C_NEO_MALI" "CNEOMOUT" "C_CEL_ESCA" "CRESTIPAN" "C_REJ_AUS"
[67] "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT" "C_REJ_OESP" "C_EPI_ESC"
[73] "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT" "C_ADQ_ACEL" "C_ADQ_SANG" "C_ADQ_PIO"
[79] "C_ADQ_ART" "C_ADQ_CONT" "C_ADQ_INT" "C_ADQ_OUT" "C_ADQ_OUTR" "C_ATI_ESC"
[85] "C_ATI_GLA" "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "C_OBS_GER" "D_COL_US"
[91] "D_RECEBE" "D_LIBERA" "C_PAT_CNS" "C_PAT_CIC" "C_FXET" "C_ID_FIEXT"
[97] "C_FCOM" "C_ID_COMP" "C_FLG_EXT" "C_FLG_POSI" "NSEQUENCIA" "H_EVENTO"
[103] "D_COL_US2" "DIDDTNASC2" "ano" "mes"
```

```
# Criar nova variável (codmun6) extraíndo os 6 primeiros dígitos da variável código do município IBGE
library(stringr)
Banco2008total1$codmun6 <- as.numeric(substr(Banco2008total1$C_IBGE,1,6))
```

```
> names(Banco2008total1)
[1] "C_CNES" "C_COMP" "C_EXAME" "C_ID_NOME" "CIDNOMEFON" "CIDNOMEMAE"
[7] "IDNOMMAEFO" "C_ID_APEL" "DIDDTNASC" "C_ID_IDAD" "C_ID_IDENT" "C_ID_EMIS"
[13] "C_ID_UFIDE" "C_ID_CIC" "C_ID_ESCO" "IDENDERECO" "CIDNUMERO" "CIDCOMPLEM"
[19] "CIDBAIRRO" "C_ID_UF" "C_IBGE" "C_ID_CEP" "C_ID_REFE" "C_ID_FONE"
[25] "C_ID_SUS" "C_US_UF" "C_US_IBGE" "C_US_CNES" "C_US_NOME" "C_PRON"
[31] "C_ANM_PREV" "C_ANM_PANO" "C_ANM_DIU" "C_ANM_GRAV" "C_ANM_CONC" "C_ANM_HORM"
[37] "C_ANM_RADI" "C_ANM_DMES" "C_ANM_RSEX" "C_ANM_MENO" "C_CLI_INSP" "C_CLI_SINA"
[43] "C_RES_NORM" "C_BEM_INFL" "CBEMMETAS" "CBEMREPAR" "CBEMATROF" "CBEMRADIA"
[49] "C_BEM_OUT" "C_BEM_OUTR" "C_MIC_LACT" "C_MIC_COCO" "C_MIC_BACI" "C_MIC_CHLA"
[55] "C_MIC_ACTI" "C_MIC_CAND" "C_MIC_TRIC" "C_MIC_HERP" "C_MIC_GARD" "C_MIC_OUT"
[61] "C_MIC_OUTR" "C_NEO_MALI" "CNEOMOUT" "C_CEL_ESCA" "CRESTIPAN" "C_REJ_AUS"
[67] "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT" "C_REJ_OESP" "C_EPI_ESC"
[73] "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT" "C_ADQ_ACEL" "C_ADQ_SANG" "C_ADQ_PIO"
[79] "C_ADQ_ART" "C_ADQ_CONT" "C_ADQ_INT" "C_ADQ_OUT" "C_ADQ_OUTR" "C_ATI_ESC"
[85] "C_ATI_GLA" "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "C_OBS_GER" "D_COL_US"
[91] "D_RECEBE" "D_LIBERA" "C_PAT_CNS" "C_PAT_CIC" "C_FXET" "C_ID_FIEXT"
[97] "C_FCOM" "C_ID_COMP" "C_FLG_EXT" "C_FLG_POSI" "NSEQUENCIA" "H_EVENTO"
[103] "D_COL_US2" "DIDDTNASC2" "ano" "mes" "codmun6"
```

```
# Ler arquivo contendo todos os municípios e respectivas micro e macrorregiões de saúde de MG -> arquivo no
formato csv, considerando a primeira linha como cabeçalho
tab_mun_cir_mac <- read.csv2("tab_mun_cir_mac.csv", header = T)
tab_mun_cir_mac$codmun6 <- as.numeric(tab_mun_cir_mac$MUN_COD)
```

```
# Incluir macro e micro no Banco fazendo merge com a planilha - variável chave: codmun6 do primeiro banco
Banco2008total2 <- merge(Banco2008total1,tab_mun_cir_mac,by="codmun6",all.x=TRUE)
```

```
> names(Banco2008total2)
[1] "codmun6" "C_CNES" "C_COMP" "C_EXAME" "C_ID_NOME" "CIDNOMEFON"
[7] "CIDNOMEMAE" "IDNOMMAEFO" "C_ID_APEL" "DIDDTNASC" "C_ID_IDAD" "C_ID_IDENT"
[13] "C_ID_EMIS" "C_ID_UFIDE" "C_ID_CIC" "C_ID_ESCO" "IDENDERECO" "CIDNUMERO"
[19] "CIDCOMPLEM" "CIDBAIRRO" "C_ID_UF" "C_IBGE" "C_ID_CEP" "C_ID_REFE"
[25] "C_ID_FONE" "C_ID_SUS" "C_US_UF" "C_US_IBGE" "C_US_CNES" "C_US_NOME"
[31] "C_PRON" "C_ANM_PREV" "C_ANM_PANO" "C_ANM_DIU" "C_ANM_GRAV" "C_ANM_CONC"
[37] "C_ANM_HORM" "C_ANM_RADI" "C_ANM_DMES" "C_ANM_RSEX" "C_ANM_MENO" "C_CLI_INSP"
[43] "C_CLI_SINA" "C_RES_NORM" "C_BEM_INFL" "CBEMMETAS" "CBEMREPAR" "CBEMATROF"
[49] "CBEMRADIA" "C_BEM_OUT" "C_BEM_OUTR" "C_MIC_LACT" "C_MIC_COCCO" "C_MIC_BACI"
[55] "C_MIC_CHLA" "C_MIC_ACTI" "C_MIC_CAND" "C_MIC_TRIC" "C_MIC_HERP" "C_MIC_GARD"
[61] "C_MIC_OUT" "C_MIC_OUTR" "C_NEO_MALI" "CNEOMOUT" "C_CEL_ESCA" "CRESTIPAN"
[67] "C_REJ_AUS" "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT" "C_REJ_OESP"
[73] "C_EPI_ESC" "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT" "C_ADQ_ACEL" "C_ADQ_SANG"
[79] "C_ADQ_PIO" "C_ADQ_ART" "C_ADQ_CONT" "C_ADQ_INT" "C_ADQ_OUT" "C_ADQ_OUTR"
[85] "C_ATI_ESC" "C_ATI_GLA" "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "C_OBS_GER"
[91] "D_COL_US" "D_RECEBE" "D_LIBERA" "C_PAT_CNS" "C_PAT_CIC" "C_FXET"
[97] "C_ID_FIEXT" "C_FCOM" "C_ID_COMP" "C_FLG_EXT" "C_FLG_POSI" "NSEQUENCIA"
[103] "H_EVENTO" "D_COL_US2" "DIDDTNASC2" "ano" "mes" "MUN_COD"
[109] "MUNICIPIO" "CIR_COD" "CIR" "MACRO_COD" "MACRO"
```

```
# Criar variável estado extraindo os 2 primeiros dígitos da variável codmun6
```

```
library(stringr)
```

```
Banco2008total2$estado <- as.numeric(substr(Banco2008total2$codmun6,1,2))
```

```
# Frequência variável estado
```

```
table(Banco2008total2$estado,useNA = "ifany")
```

```
> table(Banco2008total2$estado,useNA = "ifany")
```

```
 11  21  22  24  27  29  31  32  33  35  41
 1  64  1  1  3  1 1256790  1  7  9  1
42  50  52
 2  1  39
```

```
# Criar novo banco apenas com mulheres residentes em MG
```

```
Banco2008total3 <- Banco2008total2[Banco2008total2$estado=="31",]
```

Banco2008total1	1256921 obs. of 107 variables
Banco2008total2	1256921 obs. of 114 variables
Banco2008total3	1256790 obs. of 114 variables
tab_mun_cir_mac	853 obs. of 7 variables

```
# Criar variável identificadora
```

```
Banco2008total4 <- Banco2008total3
```

```
Banco2008total4$FONTE<-"C08"
```

```
Banco2008total4$REGISTRO<-1:1256790
```

```
Banco2008total4$ID<-paste(Banco2008total4$FONTE, sep=" ", Banco2008total4$REGISTRO)
```

```
names(Banco2008total4)
```

```
names(Banco2008total4)
```

```
[1] "codmun6" "C_CNES" "C_COMP" "C_EXAME" "C_ID_NOME" "CIDNOMEFON"
[7] "CIDNOMEMAE" "IDNOMMAEFO" "C_ID_APEL" "DIDDTNASC" "C_ID_IDAD" "C_ID_IDENT"
[13] "C_ID_EMIS" "C_ID_UFIDE" "C_ID_CIC" "C_ID_ESCO" "IDENDERECO" "CIDNUMERO"
[19] "CIDCOMPLEM" "CIDBAIRRO" "C_ID_UF" "C_IBGE" "C_ID_CEP" "C_ID_REFE"
[25] "C_ID_FONE" "C_ID_SUS" "C_US_UF" "C_US_IBGE" "C_US_CNES" "C_US_NOME"
```

```
[31] "C_PRON" "C_ANM_PREV" "C_ANM_PANO" "C_ANM_DIU" "C_ANM_GRAV" "C_ANM_CONC"
[37] "C_ANM_HORM" "C_ANM_RADI" "C_ANM_DMES" "C_ANM_RSEX" "C_ANM_MENO" "C_CLI_INSP"
[43] "C_CLI_SINA" "C_RES_NORM" "C_BEM_INFL" "CBEMMETAS" "CBEMREPAR" "CBEMATROF"
[49] "CBEMRADIA" "C_BEM_OUT" "C_BEM_OUTR" "C_MIC_LACT" "C_MIC_COCO" "C_MIC_BACI"
[55] "C_MIC_CHLA" "C_MIC_ACTI" "C_MIC_CAND" "C_MIC_TRIC" "C_MIC_HERP" "C_MIC_GARD"
[61] "C_MIC_OUT" "C_MIC_OUTR" "C_NEO_MALI" "CNEOMOUT" "C_CEL_ESCA" "CRESTIPAN"
[67] "C_REJ_AUS" "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT" "C_REJ_OESP"
[73] "C_EPI_ESC" "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT" "C_ADQ_ACEL" "C_ADQ_SANG"
[79] "C_ADQ_PIO" "C_ADQ_ART" "C_ADQ_CONT" "C_ADQ_INT" "C_ADQ_OUT" "C_ADQ_OUTR"
[85] "C_ATI_ESC" "C_ATI_GLA" "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "C_OBS_GER"
[91] "D_COL_US" "D_RECEBE" "D_LIBERA" "C_PAT_CNS" "C_PAT_CIC" "C_FXET"
[97] "C_ID_FIEXT" "C_FCOM" "C_ID_COMP" "C_FLG_EXT" "C_FLG_POSI" "NSEQUENCIA"
[103] "H_EVENTO" "D_COL_US2" "DIDDTNASC2" "ano" "mes" "MUN_COD"
[109] "MUNICIPIO" "CIR_COD" "CIR" "MACRO_COD" "MACRO" "estado"
[115] "FONTE" "REGISTRO" "ID"
```

```
# Selecionando apenas variáveis de interesse (reduzir tamanho do banco)
```

```
Banco2008total5 <- subset(Banco2008total4, select =
c(codmun6,C_CNES,C_EXAME,C_ID_NOME,CIDNOMEMAE,DIDDTNASC,C_ID_IDAD,C_ID_CIC,C_ID_ESCO,IDEND
ERECO,CIDBAIRRO,C_ID_CEP,C_ID_SUS,C_US_CNES,C_ANM_PREV,C_ANM_PANO,C_NEO_MALI,CNEOMOUT,C_
CEL_ESCA,C_REJ_AUS,C_REJ_DAN,C_REJ_ALH,C_REJ_AESP,C_REJ_OUT,C_REJ_OESP,C_EPI_ESC,C_EPI_GLA,C_EP
I_MET,C_ADQ_MAT,C_ATI_ESC,C_ATI_GLA,C_ATI_IND,C_GLA_ADEN,C_CEL_ENDO,D_RECEBE,D_LIBERA,D_COL
_US2,DIDDTNASC2,ano,MUNICIPIO,CIR_COD,CIR,MACRO_COD,MACRO,estado,mes,ID))
```

Banco2008total4	1256790 obs. of 117 variables
Banco2008total5	1256790 obs. of 47 variables

```
# Transformar Banco2008total5 em db
```

```
db<-Banco2008total5
```

```
names(db)
```

```
> names(db)
```

```
[1] "codmun6" "C_CNES" "C_EXAME" "C_ID_NOME" "CIDNOMEMAE" "DIDDTNASC"
[7] "C_ID_IDAD" "C_ID_CIC" "C_ID_ESCO" "IDENDERECO" "CIDBAIRRO" "C_ID_CEP"
[13] "C_ID_SUS" "C_US_CNES" "C_ANM_PREV" "C_ANM_PANO" "C_NEO_MALI" "CNEOMOUT"
[19] "C_CEL_ESCA" "C_REJ_AUS" "C_REJ_DAN" "C_REJ_ALH" "C_REJ_AESP" "C_REJ_OUT"
[25] "C_REJ_OESP" "C_EPI_ESC" "C_EPI_GLA" "C_EPI_MET" "C_ADQ_MAT" "C_ATI_ESC"
[31] "C_ATI_GLA" "C_ATI_IND" "C_GLA_ADEN" "C_CEL_ENDO" "D_RECEBE" "D_LIBERA"
[37] "D_COL_US2" "DIDDTNASC2" "ano" "MUNICIPIO" "CIR_COD" "CIR"
[43] "MACRO_COD" "MACRO" "estado" "mes" "ID"
```

```
# Renomear variáveis
```

```
names(db)[1] <- "COD_MUN"
names(db)[2] <- "CNES_PREST"
names(db)[3] <- "NUM_EXAME"
names(db)[4] <- "NOME"
names(db)[5] <- "MAE"
names(db)[6] <- "DT_NASC_BANCO"
names(db)[7] <- "IDADE_BANCO"
names(db)[8] <- "CPF"
names(db)[9] <- "ESCOLARID"
names(db)[10] <- "ENDERECO"
names(db)[11] <- "BAIRRO"
names(db)[12] <- "CEP"
names(db)[13] <- "CNSUS"
```

```

names(db)[14] <- "CNES_US"
names(db)[15] <- "EX_ANT"
names(db)[16] <- "ANO_EX_ANT"
names(db)[35] <- "DT_RECEB"
names(db)[36] <- "DT_LIBER"
names(db)[37] <- "DT_EXAME"
names(db)[38] <- "DT_NASC"
names(db)[39] <- "ANO"
names(db)[45] <- "ESTADO"
names(db)[46] <- "MES"

> names(db)
[1] "COD_MUN"    "CNES_PREST" "NUM_EXAME"  "NOME"      "MAE"
[6] "DT_NASC_BANCO" "IDADE_BANCO" "CPF"        "ESCOLARID" "ENDERECO"
[11] "BAIRRO"    "CEP"        "CNSUS"     "CNES_US"   "EX_ANT"
[16] "ANO_EX_ANT" "C_NEO_MALI" "CNEOMOUT"  "C_CEL_ESCA" "C_REJ_AUS"
[21] "C_REJ_DAN"  "C_REJ_ALH"  "C_REJ_AESP" "C_REJ_OUT"  "C_REJ_OESP"
[26] "C_EPI_ESC"  "C_EPI_GLA"  "C_EPI_MET"  "C_ADQ_MAT"  "C_ATI_ESC"
[31] "C_ATI_GLA"  "C_ATI_IND"  "C_GLA_ADEN" "C_CEL_ENDO" "DT_RECEB"
[36] "DT_LIBER"   "DT_EXAME"   "DT_NASC"    "ANO"        "MUNICIPIO"
[41] "CIR_COD"    "CIR"        "MACRO_COD"  "MACRO"      "ESTADO"
[46] "MES"       "ID"

# Extrair a variável ano de nascimento da data de nascimento e criar a variável idade
library(lubridate)
db$ANO_NASC <- year(db$DT_NASC)
db$IDADE2 <- 2008 - (db$ANO_NASC)

names(db)[49] <- "IDADE"

> names(db)
[1] "COD_MUN"    "CNES_PREST" "NUM_EXAME"  "NOME"      "MAE"
[6] "DT_NASC_BANCO" "IDADE_BANCO" "CPF"        "ESCOLARID" "ENDERECO"
[11] "BAIRRO"    "CEP"        "CNSUS"     "CNES_US"   "EX_ANT"
[16] "ANO_EX_ANT" "C_NEO_MALI" "CNEOMOUT"  "C_CEL_ESCA" "C_REJ_AUS"
[21] "C_REJ_DAN"  "C_REJ_ALH"  "C_REJ_AESP" "C_REJ_OUT"  "C_REJ_OESP"
[26] "C_EPI_ESC"  "C_EPI_GLA"  "C_EPI_MET"  "C_ADQ_MAT"  "C_ATI_ESC"
[31] "C_ATI_GLA"  "C_ATI_IND"  "C_GLA_ADEN" "C_CEL_ENDO" "DT_RECEB"
[36] "DT_LIBER"   "DT_EXAME"   "DT_NASC"    "ANO"        "MUNICIPIO"
[41] "CIR_COD"    "CIR"        "MACRO_COD"  "MACRO"      "ESTADO"
[46] "MES"       "ID"      "ANO_NASC"   "IDADE"

# Transformar db em db1
db1 <- db

```

db	1256790 obs. of 49 variables
db1	1256790 obs. of 49 variables

```

# Transformar as variáveis no formato caracter
db1$NOME <- as.character(db1$NOME)
db1$MAE <- as.character(db1$MAE)
db1$MUNICIPIO <- as.character(db1$MUNICIPIO)
db1$ENDERECO <- as.character(db1$ENDERECO)

# Transformar todos os campos em maiusculo
cl <- sapply(db1,class)
wc <- which((cl=="factor")|(cl=="character"))

```

```

for(j in wc){db1[,j] <- toupper(db1[,j])}

# Retirar expressões regulares nos campos NOME e MAE
aux <- db1$NOME
aux[-grep("[a-zA-Z]",aux)] <- ""
db1$NOME <- aux

aux <- db1$MAE
aux[-grep("[a-zA-Z]",aux)] <- ""
db1$MAE <- aux

# Manter apenas um espaço entre as palavras
library(stringr)
db1$NOME <- str_replace_all(string=db1$NOME,pattern=" ",repl=" ")
db1$MAE <- str_replace_all(string=db1$MAE,pattern=" ",repl=" ")

#Retirar números do campo NOME
db1$NOME <- gsub("[0-9]", "", db1$NOME)

#retirar traço do NOME
db1$NOME <- gsub("-", "", db1$NOME)

#Substituir Mª por Maria
db1$NOME <- gsub("[ª]", "ARIA", db1$NOME)

#RETIRAR PONTUAÇÃO
db1$NOME <- gsub("[,]", "", db1$NOME)
db1$NOME <- gsub("[?]", "", db1$NOME)
db1$NOME <- gsub("[*]", "", db1$NOME)
db1$NOME <- gsub("[;]", "", db1$NOME)
db1$NOME <- gsub("[!]", "", db1$NOME)
db1$NOME <- gsub("[/]", "", db1$NOME)
db1$NOME <- gsub("[\\]", "", db1$NOME)
db1$NOME <- gsub("[.]", "", db1$NOME)
db1$NOME <- gsub("[|]", "", db1$NOME)
db1$NOME <- gsub("[()]", "", db1$NOME)
db1$NOME <- gsub("[']]", "", db1$NOME)
db1$NOME <- gsub("[`]", "", db1$NOME)
db1$NOME <- gsub("[{}]", "", db1$NOME)
db1$NOME <- gsub("[[]]", "", db1$NOME)
db1$NOME <- gsub("[@]", "", db1$NOME)
db1$NOME <- gsub("[#]", "", db1$NOME)

#Retirar números do campo MAE
db1$MAE <- gsub("[0-9]", "", db1$MAE)

#retirar traço do MAE
db1$MAE <- gsub("-", "", db1$MAE)

#Substituir Mª por Maria
db1$MAE <- gsub("[ª]", "ARIA", db1$MAE)

#RETIRAR TRAÇO E INTERROGAÇÃO DO NOME da mãe
db1$MAE <- gsub("[,]", "", db1$MAE)
db1$MAE <- gsub("[?]", "", db1$MAE)
db1$MAE <- gsub("[*]", "", db1$MAE)

```



```

db1$MAE <- gsub("[(:)]", "", db1$MAE)
db1$MAE <- gsub("[(!)]", "", db1$MAE)
db1$MAE <- gsub("[(/)]", "", db1$MAE)
db1$MAE <- gsub("[(\)]", "", db1$MAE)
db1$MAE <- gsub("[.)]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[(')]", "", db1$MAE)
db1$MAE <- gsub("[(`)]", "", db1$MAE)
db1$MAE <- gsub("[(`)]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)
db1$MAE <- gsub("[[[]]", "", db1$MAE)

```

```
# Substituição de acentos e cedilhas
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("Ã", "A", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("Â", "A", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("Á", "A", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("À", "A", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("´A", "A", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("A´", "A", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("ã", "a", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("â", "a", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("á", "a", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("à", "a", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
  db1 [,j] <- gsub("Ê", "E", db1 [,j])
}
```

```
for (j in 1:ncol(db1)) {
```



```
db1$MAE[which(db1$MAE=="NAO CADSTRADA")] <- ""
db1$MAE[which(db1$MAE=="NAO CADSTRADO")] <- ""
db1$MAE[which(db1$MAE=="NAO COLOCOU")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NA FOLHA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NA REQUISICAO")] <- ""
db1$MAE[which(db1$MAE=="NAO COSTA")] <- ""
db1$MAE[which(db1$MAE=="NAO DECLARO")] <- ""
db1$MAE[which(db1$MAE=="NAO DECLARADA")] <- ""
db1$MAE[which(db1$MAE=="NAO DECLARADO")] <- ""
db1$MAE[which(db1$MAE=="NAO ENFORMANDO")] <- ""
db1$MAE[which(db1$MAE=="NAO ESTAR PREENCHIDO")] <- ""
db1$MAE[which(db1$MAE=="NAO HA")] <- ""
db1$MAE[which(db1$MAE=="NAO IDENTIFICADA")] <- ""
db1$MAE[which(db1$MAE=="NAO IDENTIFICADA NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO IDENTIFICADO")] <- ""
db1$MAE[which(db1$MAE=="NAO IDENTIFICAÇÃO")] <- ""
db1$MAE[which(db1$MAE=="NAO INDICADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INF")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMOU")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMOU NO QUESTIONARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFRMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO MENCIONADO")] <- ""
db1$MAE[which(db1$MAE=="NAO OSHIRO")] <- ""
db1$MAE[which(db1$MAE=="NAO PEGAMOS")] <- ""
db1$MAE[which(db1$MAE=="NAO SABE")] <- ""
db1$MAE[which(db1$MAE=="NAO SEI")] <- ""
db1$MAE[which(db1$MAE=="NAO TEM")] <- ""
db1$MAE[which(db1$MAE=="NAO TEM MAE")] <- ""
db1$MAE[which(db1$MAE=="NAO TEM NO CADASTRO")] <- ""
db1$MAE[which(db1$MAE=="NAO TEM NO CADSTRO")] <- ""
db1$MAE[which(db1$MAE=="NAO TENHO")] <- ""
db1$MAE[which(db1$MAE=="NAO TENHO O NOME DA MAE")] <- ""
db1$MAE[which(db1$MAE=="NAO VEIO")] <- ""
db1$MAE[which(db1$MAE=="ORFAO")] <- ""
db1$MAE[which(db1$MAE=="ORFAOI")] <- ""
db1$MAE[which(db1$MAE=="PPPPPPPPPPPPPP")] <- ""
db1$MAE[which(db1$MAE=="S/ INFORMACAO")] <- ""
db1$MAE[which(db1$MAE=="SEM CADASTRO")] <- ""
db1$MAE[which(db1$MAE=="SEM IDENTIFICACAO")] <- ""
db1$MAE[which(db1$MAE=="SEM IDENTIFICACAO NA IDENT")] <- ""
db1$MAE[which(db1$MAE=="SEM INFORMACAO")] <- ""
db1$MAE[which(db1$MAE=="SEM NOME")] <- ""
db1$MAE[which(db1$MAE=="XX")] <- ""
db1$MAE[which(db1$MAE=="XXX")] <- ""
db1$MAE[which(db1$MAE=="XXXXX")] <- ""
db1$MAE[which(db1$MAE=="ZXV TPUOB")] <- ""
db1$MAE[which(db1$MAE=="ZXXX")] <- ""
db1$MAE[which(db1$MAE=="X")] <- ""
db1$MAE[which(db1$MAE=="x")] <- ""
db1$MAE[which(db1$MAE=="P")] <- ""
db1$MAE[which(db1$MAE=="N")] <- ""
db1$MAE[which(db1$MAE=="C")] <- ""
```



```

db1$MAE[which(db1$MAE=="-")] <- ""
db1$MAE[which(db1$MAE=="=")] <- ""
db1$MAE[which(db1$MAE=="-=")] <- ""
db1$MAE[which(db1$MAE=="?????")] <- ""
db1$MAE[which(db1$MAE=="?????-")] <- ""
db1$MAE[which(db1$MAE=="0")] <- ""
db1$MAE[which(db1$MAE=="0-")] <- ""
db1$MAE[which(db1$MAE=="-0")] <- ""
db1$MAE[which(db1$MAE=="*SEM INFORMACAO NA FICHA*")] <- ""
db1$MAE[which(db1$MAE=="A MESMA")] <- ""
db1$MAE[which(db1$MAE=="*_**")] <- ""
db1$MAE[which(db1$MAE=="*****")] <- ""
db1$MAE[which(db1$MAE=="*****")] <- ""
db1$MAE[which(db1$MAE=="*****")] <- ""
db1$MAE[which(db1$MAE=="*****")] <- ""
db1$MAE[which(db1$MAE=="14111987")] <- ""
db1$MAE[which(db1$MAE=="NAO VEIO DADO NENMHUM DA PACIENTE")] <- ""
db1$MAE[which(db1$MAE=="NAO VEIO NOME")] <- ""
db1$MAE[which(db1$MAE=="NAO FOI INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO IN FORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO IM FORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMOU")] <- ""
db1$MAE[which(db1$MAE=="NAO MENCIONADO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NO FORMULARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO CITADA")] <- ""
db1$MAE[which(db1$MAE=="NAO CITADO")] <- ""
db1$MAE[which(db1$MAE=="NOME ILEGIVEL")] <- ""
db1$MAE[which(db1$MAE=="NAO VISUALIZADA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NO FORMULARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO FOI PREENCHIDO")] <- ""
db1$MAE[which(db1$MAE=="DADO NAO FORNECIDO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NOME")] <- ""
db1$MAE[which(db1$MAE=="ND CONSTA")] <- ""
db1$MAE[which(db1$MAE=="NAO IMFORMADA")] <- ""
db1$MAE[which(db1$MAE=="NAO SABE INFORMAR")] <- ""
db1$MAE[which(db1$MAE=="ACOMPANHANTE NAO SABE INFORMAR")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="MAO NAO IDENTIFICADA")] <- ""
db1$MAE[which(db1$MAE=="NAO MENSIONADO")] <- ""
db1$MAE[which(db1$MAE=="DADO NAO FORNECIDO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NO LAUDO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NO FORMULARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADA.")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NO PRONTUARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO COSNTA INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMANDA")] <- ""
db1$MAE[which(db1$MAE=="NAO IMFORMADA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA ( PAI - ALEXANDRE BROGIO)")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO.")] <- ""
db1$MAE[which(db1$MAE=="NAO DESCRITO")] <- ""
db1$MAE[which(db1$MAE=="NAO REGISTRADO")] <- ""
db1$MAE[which(db1$MAE=="DADO NAO FORNECIDO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA INFORMACAO FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA INFORMACAO DE FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA INFORMACAO NA FICHA")] <- ""

```



```

db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,")] <- ""
db1$NOME[which(db1$NOME==" .")] <- ""
db1$NOME[which(db1$NOME=="...")] <- ""
db1$NOME[which(db1$NOME=="....")] <- ""
db1$NOME[which(db1$NOME==".....")] <- ""
db1$NOME[which(db1$NOME=="*****")] <- ""
db1$NOME[which(db1$NOME=="A")] <- ""
db1$NOME[which(db1$NOME=="CAMP")] <- ""
db1$NOME[which(db1$NOME==" CAMPO NAO PREENCHIDO")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="-")] <- ""
db1$NOME[which(db1$NOME=="AO CADASTRADA")] <- ""
db1$NOME[which(db1$NOME=="DESCONHECIDA")] <- ""
db1$NOME[which(db1$NOME=="DESCONHECIDO")] <- ""
db1$NOME[which(db1$NOME=="IGNORADA")] <- ""
db1$NOME[which(db1$NOME=="IGNORADO")] <- ""
db1$NOME[which(db1$NOME=="LLLLL")] <- ""
db1$NOME[which(db1$NOME=="LLLLLLLLL")] <- ""
db1$NOME[which(db1$NOME=="LLLLLLLLLLLL")] <- ""
db1$NOME[which(db1$NOME=="NOME")] <- ""
db1$NOME[which(db1$NOME=="NOME NAO IDENTIFICADA")] <- ""
db1$NOME[which(db1$NOME=="N DECLARADO")] <- ""
db1$NOME[which(db1$NOME=="NAO")] <- ""
db1$NOME[which(db1$NOME=="NAO TEM")] <- ""
db1$NOME[which(db1$NOME=="NAO TEM")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTOR")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRAMOS")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRAOU")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTROU")] <- ""
db1$NOME[which(db1$NOME=="NAO CADADASTRADA")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRADA")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRADO")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRAR")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRDA")] <- ""
db1$NOME[which(db1$NOME=="NAO CADASTRO")] <- ""
db1$NOME[which(db1$NOME=="NAO CADATRADA")] <- ""
db1$NOME[which(db1$NOME=="NAO CADATRASDA")] <- ""
db1$NOME[which(db1$NOME=="NAO CADSTRADA")] <- ""
db1$NOME[which(db1$NOME=="NAO CADSTRADO")] <- ""
db1$NOME[which(db1$NOME=="NAO COLOCOU")] <- ""
db1$NOME[which(db1$NOME=="NAO CONSTA")] <- ""
db1$NOME[which(db1$NOME=="NAO CONSTA NA FICHA")] <- ""
db1$NOME[which(db1$NOME=="NAO CONSTA NA FOLHA")] <- ""
db1$NOME[which(db1$NOME=="NAO CONSTA NA REQUISICAO")] <- ""

```

```

db1$NOME[which(db1$NOME=="NAO COSTA")] <- ""
db1$NOME[which(db1$NOME=="NAO DECLARO")] <- ""
db1$NOME[which(db1$NOME=="NAO DECLARADA")] <- ""
db1$NOME[which(db1$NOME=="NAO DECLARADO")] <- ""
db1$NOME[which(db1$NOME=="NAO ENFORMANDO")] <- ""
db1$NOME[which(db1$NOME=="NAO ESTAR PREENCHIDO")] <- ""
db1$NOME[which(db1$NOME=="NAO HA")] <- ""
db1$NOME[which(db1$NOME=="NAO IDENTIFICADA")] <- ""
db1$NOME[which(db1$NOME=="NAO IDENTIFICADA NA FICHA")] <- ""
db1$NOME[which(db1$NOME=="NAO IDENTIFICADO")] <- ""
db1$NOME[which(db1$NOME=="NAO IDENTIFICAÇÃO")] <- ""
db1$NOME[which(db1$NOME=="NAO INDICADO")] <- ""
db1$NOME[which(db1$NOME=="NAO INF")] <- ""
db1$NOME[which(db1$NOME=="NAO INFORMA")] <- ""
db1$NOME[which(db1$NOME=="NAO INFORMADA")] <- ""
db1$NOME[which(db1$NOME=="NAO INFORMADO")] <- ""
db1$NOME[which(db1$NOME=="NAO INFORMOU")] <- ""
db1$NOME[which(db1$NOME=="NAO INFORMOU NO QUESTIONARIO")] <- ""
db1$NOME[which(db1$NOME=="NAO INFRMADO")] <- ""
db1$NOME[which(db1$NOME=="NAO MENCIONADO")] <- ""
db1$NOME[which(db1$NOME=="NAO OSHIRO")] <- ""
db1$NOME[which(db1$NOME=="NAO PEGAMOS")] <- ""
db1$NOME[which(db1$NOME=="NAO SABE")] <- ""
db1$NOME[which(db1$NOME=="NAO SEI")] <- ""
db1$NOME[which(db1$NOME=="NAO TEM")] <- ""
db1$NOME[which(db1$NOME=="NAO TEM NOME")] <- ""
db1$NOME[which(db1$NOME=="NAO TEM NO CADASTRO")] <- ""
db1$NOME[which(db1$NOME=="NAO TEM NO CADSTRO")] <- ""
db1$NOME[which(db1$NOME=="NAO TENHO")] <- ""
db1$NOME[which(db1$NOME=="NAO TENHO O NOME DA NOME")] <- ""
db1$NOME[which(db1$NOME=="NAO VEIO")] <- ""
db1$NOME[which(db1$NOME=="ORFAO")] <- ""
db1$NOME[which(db1$NOME=="ORFAOI")] <- ""
db1$NOME[which(db1$NOME=="PPPPPPPPPPPPPP")] <- ""
db1$NOME[which(db1$NOME=="S/ INFORMACAO")] <- ""
db1$NOME[which(db1$NOME=="SEM CADASTRO")] <- ""
db1$NOME[which(db1$NOME=="SEM IDENTIFICACAO")] <- ""
db1$NOME[which(db1$NOME=="SEM IDENTIFICACAO NA IDENT")] <- ""
db1$NOME[which(db1$NOME=="SEM INFORMACAO")] <- ""
db1$NOME[which(db1$NOME=="SEM NOME")] <- ""
db1$NOME[which(db1$NOME=="XX")] <- ""
db1$NOME[which(db1$NOME=="XXX")] <- ""
db1$NOME[which(db1$NOME=="XXXXX")] <- ""
db1$NOME[which(db1$NOME=="ZXV TPUOB")] <- ""
db1$NOME[which(db1$NOME=="ZXXX")] <- ""

```

```

db1$NOME<-as.character(db1$NOME)
db1$MAE<-as.character(db1$MAE)
db1$CNSUS<-as.character(db1$CNSUS)
db1$CPF<-as.character(db1$CPF)
db1$ENDERECO<-as.character(db1$ENDERECO)
db1$MUNICIPIO<-as.character(db1$MUNICIPIO)

```

```

db1$MAE[is.na(db1$MAE)]<-"C08"
db1$NOME[is.na(db1$NOME)]<-"C08"
db1$CNSUS[is.na(db1$CNSUS)]<-"C08"

```

```
db1$CPF[is.na(db1$CPF)]<-"C08"
db1$ENDERECO[is.na(db1$ENDERECO)]<-"C08"
db1$MUNICIPIO[is.na(db1$MUNICIPIO)]<-"C08"
```

```
db1$NOME[which(db1$NOME=="")] <- "C08"
db1$MAE[which(db1$MAE=="")] <- "C08"
db1$CPF[which(db1$CPF=="")] <- "C08"
db1$ENDERECO[which(db1$ENDERECO=="")] <- "C08"
db1$CNSUS[which(db1$CNSUS=="")] <- "C08"
db1$MUNICIPIO[which(db1$MUNICIPIO=="")] <- "C08"
```

```
db1$NOME[which(db1$NOME=="NAO VEIO DADO NENMHUM DA PACIENTE")] <- ""
db1$NOME[which(db1$NOME=="NAO VEIO NOME")] <- ""
db1$NOME[which(db1$NOME=="NUMERO PRONTUARIO 3513")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO 074")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO 1344")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO 1647")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO 2864")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO 3057")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO Nº 3504")] <- ""
db1$NOME[which(db1$NOME=="PRONTUARIO NUMERO 2434")] <- ""
db1$NOME[which(db1$NOME=="")] <- "C08"
```

```
db1$MAE[which(db1$MAE=="NAO CONTA INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMANDO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMOU.")] <- ""
db1$MAE[which(db1$MAE=="NAO MENCIONADA")] <- ""
db1$MAE[which(db1$MAE=="NAO MENCIONADO.")] <- ""
db1$MAE[which(db1$MAE=="NAO NFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO ONFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO ONFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO SOUBE INFORMAR")] <- ""
db1$MAE[which(db1$MAE=="NAON INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAOO INFORMADA")] <- ""
db1$MAE[which(db1$MAE=="ILEGIVEL")] <- ""
db1$MAE[which(db1$MAE=="SEM DADOS NA REQUISICAO")] <- ""
db1$MAE[which(db1$MAE=="SEM IN FORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="SEM INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="SEM INFORMES")] <- ""
db1$MAE[which(db1$MAE=="SEM NOME NO LAUDO")] <- ""
db1$MAE[which(db1$MAE=="")] <- "C08"
```

```
db1$MAE[which(db1$MAE=="PRONTUARIO N")] <- ""
db1$MAE[which(db1$MAE=="N")] <- ""
db1$MAE[which(db1$MAE=="M")] <- ""
db1$MAE[which(db1$MAE=="x")] <- ""
db1$MAE[which(db1$MAE=="x")] <- ""
db1$MAE[which(db1$MAE=="")] <- ""
```

```
db1$NOME[which(db1$NOME=="NAO VEIO DADO NENMHUM DA PACIENTE")] <- ""
db1$NOME[which(db1$NOME=="NAO VEIO NOME")] <- ""
```

```

db1$MAE[which(db1$MAE=="NAO CONTA INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO IMFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMANDO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMOU.")] <- ""
db1$MAE[which(db1$MAE=="NAO MENCIONADA")] <- ""
db1$MAE[which(db1$MAE=="NAO MENCIONADO.")] <- ""
db1$MAE[which(db1$MAE=="NAO NFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO ONFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO ONFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO SOUBE INFORMAR")] <- ""
db1$MAE[which(db1$MAE=="NAON INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAOO INFORMADA")] <- ""
db1$MAE[which(db1$MAE=="ILEGIVEL")] <- ""
db1$MAE[which(db1$MAE=="SEM DADOS NA REQUISICAO")] <- ""
db1$MAE[which(db1$MAE=="SEM IN FORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="SEM INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="SEM INFORMES")] <- ""
db1$MAE[which(db1$MAE=="SEM NOME NO LAUDO")] <- ""
db1$MAE[which(db1$MAE=="")] <- "C08"
db1$MAE[which(db1$MAE=="ANAO CONSTA")] <- ""
db1$MAE[which(db1$MAE=="PAIS NAO DECLARADOS")] <- ""
db1$MAE[which(db1$MAE=="C")] <- ""
db1$NOME[which(db1$NOME=="C")] <- ""
db1$ENDERECO[which(db1$ENDERECO=="C")] <- ""

```

Substituir campos em branco pelo código C08

```

db1$NOME[which(db1$NOME=="")] <- "C08"
db1$MAE[which(db1$MAE=="")] <- "C08"
db1$ENDERECO[which(db1$ENDERECO=="")] <- "C08"
db1$MAE[is.na(db1$MAE)] <- "C08"
db1$NOME[is.na(db1$NOME)] <- "C08"
db1$CNSUS[is.na(db1$CNSUS)] <- "C08"
db1$CPF[is.na(db1$CPF)] <- "C08"
db1$ENDERECO[is.na(db1$ENDERECO)] <- "C08"
db1$MUNICIPIO[is.na(db1$MUNICIPIO)] <- "C08"
db1$CNES_US[is.na(db1$CNES_US)] <- "C08"

```

```
write.dbf(db1, file="db1")
```

db1 1256790 obs. of 49 variables

##NOVA INSPEÇÃO NO BANCO E NOVAS FORMATAÇÕES

```

db1$NOME[which(db1$NOME=="NAO CONSTA NO PEDIDO")] <- ""
db1$MAE[which(db1$MAE==" }æ CONSTA")] <- ""
db1$MAE[which(db1$MAE=="Ñ CONSTA")] <- ""
db1$MAE[which(db1$MAE=="NÃO CONSTA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONHECEU")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA INFORMACAO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA INFORMACAO DA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NA IDENTIDADE")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NA LISTA")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTA NO DOCUMENTO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONSTO NO FORMULARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO CONTA NO FORMULARIO")] <- ""

```

```

db1$MAE[which(db1$MAE=="NAO COSNTA NO FORMULARIO")] <- ""
db1$MAE[which(db1$MAE=="NAO COSNTA NO LAUDO")] <- ""
db1$MAE[which(db1$MAE=="NAO COSTA INFORMACAO NA FICHA")] <- ""
db1$MAE[which(db1$MAE=="NAO COSTA NO DOCUMENTO")] <- ""
db1$MAE[which(db1$MAE=="NAO ENFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO I NFORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO IDENTIFICO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFOMADA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORADMO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORAMA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORAMAOD")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORAMDA")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMACAO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADDO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADO NA REQUISICAO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORMADOS")] <- ""
db1$MAE[which(db1$MAE=="NAO INFORNMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFPRMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFROAMDO")] <- ""
db1$MAE[which(db1$MAE=="NAO INFROMADA")] <- ""
db1$MAE[which(db1$MAE=="NAO INOFMRADA")] <- ""
db1$MAE[which(db1$MAE=="NAO INORMADO")] <- ""
db1$MAE[which(db1$MAE=="NAO INORMOU")] <- ""
db1$MAE[which(db1$MAE=="NAO POSSUI REGISTRO")] <- ""
db1$MAE[which(db1$MAE=="NAO REFERIDO")] <- ""
db1$MAE[which(db1$MAE=="NAO REFERIDO")] <- ""
db1$MAE[which(db1$MAE=="NAO TEM MAE NO REGISTRO")] <- ""
db1$MAE[which(db1$MAE=="NAO TEM NO LAUDO")] <- ""
db1$MAE[which(db1$MAE=="NAO VISUALIZADO")] <- ""
db1$MAE[which(db1$MAE=="NAO VISUALIZADO DEVIDO A DERRAME DE ALCOOL")] <- ""
db1$MAE[which(db1$MAE=="NAOI INFORMADO")] <- ""
db1$MAE[which(db1$MAE=="SEM DADOS")] <- ""
db1$MAE[which(db1$MAE=="SEM IDENTIFICACAO DA MAE")] <- ""
db1$MAE[which(db1$MAE=="SEM INFORMCAO")] <- ""
db1$MAE[which(db1$MAE=="SEM MAE")] <- ""
db1$MAE[which(db1$MAE=="SEM NOME LAUDO")] <- ""
db1$MAE[which(db1$MAE=="SEM NOME MAE")] <- ""
db1$MAE[which(db1$MAE=="SEM NONE DA MAE")] <- ""
db1$MAE[which(db1$MAE=="PAI FABIO CESAR MARKOWISKI")] <- ""
db1$NOME[is.na(db1$NOME)]<-"C08"
db1$MAE[is.na(db1$MAE)]<-"C08"
db1$NOME[which(db1$NOME=="")] <- "C08"
db1$MAE[which(db1$MAE=="")] <- "C08"

```

```
write.dbf (db1, file="db1")
```

APÊNDICE 2 – PADRONIZAÇÃO DA BASE 2008 NO RECLINK III

Padronização - configura opções
 Seleccione um arquivo de entrada:





Arquivo de entrada:
 [C:\Camila\fb1.dbf]




Adiciona todos    

 Volta  Próximo  Cancela

Padronização - configura opções
 Configure os campos de entrada e saída e as opções de transformação:





Entrada	Processo	Saída
ID	NADA	ID
NOME	CNOMES	NOME
NOME	PBLOCO	PNOME
NOME	UBLOCO	UNOME
MAE	CNOMES	MAE
MAE	PBLOCO	PMAE




Adiciona todos    

 Volta  Próximo  Cancela

Padronização - configura opções
 Configure os campos de entrada e saída e as opções de transformação:





Entrada	Processo	Saída
MAE	UBLOCO	UMAE
DT_NASC	CDATAA	DT_NASC
IDADE	NADA	IDADE
CNSUS	NADA	CNSUS
ESCOLARID	NADA	ESCOLARID
CPF	NADA	CPF




Adiciona todos    

 Volta  Próximo  Cancela

Padronização - configura opções
 Configure os campos de entrada e saída e as opções de transformação:

Entrada	Processo	Saída
ENDERECO	CNOMES	ENDERECO
BAIRRO	CNOMES	BAIRRO
CEP	NADA	CEP
COD_MUN	NADA	COD_MUN
CNES_US	NADA	CNES_US
DT_EXAME	CDATAA	DT_EXAME

Adiciona todos    

 Volta  Próximo  Cancela

Padronização - configura opções

Configure os campos de entrada e saída e as opções de transformação:

Entrada	Processo	Saída
CNES_PREST	NADA	CNES_PREST
DT_RECEB	CDATA2	DT_RECEB
DT_LIBER	CDATA2	DT_LIBER
MUNICIPIO	NADA	MUNICIPIO
CJR_COD	NADA	CJR_COD
MACRO_COD	NADA	MACRO_COD

Adiciona todos Adiciona Remove Limpa

Volta Próximo Cancela

Padronização - configura opções

Configure os campos de entrada e saída e as opções de transformação:

Entrada	Processo	Saída
CJR_COD	NADA	CJR_COD
MACRO_COD	NADA	MACRO_COD
ESTADO	NADA	ESTADO
MES	NADA	MES
ANO	NADA	ANO
ANO_NASC	NADA	ANO_NASC

Adiciona todos Adiciona Remove Limpa

Volta Próximo Cancela

Padronização - configura opções

Configure as opções restantes:

Exclusões:

Pontuação:

Volta Próximo Cancela

```

## Ler banco padronizado
db1PAD <- read.dbf("db1PAD.dbf")

# Transformar as variáveis em caractere
db1PAD$NOME <- as.character(db1PAD$NOME)
db1PAD$MAE <- as.character(db1PAD$MAE)
db1PAD$ENDERECO <- as.character(db1PAD$ENDERECO)

db1PAD$MAE[is.na(db1PAD$MAE)] <- "C08"
db1PAD$NOME[is.na(db1PAD$NOME)] <- "C08"
db1PAD$ENDERECO[is.na(db1PAD$ENDERECO)] <- "C08"

db1PAD$NOME[which(db1PAD$NOME=="C")] <- "C08"
db1PAD$MAE[which(db1PAD$MAE=="C")] <- "C08"
db1PAD$ENDERECO[which(db1PAD$ENDERECO=="C")] <- "C08"

write.dbf (db1PAD, file="db1PAD.dbf")

#Excluir os registros sem nome
db2 <- subset(db1PAD,(NOME!="C08"))

```

db1PAD	1256790 obs. of 28 variables
db2	1256785 obs. of 28 variables

```

# Transformar db2 em db3
db3 <- db2

```

APÊNDICE 3 - RETIRADA DE DUPLICIDADES DA BASE 2008

MÉTODO DETERMINÍSTICO

```
# Retirar possíveis duplicidades - Considerando todas as variáveis
aux <- which(duplicated(db3))

# Retirar possíveis duplicidades - Considerando variáveis nome, mae, dtnasc e dtexame
aux <- which(duplicated(db3[,c("NOME", "MAE", "DT_NASC", "DT_EXAME", "CNES_US", "COD_MUN")]))
dup2008 <- db3[aux,]
if(length(aux)>0){db4 <- db3[-aux,]}else{db4 <- db3}
```

db3	1256785 obs. of 28 variables
db4	1246393 obs. of 28 variables
dup2008	10392 obs. of 28 variables
values	
aux	int [1:10392] 1367 1566 1573 1814 2994 6105 6154...

```
# Ordenar o banco pela data do exame
library(dplyr)
db4 <- arrange(db4,DT_EXAME)
write.dbf(db4, "db4.dbf")

write.dbf(db4, "db4.dbf")
write.dbf(dup2008, "dup2008.dbf")
```

MÉTODO PROBABILÍSTICO

1ª ETAPA

Duplicidade - configura opções

Selecione um arquivo de entrada e o nome para o arquivo mestre:

Arquivo de origem:

Arquivo mestre:

Duplicidade - conf

Configure os campos de bloco:

Entrada	Processo	Saída
PNOME	NADA	PNOME
UNOME	NADA	UNOME
DT_EXAM	NADA	DT_EXAM
ANO_NASC	NADA	ANO_NASC
CNES_LUS	NADA	CNES_LUS

Duplicidade - configura opções

Configure os campos de bloco:

Entrada	Processo	Saída
PNOME	NADA	PNOME
UNOME	NADA	UNOME
DT_EXAM	NADA	DT_EXAM
ANO_NASC	NADA	ANO_NASC
CNES_LUS	NADA	CNES_LUS

Duplicidade - conf

Configure campos de comparação:

Entrada	Processo	Saída
NOME	CNOMES	CNOMES
MAE	CNOMES	CNOMES
DT_NASC		DT_NASC

Duplicidade - conf

Configure campos de conversão:

Entrada	Processo	Saída
DT_NASC		DT_NASC

Duplicidade - conf

Configure campos de conversão:

Entrada	Processo	Saída
DT_NASC		DT_NASC

Duplicidade - configura opções

Configure campos de comparação:

Entrada	Padroniza	Compara	Correto	Incorreto	Limiar	Saida
NOME	CNOMES	Aprox.	92	1	85	NOME
MAE	CNOMES	Aprox.	92	1	85	MAE
DT_MASC	NADA	Caractere	90	5	65	DT_MASC
COD_MUN	NADA	Caractere	90	5	65	COD_MUN

Duplicidade - configura opções

Configure as opções restantes:

Exclusões: DE, DA, D

Pontuação: () / * , ; -

Campo único: UNICO

Campo total: TOTAL

Calcula escores

Score mínimo: 0.0

LIMITES

Máximo: 21.3868739149987
 Mínimo: -13.7545682670464

Duplicidade - configura opções

Configure as opções restantes:

Exclusões: DE, DA, DO, DOS, DAS

Pontuação: () / * , ; -

Campo único: UNICO

Campo total: TOTAL

Calcula escores

Score mínimo: 20.0

Duplicidade - Executa

Arquivo mestre: C:\Camila\db4\MESTRE.dbf

Arquivo de atualização: C:\Camila\db4.dbf

Reg. inicial: 1

Reg. final: 1246393

Processando...

Verificando arquivo mestre...

Atualizando campo único...

Processando...

Registro corrente: 1246393

Identificador corrente:

Processamento completo: 000:25:03

```
db4 <- read.dbf("db4.dbf")
db4MESTRE <- read.dbf("db4MESTRE.dbf")
```

```
aux <- which(duplicated(db4[,c("UNICO")]))
dup2008_1 <- db4[aux,]
if(length(aux)>0){db4MESTRE1 <- db4[-aux,]}else{db4MESTRE1 <- db4}
```

```
write.dbf(dup2008_1,"dup2008_1.dbf")
write.dbf(db4MESTRE1,"db4MESTRE1.dbf")
```

db4	1246393 obs. of 29 variables
db4MESTRE	1245883 obs. of 11 variables
db4MESTRE1	1245883 obs. of 29 variables
dup2008_1	510 obs. of 29 variables

```
#Retirar variáveis excedentes
excluir<-c("UNICO")
db4MESTRE2 <- db4MESTRE1[,!(names(db4MESTRE1)%in% excluir)]
write.dbf(db4MESTRE2,"db4MESTRE2.dbf")
```

db4MESTRE	1245883 obs. of 11 variables
db4MESTRE1	1245883 obs. of 29 variables
db4MESTRE2	1245883 obs. of 28 variables

2ª ETAPA

Duplicidade - configura opções

Selecione um arquivo de entrada e o nome para o arquivo mestre:

Arquivo de origem:

Arquivo mestre:

Duplicidade - configura opções

Configure os campos de blocagem:

Entrada	Processo	Saída
PMAE	NADA	PMAE
UMAE	NADA	UMAE
DT_EXAME	NADA	DT_EXAME
ANO_INASC	NADA	ANO_INASC
CNES_US	NADA	CNES_US
CNES_PREST	NADA	CNES_PREST

Duplicidade - configura opções

Configure campos de comparação:

Entrada	Padroniza	Compara	Correto	Incorreto	Limiar	Saída
NOME	CNOMES	Aprox.	92	1	85	NOME
MAE	CNOMES	Aprox.	92	1	85	MAE
DT_NASC	NADA	Caractere	90	5	65	DT_NASC
COD_MUN	NADA	Caractere	90	5	65	COD_MUN

Duplicidade - configura opções

Configure as opções restantes:

Exclusões: DE, DA, D

Pontuação: () / ' , ; -

Calcula escores

Score mínimo: 0,0

Limite mínimo: 0,0

OK

LIMITES

Máximo: 21.3869739149987
Mínimo: -13.754562670464

OK

Duplicidade - configura opções

Configure as opções restantes:

Exclusões: DE, DA, DO, DOS, DÀS

Pontuação: () / ' , ; -

Calcula escores

Score mínimo: 20,0

Campo único: UNICO

Campo total: TOTAL

Duplicidade - Executa

Arquivo mestre: C:\Camila\bd\MESTRE3.dbf

Arquivo de atualização: C:\Camila\bd\MESTRE2.dbf

Reg. inicial: 1

Reg. final: 1245883

Processando...

Verificando arquivo mestre... ✓

Acrescentando campo único. ✓

Processando... ✓

Registro corrente: 1245883

Identificador corrente:

Processamento completo: 000:26:15

Retorna

```
db4MESTRE2 <- read.dbf("db4MESTRE2.dbf")
db4MESTRE3 <- read.dbf("db4MESTRE3.dbf")
```

```
aux <- which(duplicated(db4MESTRE2[,c("UNICO")]))
```



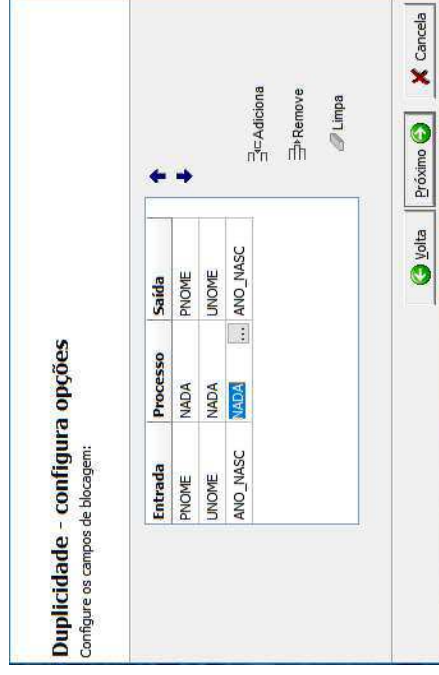
```
dup2008_2<- db4MESTRE2[aux,]  
if(length(aux)>0){db5 <- db4MESTRE2[-aux,]}else{db5 <- db4MESTRE2}  
  
write.dbf(db5, "db5.dbf")  
write.dbf(dup2008_2, "dup2008_2.dbf")
```

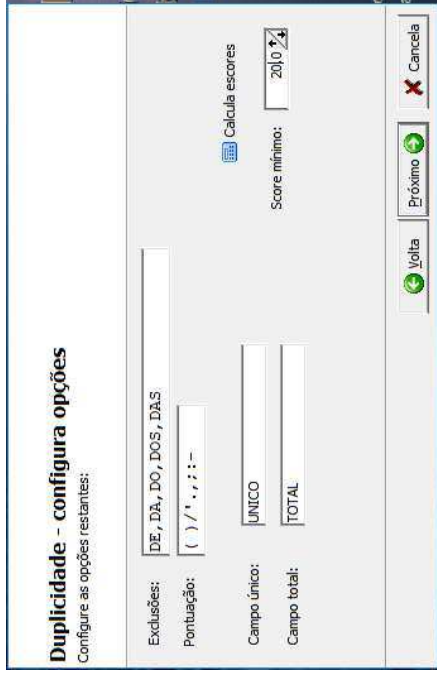
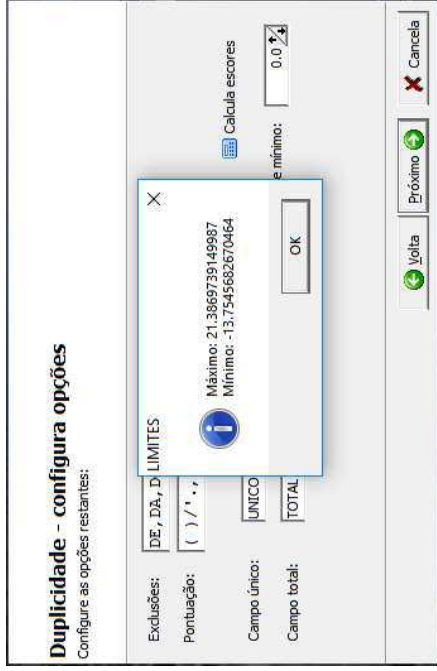
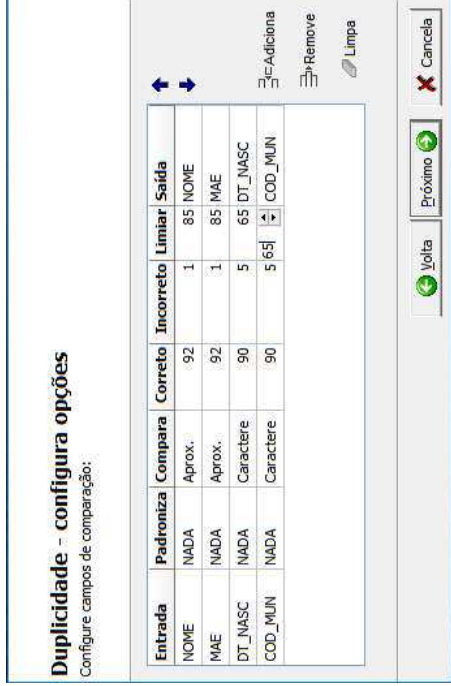
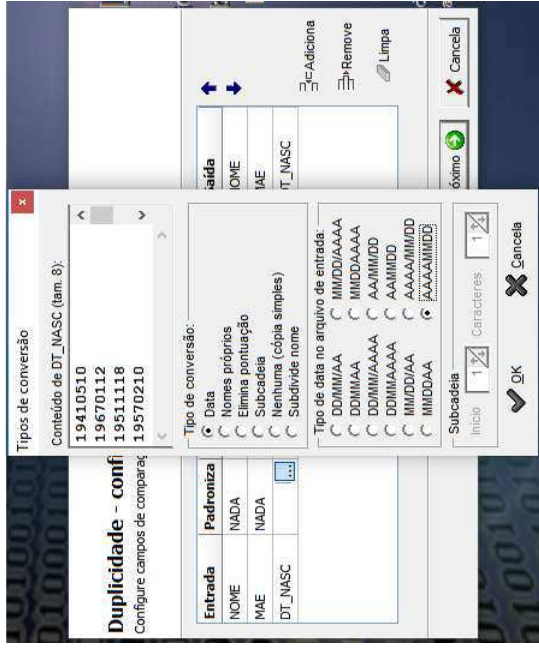
db4MESTRE2	1245883 obs. of 29 variables
db4MESTRE3	1245753 obs. of 12 variables
db5	1245753 obs. of 29 variables
dup2008_2	130 obs. of 29 variables

APÊNDICE 4 – RETIRADA DE EXAMES DE SEGUIMENTO DA BASE 2008 (REPETIÇÕES EM 2008) E OUTRAS LIMPEZAS

```
## EXCLUIR SEGUIMENTO – MÉTODO DETERMINÍSTICO
aux <- which(duplicated(db5[,c("NOME","MAE","DT_NASC","COD_MUN")]))
seguim_2008 <- db5 [aux,]
if(length(aux)>0){db6 <- db5 [-aux,]}else{ db6 <- db5}
names(db6)
excluir<-c("UNICO")
db6 <- db6[!(names(db6)%in% excluir)]
write.dbf(db6,"db6.dbf")
write.dbf(seguim_2008,"seguim_2008.dbf")
db5 1245753 obs. of 29 variables
db6 1222456 obs. of 28 variables
seguim_2008 23297 obs. of 29 variables
```

EXCLUIR SEGUIMENTO – MÉTODO PROBABILÍSTICO
1ª etapa:





```
db6 <- read.dbf("db6.dbf")
db6MESTRE <- read.dbf("db6MESTRE.dbf")
aux <- which(duplicated(db6[,c("UNICO")]))
seguim_2008_1 <- db6[aux,]
```

```
if(length(aux)>0){db6MESTRE1 <- db6[-aux,]}else{db6MESTRE1 <- db6}
```

```
write.dbf(seguint_2008_1,"seguint_2008_1.dbf")
write.dbf(db6MESTRE1,"db6MESTRE1.dbf")
```

db6	1222456 obs. of 29 variables
db6MESTRE	1213228 obs. of 9 variables
db6MESTRE1	1213228 obs. of 29 variables
seguint_2008_1	9228 obs. of 29 variables

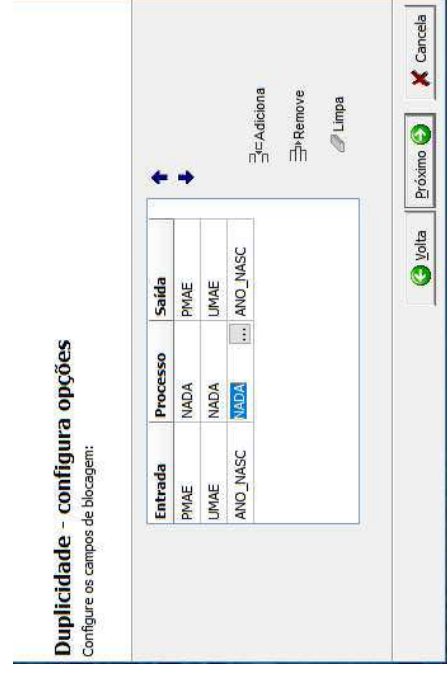
```
excluir<-c("UNICO")
```

```
db6MESTRE2 <- db6MESTRE1[,!(names(db6MESTRE1)%in% excluir)]
```

```
write.dbf(db6MESTRE2,"db6MESTRE2.dbf")
```

db6MESTRE1	1213228 obs. of 29 variables
db6MESTRE2	1213228 obs. of 28 variables

2ª etapa



Duplicidade - configura opções

Configure campos de comparação:

Entrada	Padroniza	Compara	Correto	Incorreto	Limiar	Saida
NOME	NADA	Aprox.	92	1	85	NOME
MAE	NADA	Aprox.	92	1	85	MAE
DT_NASC	NADA	Caractere	90	5	65	DT_NASC
COD_MUN	NADA	Caractere	90	5	65	COD_MUN

Duplicidade - configura opções

Configure as opções restantes:

Exclusões: DE, DA, DO, DOS, DAS

Pontuação: () / ' , , : ; =

Score mínimo: 20,0

Campo único: UNICO
 Campo total: TOTAL

Duplicidade - Executa

Arquivo mestre: C:\Camila\db6MESTRE3.dbf

Arquivo de atualização: C:\Camila\db6MESTRE2.dbf

Reg. inicial: 1
 Reg. final: 1213228

✓
 ✓
 ✓

Registro corrente: 1213228
 Identificador corrente:

Processo

Processamento completo: 001:58:52

```
db6MESTRE2 <- read.dbf("db6MESTRE2.dbf")
db6MESTRE3 <- read.dbf("db6MESTRE3.dbf")
aux <- which(duplicated(db6MESTRE2[,c("UNICO")]))
seguim_2008_2 <- db6MESTRE2[aux,]
```

```
if(length(aux)>0){db7 <- db6MESTRE2[-aux,]}else{db7 <- db6MESTRE2}
write.dbf(seguint_2008_2,"seguint_2008_2.dbf")
write.dbf(db7,"db7.dbf")
```

db6MESTRE2	1213228 obs. of 29 variables
db6MESTRE3	1212311 obs. of 9 variables
db7	1212311 obs. of 29 variables
seguint_2008_2	917 obs. of 29 variables

```
excluir<-c("UNICO")
db8 <- db7[,!(names(db7)%in% excluir)]
write.dbf(db8,"db8.dbf")
```

NOVAS LIMPEZAS – Excluir mulheres com menos de 12 anos

names(db8)

```
[1] "ID" "NOME" "PNOME" "UNOME" "MAE" "PMAE"
[7] "UMAE" "DT_NASC" "IDADE" "CNSUS" "ESCOLARID" "CPF"
[13] "ENDERECO" "BAIRRO" "CEP" "COD_MUN" "CNES_US" "DT_EXAME"
[19] "CNES_PREST" "DT_RECEB" "DT_LIBER" "MUNICIPIO" "CIR_COD" "MACRO_COD"
[25] "ESTADO" "MES" "ANO" "ANO_NASC"
```

```
> table(db8$ANO_NASC, useNA = "ifany")
```

1661	1747	1764	1769	1858	1870	1889	1900	1903	1904	1905	1906	1907	1908	1909
1	1	1	1	1	1	1	2	2	1	1	3	4	7	690
1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924
28	11	13	13	28	34	36	45	63	101	128	133	238	284	314
1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939
425	554	675	814	1029	1560	1514	1933	2010	2419	3058	3468	3586	4420	5182
1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954
5944	5565	6831	7394	8276	8418	9557	10660	12175	12077	14552	15106	16896	17592	19224
1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
19993	21370	22936	24164	24409	25185	24940	26795	27969	27891	28765	28550	28061	30028	28844
1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
29208	28753	29751	29002	30130	30480	31398	32520	32571	32336	32635	32110	30097	27913	1913
1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
26352	26049	24002	22785	19639	14891	10653	6304	3335	1394	414	195	111	81	42
2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2031	2035	2038	2039	2041
27	31	17	14	6	47	7	21	805	36	1	4	1	1	1
2042	2043	2044	2045	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	
1	4	1	1	1	5	3	1	2	5	6	2	6	5	

```

# Transformando as variáveis DTNASC e DTEXAME no formato data
library(zoo)
db8$DT_NASC2 <- as.Date(db8$DT_NASC, "%Y%m%d")
str(db8$DT_NASC)
str(db8$DT_NASC2)

# Extrair a variável ano de nascimento da data de nascimento e criar a variável idade
library(lubridate)
db8$ANO_NASC_2 <- year(db8$DT_NASC2)
db8$IDADE2 <- 2008 - (db8$ANO_NASC_2)

# Excluir registros sem DTNASC ou com idade abaixo de 12 anos
db9 <- db8[db8$ANO_NASC_2 > "1661" & db8$ANO_NASC_2 < "1997",]
dbex <- db8[db8$ANO_NASC_2 > "1996" & db8$ANO_NASC_2 < "2058",]
write.dbf(dbex, "dbex.dbf")
write.dbf(db9, "db9.dbf")

```

db8	1212311 obs. of 31 variables
db9	1211014 obs. of 31 variables
dbex	1296 obs. of 31 variables

APÊNDICE 5 – RELACIONAMENTO PROBABILÍSTICO ENTRE O BANCO 2008 E O BANCO CONTENDO EXAMES ALTERADOS DE 2007

db7	36228 obs. of 26 variables
db9	1211014 obs. of 26 variables

Passo 2

Relacionamento - configura opções
 Seleccione um nome para o arquivo de configuração

Arquivo de configuração:

Relacionamento - configura opções
 Seleccione um arquivo de comparação e outro de referência:

Arquivo de comparação (será lido integralmente):

Arquivo de referência (será lido por blocos):

Relacionamento - configura opções

Configure os campos de blocagem:

Nome do índice:

Comparação	Referência
PNOME	PNOME
UNOME	UNOME
ANO_NASC	ANO_NASC

Relacionamento - configura opções

Configure campos de comparação:

Comparação	Correto	Incorreto	Limiar	Referência
NOME Aprox.	92	1	85	NOME
DT_NASC Carattere	90	5	65	DT_NASC

Relacionamento - configura opções

Configure as opções restantes:

Ignora score baixo

LIMITES

Máximo: 10.6934869574993
 Mínimo: -6.8772841335232

mínimo:

Visualiza Arquivo

F:\Dados Bancos\autorado Passo2_Relac_0708.dbf

COMPREC	REFREC	SCORE	MATCH
1	300978	3.27563444	-
1	906851	3.27563444	-
2	18133	2.99200131	-
2	951386	9.32134694	-
2	1014052	0.01932775	-
2	1070665	3.27563444	-
3	174640	0.01932775	-
3	233132	0.01932775	-
3	313565	0.01932775	-
3	328205	0.01932775	-
3	947026	0.01932775	-
3	573348	10.6934869	-
3	772809	0.01932775	-

Total de registros: 104989

Passo 9

Relacionamento - configura opções

Selecione um nome para o arquivo de configuração

Arquivo de configuração:
 

Relacionamento - configura opções

Selecione um arquivo de comparação e outro de referência:

Arquivo de comparação (será lido integralmente):
 

Arquivo de referência (será lido por blocos):
 

Relacionamento - configura opções

Configure os campos de blocagem:

Nome do índice

Comparação	Referência
PNOME	PNOME
ANO_NASC	ANO_NASC

Relacionamento - configura opções

Configure campos de comparação:

Comparação	Compara	Correto	Incorreto	Limiar	Referência
NOME	Aprox.	92	1	85	NOME
MAE	Aprox.	92	1	85	MAE
DT_NASC	Caractere	90	5	65	DT_NASC

Visualiza Arquivo

F:\Dados\Bancosdoutorado\Passo9_Relec_0708.dbf

COMPREC	REFREC	SCORE	MATCH
2	951386	5.69199032	-
3	483337	6.02164908	-
3	573948	7.06413033	-
3	1123328	6.02164908	-
4	751720	7.06413033	-
6	178852	5.08971166	-
6	652048	5.20620384	-
6	982880	7.06413033	-
7	233325	16.9452338	-
9	573791	17.2170489	-
10	274945	6.07652100	-
10	709085	5.08971166	-
10	884132	5.12184743	-

Total de registros: 28507

Importa Retorna

Relacionamento - configura opções

Configure as opções restantes:

Ignora score baixo

mínimo:

LIMITES

Máximo: 17.2170489135563
Mínimo: -10.5066407536028

OK

Volta Próximo Cancela

Passo 12

Relacionamento - configura opções

Selecione um arquivo de comparação e outro de referência:

Arquivo de comparação (será lido integralmente):
F:\Dados\Bancosdoutorado\2007EXAMES_ALTERADOS\pb7.dbf

Arquivo de referência (será lido por blocos):
F:\Dados\Bancosdoutorado\2008\pb9.dbf

Volta Próximo Cancela

Relacionamento - configura opções

Selecione um nome para o arquivo de configuração:

Arquivo de configuração:
F:\Dados\Bancosdoutorado\Passo12_Relec_0708.dfn

Volta Próximo Cancela

Relacionamento - configura opções

Configure os campos de bloqueio:

Nome do índice
Passo 12

Comparação	Referência
P.NOME	P.NOME
U.NOME	U.NOME

Relacionamento - configura opções

Configure campos de comparação:

Comparação	Correto	Incorreto	Limiar	Referência
NOME	Aprox. 92	1	85	NOME
MAE	Aprox. 92	1	85	MAE
DT_NASC	Caractere 90	5	65	DT_NASC

Relacionamento - configura opções

Configure as opções restantes:

Ignora score baixo

mínimo:

LIMITES

Visualiza Arquivo

Total de registros: 37034

COMPREC	REFREC	SCORE	MATCH
1	212064	6.02164908	-
1	374133	5.33495835	-
2	221274	6.02164908	-
2	787762	6.02164908	-
2	951386	5.69199032	-
3	573348	7.06413033	-
4	84389	5.25417120	-
4	751720	7.06413033	-
5	513084	6.02164908	-
6	134194	6.02164908	-
6	556438	5.20620384	-
6	982580	7.06413033	-
7	233325	16.9452338	-

Passo 16

Relacionamento - configura opções

Selecione um nome para o arquivo de configuração

Arquivo de configuração:
 F:\Dados\Bancosdoutorado\Passo 16_Relac_0708.dfn

Volta Próximo Cancela

Relacionamento - configura opções

Selecione um arquivo de comparação e outro de referência:

Arquivo de comparação (será lido integralmente):
 F:\Dados\Bancosdoutorado\2007\EXAMES_ALTERADOS\db7.dbf

Arquivo de referência (será lido por blocos):
 F:\Dados\Bancosdoutorado\2008\db9.dbf

Volta Próximo Cancela

Relacionamento - configura opções

Configure os campos de blocagem:

Nome do índice
 Passo 16

Comparação	Referência
ANO_NASC	ANO_NASC

Adiciona Remove Limpa

Volta Próximo Cancela

Relacionamento - configura opções

Configure campos de comparação:

Comparação	Compara	Correto	Incorreto	Limiar	Referência
NOME	Aprox.	92	1	85	NOME
MAE	Aprox.	92	1	85	NOME
DT_NASC	Caractere	90	5	65	DT_NASC

Adiciona Remove Limpa

Volta Próximo Cancela

Relacionamento - Executa

Arquivo de configuração:
 F:\Dados\Bancosdoutorado\Passo16_Relac_0708.ch

Arquivo de comparação (será lido integralmente):
 F:\Dados\Bancosdoutorado\db7.dbf

Arquivo de referência (será lido por blocos):
 F:\Dados\Bancosdoutorado\db9.dbf

Reg. inicial: 1
 Registro - comparação: 36228

Reg. final: 36228
 Registro - referência: 12

Processa

Processamento completo: 0:16:32:20

Retorna

Relacionamento - configura opções

Configure as opções restantes:

Ignora score baixo

LIMITES

Máximo: 17.2170489135563
 Mínimo: -10.5066407536028

mínimo: 0.0

Volta Próximo Cancela

Visualiza Arquivo

F:\Dados\Bancosdoutorado\Relac_0708\Passo16_Relac_0708\Pas: Novo

COMPREC	REFREC	SCORE	MATCH
1	340245	17.2170489	X
3	708138	6.02164908	!
3	769946	6.02164908	!
3	1170474	17.2170489	X
4	1170798	7.06413033	X
5	1181151	17.2170489	X
6	1183717	17.2170489	X
7	844696	6.02164908	!
7	1193951	6.23224390	!
8	9690	6.02164908	!
8	182522	6.02164908	!
8	234498	6.02164908	!
8	424843	5.61095228	!

Total de registros: 62416

Importa Retorna

APÊNDICE 6 – SCRIPT PARA RODAR A CURVA PRECISION-RECALL E CALCULAR ACURÁCIA

```

#A2
A2_0708_conferido <- read.csv2("A2_0708_conferido.csv")
A2 <- A2_0708_conferido
str(A2)

#Retirar variáveis excedentes
excluir<-c("X", "X.1")
A2 <- A2[,!(names(A2))%in% excluir]
write.dbf(A2, "A2.dbf")

str(A2)

A2$SCORE <- as.character(A2$SCORE_x)
A2$ESCORE <- as.numeric(A2$SCORE)

names(A2)

# ajustando o banco (separando os escores dos pares verdadeiros e falsos em bancos diferentes)
A2$par <- 0
A2$par[A2$MATCH=="X"] <- 1
names(A2)

# fazendo a curva precision-recall
library("PRROC", lib.loc="C:/Program Files/R/R-3.3.1/library")
pr <- pr.curve(scores.class0 = A2[A2$MATCH=="X",17], scores.class1 = A2[A2$MATCH=="!",17], curve=TRUE, max.compute=T)
print(pr)
plot(pr,scale.color = rev(gray.colors(100)))
curva.pr <- as.data.frame(pr$curve)
names(curva.pr) <- c("Recall", "Precision", "Escore")
curva.pr$F1 = 2 * (curva.pr$Recall*curva.pr$Precision / (curva.pr$Recall+curva.pr$Precision))

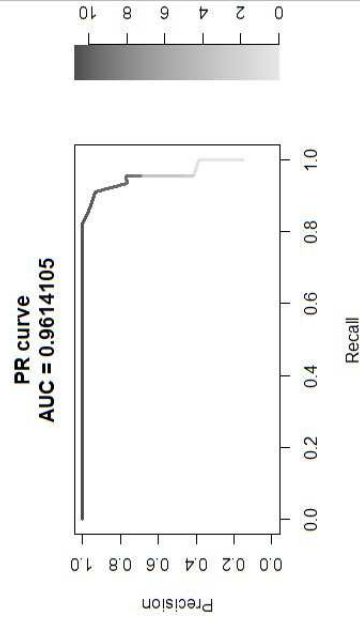
```

```
maximoF1 = max(curva.pr$F1)
best.cutoff = curva.pr$Escore[which(curva.pr$F1==maximoF1)]

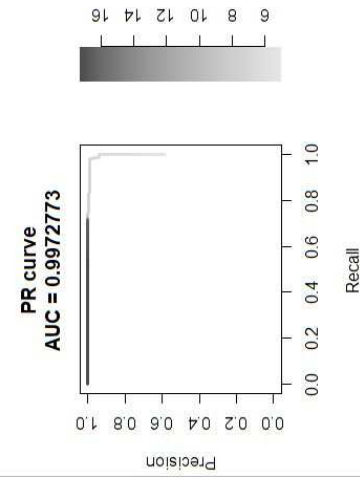
#Acurácia do ponto de corte
library(ROCR)
pred <- prediction(predictions = A2$SCORE, labels = A2$par)
perf1 <- performance(pred, "acc")
plot(perf1, main="AcurÃ¡cia")
acuracia <- as.data.frame(cbind(perf1@x.values[[1]],perf1@y.values[[1]]))
names(acuracia) <- c("Escore", "Acuracia")
max.acuracia = max(acuracia$Acuracia)
best.cutoff <- acuracia$Escore[which(acuracia$Acuracia==max.acuracia)]
print(best.cutoff)
```

APÊNDICE 7 – CURVAS PRECISION-RECALL REFERENTES AO RELACIONAMENTO ENTRE O BANCO 2008 E O BANCO 2007 CONTENDO MULHERES COM EXAME ALTERADO EM 2007

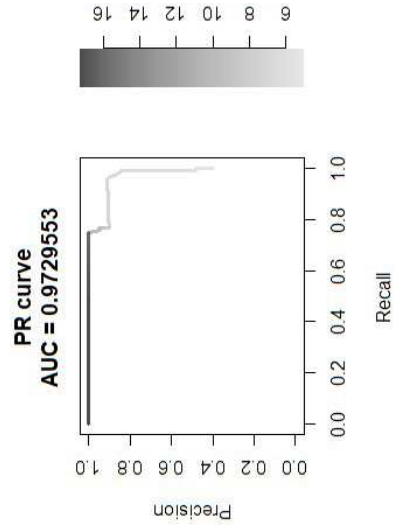
Passo 2 (amostra A2)



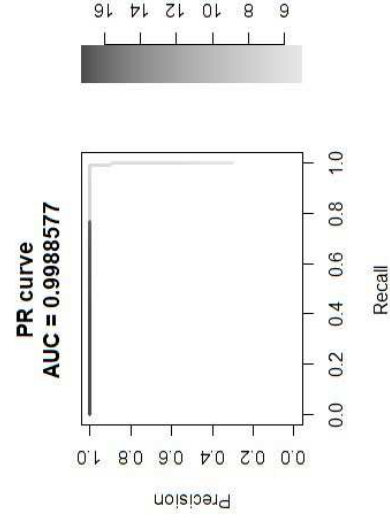
Passo 9 (amostra A9)



Passo 12 (amostra A12)



Passo 16 (amostra A16)



APÊNDICE 8 – DESCRIÇÃO DOS PASSOS E DOS DADOS REFERENTES AO RELACIONAMENTO ENTRE O BANCO REFERÊNCIA E OS BANCOS DE COMPARAÇÃO

Relacionamento 2008-Seguimento 2008						
Passo	Pares formados	Amostra de 300 pares			Ponto de corte	Pares verdadeiros
		Precision; Recall	Medida F	Acurácia		
2	103.672	0,97; 0,97	0,97	0,99	9,76	24.200
9	34.712	0,99; 1,00	0,99	0,99	7,06	24.291
12	41.266	1,00; 0,99	0,99	0,99	16,07	23.640
16	71.291	0,99; 1,00	0,99	0,99	7,06	25.252
Relacionamento 2008-2009						
Passo	Pares formados	Amostra de 300 pares			Ponto de corte	Pares verdadeiros
		Precision; Recall	Medida F	Acurácia		
2_REF1	897.756	0,94; 0,97	0,95	0,99	9,78	112.058
2_REF2	989.230	0,91; 0,94	0,92	0,98	10,17	112.951
2_REF3	913.100	1,00; 0,86	0,92	0,98	10,38	76.754
2_REF4	768.628	1,00; 1,00	1,00	1,00	10,28	46.682
9	805.115	0,99; 0,98	0,99	0,99	6,71	386.597
12	1.069.803	0,90; 0,99	0,95	0,97	6,72	397.940
15	934.111	0,99; 0,99	0,99	0,99	6,78	364.830
Relacionamento 2008-2010						
Passo	Pares formados	Amostra de 300 pares			Ponto de corte	Pares verdadeiros
		Precision; Recall	Medida F	Acurácia		
2_REF1	871.506	1,00; 0,93	0,96	0,99	10,19	88.709
2_REF2	971.203	1,00; 0,97	0,98	0,99	10,00	111.562
2_REF3	916.484	0,94; 0,94	0,94	0,97	10,12	104.849
2_REF4	798.427	1,00; 0,95	0,97	0,99	10,17	93.331

9	850.984	1,00; 0,96	0,98	0,98	6,61	439.292
12	1.103.512	0,92; 0,99	0,96	0,96	7,06	435.824
15	871.506	0,97; 0,99	0,98	0,98	6,75	411.021
Relacionamento 2008-2011						
Amostra de 300 pares						
Passo	Pares formados	Precision; Recall	Medida F	Acurácia	Ponto de corte	Pares verdadeiros
2_REF1	888.819	0,97; 1,00	0,98	0,99	10,17	95.448
2_REF2	992.958	0,97; 0,97	0,97	0,99	10,17	107.550
2_REF3	933.311	1,00; 0,95	0,97	0,99	10,19	92.350
2_REF4	812.897	0,95; 1,00	0,97	0,99	9,78	91.642
9	850.185	0,97; 1,00	0,98	0,98	6,24	464.207
12	1.105.403	0,98; 0,95	0,97	0,98	6,87	423.419
15	971.186	0,99; 0,98	0,99	0,99	6,72	399.180
Relacionamento 2008-2012						
Amostra de 300 pares						
Passo	Pares formados	Precision; Recall	Medida F	Acurácia	Ponto de corte	Pares verdadeiros
2_REF1	758.018	1,00; 1,00	1,00	1,00	10,17	79.524
2_REF2	844.553	1,00; 0,88	0,94	0,98	10,35	79.132
2_REF3	797.750	0,92; 1,00	0,96	0,99	9,84	86.364
2_REF4	692.942	0,94; 0,97	0,96	0,99	9,87	74.844
9	720.161	0,98; 0,97	0,97	0,98	6,72	355.837
12	933.859	0,94; 0,98	0,96	0,97	6,65	357.523
15	821.750	1,00; 0,98	0,99	0,99	6,80	326.404

APÊNDICE 9 – SCRIPT UTILIZADO PARA EMPILHAR OS PASSOS DO RELACIONAMENTO 2008-2012, IDENTIFICAR O NÚMERO TOTAL DE PARES E CONSTRUIR O BANCO SEGUIMENTO 2008-2012

```

# Empilhando os passos

setwd("F:/Dados/Pos_Relac/Passo2")
library(foreign)
library(readxl)

# Passo2
P2_08<-read.dbf("P2_08.dbf")
P2_REF1_09<-read_xlsx("P2_REF1_09.xlsx")
P2_REF1_10<-read.dbf("P2_REF1_10_v.dbf")
P2_REF1_11<-read_xlsx("P2_REF1_11.xlsx")
P2_REF1_12<-read_xlsx("P2_REF1_12.xlsx")
P2_REF2_09<-read_xlsx("P2_REF2_09.xlsx")
P2_REF2_10<-read.dbf("P2_REF2_10_v.dbf")
P2_REF2_11<-read.dbf("P2_REF2_11_v.dbf")
P2_REF2_12<-read.dbf("P2_REF2_12_v.dbf")
P2_REF3_09<-read_xlsx("P2_REF3_09.xlsx")
P2_REF3_10<-read.dbf("P2_REF3_10_v.dbf")
P2_REF3_11<-read.dbf("P2_REF3_11_v.dbf")
P2_REF3_12<-read_xlsx("P2_REF3_12.xlsx")
P2_REF4_09<-read.dbf("P2_REF4_09_v.dbf")
P2_REF4_10<-read.dbf("P2_REF4_10_v.dbf")
P2_REF4_11<-read_xlsx("P2_REF4_11.xlsx")
P2_REF4_12<-read_xlsx("P2_REF4_12.xlsx")

names(P2_08)
P2_08 <- subset(P2_08, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULT,C_AMO
STRA,C_EPITELIO,C_MICROBIO))

names(P2_REF1_09)

```

```

P2_REF1_09 <- subset(P2_REF1_09, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF1_09)
names(P2_REF1_09)[15] <- "C_RESULT"

names(P2_REF1_10)
P2_REF1_10 <- subset(P2_REF1_10, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF1_10)
names(P2_REF1_10)[15] <- "C_RESULT"

names(P2_REF1_11)
P2_REF1_11 <- subset(P2_REF1_11, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF1_11)
names(P2_REF1_11)[15] <- "C_RESULT"

names(P2_REF1_12)

P2_REF1_12 <- subset(P2_REF1_12, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF1_12)
names(P2_REF1_12)[15] <- "C_RESULT"

names(P2_REF2_09)
P2_REF2_09 <- subset(P2_REF2_09, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF2_09)
names(P2_REF2_09)[15] <- "C_RESULT"

names(P2_REF2_10)

```

```

P2_REF2_10 <- subset(P2_REF2_10, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF2_10)
names(P2_REF2_10)[15] <- "C_RESULT"

names(P2_REF2_11)
P2_REF2_11 <- subset(P2_REF2_11, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF2_11)
names(P2_REF2_11)[15] <- "C_RESULT"

names(P2_REF2_12)
P2_REF2_12 <- subset(P2_REF2_12, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF2_12)
names(P2_REF2_12)[15] <- "C_RESULT"

names(P2_REF3_09)
P2_REF3_09 <- subset(P2_REF3_09, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF3_09)
names(P2_REF3_09)[15] <- "C_RESULT"

names(P2_REF3_10)
P2_REF3_10 <- subset(P2_REF3_10, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF3_10)
names(P2_REF3_10)[15] <- "C_RESULT"

names(P2_REF3_11)

```

```

P2_REF3_11 <- subset(P2_REF3_11, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF3_11)
names(P2_REF3_11)[15] <- "C_RESULT"

names(P2_REF3_12)
P2_REF3_12 <- subset(P2_REF3_12, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF3_12)
names(P2_REF3_12)[15] <- "C_RESULT"

names(P2_REF4_09)
P2_REF4_09 <- subset(P2_REF4_09, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF4_09)
names(P2_REF4_09)[15] <- "C_RESULT"

names(P2_REF4_10)
P2_REF4_10 <- subset(P2_REF4_10, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF4_10)
names(P2_REF4_10)[15] <- "C_RESULT"

names(P2_REF4_11)
P2_REF4_11 <- subset(P2_REF4_11, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF4_11)
names(P2_REF4_11)[15] <- "C_RESULT"

names(P2_REF4_12)

```

```

P2_REF4_12 <- subset(P2_REF4_12, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P2_REF4_12)
names(P2_REF4_12)[15] <- "C_RESULT"

P2_08_dt <- data.table(P2_08)
P2_REF1_09_dt <- data.table(P2_REF1_09)
P2_REF1_10_dt <- data.table(P2_REF1_10)
P2_REF1_11_dt <- data.table(P2_REF1_11)
P2_REF1_12_dt <- data.table(P2_REF1_12)
P2_REF2_09_dt <- data.table(P2_REF2_09)
P2_REF2_10_dt <- data.table(P2_REF2_10)
P2_REF2_11_dt <- data.table(P2_REF2_11)
P2_REF2_12_dt <- data.table(P2_REF2_12)
P2_REF3_09_dt <- data.table(P2_REF3_09)
P2_REF3_10_dt <- data.table(P2_REF3_10)
P2_REF3_11_dt <- data.table(P2_REF3_11)
P2_REF3_12_dt <- data.table(P2_REF3_12)
P2_REF4_09_dt <- data.table(P2_REF4_09)
P2_REF4_10_dt <- data.table(P2_REF4_10)
P2_REF4_11_dt <- data.table(P2_REF4_11)
P2_REF4_12_dt <- data.table(P2_REF4_12)

# Empilhar passos 2 de cada relacionamento
Passo2 <-
rbind(P2_08_dt,P2_REF1_09_dt,P2_REF1_10_dt,P2_REF1_11_dt,P2_REF1_12_dt,P2_REF2_09_dt,P2_REF2_10_dt,P2_REF2_11_dt,P2_REF2_12_dt,P2_REF3_09_dt,P2_REF3_10_dt,P2_REF3_11_dt,P2_REF3_12_dt,P2_REF4_09_dt,P2_REF4_10_dt,P2_REF4_11_dt,P2_REF4_12_dt)
write.dbf(Passo2, file="Passo2")

names(Passo2)[1] <- "SCORE2"
names(Passo2)

# Passo 9
setwd("F:/Dados/Pos_Relac/Passo9")
P9_08 <- read.dbf("P9_08.dbf")
P9_08_09 <- read.dbf("P9_08_09_v.dbf")

```

```

P9_08_10<-read.dbf("P9_08_10_v.dbf")
P9_08_11<-read.dbf("P9_08_11_v.dbf")
P9_08_12<-read.dbf("P9_08_12_v.dbf")

names(P9_08)
P9_08 <- subset(P9_08, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULT,C_AMO
STRA,C_EPITELIO,C_MICROBIO))

names(P9_08_09)
P9_08_09 <- subset(P9_08_09, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P9_08_09)
names(P9_08_09) [15] <- "C_RESULT"

names(P9_08_10)
P9_08_10 <- subset(P9_08_10, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P9_08_10)
names(P9_08_10) [15] <- "C_RESULT"

names(P9_08_11)
P9_08_11 <- subset(P9_08_11, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P9_08_11)
names(P9_08_11) [15] <- "C_RESULT"

names(P9_08_12)
P9_08_12 <- subset(P9_08_12, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P9_08_12)
names(P9_08_12) [15] <- "C_RESULT"

```



```

P9_08_dt <- data.table(P9_08)
P9_08_09_dt <- data.table(P9_08_09)
P9_08_10_dt <- data.table(P9_08_10)
P9_08_11_dt <- data.table(P9_08_11)
P9_08_12_dt <- data.table(P9_08_12)

# Empilhar passos 9 de cada relacionamento
Passo9 <- rbind(P9_08_dt, P9_08_09_dt, P9_08_10_dt, P9_08_11_dt, P9_08_12_dt)
write.dbf(Passo9, file="Passo9")

names(Passo9)[1] <- "SCORE9"
names(Passo9)

# Passo 12
setwd("C:/Users/camila/Desktop/Doutorado/Pos_Relac/Passo12")
P12_08 <- read.dbf("P12_08.dbf")
P12_08_09 <- read.dbf("P12_08_09_v.dbf")
P12_08_10 <- read.dbf("P12_08_10_v.dbf")
P12_08_11 <- read.xlsx("P12_08_11.xlsx")
P12_08_12 <- read.dbf("P12_08_12_v.dbf")

names(P12_08)
P12_08 <- subset(P12_08, select =
c(SCORE_R_ID, C_ID, C_IDADE, C_DT_NASC, C_COD_MUN, C_CIR, C_MACRO, C_CNES_US, C_DT_EXAME, C_CNES_PRE, C_DT_RECEB, C_DT_LIBER, C_ANO_EX_A, C_RESULT, C_AMO
STRA, C_EPITELIO, C_MICROBIO))

names(P12_08_09)
P12_08_09 <- subset(P12_08_09, select =
c(SCORE_R_ID, C_ID, C_IDADE, C_DT_NASC, C_COD_MUN, C_CIR, C_MACRO, C_CNES_US, C_DT_EXAME, C_CNES_PRE, C_DT_RECEB, C_DT_LIBER, C_ANO_EX_A, C_RESULTAD, C_A
MOSTRA, C_EPITELIO, C_MICROBIO))
names(P12_08_09)
names(P12_08_09)[15] <- "C_RESULT"

names(P12_08_10)
P12_08_10 <- subset(P12_08_10, select =
c(SCORE_R_ID, C_ID, C_IDADE, C_DT_NASC, C_COD_MUN, C_CIR, C_MACRO, C_CNES_US, C_DT_EXAME, C_CNES_PRE, C_DT_RECEB, C_DT_LIBER, C_ANO_EX_A, C_RESULTAD, C_A
MOSTRA, C_EPITELIO, C_MICROBIO))

```

```

names(P12_08_10)
names(P12_08_10)[15] <- "C_RESULT"

names(P12_08_11)
P12_08_11 <- subset(P12_08_11, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P12_08_11)
names(P12_08_11)[15] <- "C_RESULT"

names(P12_08_12)
P12_08_12 <- subset(P12_08_12, select =
c(SCORE_R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P12_08_12)
names(P12_08_12)[15] <- "C_RESULT"

P12_08_dt <- data.table(P12_08)
P12_08_09_dt <- data.table(P12_08_09)
P12_08_10_dt <- data.table(P12_08_10)
P12_08_11_dt <- data.table(P12_08_11)
P12_08_12_dt <- data.table(P12_08_12)

# Empilhar passos 12 de cada relacionamento
Passo12<-rbind(P12_08_dt,P12_08_09_dt,P12_08_10_dt,P12_08_11_dt,P12_08_12_dt)
write.dbf(Passo12, file="Passo12")

names(Passo12)[1] <- "SCORE12"
names(Passo12)

# Passo 15
setwd("C:/Users/camila/Desktop/Doutorado/Pos_Relac/Passo15")
P16_08<-read.dbf("P16_08.dbf")
P15_08_09<-read.dbf("P15_08_09_v.dbf")
P15_08_10<-read.csv2("P15_08_10_v.csv")
P15_08_10<-read_xlsx("P15_08_10_v.xlsx")

```

```

P15_08_11<-read.dbf("P15_08_11_v.dbf")
P15_08_12<-read.dbf("P15_08_12_v.dbf")

names(P15_08)
P16_08 <- subset(P16_08, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULT,C_AMO
STRA,C_EPITELIO,C_MICROBIO))

names(P15_08_09)
P15_08_09 <- subset(P15_08_09, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P15_08_09)
names(P15_08_09) [15] <- "C_RESULT"

names(P15_08_10)
P15_08_10 <- subset(P15_08_10, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P15_08_10)
names(P15_08_10) [15] <- "C_RESULT"

names(P15_08_11)
P15_08_11 <- subset(P15_08_11, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P15_08_11)
names(P15_08_11) [15] <- "C_RESULT"
names(P15_08_12)
P15_08_12 <- subset(P15_08_12, select =
c(SCORE,R_ID,C_ID,C_IDADE,C_DT_NASC,C_COD_MUN,C_CIR,C_MACRO,C_CNES_US,C_DT_EXAME,C_CNES_PRE,C_DT_RECEB,C_DT_LIBER,C_ANO_EX_A,C_RESULTAD,C_A
MOSTRA,C_EPITELIO,C_MICROBIO))
names(P15_08_12)
names(P15_08_12) [15] <- "C_RESULT"

P16_08_dt <- data.table(P16_08)
P15_08_09_dt <- data.table(P15_08_09)

```

```

P15_08_10_dt <- data.table(P15_08_10)
P15_08_11_dt <- data.table(P15_08_11)
P15_08_12_dt <- data.table(P15_08_12)

# Empilhar passos 15 de cada relacionamento
Passo15<-rbind(P16_08_dt,P15_08_09_dt,P15_08_10_dt,P15_08_11_dt,P15_08_12_dt)
write.dbf(Passo15, file="Passo15")

names(Passo15)[1] <- "SCORE15"
names(Passo15)


```

Passo12	1638346 obs. of 18 variables
Passo15	1526687 obs. of 18 variables
Passo2	1477950 obs. of 18 variables
Passo9	1670224 obs. of 18 variables

```

names(Passo15)
P15 <-
Passo15[,c("SCORE15", "R_ID", "C_ID", "C_IDADE", "C_COD_MUN", "C_CIR", "C_MACRO", "C_DT_EXAME", "C_DT_LIBER", "C_RESULT", "C_AMOSTRA", "C_EPITELIO", "C_MICROBIO", "O")]
str(P15)

names(Passo12)
P12 <-
Passo12[,c("SCORE12", "R_ID", "C_ID", "C_IDADE", "C_COD_MUN", "C_CIR", "C_MACRO", "C_DT_EXAME", "C_DT_LIBER", "C_RESULT", "C_AMOSTRA", "C_EPITELIO", "C_MICROBIO", "O")]

names(Passo9)
P9 <-
Passo9[,c("SCORE9", "R_ID", "C_ID", "C_IDADE", "C_COD_MUN", "C_CIR", "C_MACRO", "C_DT_EXAME", "C_DT_LIBER", "C_RESULT", "C_AMOSTRA", "C_EPITELIO", "C_MICROBIO", "O")]

names(Passo2)
P2 <-
Passo2[,c("SCORE2", "R_ID", "C_ID", "C_IDADE", "C_COD_MUN", "C_CIR", "C_MACRO", "C_DT_EXAME", "C_DT_LIBER", "C_RESULT", "C_AMOSTRA", "C_EPITELIO", "C_MICROBIO", "O")]

```

```

P12 1638346 obs. of 13 variables
P15 1526687 obs. of 13 variables
P2 1477950 obs. of 13 variables
P9 1670224 obs. of 13 variables

```

```

# Transformar variáveis
P2$SCORE2 <- as.character(P2$SCORE2)
P2$SCORE2 <- as.numeric(P2$SCORE2)

P9$SCORE9 <- as.character(P9$SCORE9)
P9$SCORE9 <- as.numeric(P9$SCORE9)

P12$SCORE12 <- as.character(P12$SCORE12)
P12$SCORE12 <- as.numeric(P12$SCORE12)

P15$SCORE15 <- as.character(P15$SCORE15)
P15$SCORE15 <- as.numeric(P15$SCORE15)

# Transformar variável no formato data
library(zoo)
P2$C_DT_EXAME <- as.character(P2$C_DT_EXAME)
P2$C_DT_EXAME <- as.Date(P2$C_DT_EXAME, "%Y%m%d")

P2$C_DT_LIBER <- as.character(P2$C_DT_LIBER)
P2$C_DT_LIBER <- as.Date(P2$C_DT_LIBER, "%Y%m%d")

P9$C_DT_EXAME <- as.character(P9$C_DT_EXAME)
P9$C_DT_EXAME <- as.Date(P9$C_DT_EXAME, "%Y%m%d")

P9$C_DT_LIBER <- as.character(P9$C_DT_LIBER)
P9$C_DT_LIBER <- as.Date(P9$C_DT_LIBER, "%Y%m%d")

P12$C_DT_EXAME <- as.character(P12$C_DT_EXAME)
P12$C_DT_EXAME <- as.Date(P12$C_DT_EXAME, "%Y%m%d")

```

```

P12$C_DT_LIBER <- as.character(P12$C_DT_LIBER)
P12$C_DT_LIBER <- as.Date(P12$C_DT_LIBER, "%Y%m%d")

P15$C_DT_EXAME <- as.character(P15$C_DT_EXAME)
P15$C_DT_EXAME <- as.Date(P15$C_DT_EXAME, "%Y%m%d")

P15$C_DT_LIBER <- as.character(P15$C_DT_LIBER)
P15$C_DT_LIBER <- as.Date(P15$C_DT_LIBER, "%Y%m%d")

## VERIFICAR O TOTAL DE PARES VERDADEIROS FORMADOS
# Criando a variável pares
P2$pares <- as.factor(paste(P2$R_ID,sep=" ",P2$C_ID))
P9$pares <- as.factor(paste(P9$R_ID,sep=" ",P9$C_ID))
P12$pares <- as.factor(paste(P12$R_ID,sep=" ",P12$C_ID))
P15$pares <- as.factor(paste(P15$R_ID,sep=" ",P15$C_ID))

names(P2)

# juntando todos os bancos (apenas as variáveis pares e SCORE)
Passo2s <- P2[,c(1,14)]
Passo9s <- P9[,c(1,14)]
Passo12s <- P12[,c(1,14)]
Passo15s <- P15[,c(1,14)]

Passostotal_s <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="pares",all=T),
                        list(Passo2s,Passo9s,Passo12s,Passo15s))

write.dbf(Passostotal_s, file="Passostotal_s")

```

Passo12s	1638346	obs. of 2 variables
Passo15s	1526687	obs. of 2 variables
Passo2s	1477950	obs. of 2 variables
Passo9s	1670224	obs. of 2 variables
Passostotal_s	3431089	obs. of 5 variables

```

library(arsenal)

```

```

comparedf(P2,P12,by="pares")[1]
comparedf(P2,P9,by="pares")[1]
comparedf(P2,P15,by="pares")[1]
comparedf(P9,P12,by="pares")[1]
comparedf(P9,P15,by="pares")[1]
comparedf(P12,P15,by="pares")[1]
> library(arsenal)
> comparedf(P2,P12,by="pares") [1]
$frame.summary
  version arg ncol  nrow  by  attrs  unique n.shared
1     X   P2   14 1477950 pares 4 attributes 83850 unique obs 1394100
2     Y   P12  14 1638346 pares 4 attributes 244246 unique obs 1394100
> comparedf(P2,P9,by="pares") [1]
$frame.summary
  version arg ncol  nrow  by  attrs  unique n.shared
1     X   P2   14 1477950 pares 4 attributes 1477950 unique obs 0
2     Y   P9   14 1670224 pares 4 attributes 1670224 unique obs 0
> comparedf(P2,P15,by="pares") [1]
$frame.summary
  version arg ncol  nrow  by  attrs  unique n.shared
1     X   P2   14 1477950 pares 4 attributes 1477950 unique obs 0
2     Y   P15  14 1526687 pares 4 attributes 1526687 unique obs 0
> comparedf(P9,P12,by="pares") [1]
$frame.summary
  version arg ncol  nrow  by  attrs  unique n.shared
1     X   P9   14 1670224 pares 4 attributes 1670224 unique obs 0
2     Y   P12  14 1638346 pares 4 attributes 1638346 unique obs 0
> comparedf(P9,P15,by="pares") [1]
$frame.summary
  version arg ncol  nrow  by  attrs  unique n.shared
1     X   P9   14 1670224 pares 4 attributes 182206 unique obs 1488018
2     Y   P15  14 1526687 pares 4 attributes 38669 unique obs 1488018
> comparedf(P12,P15,by="pares") [1]
$frame.summary
  version arg ncol  nrow  by  attrs  unique n.shared
1     X   P12  14 1638346 pares 4 attributes 1638346 unique obs 0
2     Y   P15  14 1526687 pares 4 attributes 1526687 unique obs 0
names(P2)[1] <- "SCORE"

```

```

names(P9)[1] <- "SCORE"
names(P12)[1] <- "SCORE"
names(P15)[1] <- "SCORE"

P_2_12<-rbind(P2,P12)
P_9_15<-rbind(P9,P15)

## Retirar duplicidade
aux <- which(duplicated(P_2_12[,c("pares")]))
dup_P2_12 <- P_2_12[aux,]
if(length(aux)>0){P_2_12_v <- P_2_12[-aux,]}else{P_2_12_v <- P_2_12}

aux <- which(duplicated(P_9_15[,c("pares")]))
dup_P9_15 <- P_9_15[aux,]
if(length(aux)>0){P_9_15_v <- P_9_15[-aux,]}else{P_9_15_v <- P_9_15}

```

```

1 dup_P2_12 1394100 obs. of 14 variables
2 dup_P9_15 1488018 obs. of 14 variables
3 P_2_12 3116296 obs. of 14 variables
4 P_2_12_v 1722196 obs. of 14 variables
5 P_9_15 3196911 obs. of 14 variables
6 P_9_15_v 1708893 obs. of 14 variables

```

```
comparedf(P_2_12_v,P_9_15_v,by="pares")[1]
```

```
P_total<-rbind(P_2_12_v,P_9_15_v)
```

```

> comparedf(P_2_12_v,P_9_15_v,by="pares") [1]
$frame.summary
  version  arg ncol  nrow  by  attrs  unique  n.shared
1 X P_2_12_v  14 1722196 pares 4 attributes 1722196 unique obs  0
2 Y P_9_15_v  14 1708893 pares 4 attributes 1708893 unique obs  0

```

```

1 P_2_12_v 1722196 obs. of 14 variables
2 P_9_15_v 1708893 obs. of 14 variables
3 P_total 3431089 obs. of 14 variables

```



```

#Identificar as datas das repetições dos exames e combinar os bancos em um novo banco

library(foreign)
setwd("F:/Dados/Pos_Relac")

referencia <- read.dbf("REF_2008_3.dbf")

# Transformar o banco no formato data table (formato data frame dando erro)
referencia_dt <- data.table(referencia)
names(referencia_dt)

# Transformar variáveis contendo data no formato data
referencia_dt$DT_EXAME<- as.character(referencia_dt$DT_EXAME)
library(zoo)
referencia_dt$DT_EXAME<- as.Date(referencia_dt$DT_EXAME, "%Y%m%d")

referencia_dt$DT_LIBER<- as.character(referencia_dt$DT_LIBER)
referencia_dt$DT_LIBER<- as.Date(referencia_dt$DT_LIBER, "%Y%m%d")

str(referencia_dt)

names(P_total)

REF <- referencia_dt[,c("ID", "IDADE", "COD_MUN", "CIR", "MACRO", "DT_EXAME", "DT_LIBER", "RESULTADO", "AMOSTRA", "EPITELIO", "MICROBIOL")]
names(REF)

Passos_final <- P_total[,c("R_ID", "C_IDADE", "C_COD_MUN", "C_CIR", "C_MACRO", "C_DT_EXAME", "C_DT_LIBER", "C_RESULT", "C_AMOSTRA", "C_EPITELIO", "C_MICROBIO")]
names(Passos_final)
names(Passos_final) <- c("ID", "IDADE", "COD_MUN", "CIR", "MACRO", "DT_EXAME", "DT_LIBER", "RESULTADO", "AMOSTRA", "EPITELIO", "MICROBIOL")

```

P_total	3431089 obs. of 14 variables
Passos_final	3431089 obs. of 11 variables
REF	1162411 obs. of 11 variables
referencia_dt	1162411 obs. of 27 variables

```

# empilhar os bancos de dados
db <- rbind(Passos_final,REF)

# Ordenar o banco pela data de exame
db <- db[order(db$DT_EXAME),]

# identificar e excluir duplicidades
aux <- which(duplicated(db))
dbdup <- db[aux,]
if(length(aux)>0){db1 <- db[-aux,]}else{db1 <- db}
rm(aux)

```

	db	4593500 obs. of 11 variables
	Passos_final	3431089 obs. of 11 variables
	REF	1162411 obs. of 11 variables

```

# frequencias de repetições do exame
rep <- as.data.frame(table(db1$ID))
table(rep$Freq)

```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	304375	310214	265447	169893	72877	19732	7751	3854	2001	1319	832	607	484	419	304	294	229	194
	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	179	148	129	109	99	84	66	63	62	56	43	35	31	30	29	20	24	23
	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
	18	17	21	20	23	13	20	15	11	15	9	9	10	4	9	4	8	6
	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
	4	3	4	3	4	5	4	5	3	6	8	6	1	3	4	4	2	3
	73	74	75	76	77	78	79	80	81	82	84	85	88	133	134	135	136	138
	1	2	3	2	1	3	1	3	1	2	1	1	1	1	1	1	1	1
	139	145	146	147	148	149	151	153	154	155	157	159	164	169				
	2	1	1	1	1	1	1	2	2	1	1	1	1	1				

```

# identificar as primeiras datas de exame
aux <- which(duplicated(db1[,c("ID")]))
dbdup1 <- db1[aux,]
if(length(aux)>0){data0 <- db1[-aux,]}else{data0 <- db1}
names(data0)
names(data0) <- c("ID", "IDADE0", "COD_MUN0", "CIR0", "MACRO0", "DT_EXAME0", "DT_LIBERO", "AMOSTRA0", "RESULTADO0", "EPITELIO0", "MICROBIOLO")
rm(aux)

```

```

# identificar as primeiras repetições do exame
aux <- which(duplicated(dbdup1[,c("ID")]))
dbdup2 <- dbdup1[aux,]
if(length(aux)>0){data1 <- dbdup1[-aux,]}else{data1 <- dbdup1}
names(data1) <- c("ID", "IDADE1", "COD_MUN1", "CIR1", "MACRO1", "DT_EXAME1", "RESULTADO1", "AMOSTRA1", "EPITELIO1", "MICROBIO1")
rm(aux)

# identificar as segundas repetições do exame
aux <- which(duplicated(dbdup2[,c("ID")]))
dbdup3 <- dbdup2[aux,]
if(length(aux)>0){data2 <- dbdup2[-aux,]}else{data2 <- dbdup2}
names(data2) <- c("ID", "IDADE2", "COD_MUN2", "CIR2", "MACRO2", "DT_EXAME2", "DT_LIBER2", "RESULTADO2", "AMOSTRA2", "EPITELIO2", "MICROBIO2")
rm(aux)

# identificar as terceiras repetições do exame
aux <- which(duplicated(dbdup3[,c("ID")]))
dbdup4 <- dbdup3[aux,]
if(length(aux)>0){data3 <- dbdup3[-aux,]}else{data3 <- dbdup3}
names(data3) <- c("ID", "IDADE3", "COD_MUN3", "CIR3", "MACRO3", "DT_EXAME3", "DT_LIBER3", "RESULTADO3", "AMOSTRA3", "EPITELIO3", "MICROBIO3")
rm(aux)

# identificar as quartas repetições do exame
aux <- which(duplicated(dbdup4[,c("ID")]))
dbdup5 <- dbdup4[aux,]
if(length(aux)>0){data4 <- dbdup4[-aux,]}else{data4 <- dbdup4}
names(data4) <- c("ID", "IDADE4", "COD_MUN4", "CIR4", "MACRO4", "DT_EXAME4", "DT_LIBER4", "RESULTADO4", "AMOSTRA4", "EPITELIO4", "MICROBIO4")
rm(aux)

# combinar todas as datas de cada mulher
dados <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="ID",all.x=T),
list(data0,data1,data2,data4,data4))
write.dbf(dados, "Relac_total.dbf")

```

Banco_seguintotal	1162411 obs. of 51 variables
data0	1162411 obs. of 11 variables
data1	858036 obs. of 11 variables
data2	547822 obs. of 11 variables
data3	282375 obs. of 11 variables
data4	112482 obs. of 11 variables
db	4593500 obs. of 11 variables
db1	3095034 obs. of 11 variables
dbdup	1498466 obs. of 11 variables
dbdup1	1932623 obs. of 11 variables
dbdup2	1074587 obs. of 11 variables
dbdup3	526765 obs. of 11 variables
dbdup4	244390 obs. of 11 variables
dbdups	131908 obs. of 11 variables
rep	1162411 obs. of 2 variables

ID	IDADE	COD_MUN	CIR	MACRO	DT_EXAME	DT_LIBER	RESULTADO	AMOSTRA	EPITELIO	MICROBIOL
4593492	C08 592006	64	313420	ITUIUTABA	2012-12-27	2012-12-28	0	0	6	0
4593493	C08 403696	22	312090	CURVELO	2012-12-28	2012-12-28	0	0	3	3
4593494	C08 395677	33	311970	SAO JOAO DEL REI	2012-12-28	2012-12-28	0	0	6	0
4593495	C08 875045	53	314900	CARANICOLA	2012-12-28	2012-12-31	0	0	3	0
4593496	C08 1003106	37	315733	SAO JOAO DEL REI	2012-12-28	2012-12-28	0	0	6	0
4593497	C08 395677	33	311970	SAO JOAO DEL REI	2012-12-28	2012-12-28	0	0	6	0
4593498	C08 403696	22	312090	CURVELO	2012-12-28	2012-12-28	0	0	3	3
4593499	C08 875045	53	314900	CARANICOLA	2012-12-28	2012-12-31	0	0	3	0
4593500	C08 1003106	37	315733	SAO JOAO DEL REI	2012-12-28	2012-12-28	0	0	6	0