

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

Daniela de Almeida Pereira

**AVALIAÇÃO DO RASTREAMENTO DO CÂNCER DE MAMA EM MINAS GERAIS A
PARTIR DE DADOS DO SISMAMA**

JUIZ DE FORA
2021

Daniela de Almeida Pereira

**AVALIAÇÃO DO RASTREAMENTO DO CÂNCER DE MAMA EM MINAS GERAIS A
PARTIR DE DADOS DO SISMAMA**

Tese apresentada ao Programa de Saúde Coletiva da Faculdade de Medicina da Universidade Federal de Juiz de Fora, como requisito parcial a obtenção do título de doutor.

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)

Pereira, Daniela de Almeida.

Avaliação do rastreamento do câncer de mama em Minas Gerais a partir de dados do SISMAMA / Daniela de Almeida Pereira. -- 2021. 222 f. : il.

Orientador: 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. Sistema de Informação em Saúde. 2. Relacionamento de bases de dados. 3. Cobertura de Mamografia. 4. Rastreamento. 5. Fatores Socioeconômicos. I. Bustamante Teixeira, Maria Teresa, orient. II. Nogueira, Mário Círio, coorient. III. Título.

Daniela de Almeida Pereira

Avaliação do Rastreamento do Câncer de Mama em Minas Gerais a partir de Dados do SISMAMA.

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.

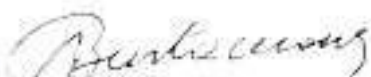
Área de Concentração: Saúde Coletiva.

Aprovada em 22 de junho de 2021

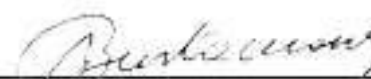
BANCA EXAMINADORA



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



Prof. Dr. Mário Cirio Nogueira – Coorientador
Universidade Federal de Juiz de Fora



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



Profa. Dra. Jeane Glauca Tomazelli

Instituto Nacional do Câncer



Profa. Dra. Joan Leigh Warren

National Cancer Institute - EUA



Profa. Dra. Isabel Cristina Gonçalves Leite

Universidade Federal de Juiz de Fora

Dedico este trabalho, com muito amor e gratidão:

À minha mãe Maura, que não teve a oportunidade de cursar o ensino superior, mas sempre incentivou e se esforçou para que todos os filhos estudassem e crescessem profissionalmente.

À minha filha Rebeca, que me mostrou que sou mais forte do que eu imaginava e me ajudou a crescer e a ser quem eu sou hoje.

AGRADECIMENTOS

A Deus pela oportunidade de fazer doutorado, por cuidar de mim em todas as viagens de ida e volta para Juiz de Fora, por colocar pessoas especiais no meu caminho e por tantas bênçãos durante toda essa jornada e que aqui nem dá para descrever.

À minha mãe, Maura, que me acompanhou nesta jornada, pelo apoio e orações, sem você eu não teria forças para chegar até o fim! Aos meus irmãos Deyliane e Daniel, pelo apoio. E o Dênis pelo apoio e suporte em informática e tecnologia.

À minha filha Rebeca, que mesmo tão pequena soube entender e compreender que eu precisava de tempo para me dedicar ao desenvolvimento desse trabalho, e com seu jeitinho tão meigo e carinhoso cuidou de mim, sempre fazia um cafezinho ou chazinho e ao trazer para mim dizia: “mamãe, fiz com muito amor para você estudar”. Te amo minha filha!

À minha tia Ana por me receber em sua casa em Juiz de Fora com muito carinho e hospitalidade. À minha tia Lúcia pelo apoio e incentivo em todos os momentos que precisei, principalmente nos cuidados com a Rebeca.

Às professoras do Centro Municipal de Educação Infantil Raios de Sol, Marcinha, Silvania, Edivane, Gena e Sandra, por cuidarem da Rebeca com tanto carinho nos anos iniciais do meu doutorado e por sempre me enviar fotos, vídeos e mensagens. Vocês não imaginam o quanto isso me confortava e alegrava, pois sabia que na minha ausência minha filha estava bem!

À minha orientadora, Maria Teresa Bustamante, por todo ensinamento, dedicação e por acreditar em mim e ser sempre essa pessoa tão amável e querida, como sempre dizemos: “Uma grande mãe”! Ao meu co-orientador Mário Círio, pela dedicação, paciência e disponibilidade para nos ensinar estatística, sobre o R e a construir os scripts e mapas, sem você não teríamos conseguido!

À minha amiga, companheira nesta jornada, presente de Deus na minha vida, Camila, por compartilhar seu conhecimento, pelas trocas de ideias e ajuda mútua, juntas conseguimos avançar e ultrapassar todos os obstáculos! À Vívian Assis Fayer, pela parceria, disponibilidade e paciência para me ensinar sobre o RecLink.

Às professoras Gulnar Azevedo e Silva e Maria Teresa Bustamante pela oportunidade de fazer parte do grupo de pesquisa Newton. Como eu cresci e estou crescendo ao lado de vocês!

Aos membros da banca Gulnar Azevedo e Silva, Jeane Gláucia Tomazelli, Isabel Cristina Gonçalves Leite e Joan Leigh Warren, pela disponibilidade, reflexões e contribuições desde a qualificação.

À toda equipe do Nates, por sempre me receberem com carinho, por se preocuparem com meu bem-estar no deslocamento de ida para Juiz de Fora e retorno para casa. Estar ao lado de vocês sempre fez muito bem para mim!

À Universidade Federal de Viçosa, à ex-chefe da Divisão de Saúde Nathália Dias Pereira Alves Oliveira, a atual chefe Marisa Dibbern Lopes Correia e aos funcionários do serviço de capacitação pela liberação para cursar o doutorado, apoio, incentivo e orientações em todas as etapas do processo administrativo.

Aos amigos da Divisão de Saúde da Universidade Federal de Viçosa pelas palavras de incentivo, apoio e carinho.

À Fundação de Amparo à Pesquisa de Minas Gerais – FAPEMIG pelo financiamento que possibilitou a operacionalização do estudo.

Por fim, agradeço à vida, que com suas dificuldades e desafios, me ensinou a ser forte e a lutar por meus sonhos, a reconhecer que existem dias ruins, mas não podemos deixar que eles nos destruam e nos tirem o foco, porque existem também dias maravilhosos, dias em que os sonhos se realizam, e hoje é um desses dias!

As lutas concernentes à saúde e às desigualdades em saúde não são uma simples questão de “fatos” científicos. As lutas pela saúde são lutas políticas. Os cientistas desempenham um papel nelas, de um lado ou de outro. Uma das maneiras pelas quais eles podem contribuir é documentar os problemas existentes em relação à saúde e às desigualdades em saúde. Muitas vezes, tais problemas permanecem ocultos por falta dos dados relevantes. Revelar as desigualdades anteriormente obscurecidas é o primeiro passo no caminho da ação para reduzi-las.

(McIntyre; Mooney. 2014, p. 48)

RESUMO

O objetivo desta tese foi avaliar o rastreamento do câncer de mama em Minas Gerais a partir de dados identificados do SISMAMA e sua conformidade com as diretrizes nacionais. Foram realizados três estudos, cujos resultados são apresentados em formato de artigos. O relacionamento probabilístico de banco de dados foi uma das principais técnicas utilizadas no desenvolvimento desta tese. De acordo com o tamanho das bases vinculadas, o procedimento pode resultar na formação de muitos pares, o que inviabiliza a conferência manual para identificação de um par verdadeiro. Assim, no primeiro estudo, foi criado um método automatizado para seleção de escore e identificação de pares em relacionamento probabilístico. Para isso, as bases de exames alterados e referência foram relacionadas por meio de 16 passos. Cada passo foi inspecionado manualmente para se obter um padrão-ouro e selecionaram-se amostras que foram inspecionadas e avaliadas para calcular a acurácia do método. A criação do método agilizou e facilitou a etapa de identificação dos pares verdadeiros, sendo constatada uma boa qualidade das medidas de avaliação dos relacionamentos realizados.

No segundo estudo, foi estimada a cobertura do rastreamento mamográfico no estado e suas microrregiões de saúde e avaliados os fatores associados. Para calcular a cobertura do rastreamento foram removidas as duplicidades e repetições pelos métodos determinístico e probabilístico. A associação entre a cobertura e as demais variáveis foi analisada por modelos de regressão linear. A cobertura estimada para a população alvo do rastreamento mamográfico no estado foi de 36,3%. Nas microrregiões a cobertura variou de 2,66% a 61,70%. Foi observado que as maiores coberturas ocorreram nas microrregiões com melhor nível de desenvolvimento socioeconômico e menor vulnerabilidade na saúde (Centro e Sul); e as menores nas microrregiões menos desenvolvidas e com maior vulnerabilidade na saúde, Norte e Nordeste. Foi encontrada associação positiva entre cobertura e renda, número de mamografias e capacidade de utilizá-las. Em contraste, a análise mostrou uma associação negativa com o Índice de Gini, demonstrando que quanto maior a renda, menor desigualdade de renda, maior número de mamógrafos e capacidade de utilização desse equipamento, maior a cobertura de mamografia.

No terceiro estudo o sobrerastreamento foi avaliado a partir de uma coorte não concorrente de mulheres rastreadas com registro no SISMAMA, por meio do relacionamento probabilístico entre a base referência e seguimento. Foram estimados o percentual de

sobrerrastreio por faixa etária (realização de exames em mulheres com idade <50 anos ou >=70 anos) e o sobrerrastreamento mamográfico por periodicidade excessiva (repetição do exame em um período inferior a 18 meses). A associação entre o sobrerrastreamento e as variáveis sociodemográficas e assistenciais foram analisadas em um estudo ecológico que tomou como unidades de análise as 77 microregiões de saúde do estado, utilizando modelos de regressão. Verificou-se que sobrerrastreamento por faixa etária atingiu 47,1% das mulheres e por periodicidade 27,8%. Os fatores associados ao sobrerrastreamento por periodicidade foram maior renda, número de mamógrafos e menor desigualdade de renda. Não foi encontrada associação entre o sobrerrastreamento por faixa etária e as variáveis estudadas. Ressalta-se, que foi observado que microrregiões com baixa cobertura e menor número de recursos humanos e equipamentos, realizam sobrerrastreamento, o que demonstra a necessidade de adequações.

A partir do desenvolvimento dessa tese conclui-se que em Minas Gerais, a cobertura de mamografia na faixa etária alvo, 50 a 69 anos no período do estudo era baixa, e que as recomendações de faixa etária e periodicidade não eram adotadas. Dessa forma, as ações de rastreamento no estado precisam de melhorias e adequações e estas devem focar na distribuição de recursos de acordo com a necessidade e característica de cada região, na organização e otimização das ações de rastreamento, na conscientização da população em geral e capacitação dos profissionais, sobretudo acerca das evidências que embasaram a elaboração das diretrizes adotadas no Brasil.

ABSTRACT

This thesis aims to evaluate breast cancer screening in Minas Gerais according to data from SISMAMA and national guidelines. Three studies were conducted, and results are presented as articles. Probabilistic data linkage was one of the main techniques used in the thesis. According to size of linked databases, the procedure may lead to multiple pairs, hindering manual conference to identifying a real pair. Thus, the first study created an automated method to select scores and identify pairs in probabilistic data linkage. Sixteen steps linked databases of altered exams and references. Every step was manually verified to obtain a gold standard, and samples were selected to calculate accuracy of the method. The method simplified and facilitated identification of real pairs and found good quality of assessment measures from linkages performed.

The second study estimated mammographic screening in the state and health micro-regions and associated factors were evaluated. Duplicates and repetitions were removed by deterministic and probabilistic methods to calculate screening coverage. Associations between coverage and other variables were analyzed using linear regression model. Estimated coverage for the population of mammographic screening was 36.3%. In micro-regions, coverage varied from 2.66% to 61.70%. Higher coverage occurred in micro-regions with better socioeconomic development levels and low health vulnerability (central and south) ; north and northeast micro-regions presented lower coverage since they are less developed and have higher health vulnerability. A positive association was found between coverage and income, number of mammograms, and capacity of using mammograms. In contrast, analysis showed a negative association with Gini Index, indicating the higher the mammography coverage, the higher the income, number of mammograms, and capacity of using this equipment, and the lower the income inequality.

In the third study, over-screening was evaluated in a non-concurrent cohort of registers in SISMAMA using probabilistic data linkage between the reference database and follow-up. Percentage of over-screening was estimated by age (exams in women aging < 50 or ≥ 70 years) and excessive periodicity (repetition of the exam in a period lower than 18 months). Associations between over-screening and sociodemographic and assistance variables were analyzed using regression model in an ecological study using 77 health micro-regions of the state. Over-screening by age and periodicity reached 47.1% and 27.8% of women, respectively. The latter was associated with income, number of mammograms, and

low-income inequality. No associations were found between over-screening by age and variables studied. Micro-regions with low coverage, number of human resources, and equipments performed over-screening, demonstrating the need for adequacies.

We conclude that mammography coverage between 50-69 years in Minas Gerais was low in the period studied , and recommendations by age and periodicity were not adopted. Screening actions require improvements and adequacies focusing on resource distribution according to characteristics and needs of each region, organized and optimized screening actions, awareness of the general population, and professional training regarding evidence that support guidelines adopted in Brazil.

LISTA DE QUADROS

Quadro 1 - Programas de Rastreamento do Câncer de Mama nos Países do Mundo.....	32
Quadro 2 - Condutas clínicas segundo resultado do exame mamográfico.....	42
Quadro 3 - Síntese das recomendações de mamografia por faixa etária.....	43
Quadro 4 - Variáveis selecionadas no SISMAMA.....	62
Quadro 5 - Combinação das variáveis para retirada de duplicidades e repetições.....	67
Quadro 6 - Critérios para classificação manual dos pares formados no relacionamento probabilístico.....	68
Quadro 7 - Cálculo dos indicadores socioeconômicos e de saúde, 2010.....	73
Quadro 8 - Pesos atribuídos aos indicadores selecionados, IVS 2012.....	75

LISTA DE TABELAS

Tabela 1 -	Passos, Variáveis de blocagem e pareamento, escores, pares formados e pares verdadeiros do relacionamento dos bancos do SISMAMA 2009 e 2010.....	68
Tabela 2 -	População feminina de Minas Gerais e cobertura de mamografia, 2010.....	70
Tabela 3 -	Passos, Variáveis de blocagem e pareamento para relacionamento probabilístico dos bancos do SISMAMA.....	71

ARTIGO 1

Tabela 1 -	Passos, Variáveis de blocagem e pareamento, escores, pares formados e pares verdadeiros do relacionamento dos bancos do SISMAMA 2009 e 2010.....	82
Tabela 2 -	Combinação de pares do relacionamento dos bancos do SISMAMA 2009 e 2010.....	84
Tabela 3 -	Ponto de corte dos passos e amostras do Relacionamento dos bancos do SISMAMA 2009 e 2010.....	85
Tabela 4 -	Recall e precision do ponto de corte dos passos e amostras do relacionamento dos bancos do SISMAMA 2009.....	86
Tabela 5 -	Acurácia, medida F e área sob a curva <i>precision-recall</i> (AUPRC) do ponto de corte dos passos e amostras do relacionamento dos bancos do SISMAMA 2009 e 2010.....	87

ARTIGO 2

Tabela 1 -	Caracterização da cobertura de mamografias em mulheres nas faixas etárias de 50-59, 60-69 e 50-69 anos e dos indicadores socioeconômicos e de saúde das 77 microrregiões de saúde de Minas Gerais, Brasil, 2010.....	98
Tabela 2 -	Modelos de Regressão Simples e Múltiplos entre as coberturas de mamografias e indicadores sociodemográficos e assistenciais das microrregiões de saúde de Minas Gerais, 2010.....	101

ARTIGO 3

Tabela 1 - Sobrerrastreamento por faixa etária* no estado de Minas Gerais e macrorregiões, janeiro a junho de 2010.....	121
Tabela 2 – Sobrerrastreamento por periodicidade nas macrorregiões de saúde de Minas Gerais, 2010.....	123
Tabela 3 – Distribuição do tempo em meses entre a data do primeiro exame realizado em 2010 e a data da primeira repetição.....	123
Tabela 4 - Modelos de Regressão Simples e Múltiplo entre o sobrerrastreamento e indicadores sociodemográficos e assistenciais das microrregiões de saúde de Minas Gerais, 2010.....	125

LISTA DE ILUSTRAÇÕES

Figura 1 - Principais causas de mortalidade por câncer em 2020, sexo feminino, em todas as idades.....	24
Figura 2 - Taxas de incidência e mortalidade do câncer de mama padronizadas por idade, nos continentes, em todas as idades, em 2018 e em 2020.....	25
Figura 3 - Principais ponto de atenção para o rastreamento do câncer de mama.....	40
Figura 4 - Fluxo do SISCAN.....	52
Figura 5 - Macrorregiões e Microrregiões de Saúde de Minas Gerais.....	61
Figura 6 - Processo de limpeza da base exames alterados (ano 2009).....	64
Figura 7 - Processo de limpeza da base referência (2010).....	64
Figura 8 - Processo de limpeza da base seguimento (2010-2012).....	65
Figura 9 - Processo de limpeza da base cobertura (ano 2010).....	65

ARTIGO 2

Figura 1 - Macrorregiões e Microrregiões de Saúde de Minas Gerais.....	95
Figura 2 - Distribuição e Lisa Map das coberturas de mamografias nas microrregiões de saúde de Minas Gerais, 2010.....	99
Figura 3 - Macrorregiões de saúde Indicadores socioeconômicos das microrregiões de saúde de Minas Gerais, 2010.....	102
Figura 4 - Indicadores assistenciais das microrregiões de saúde de Minas Gerais, 2010.....	103

ARTIGO 3

Figura 1 - Processo de limpeza das bases referência e seguimento para relacionamento probabilístico e da base de sobrerastreamento.....	117
Figura 2 - Mapas temáticos sobrerastreamento por por faixa etária, periodicidade e cobertura de mamografia (esquerda), índice de Moran local - LisaMap (direita), microrregiões de saúde de Minas Gerais, 2010.....	122
Figura 3 - Curva do sobrerastreamento por periodicidade (faixa etária de 50 a 69 anos)	124
Figura 4 - Mapas temáticos dos indicadores sociodemográficos e assistenciais, microrregiões de saúde de Minas Gerais, 2010.....	126

LISTA DE ABREVIATURAS E SIGLAS

APS – Atenção Primária à Saúde

DATASUS - Departamento de Informática do Sistema Único de Saúde

DCNT – Doenças Crônicas não Transmissíveis

IARC – Agência Internacional de Pesquisa sobre o Câncer

IBGE – Instituto Brasileiro de Geografia e Estatística

IDH – Índice de Desenvolvimento Humano

IVS – Índice de Vulnerabilidade na Saúde

PROADESS - Programa de Avaliação do Desempenho do Sistema de Saúde

SIS – Sistemas de Informação em Saúde

SISCAN - Sistema de Informação do Câncer

SISCOLO - Sistema Informação do Câncer de Colo do útero

SISMAMA – Sistema de Informação do Câncer de Mama

SUS - Sistema Único de Saúde

SUMÁRIO

1	INTRODUÇÃO	19
2	REFERENCIAL TEÓRICO	23
	2.1 Magnitude do Câncer de Mama.....	23
	2.2 Programas de Rastreamento do Câncer de Mama no Mundo	27
	2.3 Controle do Câncer de Mama no Brasil.....	38
	2.4 Desigualdade no Rastreamento do Câncer de Mama	45
	2.5 Sistema de Informação e Controle do Câncer de Mama (SISMAMA)	50
	2.6 Relacionamento de Sistemas de Informação em Saúde	53
3	JUSTIFICATIVA.....	57
4	OBJETIVOS	59
	4.1 Objetivo Geral.....	59
	4.2 Objetivos Específicos.....	59
5	METODOLOGIA	60
	5.1 Tipo de Estudo	60
	5.2 População de Estudo e Fonte dos Dados	60
	5.3 Construção da base de dados	61
	5.3.1 Limpeza das bases e padronização das variáveis	63
	5.3.2 Remoção de duplicidades e de exames repetidos	66
	5.3.3 Exclusão de exames com resultados alterados	67
	5.3.4 Construção da base cobertura, referência e seguimento.....	69
	5.3.5 Acurácia de um método para seleção de pares em relacionamento probabilístico de banco de dados.....	69
	5.4 Análise dos dados	70

5.4.1 Cálculo da cobertura do rastreamento	70
5.4.2 Avaliação do sobrerastreamento por faixa etária e periodicidade excessiva	71
5.4.4 Associação da cobertura e do sobrerastreamento com variáveis sociodemográficas e assistenciais	72
5.5 Aspectos Éticos	76
6 RESULTADOS.....	78
6.1 Artigo 1: Avaliação de Método para Classificação Automatizada de Pares em Relacionamentos Probabilísticos de Bancos de Dados	78
6.2 Artigo 2: Cobertura do rastreamento mamográfico em Minas Gerais a partir de dados identificados do SISMAMA	92
6.3 Artigo 3: Sobrerastreamento mamográfico por faixa etária e periodicidade e fatores associados	113
CONSIDERAÇÕES FINAIS	135
REFERÊNCIAS	138
APÊNDICES	153

1 INTRODUÇÃO

O câncer de mama é a neoplasia mais comumente diagnosticada no sexo feminino e é considerada a principal causa de morte de mulheres por câncer. Em diversos países tem sido observado aumento da sua incidência, porém naqueles de alta renda, desde o início de 2000 tem ocorrido estabilidade ou queda, provavelmente devido a reduções no uso da terapia hormonal na menopausa. Ainda nesses países, melhorias na detecção precoce por meio do rastreamento populacional e intervenções terapêuticas mais eficazes contribuíram para o declínio das taxas de mortalidade na década de 1990. Por outro lado, países de baixa renda experienciam o aumento das taxas de incidência resultante das mudanças no estilo de vida e no ambiente sociocultural das mulheres, que associadas a infraestrutura de saúde deficiente e à falta de organização dos serviços para prevenção, detecção e diagnóstico precoces contribuem para o aumento de casos identificados em estágio avançado e da mortalidade por câncer de mama (SUNG et al., 2021).

No Brasil, entre 1990 e 2017, foi observado em quase metade dos estados localizados nas regiões mais desenvolvidas, Sul, sudeste e Centro-Oeste, redução das taxas de mortalidade por câncer de mama. Por outro lado, alguns estados tiveram aumento da mortalidade, principalmente em uma região menos desenvolvida, Nordeste (GUERRA et al., 2020). Em 2020, foram estimados para o país 20.725 óbitos de mulheres por câncer de mama e 88.492 casos novos (SUNG et al., 2021).

Por se tratar de um tumor que tem grande chance de cura, se diagnosticado precocemente e tratado adequadamente, em tempo hábil, o câncer de mama ocupa importante espaço no campo da ciência, principalmente, no que diz respeito às suas formas de prevenção (WHO, 2014). Nesse sentido, cada país, de acordo com o seu contexto e questões demográficas, epidemiológicas e econômicas define seu programa de controle do câncer. A finalidade desse tipo de programa é reduzir a incidência, a mortalidade e a morbidade por câncer através de ações sistemáticas e integradas que fazem parte do cuidado integral, a saber: prevenção primária (redução ou eliminação dos fatores de risco), detecção precoce, tratamento, reabilitação e cuidados paliativos. No caso do câncer de mama a detecção precoce consiste em diagnóstico precoce, ou seja, na identificação, o mais precocemente possível, do câncer de mama em indivíduos sintomáticos; e rastreamento, definido como a identificação do câncer de mama em indivíduos assintomáticos (INCA, 2015a).

Nos países de alta renda, que implementaram ações de diagnóstico precoce e rastreamento populacional organizado, os cânceres de mama são diagnosticados em estágios localizados, o que contribui para um bom prognóstico e para que a paciente se beneficie das terapias disponíveis para tratamento. Por outro lado, em países de baixa e média renda, como por exemplo o Brasil, tem sido observado que os cânceres de mama são diagnosticados em estágios avançados, o que reduz o prognóstico e aumenta a morbidade relacionada ao tratamento, comprometendo a qualidade de vida dos pacientes. Nesses países, a implementação das ações de detecção precoce de forma adequada pode modificar em curto e médio prazos (de 5 a 10 anos) o estágio do câncer no momento do diagnóstico (INCA, 2015a).

No Brasil, as estratégias para o controle do câncer de mama vêm sendo realizadas desde a década de 1980, porém com ações isoladas. O primeiro passo para introdução formal de um programa de controle do câncer de mama no país foi em 2004, quando foi lançado o Documento de Consenso do Controle do Câncer de Mama, porém as recomendações desse documento tinham como foco indivíduos assintomáticos, não contemplando aqueles que tivessem sinais e sintomas suspeitos (INCA, 2004).

Em 2015, o programa foi reestruturado com diretrizes elaboradas a partir de evidências e recomendações de detecção precoce direcionadas para indivíduos assintomáticos e sintomáticos (INCA, 2015a). A mamografia de rastreamento é considerada um dos métodos indicados e mais efetivos na detecção precoce, garantia de diagnóstico e tratamento em tempo oportuno (WHO, 2014). Assim, o Ministério da Saúde recomenda a mamografia de rastreamento a cada dois anos, para mulheres assintomáticas, na faixa etária de 50 a 69 anos e que indivíduos sintomáticos tenham prioridade de acesso à confirmação diagnóstica e garantia da integralidade e continuidade do cuidado. Contudo, no Brasil, o rastreamento é considerado oportunístico, ou seja, o exame só é solicitado quando a mulher procura o serviço de saúde ou em datas específicas, como por exemplo, o “outubro rosa”. Ademais, não há acompanhamento para solicitação de exame subsequente ou busca ativa daquelas que faltaram ao exame agendado (INCA, 2015a).

Desde a introdução no país do programa de rastreamento em 2004, o que se tem observado é que a cobertura de mulheres na faixa etária alvo não atinge a meta de 70-75% considerada ideal pela Organização Mundial de Saúde (NOGUEIRA et al., 2019); que quase metade das mamografias são realizadas em mulheres com menos de 50 anos, com

periodicidade anual (AZEVEDO e SILVA et al., 2014; CORREA et al., 2017); que o sobrerastreamento por periodicidade teve uma taxa de 150 mil mulheres/ano em 2010 (RODRIGUES et al., 2019); e que aproximadamente 40% dos casos de câncer de mama são diagnosticados em estágios tardios e esse percentual tem se mantido constante (SANTOS-SILVA et al., 2019).

O rastreamento oportunístico é oneroso e produz poucos benefícios, e a não observância da periodicidade pode levar à superestimação da cobertura, gerando falsa impressão de população coberta (RODRIGUES et al., 2019). No Brasil, utiliza-se como *proxy* da cobertura de mamografia a razão expressa entre o número de exames realizados (extraídos do SISMAMA e, algumas vezes, do SIA), e a metade da população feminina no período, que pode fornecer valores superestimados, gerando a falsa sensação de população coberta (INCA, 2021). Para a cobertura é ideal que no numerador seja utilizado o número de mulheres rastreadas e não de exames. Além disso, o método utilizado, só estima a cobertura de mulheres dependentes do SUS, para estimar a verdadeira cobertura devem ser realizado estudos de base populacional, como a Pesquisa Nacional de Saúde (PNS).

Diferenças regionais e socioeconômicas podem influenciar a forma como os serviços de saúde são ofertados. Nas áreas de menor nível socioeconômico e alta vulnerabilidade na saúde, são observadas menor cobertura do rastreamento mamográfico, menor disponibilidade de equipamentos e recursos humanos e maiores distâncias percorridas pelas mulheres para realização de exames. Por outro lado, nas regiões mais desenvolvidas concentram-se grande parte dos serviços e recursos para o rastreamento do câncer de mama, o que resulta em maior cobertura para mulheres residentes nesses locais e vazio assistencial para aquelas que residem nas áreas desprovidas (TOMAZELLI e AZEVEDO e SILVA, 2017; RODRIGUES et al., 2019; NOGUEIRA, et al., 2019).

O objetivo dos programas de rastreamento do câncer de mama é evitar o óbito por câncer, reduzir a incidência de estágios avançados da doença e ofertar tratamento adequado em tempo oportuno (WENDER et al., 2019). Desse modo, para que as ações de rastreamento tenham sucesso e impacto sobre a mortalidade é preciso que as metas de cobertura da população alvo sejam realmente atingidas, que o aprazamento entre exames seja realizado conforme as recomendações e as desigualdades de acesso ao diagnóstico precoce e tratamento sejam eliminadas e/ou reduzidas.

Diante do descrito, verifica-se a importância de investigações que avaliem as estratégias para redução do câncer de mama e recomendações do Ministério da Saúde, por

exemplo, cobertura, idade para rastreamento, periodicidade e associação com fatores socioeconômicos. Logo, o exame minucioso do SISMAMA em Minas Gerais é oportuno para compreender tais associações e interações, pois o estado possui características geográficas, estruturais e econômicas muito parecidas com o Brasil (DOMINGUES; MAGALHÃES; FARIA, 2009). Assim, a tese foi estruturada em seis sessões: referencial teórico, justificativa, objetivos, métodos, resultados e discussão e considerações finais. Os resultados e discussões foram organizados na forma de três artigos.

No Artigo 1 foi testada e avaliada a acurácia de um método para seleção de escore em relacionamento probabilístico de banco de dados, de forma a viabilizar a automatização da identificação de pares verdadeiros dispensando a etapa de inspeção manual. A criação do método contribuiu com a automatização do relacionamento das bases de dados utilizadas no estudo, sendo constatada uma boa qualidade das medidas de avaliação dos relacionamentos realizados, o que demonstra que é possível utilizá-lo no âmbito dos serviços de saúde, especialmente para a vigilância e gestão em saúde.

No Artigo 2 foi estimada a cobertura do rastreamento mamográfico em Minas Gerais e avaliada a associação com fatores sociodemográficos e assistenciais. Por meio dessa avaliação foi possível identificar as regiões que necessitam de maior atenção e planejamento para que a sua cobertura seja ampliada e de que forma os fatores socioeconômicos estão associados às coberturas encontradas nas microrregiões de saúde.

No Artigo 3 foi estimado o sobrerastreamento mamográfico por faixa etária e periodicidade e analisou-se possíveis fatores sociodemográficos e assistenciais associados, sendo observado que as recomendações do Ministério da Saúde não têm sido adotadas pelas microrregiões de saúde de Minas Gerais e que o sobrerastreamento por periodicidade está associado às desigualdades regionais.

2 REFERENCIAL TEÓRICO

2.1 Magnitude do Câncer de Mama

O câncer de mama é uma neoplasia maligna decorrente da multiplicação desordenada e anormal de células da mama. Existem vários tipos de câncer de mama e diferentes formas de evolução da doença. Alguns tipos são mais agressivos e desenvolvem mais rápidos, outros crescem lentamente e outros podem ficar latentes durante toda a vida da mulher (BRASIL, 2013). Os tumores que formam nódulos, levam em média 10 anos desde a sua formação até a fase em que ele pode ser palpado no exame físico, a partir de 1 cm de diâmetro. É estimado que o tumor duplica de tamanho a cada período de 3-4 meses. Se não tratado, entre 3-4 anos da sua descoberta pela palpação, ocorre o óbito (BRASIL, 2006a). Os principais sinais e sintomas que o câncer de mama apresenta são: nódulos fixos, geralmente indolores, pele da mama avermelhada, alterações no mamilo, alterações na pele, como abaulamentos ou retrações com aspecto semelhante à casca de laranja, pequenos nódulos na axila ou pescoço e saída espontânea de líquido anormal pelos mamilos (SILVA; RIUL, 2011).

Os fatores de risco para o câncer de mama são: idade, história de menarca precoce, menopausa tardia, primeira gravidez após os 30 anos ou nuliparidade, menor tempo de aleitamento materno, uso de anticoncepcionais orais, terapia de reposição hormonal pós-menopausa, uso de bebida alcoólica, sobrepeso e obesidade pós-menopausa e exposição à radiação ionizante e mutações nos genes BRCA1 e BRCA2 (ANOTHAISINTAWEE et al., 2013; INUMARU; SILVEIRA; NAVES, 2011).

No mundo, o câncer de mama é a quinta causa de morte por câncer em geral e em 110 países é a causa mais frequente de óbitos de mulheres por câncer (SUNG et al., 2021) (figura 1).

Em 2020, com exceção do câncer de pele não melanoma, o câncer de mama representou 24,5% do total de casos de cânceres femininos, com 2,3 milhões de casos novos aproximadamente e uma taxa de incidência de 47,8/100 mil (GLOBOCAN, 2020).

Nos países com alto IDH, as taxas de incidência dessa neoplasia aumentaram durante as décadas de 1980 e 1990 devido à alta prevalência de fatores de risco reprodutivos e hormonais, estilo de vida e detecção por meio de mamografia, porém durante o início dos anos 2000 as taxas se estabilizaram ou reduziram, provavelmente devido a redução do uso

da terapia hormonal na menopausa e a um platô na participação nos programas de rastreamento. Por outro lado, tem-se observado que países com menor IDH estão experimentando o processo vivido anteriormente pelos países de alta renda. Com o crescimento da economia desses países, a proporção de mulheres no mercado de trabalho aumentou, o que influencia o estilo de vida e comportamento dessas mulheres e contribui com o aumento da prevalência dos fatores de risco de câncer de mama (SUNG et al., 2021).

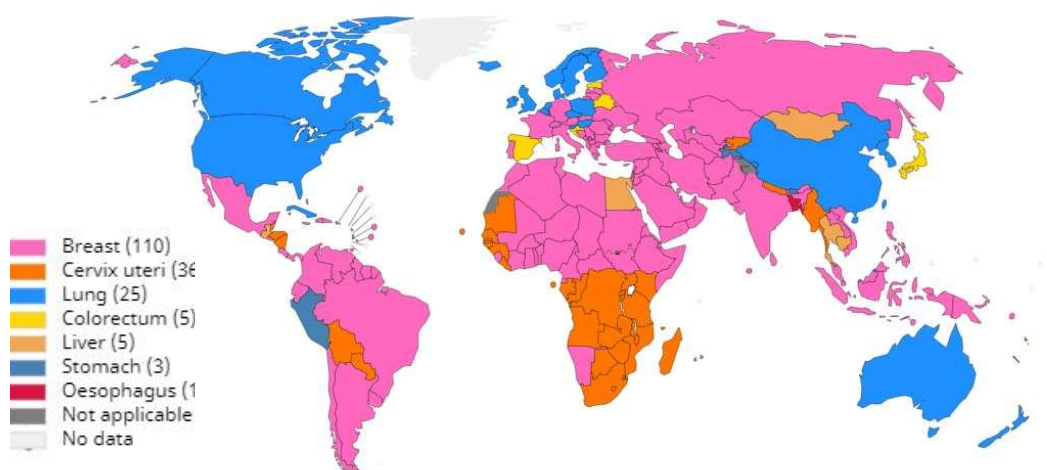


Figura 1 – Principais causas de mortalidade por câncer em 2020, sexo feminino, em todas as idades. Fonte: GLOBOCAN, 2020.

Países de alta renda tem maiores taxas de incidência (55,9/100.000) que países de média e baixa renda (29,7/100.000). Em contrapartida, países de média e baixa renda apresentam as maiores taxas de mortalidade e os de alta renda as menores (15,0 e 12,8/100.000, respectivamente) (SUNG et al., 2021). Ao analisar as taxas de incidência e mortalidade nas regiões do mundo (figura 2), percebe-se alta incidência nos países da América do Norte e baixa incidência nos países da Ásia. A mortalidade é maior nos países da África e menor nos países da Ásia. Destaca-se a mudança ocorrida desde 2018, quando países da Oceania apresentavam alta incidência (GLOBOCAN, 2018; GLOBOCAN, 2020).

Desde os anos 1990 as taxas de mortalidade por câncer de mama têm decrescido nos países de alta renda devido a melhorias na detecção precoce por meio do rastreamento populacional e intervenções terapêuticas mais eficazes (TORRE et al., 2017). Embora a incidência do câncer de mama aumente com níveis mais altos de renda, a carga de mortalidade é maior para as mulheres que residem nos países mais pobres e está relacionada principalmente à infraestrutura de saúde deficiente, falta de conscientização da população,

atraso no diagnóstico e baixos níveis de educação (BELLANGER et al., 2018; BRINTON et al., 2017).

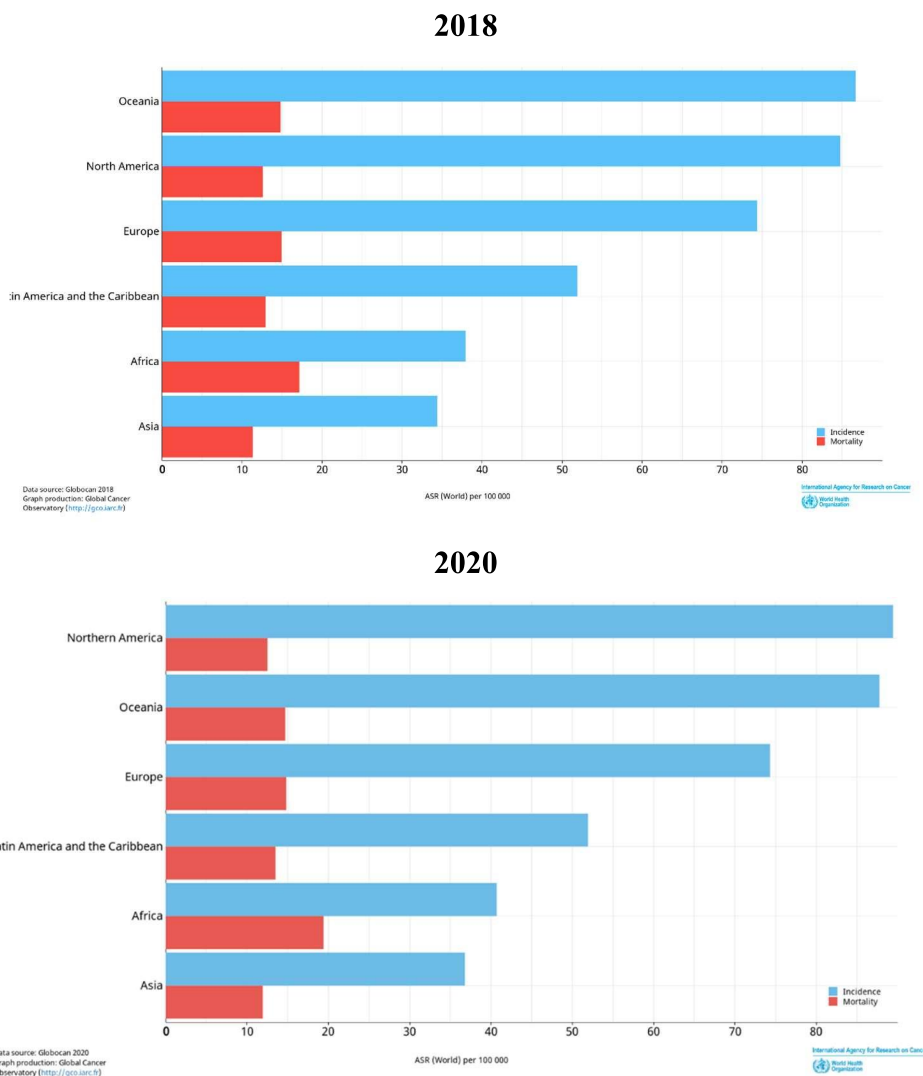


Figura 2 – Taxas de incidência e mortalidade do câncer de mama padronizadas por idade, nos continentes, em todas as idades, em 2018 e em 2020.

Fonte: GLOBOCAN, 2018 e 2020.

Canadá e Estados Unidos, países da América do Norte, apresentaram em 2018 incidência de 83,8 e 84,9/100.000 mulheres, respectivamente, com mortalidade menor do que 12,8/100.000 mulheres nos dois países (GLOBOCAN, 2018). Em 2020, a incidência aumentou nos Estados Unidos e teve uma redução discreta no Canadá (90,3 e 82,1/100.000 respectivamente), enquanto a mortalidade decresceu um pouco nos Estados Unidos (12,4/100.000), mas não no Canadá (13,3/100.000) (GLOBOCAN, 2020).

As taxas de incidência na Austrália e Nova Zelândia, países da Oceania, eram menores que 94,2/100.000 em 2018 (BRAY et al., 2018) e em 2020 apresentam taxa 96,0 e 93,0/100.000, respectivamente. A mortalidade entre 2018 e 2020 nos dois países foi de 12,3 e 14,2/100.000 para 11,7 e 14,1/100.000, respectivamente (GLOBOCAN, 2018; GLOBOCAN, 2020).

Na Europa, a Bélgica, França, Suíça e Reino Unido continuam entre os países que apresentam maior incidência (>87,0/100.000) e Ucrânia e República da Moldávia as menores (<45,0/100.000). República da Moldávia, Hungria e Polônia apresentam as maiores taxas de mortalidade (>17,5/100.000) e Malta, Finlândia e Bielo-Rússia as menores (<13,0/100.000) (GLOBOCAN, 2020).

Na África, a incidência é maior em Maurício (66,2/100.000) e menor em Camarões (43,6/100.000), enquanto a Costa do Marfim apresenta a maior taxa de mortalidade (>25,2/100.000) e França, La Réunion a menor (12,6/100.000) (GLOBOCAN, 2020).

Israel e Cingapura são os países com maior incidência na Ásia (77,3 e 77,9/100.000, respectivamente). República Árabe da Síria e Geórgia tem as maiores taxas de mortalidade (26,2 e 23,5/100.000, respectivamente), e República da Coreia e Japão, as menores (6,4 e 9,9/100.000 respectivamente) (GLOBOCAN, 2020).

Localizado na América Latina, o Brasil apresenta incidência de 61,9/100.000 mulheres, inferior aos países da América do Norte e Europa e superior à maioria dos países da África. De 2002 a 2012, no país, a taxa de mortalidade padronizada por idade utilizando a população mundial foi de 11,5/100.000 (CARIOLI et al., 2018), passando para 13,0/100.000 em 2018 (GLOBOCAN, 2018) e 13,8/100.000 em 2020 (GLOBOCAN, 2020), demonstrando pequenas elevações ao longo dos anos.

De 1990 a 2017, todos os estados brasileiros apresentaram aumento na taxa de incidência do câncer de mama, sendo o aumento maior observado em alguns estados das regiões menos desenvolvidas, Norte e Nordeste. Na comparação entre regiões, Sudeste e Sul, regiões mais desenvolvidas, apresentaram as maiores taxas de incidências, Centro-Oeste taxas intermediárias e Norte e Nordeste as menores taxas. Quanto à mortalidade, a maioria dos estados brasileiros apresentaram taxas estáveis, alguns tiveram aumento da mortalidade, principalmente na região Nordeste. Nas regiões Sul, Sudeste e Centro-Oeste, quase metade dos estados tiveram redução nas taxas de mortalidade (GUERRA et al., 2020).

Em Minas Gerais, região Sudeste, de 2008 a 2012 e de 2013 a 2017, ocorreram no estado 6.211 e 7.571 óbitos por câncer de mama, respectivamente. A taxa de mortalidade

padronizada foi de 10,3 e 11,8/100.000, respectivamente. As microrregiões de saúde que apresentaram as maiores taxas de mortalidade padronizadas no primeiro período, em ordem decrescente foram: Além Paraíba (17,0); Belo Horizonte/Nova Lima/Caeté (13,6); Juiz de Fora/Lima Duarte/Bom Jardim (13,4). No segundo período foram: Curvelo (16,8), Coronel Fabriciano/Timóteo (15,5) e Belo Horizonte/Nova Lima/Caeté (15,1). As menores taxas no primeiro período foram encontradas, em ordem crescente, nas microrregiões de Brasília de Minas/São Francisco (2,9); Pedra Azul (3,3) e Padre Paraíso (3,3). E no segundo período nas microrregiões Coração de Jesus (1,8), Itaobim (4,0) e Pedra Azul (4,5) (DUARTE, 2016; DUARTE et al., 2020).

Diferenças nas taxas de incidência e mortalidade por câncer de mama nos países e suas regiões, podem ser explicadas não só por fatores biológicos, comportamentais e sociais, mas também pelas estratégias governamentais adotadas para prevenção, detecção e diagnóstico precoces. Estudo sobre o papel da política de saúde na carga do câncer de mama no Brasil, no período de 2004 a 2014, demonstrou que o país investe principalmente em tratamento, que é mais oneroso e pouco em ações de promoção da saúde, prevenção e detecção precoce de doenças, que apresentam menor custo (FIGUEIREDO et al., 2017).

A redução da mortalidade por câncer de mama depende de diminuição na incidência e aumento no tempo de sobrevivência, que, por sua vez, depende de diagnóstico precoce e tratamento oportuno e adequado. A Organização Mundial de Saúde recomenda que os países adotem o rastreamento mamográfico como meio para detectar o câncer em fase inicial e reduzir a mortalidade por esta causa, uma vez que a doença é considerada de baixa letalidade (número de mortes menor que um terço da incidência) (INCA, 2019a).

2.2 Programas de Rastreamento do Câncer de Mama no Mundo

Rastreamento, em inglês *screening*, consiste num conjunto de exames e/ou testes aplicados em indivíduos assintomáticos a fim de identificar precocemente sinais de uma doença em desenvolvimento, ou seja, detectar uma doença antes que apareçam os sintomas, em estágio inicial. Durante o rastreamento as pessoas que apresentam alguma alteração são classificadas como suspeitas e submetidas a exames complementares para diagnóstico final, e caso haja confirmação, o tratamento é iniciado (BRASIL, 2010).

Os critérios para implantação de um programa de rastreamento são: a doença deve representar um importante problema de saúde pública (magnitude, transcendência e

vulnerabilidade); deve ter história natural ou problema clínico bem conhecido e estágio pré-clínico que permita seu diagnóstico; os benefícios da detecção e tratamento precoce devem ser maiores que a descoberta e tratamento em momento habitual; e os exames para detecção das condições devem estar disponíveis a todos indivíduos da faixa etária alvo, devem ser aceitáveis e confiáveis (BRASIL, 2010).

Existem dois tipos de programas de rastreamento: oportunístico e organizado/populacional. O primeiro ocorre quando a pessoa solicita o exame ou procura o serviço de saúde por qualquer motivo e nesse momento o profissional de saúde aproveita para solicitar exames, porém não existe um acompanhamento sistemático e nem diretrizes bem definidas, o que torna esse tipo de programa mais oneroso para o sistema de saúde e menos efetivo na redução da morbidade e mortalidade (INCA, 2015a).

Por outro lado, o rastreamento organizado consiste no convite ou convocação de uma população alvo, com periodicidade pré-determinada. Trata-se de um programa governamental, estruturado, sistematizado, definido por meio de diretrizes, que convida e monitora a pessoa desde o momento em que ela se torna elegível para rastreio até o fim do período de duração do mesmo. Além disso, oferece tratamento e acompanhamento das pessoas que apresentam alterações na rastreamento e tem sua rede de referências e fluxos bem organizadas e monitoradas (SANKARANARAYANAN, 2014).

Para o rastreamento do câncer de mama, na maioria dos países, a mamografia é considerada como primeira escolha e como um dos métodos mais efetivos na detecção precoce, garantia de diagnóstico e tratamento em tempo oportuno. Entretanto, devido à falta de recursos e problemas estruturais, nem todos os países de baixa a média renda conseguem ofertar rastreamento mamográfico populacional, optando em alguns casos pelo rastreamento oportunístico (WU; LEE, 2019). Na Índia, por exemplo, a dificuldade para implantação do rastreamento mamográfico organizado está relacionada à falta de mão de obra treinada, infraestrutura, logística, garantia de qualidade, frequência de rastreamento e custos envolvidos (MITTRA et al., 2009).

Para ser considerado efetivo um programa de rastreamento deve ser capaz de ofertar condições de diagnóstico e tratamento para população susceptível, dessa maneira, devem apresentar quatro importantes características: (1) acesso: refere-se a garantia de disponibilização do serviço à todas as pessoas da faixa etária alvo, sem exclusão em virtude da localização geográfica, raça/cor ou status socioeconômico; (2) agilidade: significa que os serviços de assistência à saúde devem eliminar todas as barreiras burocráticas que retardam

o atendimento das necessidades do usuário, visando diagnóstico e tratamento no momento apropriado; (3) melhores evidências: os programas devem ofertar procedimentos cujos benefícios superam os danos e a população deve ser esclarecida com a finalidade de adesão voluntária; (4) informações: o usuário deve ser informado sobre todos os riscos, benefícios, formas de acesso, rotinas e deve receber todas orientações que julgar pertinentes para sua participação no programa de rastreamento (BRASIL, 2010).

Inicialmente, a implementação de um programa de rastreamento mamográfico pode resultar em um aumento da incidência do câncer de mama, e este fato é esperado devido à detecção precoce de tumores assintomáticos, porém, espera-se à longo prazo declínio da incidência de tumores em estágio avançado, com consequente impacto sobre a mortalidade. Na Alemanha, por exemplo, após a introdução do programa de rastreamento para mulheres na faixa etária de 50 a 69 anos, foi constatado que a incidência de tumores em estágio avançados foram diminuindo nos rastreamentos subsequentes e nos dois primeiros anos a redução foi predominantemente em mulheres de 60 a 69 anos, indicando que o rastreamento mamográfico reduz as taxas de câncer de mama em estágio avançado na população (WEIGEL et al., 2016; SIMBRICH et al., 2016).

Os programas de rastreamento mamográfico, implantados em diversos países ou adotados na prática por alguns profissionais, divergem quanto ao tipo de ação, idade de início e periodicidade entre exames segundo as recomendações propostas pelas instituições que se dedicam ao assunto. As recomendações são elaboradas por órgãos governamentais como a Organização Mundial de Saúde - OMS, Serviço Nacional de Saúde do Reino Unido e Ministério da Saúde do Brasil e por agências independentes como a Força Tarefa de Saúde Preventiva dos Estados Unidos - USPSTF e Força-Tarefa Canadense para Assistência à Saúde Preventiva – CTFPHC, que usam como base para as orientações uma rigorosa revisão de evidências; e por associações médicas, dos EUA ou Brasil, cujas diretrizes são baseadas em revisões com menor rigor metodológico ou consenso de uma comissão de especialistas (MIGOWSKI et al., 2018c; URBAN et al., 2012; VIEIRA, 2017). Dessa maneira, existem variações entre as recomendações, principalmente com relação à faixa etária de início do rastreamento, a saber:

- Organização Mundial de Saúde (OMS): Em programas de rastreamento organizado recomenda a mamografia para mulheres de 50 a 69 anos, com periodicidade bianual (WHO, 2007);

- Serviço Nacional de Saúde do Reino Unido (National Health Service, United Kingdom - NHS England): Recomenda rastreamento mamográfico entre os 50 a 70 anos a cada 3 anos (MARMOT et al, 2013);
- Ministério da Saúde/Instituto Nacional do Câncer (INCA): Recomenda a mamografia na faixa etária de 50 a 69 anos com periodicidade bianual (INCA, 2015a);
- Força Tarefa de Saúde Preventiva dos Estados Unidos (United States Preventive Service Task Force – USPSTF): Recomenda a mamografia para mulheres na faixa etária de 50-74 anos, a cada dois anos. De 40 a 49 anos a decisão deve ser individual com potencial maior de benefício do que dano, podendo optar pelo rastreamento bianual. Para maiores de 75 anos considera as evidências insuficientes para avaliar o balanço entre benefício e danos (SIU, 2016);
- Força-Tarefa Canadense para Assistência à Saúde Preventiva (Canadian Task Force on Preventive Health Care – CTFPHC): Sugere mamografia para mulheres com idade entre 50 e 74 anos, a cada 2 a 3 anos (CTFPHC, 2017);
- Sociedade Americana do Câncer (American Cancer Society): O rastreamento deve ter início aos 45 anos e periodicidade anual até os 54 anos. A partir dos 55 anos, periodicidade bienal, devendo continuar até o momento em que as mulheres tenham estado de saúde geral bom e uma expectativa de vida de 10 anos ou mais (OEFFINGER et al., 2015);
- Colégio Americano de Ginecologia e Obstetrícia (American College of Obstetricians and Gynecologists - ACOG): Recomenda que na faixa etária de 40 a 50 anos, a decisão sobre a idade para começar o rastreamento mamográfico deve ser definida através de um processo de tomada de decisão compartilhado. A periodicidade deve ser a cada 1 ou 2 anos com base em um processo de tomada de decisão informado e compartilhado e deve incorporar as preferências da paciente. A mamografia bianual, particularmente após os 55 anos de idade, é uma opção razoável para reduzir a frequência de danos. O rastreamento deve continuar até, pelo menos, 75 anos de idade (ACOG, 2017);
- Colégio Americano de Radiologia (American College of Radiology – ACR): Recomenda rasteio anual a partir dos 40 anos para mulheres com risco médio (MONTICCILO et al., 2017);

- Colégio Brasileiro de Radiologia e Diagnóstico por Imagem, Sociedade Brasileira de Mastologia e Federação Brasileira das Associações de Ginecologia e Obstetrícia: Recomendam mamografia para mulheres na faixa etária de 40 a 69 anos, com periodicidade anual. Para mulheres com 70 anos ou mais, que tenham expectativa de vida maior que 7 anos, com base nas comorbidades ou que tenham condições de serem submetidas a investigação diagnóstica invasiva e tratamento após um resultado anormal do rastreamento, a recomendação também é o rastreamento anual (URBAN et al., 2012).

Introduzir o rastreamento organizado em locais com infraestrutura de saúde pública deficiente é um grande desafio. Nos locais onde os sistemas de saúde não são predominantemente públicos e cuidados de saúde estão amplamente disponíveis, a substituição ou redução do rastreamento oportunista é difícil (MILES et al., 2004).

Grupos mais favorecidos têm maior probabilidade de participar no rastreamento organizado e oportunista. Isso é parcialmente explicado pelo fato de que indivíduos mais favorecidos estão mais preocupados com sua saúde e quando decidem realizar uma avaliação da sua saúde, optam pelo serviço que no momento oferecem acesso mais fácil (CHOI et al., 2018).

Com base em evidências científicas e nas recomendações nacionais e internacionais cada país, de acordo com suas condições econômicas e estruturais, adota um ou vários tipos de rastreamento, conforme sintetizado no quadro 1.

A faixa etária alvo de 50 a 69 anos para rastreamento mamográfico é adotada por países da América do Norte, América do Sul e Europa. Idade inferior a 50 anos e superior a 70 anos é adotada por países de todos os continentes, exceto Oceania para menos de 50 anos e África para superior a 69 anos. Quanto à periodicidade, o rastreamento bienal é praticado na maioria dos países, exceto no Reino Unido que recomenda o exame a cada três anos, a Suécia a cada 18 meses, a Romênia anualmente e a Croácia semestralmente.

O que se observa é que a maioria dos países seguem as recomendações da Organização Mundial de Saúde, dentre eles o Ministério da Saúde no Brasil, preconizando a faixa etária de 50 a 69 anos e periodicidade bienal.

Estudos demonstram que o rastreamento mamográfico reduz a mortalidade de mulheres de 50 a 69 anos, o que demonstra a importância do rastreamento para essa faixa etária (SARKEALA et al., 2014; SARKEALA; HEINÄVAARA; ANTTILA, 2008; AUTIER et al., 2017).

Quadro 1 – Programas de Rastreamento do Câncer de Mama nos países do mundo.

CONTINENTE	PAÍS	DATA DE INÍCIO	EXAME	FAIXA ETÁRIA	PERIODICIDADE	TIPO	REFERÊNCIA
África	África do Sul	–	MMG	A partir dos 40 a 54 anos > 55 anos	Anual Bienal**	Oportunista	LIPSCHITZ, 2018
	Marrocos	2010	ECM, se alterado MMG	40 a 69 anos	Bienal**	Oportunista	MARROCOS, 2011
América Central	Costa Rica	2010	MMG	45 a 69 anos	Bienal**	Oportunista	ANDERS, 2017
	Cuba		MMG	50 a 64 anos	Bienal**	Oportunista	GONZÁLEZ, 2014
	Nicarágua		MMG	A partir 40 anos	Anualmente	Oportunista	NICARÁGUA, 2014
	Panamá		MMG	40 a 74 anos	Bienal**	Oportunista	PANAMÁ, 2015
América do Norte	Canadá	–	MMG	50 a 69 anos*	Bienal**	Organizado**	VAHABI et al., 2015
	Estados Unidos	–	MMG	50 a 74 anos	Bienal**	Oportunista	SIU, 2016
	México	–	MMG	40 a 69 anos	Bienal**	Oportunista	MÉXICO, 2011
América do Sul	Argentina	–	MMG	50 a 69 anos*	Bienal**	Oportunista	DI SIBIO, 2018
	Bolívia	–	MMG	A partir 40 anos	Anualmente	Oportunista	BOLÍVIA, 2013
	Brasil	1990	MMG	50 a 69 anos*	Bienal**	Oportunista	INCA, 2015a
	Chile	2011	MMG	50 a 64 anos	Trienal	Oportunista	OECD, 2018
	Colômbia	–	MMG	50 a 69 anos*	Bienal**	Oportunista	MURILLO et al., 2015
	Equador	–	MMG	50 a 74 anos	Bienal**	Oportunista	EQUADOR, 2019
	Paraguay	–	MMG	1 mamografia aos 35 anos e depois a partir dos 40 anos	Anualmente	Oportunista	PARAGUAY, 2016
	Peru	2017	MMG	50 a 69 anos*	Trienal	Oportunista	PERU, 2017
	Uruguai		MMG	50 a 69 anos*	Bienal**	Oportunista	URUGUAY, 2015
	Ásia	China	2009	ECM, se alterado MMG	35 e 64 anos	Trienal	Oportunista

continua...

							continuação...
Europa	Coreia	2005	MMG	40 anos	Bienal**	Organizado**	LEE; LEE, 2018
	Israel	–	MMG	50 a 74 anos	Bienal**	Oportunista	OECD, 2018
	Japão	2000	MMG	40 anos ou mais	Bienal**	Oportunista	KASAHARA et al., 2012; OECD, 2018
	Líbano	–	MMG	40 anos e enquanto a mulheres estiver em boa saúde.	Anual	Oportunista	ADIB; SAGHIR; AMMAR, 2009
	Malásia	–	MMG	50 a 74 anos	Bienal**	Oportunista	MAHMUD; ALJUNID, 2018
	Tailândia	–	MMG	40 anos	Anual	Oportunista	SRIPAIBOONKIJ et al., 2016; MUKEM et al., 2014
	Turquia	2009	MMG	40 a 69 anos	Bienal**	Organizado**	ÖZMEN et al., 2017
	Alemanha	2005	MMG	50 a 69 anos*	Bienal**	Organizado**	BIESHEUVEL; WEIGEL; HEINDEL, 2011
	Austria	1974	MMG	entre 35 e 49 anos 50 anos ou mais	Inicial Bienal**	Organizado**	VUTUC; WALDHOER; HAIDINGER, 2006
	Bélgica	–	MMG	50 a 69 anos*	Bienal**	Organizado**	VERLEYE et al., 2011
	Chipre	2006	MMG	50 a 69 anos*	Bienal**	Organizado**	HADJISAWAS et al., 2010
	Croácia	2006	MMG	50 a 69 anos*	Semestrais	Organizado**	KELAVA et al., 2012
	Dinamarca	2007-2010	MMG	50 a 69 anos*	Bienal**	Organizado**	BEAU et al., 2017
	Eslovênia	2008	MMG	50 69 anos	Bienal**	Organizado**	IOL, 2019
	Espanha	1990	MMG	45 a 69 ou 50 a 69 anos* dependendo da região	Bienal**	Organizado**	ASCUNCE et al., 2010
							continua...

							continuação...
Estônia	2002	MMG	50 a 65 anos	Bienal**	Organizado** (para mulheres com seguro saúde)	SUONIO; DEAN, 2012	
França	2004	MMG	50 a 74 anos	Bienal**	Organizado**	OECD, 2018	
Georgia	1990	MMG	50 a 74 anos	Bienal**	Oportunístico	ADAMS et al., 2015; MCNAMARA et al., 2016	
Grécia	–	MMG	40 anos ou mais	Anual ou Bienal**	Oportunista	OECD, 2018	
Holanda	1989	MMG	50 a 75 anos	Bienal**	Organizado**	AUTIER et al., 2017	
Hungria	2002	MMG	45 a 65 anos	Bienal**	Organizado**	BONCZ et al., 2018	
Irlanda	2000	MMG	50 a 69 anos*	Bienal**	Organizado**	MHCI, 2016	
Itália	1990	MMG	50 a 69 anos*	Bienal**	Organizado**	ALTOBELLI; LATTANZI, 2014	
Lituânia	2005	MMG	50 a 69 anos*	Bienal**	Organizado**	STEPONAVICIENE et al., 2019	
Luxemburgo	1992	MMG	50 a 69 anos*	Bienal**	Organizado**	SHANNOUN et al., 2008	
Noruega	–	MMG	50 a 69 anos*	Bienal**	Organizado**	VAN LUIJT; HEIJNSDIJK; KONING, 2017	
Polônia	–	MMG	50 a 69 anos*	Bienal**	Organizado**	AUGUSTYNOWICZ et al., 2018; OECD, 2018	
Portugal	1990	MMG	45 a 69 anos	Bienal**	Organizado**	ALTOBELLI; LATTANZI, 2014	
Reino Unido	1990	MMG	50 a 70 anos	Trienal	Organizado**	ALTOBELLI; LATTANZI, 2014	
Republica Tcheca	2002	MMG	45 a 69 anos	Bienal**	Oportunista	MAJEK et al., 2011	
							continua...

							...continuação...
	Romênia	–	MMG	a partir dos 40 anos 45 a 50 anos após 50 anos	Trienal Semestralmente Anual	Oportunista	ALTOBELLI; LATTANZI, 2014
	Rússia	1998	MMG	40 a 60 anos	Bienal**	Organizado**	ZAKHAROVA et al., 2011
	Suécia	–	MMG	40 a 49 anos 50 a 69 anos* 70 a 74 anos	a cada 18 meses Bienal**	Organizado**	ALTOBELLI; LATTANZI, 2014
	Suíça		MMG	50 a 70 ou 74 anos	Bienal**	Organizado**	SUIÇA, 2014
Oceania	Austrália	1989	MMG	50 a 74 anos	Bienal**	Organizado**	BECKMANN et al., 2015
	Nova Zelândia	–	MMG	50 a 64 anos	Bienal**	Oportunista	OECD, 2018

*Faixa etária de 50 a 69 anos; **Periodicidade Bienal; ***Programas Organizados

Fonte: Elaborado pela autora, 2019.

Quanto à periodicidade, uma meta-análise realizada pelo Painel Independente do Reino Unido sobre Rastreamento do Câncer de Mama concluiu que o intervalo de 3 anos entre exames reduz 20% o risco relativo de morte comparado às mulheres que não realizam o rastreamento (MARMOT et al., 2013).

O Chile tem uma das menores taxas de incidência do câncer de mama, dessa maneira, devido à limitação de recurso, estudo sobre mortalidade e custo-efetividade, concluiu que intervalos entre rastreamento de dois anos são menos eficazes no país, recomendando duas políticas para que a mulher realize dez mamografias ao longo de sua vida: (1) Idade de início aos 40 anos com intervalos de quatro anos; (2) Idade de início aos 48 anos com intervalos de três anos (KONG; MONDSCHHEIN; PEREIRA, 2018).

As principais desvantagens do rastreamento do câncer de mama são: sobrediagnóstico, alto custo, exposição à radiação ionizante, recomendação de biópsia falso-positiva, resultados falsos negativos e suas consequências (SHAH; GURAIÁ, 2017)

Por sobrediagnóstico, entende-se o diagnóstico de casos de câncer de mama que nunca se manifestariam clinicamente caso a mulher assintomática nunca tivesse realizado a mamografia. Destaca-se que não se trata de um diagnóstico incorreto, pois nesse caso, o câncer de mama sobrediagnosticado nunca iria progredir ou poderia entrar em remissão; ou progredir lentamente, assim a mulher iria a óbito por outras causas antes mesmo de apresentar sinais/sintomas da neoplasia mamária (OLVER; ROLDER, 2017).

As consequências do sobrediagnóstico estão relacionadas tanto aos aspectos psicológicos devido ao medo da doença e ansiedade com relação ao futuro; quanto ao sobretratamento, isto é, instituição de terapias desnecessárias que não produzem nenhum benefício para mulher e que podem ainda causar danos à sua saúde em função dos riscos inerentes à cada procedimento adotado (INCA, 2015a; MIGOWSKI et al., 2018c).

Revisão sistemática de estudos sobre os benefícios e malefícios do rastreamento do câncer de mama verificou que o sobrediagnóstico na mamografia variou de 1 a 60% nos estudos em geral e de 1 a 12% nos estudos com baixo risco de viés e os resultados falso-positivos foram maiores nas triagens anuais comparados às bienais (0,01 a 41% vs 3 a 17%) (MANDRIK et al., 2019).

Revisão de ensaios randomizados sobre o efeito do rastreamento mamográfico na mortalidade e morbidade concluiu que se o rastreamento reduzir 15% da mortalidade por câncer de mama e se o percentual de sobrerastreamento e sobretratamento for de 30%, ao longo de 10 anos seria observado que para cada 2000 mulheres convidadas para o

rastreamento, uma evitaria o câncer de mama e 10 mulheres saudáveis, que não seriam diagnosticadas se não houvesse rastreamento, seriam tratadas desnecessariamente (GÖTZSCHE; JORGENSEN, 2013).

Na Suíça, no ano de 2013, o conselho médico concluiu que novos programas de rastreamento mamográfico não devem ser introduzidos e os programas existentes devem ser avaliados quanto à qualidade. O conselho justifica essa recomendação com base nas conclusões do estudo do Canadian National Breast Screening que demonstrou que as mamografias não reduziram as mortes por câncer de mama em mulheres de 40 a 59 anos e que um em cada cinco cânceres detectados na mamografia não representariam ameaças à saúde das pacientes (BILLER-ADORNO; JUNI, 2014).

O Canadian National Breast Screening Study foi um ensaio controlado randomizado de mamografia de rastreamento e exame físico iniciado em 1980 no Canadá. Fizeram parte do estudo 89.835 mulheres com idade entre 40 e 59 anos. Essas mulheres foram seguidas até o ano de 2005, média de 22 anos, sendo constatado aproximadamente 22% de sobrediagnóstico resultante de mamografia (106 cânceres invasivos detectados em excesso a cada 484 mulheres rastreadas, levando a tratamentos e intervenções cirúrgicas desnecessárias). O estudo não encontrou reduções na mortalidade por câncer de mama a partir de um programa que oferece rastreamento mamográfico anualmente, contudo, os autores ressaltam que os resultados do estudo podem não ser generalizáveis para todos os países porque em alguns a incidência de tumores maiores pode ser maior que em outro, assim o rastreamento pode ser mais benéfico para aqueles que tem alta proporção de nódulos positivos (MILLER et al., 2014).

Ademais, não é possível prever individualmente o comportamento biológico de um tumor. Dessa maneira, o ideal é tratar o câncer de mama em estágio inicial, uma vez que não é possível determinar quais permanecerão clinicamente irrelevantes ou se tornarão invasivos. A única forma de comprovar o sobrediagnóstico seria identificar a mulher diagnosticada com câncer de mama que recusou tratamento e que viveu grande período de tempo após o diagnóstico, sem apresentar sinais/sintomas e morreu por outras causas, o que praticamente não acontece (GHOSH; WAHNER-ROEDLER; BRANDT, 2015).

Em síntese, o objetivo dos programas de rastreamento do câncer de mama é evitar o óbito por câncer, reduzir a incidência em estágios avançados da doença e ofertar tratamento adequado em tempo oportuno. A efetividade de um programa de rastreamento organizado pode ser medida de acordo com a diminuição da incidência de casos avançados e com o grau

de redução na mortalidade específica por câncer de mama. Além disso, a implantação de um rastreamento organizado propicia igualdade de acesso, garantia de qualidade e tem maior potencial de alcançar a cobertura populacional (WENDER et al., 2019).

2.3 Controle do Câncer de Mama no Brasil

O controle do câncer de mama no Brasil tem um histórico marcado pelas transformações sociais, econômicas e de saúde que ocorreram no país e pelo surgimento de novas tecnologias diagnósticas e terapêuticas. Até 1940 o câncer de mama era uma doença pouco conhecida, comumente visto em mulheres idosas das classes mais ricas cujo tratamento era apenas cirúrgico e não era considerado um problema de saúde pública (TEIXEIRA e NETO, 2020a).

Com a criação do Serviço Nacional de Câncer na década de 1940, surgiram as primeiras campanhas educativas que alertavam para a importância do diagnóstico precoce como a única forma de possibilitar um tratamento mais efetivo, porém o acesso à seguridade social e saúde pública abrangia determinadas categorias de trabalhadores, o que deixava a grande maioria das mulheres que precisavam de tratamento sem cobertura. O sucesso das campanhas de prevenção do câncer de colo do útero associadas ao desenvolvimento de novas tecnologias de imagens médicas contribuiu para ampliar as discussões acerca da introdução do rastreamento do câncer de mama no Brasil nas décadas de 1970 e 1980 (TEIXEIRA e NETO, 2020b). Assim, desde 1980 políticas públicas direcionadas ao controle do câncer de mama vêm sendo desenvolvidas no Brasil.

Em 1984, discussões acerca da atenção integral às mulheres culminaram na criação do Programa de Assistência Integral à Saúde da Mulher (PAISM). Paralelamente, surgiam no país discussões e reivindicações de mudanças e transformações do sistema de saúde, o Movimento de Reforma Sanitária, cujo arcabouço conceitual embasou a formulação do Sistema Único de Saúde (SUS). Assim, o PAISM incorporou como princípios e diretrizes as propostas de descentralização, hierarquização e regionalização dos serviços, bem como a integralidade e a equidade da atenção. O novo programa para a saúde da mulher incluía ações educativas, preventivas, de diagnóstico, tratamento e recuperação, englobando a assistência à mulher em clínica ginecológica, no pré-natal, parto e puerpério, no climatério, em planejamento familiar, câncer de colo de útero e de mama, além de outras necessidades

identificadas a partir do perfil populacional das mulheres, porém as ações de controle do câncer de mama ainda eram isoladas (TEIXEIRA e NETO, 2020a).

Por se tratar de um grave problema de saúde, a prevenção dessa neoplasia ganhou ainda mais notoriedade no cenário nacional, o que propiciou a criação em 1998 de um programa específico, o Programa Nacional de Combate ao Câncer de Colo Uterino e de Mama -Viva Mulher (AZEVEDO e SILVA et al., 2015). Diante da necessidade de padronização das recomendações de prevenção, detecção precoce, diagnóstico, tratamento e cuidados paliativos no câncer de mama, foi lançado, em 2004, o Documento de Consenso do Controle do Câncer de Mama, elaborado por técnicos de diferentes áreas do Ministério da Saúde, gestores, pesquisadores que atuavam na área de controle de câncer e representantes de sociedades científicas afins e de entidades de defesa dos direitos da mulher (INCA, 2004).

No decorrer dos anos, as ações foram ampliadas e melhoradas e, em 2006, foram contempladas no Pacto pela Saúde na dimensão Pacto Pela Vida, cujos objetivos e metas visavam cobertura de 60% de mamografia nas mulheres da população alvo (BRASIL, 2006b). Para monitorar as ações do rastreamento do câncer de mama no país foi implantado em 2009 o SISMAMA - Sistema de Informação do Câncer de Mama que vem sendo substituído desde 2013 por um sistema que permite uma gestão mais detalhada das ações de controle do câncer de mama, o SISCAN - Sistema de Controle do Câncer. Hoje o controle do câncer de mama é uma prioridade da agenda de saúde do país e integra o Plano de Ações Estratégicas para o Enfrentamento das Doenças Crônicas Não Transmissíveis (DCNT), lançado pelo Ministério da Saúde em 2011 e reformulado para vigência no período de 2021-2030, sendo alterada a meta de cobertura de mulheres da faixa etária alvo de 60% para 70% (BRASIL, 2020).

O controle do câncer de mama perpassa todos os níveis de atenção à saúde, mas é na Atenção Primária à Saúde (APS) que se torna possível um maior alcance das ações, devido a sua abordagem mais próxima do usuário, o que possibilita o efetivo exercício de suas funções, a saber: resolver os problemas de saúde da população, organizar fluxos e contrafluxos dos usuários dos serviços de saúde pelos diversos pontos de atenção à saúde e responsabilizar-se pelos usuários em qualquer ponto de atenção à saúde que estejam (INCA, 2015a).

Na linha de cuidados do câncer, a APS tem responsabilidade quanto as ações de promoção, prevenção, detecção precoce e cuidados paliativos, ou seja, em todos os níveis de prevenção da história natural da doença. As ações de promoção da saúde e prevenção de

doenças fazem parte da prevenção primária, cujo objetivo é evitar a ocorrência das doenças com estratégias voltadas para redução da exposição aos fatores de risco. De outro modo, a detecção precoce visa identificar a doença em estágios iniciais com a finalidade de obter melhor prognóstico, sendo classificada como prevenção secundária (SANTOS et al., 2019).

A detecção precoce do câncer de mama na atenção básica envolve duas estratégias: o rastreamento e o diagnóstico precoce. O rastreamento do câncer de mama é definido como a aplicação de teste ou exame numa população assintomática, aparentemente saudável, com o objetivo de identificar lesões sugestivas de câncer e, a partir daí, encaminhar as mulheres com resultados alterados para investigação diagnóstica e tratamento. O diagnóstico precoce é realizado nas mulheres que apresentam sinais ou sintomas da doença e tem por finalidade confirmar ou afastar a suspeita da doença (BRASIL, 2010).

Para que a APS tenha sucesso nas suas ações de controle do câncer de mama é preciso que mulheres e profissionais sejam informados e conscientizados acerca das diretrizes para controle, dos sinais e sintomas dessa neoplasia e sobre o acesso rápido e facilitado aos serviços de saúde para detecção precoce, pois a identificação do câncer de mama em estágios iniciais apresenta prognóstico favorável e chances de 100% de cura (INCA, 2015b). O fluxo das mulheres, de acordo com sua necessidade, pelos diferentes serviços de atenção à saúde partindo da atenção básica é ilustrado na figura 3.

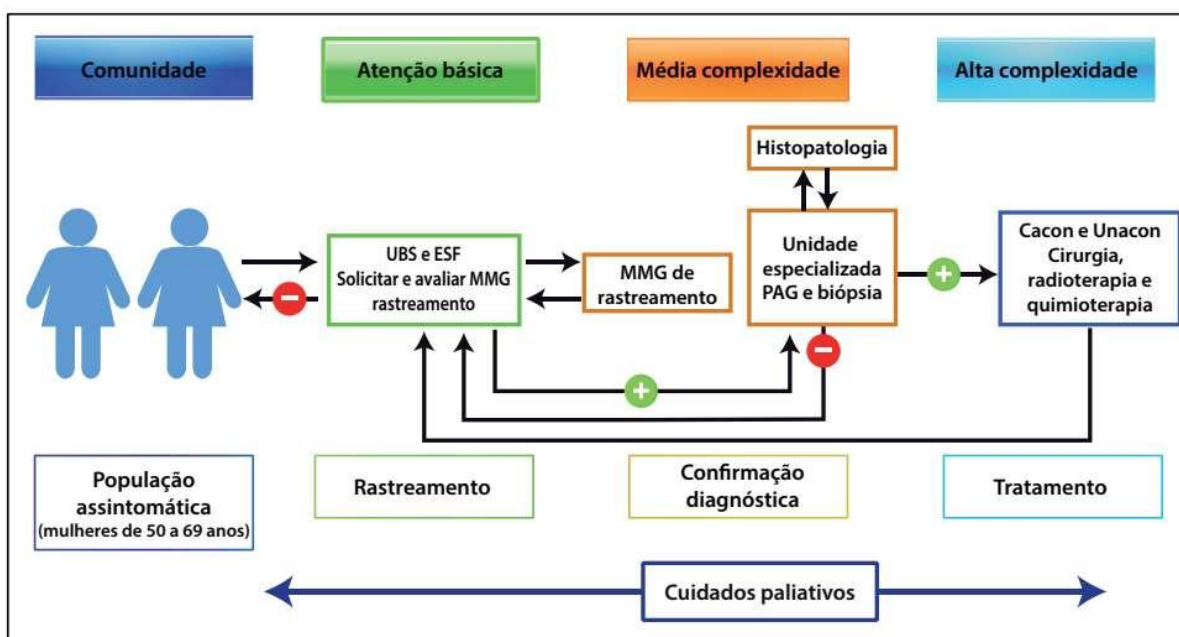


Figura 3: Principais pontos de atenção para o rastreamento do câncer de mama.
Fonte: INCA, 2015b Adaptado de Cancer Control: Knowledge into action.

Conforme a figura, mulheres assintomáticas na faixa etária de 50 a 69 são encaminhadas pela atenção básica para rastreamento por meio de mamografia. Caso o resultado apresente alteração, essa mulher é direcionada para média complexidade onde são realizados procedimentos diagnósticos e confirmatórios, e se comprovado, ela é referenciada para tratamento na alta complexidade. Esse mesmo fluxo é adotado para mulheres de qualquer faixa etária que procuram o serviço com sinais de alterações nas mamas ou que apresentam durante o exame clínico sinais sugestivos de câncer de mama. Durante todo processo de assistência nos diversos níveis de atenção à saúde, a mulher é contra referenciada, monitorada e acompanhada pela unidade de atenção básica de referência. Nas situações em que o resultado do rastreamento é favorável, a mulher é orientada pela atenção básica a seguir a rotina de rastreamento (INCA, 2015b).

Um dos pré-requisitos para que uma modalidade de rastreamento seja adotada é que existam evidências científicas sobre a sua eficácia. Desse modo, a mamografia é o método considerado como padrão-ouro para detecção precoce do câncer de mama. Nesse exame, é formada uma imagem da mama por meio de um feixe de raios X. A radiação ao atravessar as estruturas da mama produz atenuações diferenciadas dependendo da densidade e espessura do tecido, formando então, a imagem radiográfica (INCA, 2019b).

A mamografia é útil na identificação de tumores assintomáticos e impalpáveis o que contribui com a redução do número de diagnósticos de câncer de mama em estágios avançados quando utilizada para rastreamento ou diagnóstico precoce e consequentemente para diminuição dos óbitos devido essa neoplasia (INCA, 2015a). Estudo recente demonstrou que a redução dos anos de vida ajustados por incapacidade é mais um benefício do rastreamento com mamografia (MOLASSIOTIS et al., 2021).

A mamografia pode ser classificada como diagnóstica (para mulheres com sinais e sintomas de câncer de mama) ou de rastreamento (para mulheres assintomáticas). O resultado desse exame segue a classificação BI-RADS®, do American College of Radiology (ACR), com seis categorias para mamografias de rastreamento (0 a 5) e sete para as diagnósticas (0 a 6), gerando recomendações específicas que são apresentadas no quadro 2 (INCA, 2015b).

Quadro 2 – Condutas clínicas segundo resultado do exame mamográfico

Categoria BI-RADS®	Achados Mamográficos	Conduta	
0	Indefinido	Necessidade de avaliação adicional (outras incidências mamográficas, USG, etc.)	Realizar a ação necessária e classificar conforme categorias anteriores.
1	Negativo	Sem achados	Rotina do rastreamento
2	Benigno	Achados benignos	Rotina do rastreamento
3	Provavelmente Benigno	Achados provavelmente benignos	Controle radiológico por três anos (semestral no primeiro ano e anual nos segundo e terceiro anos). Confirmando estabilidade de lesão, volta à rotina. Eventualmente biópsia.
4	Suspeito (baixa, média e alta suspeição)	Achados suspeitos de malignidade	Biópsia e estudo histopatológico
5	Altamente Suspeito	Achados altamente suspeitos de malignidade	Biópsia e estudo histopatológico
6	Achado já com diagnóstico de câncer	Diagnóstico de câncer comprovado histologicamente	Seguir tratamento conforme o caso

Fonte: American College of Radiology (ACR), 2013.

No Brasil, a nova diretriz de 2015 foi construída a partir de evidências que discutiam sobre faixa etária, periodicidade, balanço entre riscos e benefícios de cada intervenção para detecção precoce do câncer de mama e que recomendam o rastreamento bianual por meio da mamografia para mulheres na faixa etária de 50 a 69 anos, sob a justificativa de que os benefícios para esse grupo são superiores aos riscos (INCA, 2015a).

No quadro 3 é apresentada uma síntese das recomendações de mamografia para cada faixa etária, conforme proposta do Ministério da Saúde.

A recomendação contra o rastreamento de mulheres com idade inferior a 50 anos é explicada pela maior densidade que as mamas jovens possuem, o que contribui para altas taxas de achados falso-positivos. Dessa maneira, a acurácia da mamografia diminui com o aumento da densidade mamária porque o tecido denso obscurece as alterações que um câncer pode apresentar. Com relação às mulheres com idade superior a 70 anos, poucos ensaios clínicos estudaram os benefícios do rastreamento para esse público. Quanto à periodicidade, as evidências demonstraram que o intervalo menor que o bienal não traz benefício adicional, pelo contrário, pode aumentar os danos associados ao rastreamento. Assim, a tomada de decisão em cada país, dentre eles o Brasil, tem se baseado na expectativa de vida das

mulheres e nos benefícios que o tratamento pode trazer (INCA 2015a, WENDER et al., 2019).

Quadro 3: Síntese das recomendações de mamografia por faixa etária.

	Questão norteadora	Qual a eficácia do rastreamento com mamografia na redução da mortalidade global e por câncer de mama, comparada à ausência de rastreamento?
RECOMENDAÇÃO	< de 50 anos	O Ministério da Saúde recomenda contra o rastreamento com mamografia em mulheres com menos de 50 anos (recomendação contrária forte: os possíveis danos claramente superam os possíveis benefícios).
	De 50 a 59 anos	O Ministério da Saúde recomenda o rastreamento com mamografia em mulheres com idade entre 50 e 59 anos (recomendação favorável fraca: os possíveis benefícios e danos provavelmente são semelhantes).
	De 60 a 69 anos	O Ministério da Saúde recomenda o rastreamento com mamografia em mulheres com idade entre 60 e 69 anos (recomendação favorável fraca: os possíveis benefícios provavelmente superam os possíveis danos).
	De 70 a 74 anos	O Ministério da Saúde recomenda contra o rastreamento com mamografia em mulheres com idade entre 70 e 74 (recomendação contrária fraca: o balanço entre possíveis danos e benefícios é incerto).
	75 anos ou mais	O Ministério da Saúde recomenda contra o rastreamento com mamografia em mulheres com 75 anos ou mais (recomendação contrária forte: os possíveis danos provavelmente superam os possíveis benefícios).
	Periodicidade: O Ministério da saúde recomenda que a periodicidade do rastreamento com mamografia nas faixas etárias recomendadas seja bienal (recomendação favorável forte: os possíveis benefícios provavelmente superam os possíveis danos quando comparada às periodicidades menores do que a bienal).	

Fonte: INCA, 2015a.

O papel de uma diretriz é orientar gestores, profissionais de saúde e usuários acerca da melhor prática a partir das melhores evidências científicas disponíveis (MIGOWSKI et al., 2018a). No país, foi observado que desde a implantação do programa de rastreamento em 2004, a prevalência de casos de câncer de mama diagnosticados em estadiamento avançado (III e IV) ainda é alta, aproximadamente 40% e apresenta desigualdades relacionadas aos fatores etnicoraciais e sociais. Além disso, esse mesmo estudo concluiu que em um ano poderiam ser evitados 7.500 óbitos caso o país tivesse uma alta cobertura de mamografia e se as mulheres diagnosticadas com estágios avançados da doença fossem diagnosticadas precocemente e as anormalidades acompanhadas e manejadas de maneira adequada (SANTOS-SILVA et al., 2019). Outros estudos, constataram que aproximadamente 50% das mamografias no país são realizadas em idades não preconizadas

e com periodicidade anual (TOMAZELLI et al., 2017; CORRÊA et al., 2017) e que o sobrerastreamento por periodicidade é de 21% (RODRIGUES et al., 2019). Todos esses aspectos demonstram que a política de controle executada no país desde 2004 não tem sido eficaz e que a diretriz de 2015 não tem sido seguida pelos gestores, profissionais de saúde e usuários.

Existem muitos desafios para que as recomendações trazidas pela nova diretriz de 2015 sejam implementadas e se consolidem nos serviços de saúde. Os principais desafios a serem superados são: natureza oportunística do rastreamento no país, dificuldade para ofertar todos os procedimentos de confirmação diagnóstica em um único serviço, baixa cobertura de mulheres da faixa etária alvo, principalmente nas regiões pouco desenvolvidas, como Norte e Nordeste e dificuldade dos profissionais em adotar as recomendações de faixa etária e periodicidade (MIGOWSKI et al., 2018c).

A existência de diretrizes clínicas brasileiras e internacionais que indicam o rastreamento antes dos 50 anos e a periodicidade anual, desconhecimento e a aceitação por parte dos profissionais de saúde das diretrizes e recomendações do Ministério da Saúde e orientações difundidas ou legitimadas por líderes de opinião são fatores que influenciam a forma de execução do rastreamento no país (SANTOS; RAMOS; MIGOWSKI, 2019; MIGOWSKI et al., 2018c).

Estudo realizado em um município de Minas Gerais sobre conhecimento, atitudes e práticas dos profissionais médicos e enfermeiros da ESF nas ações de prevenção e detecção precoce do câncer de mama, verificou que 75,6% tinham conhecimento adequado sobre a faixa etária e periodicidade recomendada para o rastreamento mamográfico (FERREIRA, 2020). No entanto, outros estudos (MELO et al., 2017; MORAES et al., 2016; FONSECA et al., 2016; JÁCOME e BARBOSA, 2011), relatam que mais da metade dos profissionais indicam mamografia a partir dos 40 anos e em São Paulo, região sudeste, 72,9% dos profissionais recomendam a periodicidade anual (MELO et al., 2017).

Portanto, para que a nova diretriz seja realmente adotada na prática é preciso que ela seja reconhecida como fruto de evidências de qualidade que discutem sobre faixa etária, periodicidade e balanço entre riscos e benefícios do rastreamento mamográfico e que o não seguimento de suas recomendações elevam as taxas de sobrerastreamento, sobrediagnóstico, sobretratamento e danos resultantes (INCA 2015a; MIGOWSKI et al., 2018a).

2.4 Desigualdade no Rastreamento do Câncer de Mama

A detecção precoce do câncer e o tratamento em tempo oportuno são fatores relevantes na sobrevida e mortalidade por esta causa. Revisão de literatura encontrou fortes evidências da associação entre sobrevida no câncer e medidas de status socioeconômico. As diferenças mais marcantes entre grupos sociais encontradas nos estudos se referem principalmente ao estágio da doença no momento do diagnóstico e a barreiras de acesso ao tratamento oportuno e eficaz. Atrasos na obtenção do diagnóstico contribui para que a neoplasia seja identificada em estágios mais avançados. Desigualdades de acesso aos cuidados em saúde e utilização dos serviços geram resultados desiguais no tratamento, sobrevida e mortalidade (WOODS; RACHET; COLEMAN, 2006). Dessa forma, grupos socialmente em desvantagem apresentam maior risco de adoecer e morrer do que os grupos mais privilegiados (TRAVASSOS e CASTRO, 2008).

Iniquidades em saúde tem sido relacionada também ao rastreamento do câncer de mama. Elas têm como características diferenças no estado de saúde dos grupos populacionais que são evitáveis, injustas e desnecessárias. Essas diferenças fazem parte de um processo histórico que culminou na estratificação social baseada na riqueza, ocupação, renda, educação, raça e etnia. Assim, os indivíduos passaram a ocupar diversas posições sociais, conquistando com isso distintas oportunidades para atender suas necessidades/interesses. (WHITEHEAD, 1992; CNDSS, 2008; VIEIRA-DA-SILVA e FILHO, 2009).

As desigualdades marcantes no câncer de mama são, em parte, atribuíveis à disponibilidade e à acessibilidade dos serviços de detecção e tratamento precoce. Diferenças na mortalidade por câncer de mama entre populações tem sido atribuída à renda, educação, região de residência e raça. Esses fatores individualmente ou em conjunto, influenciam na capacidade do indivíduo de obter acesso aos serviços de assistência à saúde (SIEGEL et al., 2018).

A mamografia de rastreamento é considerada um dos métodos indicados mais efetivo na detecção precoce do câncer de mama, garantia de diagnóstico e tratamento em tempo oportuno (WHO, 2014). Mulheres com menor renda e sem cobertura adequada de um sistema de saúde têm menos acesso à mamografia de rastreamento, maior probabilidade de apresentar câncer de mama em estágio avançado e menor chance de cura e sobrevida (RIBA et al., 2019). Na Tailândia, por exemplo, pessoas mais ricas tem maior cobertura de mamografia comparada às mais pobres e os fatores socioeconômicos que determinam essa

situação são: educação, bem-estar de saúde e índice de riqueza (CHONGTHAWONSATID, 2017).

Quanto à escolaridade, mulheres com níveis mais altos de educação comparadas àquelas de menor escolaridade, apresentam uma taxa mais alta de realização de mamografia devido à maior consciência sobre o risco do câncer de mama, mais conhecimento sobre comportamento de saúde e melhor acesso às informações e recursos para melhoria de sua saúde. Pobreza e baixa escolaridade têm sido relacionadas à dificuldade de acesso aos serviços de saúde, resultando em baixa frequência de rastreamento do câncer de mama (WILLEMS; BRACKE, 2018; LUNDQVIST et al., 2016).

Estudo verificou se a escolaridade está relacionada a diferenças no acesso ao rastreamento do câncer de mama. Foram comparados dois países: Estados Unidos que não possui programa de rastreamento organizado e subsidiado pelo serviço público; e Holanda que oferece rastreamento universal e subsidiado pelo serviço público, sendo constatado que nos Estados Unidos as mulheres menos escolarizadas e mais pobres não possuem cobertura de seguro saúde e acesso à aconselhamento médico para realização de mamografia, o que contribui para menores taxas de rastreamento. Em contrapartida, na Holanda, entre as mulheres convidadas para rastreamento, esses fatores não exercem influência, uma vez que o rastreamento é garantido para todas as mulheres da faixa etária alvo (50 a 75 anos). Isso demonstra que o rastreamento populacional organizado e bem implantado pode reduzir as desigualdades de acesso (KOÇ; O'DONNELL; OURTI, 2018).

Residir em áreas mais pobres e/ou distantes está associado à menores taxas de rastreamento. No Canadá, mesmo existindo um programa de rastreamento gratuito e universal, foi observado que a participação nos programas de rastreamento de grupos populacionais de baixo nível socioeconômico e residentes em áreas remotas é baixa (ZHA et al., 2019). Na Tailândia a situação é parecida, e mesmo quem possui cobertura pelo seguro de saúde não consegue acessar facilmente os serviços de diagnóstico precoce, uma vez que os centros de saúde e hospitais comunitários em áreas não-municipais ou rurais não fornecem mamografias, e as pessoas dessas áreas precisam ser encaminhadas a hospitais provinciais ou de grande porte (CHONGTHAWONSATID, 2017).

De outro modo, mulheres que residem próximo às áreas melhor nível socioeconômico, mesmo tendo poucos recursos financeiros, podem ter melhor acesso aos serviços de saúde (ZIDAR ET AL., 2015). Estudo realizado na Coreia demonstrou que residir em área urbana ou rural não faz diferença para o acesso à mamografia porque as áreas

rurais mais remotas estão localizadas a duas horas de uma metrópole onde existe instalações com mamógrafos, além disso, unidades móveis de mamografia estão disponíveis para essas áreas mais distantes (KWEON et al., 2017).

Estudo realizado no Quênia, África, constatou que mulheres residentes em áreas rurais enfrentam as seguintes barreiras no rastreamento do câncer de mama: falta de informação, longas distâncias nos serviços de saúde, longas filas de espera nos hospitais, restrições financeiras para custeio de transporte, rastreamento e/ou tratamento e ausência de profissionais de saúde para realizar os testes de rastreamento necessários. Características das áreas rurais como baixa densidade populacional e acesso precário aos cuidados de saúde contribuem para persistência dessas barreiras (KISIANGANI et al., 2018).

Espera-se que programas de rastreamento organizados e que ocupam lugar de destaque entre as opções de rastreamento disponíveis em um país, reduzam a desigualdade de acesso, uma vez que convidem todas as mulheres de uma determinada faixa etária a participar, independentemente de sua renda, raça, escolaridade e outros fatores socioeconômicos. Contudo, os programas de rastreamento dentro de um mesmo país/região podem assumir diferentes formas, o que permite que eles contribuam ou não com a diminuição das desigualdades socioeconômicas de acesso (CARNEY; O'NEILL, 2018).

Na Colômbia, por exemplo, a cobertura de seguro saúde atingiu 100% da população em 2010, o que teoricamente significa importante melhoria no acesso aos serviços de saúde. Porém na prática, principalmente para pacientes com câncer, o acesso oportuno aos serviços de diagnóstico e tratamento onera muito o sistema de saúde devido ao alto custo e longos períodos de espera para utilização dos serviços que dependem da autorização da seguradora. Além disso, os regimes de seguro saúde são diferenciados e acabam gerando desigualdades de acesso e cuidados. A má distribuição de renda per capita agrava ainda mais essa situação. Pessoas com melhores condições financeiras, em muitos casos, não esperam resposta da seguradora e pagam por alguns serviços, em contrapartida aquelas com baixos recursos aguardam os procedimentos administrativos de suas seguradoras, o que retarda o atendimento da sua necessidade. Nesse cenário, a universalidade não tem garantido a igualdade e equidade (ARIAS-ORTIZ; VRIES, 2018).

Na China, apesar da existência de um programa de rastreamento organizado, a cobertura populacional é limitada devido às restrições orçamentárias. Dessa maneira a cota de exames gratuitos é inferior ao tamanho da população feminina elegível, o que prejudica

ainda mais o acesso de mulheres residentes em áreas rurais vulneráveis e com baixo status socioeconômico e sem condições de pagar pelos serviços de saúde (YOU et al., 2019).

Mesmo nos países de alta renda os programas de rastreamento apresentam desigualdades socioeconômicas de acesso. Na Suécia, por exemplo, foi constatado que muitas mulheres que residem em municípios distantes da unidade onde o mamógrafo está instalado não comparecem para realizar a mamografia agendada. Dois fatores podem explicar essa situação: insuficiência de recursos para deslocamento e/ou medo de perder o emprego porque o grande tempo dispendido até o local onde o exame será realizado, pode ocasionar absenteísmo no trabalho (WEE et al., 2012; UJHELYI et al., 2018).

No Brasil, a situação não é diferente. A alta desigualdade de renda está associada a piores indicadores sociais e demográficos e à maior mortalidade por câncer de mama. As capitais das regiões Sul e Sudeste, têm apresentado decréscimo das taxas de mortalidade. Por outro lado, nas regiões que possuem desigualdades de acesso aos serviços de saúde, Norte e Nordeste, as taxas têm crescido tanto nas capitais quanto no interior (AZEVEDO e SILVA et al., 2014). Além disso, a maioria dos serviços credenciados no SUS para tratamento estão localizados nas regiões de melhor nível socioeconômico, Sul e Sudeste. Mulheres residentes nessas regiões de pele branca, com melhor escolaridade e cobertas por plano de saúde privado tem mais acesso à rastreamento, comparadas às demais (FIGUEIREDO; ADAMI, 2019).

O acesso ao exame de mamografia entre as microrregiões brasileiras varia conforme as características sociodemográficas, microrregiões com menor desigualdade de renda e melhor acesso geral aos cuidados de saúde tem cobertura mais alta (NOGUEIRA et al., 2019). A prevalência de não realização de mamografia é maior entre mulheres com baixa escolaridade, sem plano de saúde e residentes nas regiões Norte e Nordeste (TIENSOLI; FELISBINO-MENDES; MELENDEZ, 2020). Em 2010, a região Norte apresentou a menor cobertura de mamografia na faixa etária de 60 a 69 anos (8,3%). Destaca-se que nesse período, 46% das mamografias foram realizadas em faixas etárias não recomendadas para o rastreamento pelo Ministério da Saúde (AZEVEDO e SILVA et al., 2014). Ainda na região Norte, no período de agosto de 2013 e fevereiro de 2014, a proporção de mulheres na faixa etária de 50 a 69 anos que nunca realizaram mamografia excedeu os 50% e apenas 33,1% realizaram mamografia há menos de dois anos (THEME FILHA et al., 2016).

Na região Sudeste, as macrorregiões de saúde do estado de Minas Gerais localizadas em áreas menos desenvolvidas também enfrentam desvantagens no

rastreamento. As macrorregiões Norte e Nordeste apresentaram as menores razões de mamografia na faixa etária de 50 a 69 anos e a macrorregião de Jequitinhonha teve o maior número de mulheres na faixa etária de 55 a 69 anos sem exame anterior de mamografia (CORRÊA et al., 2017).

No país o acesso desigual contribui para que mulheres de menor nível socioeconômico, não brancas, com baixa escolaridade e tratadas nas regiões Norte e Nordeste sejam mais propensas a apresentar câncer de mama em estágio avançado. Destaca-se que viver nessas regiões reduz as chances de que uma mulher tenha um pedido médico de mamografia. Mesmo que a mulher tenha diagnóstico em um estágio onde as chances de cura são grandes, a longa espera para início do tratamento pode piorar o prognóstico (RENNA JUNIOR; SILVA, 2018; AZEVEDO E SILVA et al., 2017).

Estudo realizado no Ceará, estado da região Nordeste do Brasil, verificou que pacientes com câncer de mama encaminhados pela rede pública comparados àqueles encaminhados pelos serviços privados ou planos de saúde, esperaram mais tempo entre o diagnóstico e a primeira consulta médica (mediana de 94 dias versus 53 dias); e diagnóstico e início do tratamento (mediana de 71,5 dias versus 39 dias). Ademais, 33,3% dos pacientes encaminhados pelo serviço privado ou planos de saúde apresentavam estágio avançado da doença, em contraste 46,7% daqueles encaminhados pelo sistema público (FERREIRA et al., 2017).

Desigualdade na distribuição do orçamento em saúde afeta diretamente a organização e execução do programa de rastreamento do câncer de mama nas diversas regiões do Brasil. Os repasses federais são baseados em valores per capita, o que causa distribuição injusta dos recursos financeiros porque regiões mais ricas e populosas tendem a receber mais, porém, regiões pobres necessitam de mais recursos para ampliar suas ações e atender pessoas com dificuldade de deslocamento para acesso aos serviços de prevenção e/ou tratamento, contratar recursos humanos, construir instalações físicas e adquirir equipamentos necessários à rastreamento (BATTESINI; ANDRADE, SETA, 2017).

No tocante à desigualdade racial, mulheres negras, residentes em áreas de baixa renda, são as que mais utilizam os sistemas públicos de saúde e possuem maior chance de serem diagnosticadas com câncer de mama em estágios mais avançados (NOGUEIRA et al., 2018). Em Campinas, São Paulo, no período de 2001 a 2002 foi observada menor prevalência de mamografia entre mulheres negras e com menor renda familiar per capita (AMORIM et al., 2008). Essas situações demonstram que a raça/cor pode ser um fator que

influencia no acesso a bens, serviços e recursos sociais e de saúde, gerando a iniquidade de acesso. Portanto, é necessária a ampliação da cobertura para grupos mais vulneráveis socioeconomicamente.

Dentro de um sistema universal de saúde, que oferta o rastreamento do câncer de mama à todas as mulheres da faixa etária alvo, podem existir grupos vulneráveis que não tem acesso aos exames de rastreamento, sendo necessária a adoção de estratégias que melhorem e/ou eliminem essas disparidades. Estudo realizado no Paraná, Brasil, demonstrou que alta mortalidade por câncer de mama pode estar relacionada à forma como os serviços de saúde para diagnóstico e tratamento estão organizados e à qualidade da assistência prestada por eles. Reduzir a iniquidade de acesso é e continua a ser uma prioridade das políticas públicas para prevenção do câncer de mama (ROCHA-BRISCHILIARI et al., 2018).

2.5 Sistema de Informação e Controle do Câncer de Mama (SISMAMA)

Um dos elementos chave para o sucesso dos programas de rastreamento do câncer de mama é o monitoramento e avaliação das ações realizadas. Para esse fim, é necessário que haja um registro e que este seja constantemente alimentado e suas informações sejam de qualidade. Diante da necessidade de um sistema de gerenciamento específico do rastreamento do câncer de mama no Brasil, foi instituído em 31 de dezembro de 2008 através da Portaria nº 779 do Ministério da Saúde e implantado em junho de 2009, o Sistema de Informação e Controle do Câncer de Mama (SISMAMA).

O SISMAMA foi concebido pelo Departamento de Informática do Sistema Único de Saúde - DATASUS como um subsistema do Sistema de Informação Ambulatorial do SUS (SIA-SUS). De acordo com o § 2º do artigo 1º da referida Portaria, as principais características desse sistema são:

- I. Permite o gerenciamento das ações de rastreamento de câncer;
- II. Padroniza e aprimora a qualidade dos laudos mamográficos;
- III. Permite o seguimento das mulheres com exames alterados - gestão de casos positivos;
- IV. Permite a análise da distribuição na população das lesões encontradas nas mamografias, podendo indicar necessidades de auditorias e capacitações locais;
- V. Contribui para o planejamento da oferta de serviços, otimizando recursos;

- VI. Permite avaliar a indicação dos procedimentos de diagnóstico inicial e rastreamento;
- VII. Gera o Boletim de Produção Ambulatorial Individualizado - BPA-I;
- VIII. Possui estrutura semelhante ao SISCOLO, hoje implantado em todos os laboratórios de citopatologia do SUS;
- IX. Os dados gerados poderão ser tabulados com o auxílio do Tabwin/DATASUS, além da possibilidade de gerar relatórios padronizados. (BRASIL, 2009, § 2º art. 1º).

O sistema possui dois módulos específicos: (1) prestador de serviço: destinado aos serviços de radiologia e laboratório de patologia para cadastro dos exames, emissão de laudos e avaliação do desempenho dos serviços; (2) módulo de coordenação: para gestores monitorarem e gerenciarem as ações de detecção precoce do câncer de mama e para seguimento das mulheres com exames alterados (INCA, 2010).

A coleta de dados para alimentação do SISMAMA é feita por meio de quatro formulários padronizados, estes permitem a análise tanto espacial quanto temporal das informações, a saber: (1) Requisição de mamografia: disponível nas unidades básicas de saúde para solicitação de mamografia; (2) Resultado de mamografia: disponíveis nos serviços que realizam a mamografia para complementação de informações relacionadas à anamnese da paciente e registro das alterações do exame mamográfico; (3) Requisição de exame citopatológico: disponível nas unidades secundárias de referência para patologias mamárias; (4) Requisição de exame histopatológico: disponível nas unidades secundárias de referência para patologias mamárias e hospitais (INCA, 2011).

Os principais dados coletados por esses instrumentos são: identificação do usuário, data do exame, resultado do exame e no caso de citopatologia e histopatologia a adequabilidade (INCA, 2011).

Com o objetivo de fortalecer, integrar e promover maior resolutividade do SUS nas ações de prevenção do câncer, foi lançado em 2011, o Plano de Fortalecimento das Ações de Prevenção, Diagnóstico e Tratamento do Câncer do Colo do útero e de Mama (BRASIL, 2020). Um dos eixos desse plano, a melhoria dos sistemas de informação e vigilância do câncer, possibilitou a união de dois sistemas: SISMAMA e SISCOLO, sendo essa nova ferramenta instituída em 30 de dezembro de 2013 pela Portaria nº 3.394, sob a denominação Sistema de Informação do Câncer (SISCAN).

Esse sistema possibilita a identificação do usuário ao invés de exames, o que permite construir um histórico do paciente e da sua situação de seguimento do câncer de mama e colo do útero. Também proporciona o acesso dos dados em tempo real com a finalidade de promover maior agilidade no processo de detecção precoce do câncer e o manejo dos resultados alterados (INCA, 2013).

O fluxo do SISCAN é descrito na figura 4.

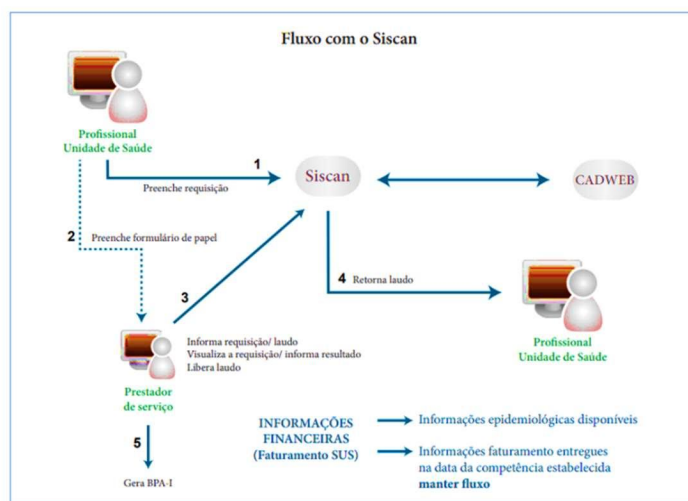


Figura 4- Fluxo do SISCAN.
Fonte: INCA, 2013.

A unidade de saúde solicita exames pelo SISCAN (1). Se a unidade de saúde não tiver acesso à internet, deverá fazer a solicitação nos formulários impressos (2), que serão digitados pelo prestador de serviço no SISCAN (3). O prestador de serviço poderá visualizar a solicitação do exame digitada pela unidade de saúde, informar o resultado e liberar o laudo (3). Após a liberação do laudo, a unidade de saúde tem acesso ao mesmo (4). Ao final de cada competência o prestador de serviço gera o boletim de produção ambulatorial individualizado (BPA-I), e encaminha o mesmo ao setor de faturamento (5). As informações epidemiológicas ficam disponíveis automaticamente na base nacional e a coordenação do local de residência e a unidade de saúde da mulher tem acesso aos dados epidemiológicos e de seguimento da mesma.

Devido as dificuldades de implantação como utilização de sistema próprio por parte de alguns serviços e ausência de ferramenta de *webservice* para migração de dados para o Siscan, principalmente na região Sudeste, o registro de mamografias ainda é realizado no Sismama/Siscan, o que dificulta a entrada regular de registros e o levantamento de dados mais fidedignos com a realidade (INCA, 2019c).

2.6 Relacionamento entre Sistemas de Informação em Saúde

A Informação em Saúde é um elemento essencial para os profissionais responsáveis pelo planejamento de políticas e ações que visam a melhorias das condições de vida e saúde da população. O conhecimento acerca do público, seu território, suas características sociais e suas inter-relações no ambiente em que vivem, permite a construção de indicadores que norteiam todo o planejamento, gerenciamento e avaliação das ações na saúde. Para atender a esse requisito foram criados os Sistemas de Informação em Saúde (FRANCO, 2005).

Sistemas de Informação em Saúde (SIS) são estruturas organizacionais, constituídas por vários subsistemas, cujas estruturas administrativas e unidades de produção perfeitamente articuladas, propiciam a obtenção, processamento, análise e transformação da informação de acordo com a necessidade de gerenciamento (BRASIL, 2005).

No Brasil os SIS são numerosos e diversificados e a maioria foi criada para atender as necessidades de gerenciamento financeiro e/ou clínico. Os SIS hoje existentes contêm informações sobre: assistência à saúde, epidemiológicas e de morbidade, rede assistencial, estatísticas vitais, demográficas e socioeconômicas, financeiras e outras (FRANCO, 2005).

Uma das limitações no uso dessas ferramentas para o estabelecimento de uma linha de cuidados, está relacionada à alta heterogeneidade dos formatos de dados, que na maioria das vezes, impedem a identificação de um indivíduo e a vinculação de seus dados nas diferentes fontes, durante longos períodos de tempo ou nas diferentes configurações de serviços, por falta de um identificador único. Alguns sistemas têm apenas o nome do indivíduo, outros acrescentam nome da mãe, ou Cartão Nacional do Sistema Único de Saúde (CNS) ou Cadastro de Pessoas Físicas (CPF) e outras informações. Além disso, os profissionais também esbarram nos erros de registros, homônimos, incompletude dos dados e mudanças de endereço. Dessa forma, uma solução para o acompanhamento longitudinal de um grupo específico de indivíduos, é o relacionamento de bases por meio do método probabilístico.

O relacionamento probabilístico consiste na ligação de determinadas variáveis e de acordo com o grau de semelhança entre elas, o software utilizado irá calcular pontuações de probabilidade. Pares com altas pontuações tem maior probabilidade de serem verdadeiros e com baixas, menor probabilidade (CAPUANI et al., 2014).

Existem diversos softwares de relacionamento probabilístico de registros, o RecLink é um deles. Ele foi desenvolvido no ano 2000 por Camargo e Coeli, e tem como

base três processos: (1) padronização dos registros: é o primeiro procedimento a ser realizado para preparação dos campos de dados, visando minimizar a ocorrência de erro durante o pareamento. Nessa etapa, os caracteres alfabéticos são convertidos em maiúsculo, são eliminadas as pontuações, caracteres especiais, espaços em branco e formatação de campos data; (2) blocagem dos registros: divisão das bases em blocos mutuamente exclusivos, limitando-se as comparações aos registros pertencentes ao mesmo bloco, por exemplo: utilização do código fonético do último nome; (3) pareamento de registros: construção, a partir da blocagem, de escores para as diferentes partes possíveis (CAMARGO, COELI, 2000).

Considerando as bases de dados em que foram empregadas ao final do relacionamento são formados os pares, que, a partir da soma dos escores ponderados de cada variável empregada no processo de pareamento, por exemplo, nome, nome da mãe ou data de nascimento, é construído um escore de cada par. As pontuações são proporcionais às razões de log-verossimilhança, onde mi é a probabilidade de que a cadeia de caracteres das variáveis dos registros dos dois bancos sejam coincidentes e essa correspondência seja verdadeira (sensibilidade) e ui é a probabilidade de que a sequência concordante trata-se de um falso par (CAMARGO, COELI, 2000).

É possível ainda que um par verdadeiro seja considerado como falso par ($1-mi$) e também que um falso par, seja considerado como par verdadeiro ($1-ui$). A probabilidade dessa discordância também é calculada. Posteriormente, com base nessas probabilidades é gerado um fator de ponderação para concordância e outro para discordância. O fator de ponderação da concordância é calculado por meio do logaritmo de base 2 da razão de verossimilhança entre as probabilidades mi e ui , $\log_2(mi/ui)$ e o de discordância como logaritmo de base 2 da razão de verossimilhança entre ($1-mi$) e ($1-ui$), $\log_2((1-mi)/(1-ui))$ (CAMARGO, COELI, 2000).

O escore total é obtido a partir da soma dos fatores de ponderação atribuídos (mi e ui). Geralmente, mi é maior que ui , dessa maneira, mi contribui positivamente para o escore final, enquanto o fator discordante ui , contribui negativamente (CAMARGO, COELI, 2000).

Diversos estudos utilizaram o RecLink no relacionamento probabilístico. Coutinho et al. (2008), para identificar nascimentos informados em um estudo com o Sistema de Informação de Nascidos Vivos (SINASC). Selig et al. (2004), para avaliar o programa de tuberculose a partir dos óbitos registrados no Sistema de Informação de Mortalidade (SIM). Capuani et al. (2014), para ligação de um banco de dados de doadores de sangue com o SIM.

Romero et al. (2016), usou para associar o SIM e o Sistema de Informação de Agravos de Notificação (SINAN).

Especificamente para o câncer, Girianelli; Thuler; Azevedo e Silva (2008), utilizaram para avaliar a qualidade do Sistema de Informação do Câncer do Colo do Útero (SISCOLO). Peres et al. (2015), para propor pontos de corte nos escores calculados no processo de linkage probabilístico, para as diversas topografias de câncer. No câncer de mama, Albrecht et al. (2013) utilizaram para verificar a associação entre a mortalidade de mulheres com câncer de mama e as características sociodemográficas e clínicas. Tomazelli, Girianelli, Azevedo e Silva (2018), no relacionamento de bases de dados do Sistema Único de Saúde (SUS) para formar uma coorte de seguimento de mulheres com mamografia suspeita de malignidade.

O relacionamento probabilístico, pode ser feito em vários passos, sendo os primeiros mais restritivos, formando menor número de pares. E os demais menos restritivos, com possibilidade de maior formação de pares. A cada passo o número total de indivíduos pode ser mantido ou no passo seguinte podem ser relacionados apenas aqueles que não formaram pares no passo anterior.

Apesar da existência de vários estudos que utilizaram o Reclink, ainda não existe um consenso quanto ao número de passos e do ponto de corte ideal para determinação de pares verdadeiros.

Girianelli, Thuler, Azevedo e Silva (2008) realizaram o relacionamento por meio de sete passos. Da Matta et al. (2008), apenas três passos. Migowski et al. (2011), utilizaram cinco passos. Tomazelli, Girianelli, Azevedo e Silva (2018), usaram quinze passos. Os três primeiros estudos retiraram dos passos seguintes os registros classificados anteriormente como pares verdadeiros. E o último manteve todos os registros do banco referência e retirou nos passos seguintes, da base seguimento, apenas as mulheres que tinham formado par anteriormente.

Quanto ao ponto de corte, os estudos utilizaram escores para o limiar superior e inferior, indicando para a inspeção os pares com escores intermediários (PERES et al., 2015; TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2018; ROMERO et al., 2016; FONSECA et al., 2010). Camargo Jr e Coeli (2000), sugeriram a definição do ponto de corte para o limiar inferior de forma arbitrária, determinando que registros com escore negativo deveriam ser considerados falsos pares e com escores positivos deveriam ser revistos.

Verifica-se que escore de relacionamento é determinado pelas variáveis utilizadas no pareamento, e cada estudo utiliza diferentes variáveis. Outro fator que limita o uso dos dados e a determinação dos escores, é a má qualidade dos dados, que em geral apresentam os seguintes aspectos: registros com nomes ausentes ou incompletos, inconsistências, duplicidades, incompletude de importantes campos como: nome da mãe, CPF, CNS, escolaridade e outros (SELIG et al., 2004; FONSECA et al., 2010).

Uma forma de solucionar esse problema, proposta por Coutinho e Coeli (2006), é a avaliação da acurácia do método probabilístico comparando os resultados do processo de relacionamento com uma fonte de informação confiável e que contenha o desfecho de interesse, sendo esta considerada padrão-ouro. Ressalta-se que nem sempre essas informações estarão disponíveis.

Camargo e Coeli (2000), recomendam a conferência manual dos pares, porém, dependendo da quantidade de dados vinculados, esse procedimento torna-se inviável, o que impede a determinação de um padrão-ouro. Tomazelli; Girianelli; Azevedo e Silva (2018), tomaram como padrão-ouro os registros que continham CNS e/ou CPF, sendo estes identificados nos passos 1 e 2. E Silveira e Artmann (2009), em uma revisão sistemática, propõem a geração de uma amostra do banco total e revisão desses pares, porém não apresentaram nenhum estudo onde esse método foi aplicado.

Desse modo, no relacionamento entre SIS, torna-se relevante a determinação do número mínimo de passos capaz de encontrar maior quantidade de pares verdadeiros e de um escore confiável para classificação dos pares. Para Migowski et al. (2011), a revisão manual de pares é o ponto crítico do processo de relacionamento porque demanda muito tempo e formuladores de políticas e ações de saúde necessitam de informações para tomada de decisões, no menor tempo possível, assim é necessário o desenvolvimento de estudos que aperfeiçoem a sistematização dessa etapa e do método de cálculo de escore visando aumentar seu poder discriminante.

Nesse estudo, a aplicação da técnica do relacionamento probabilístico foi essencial na criação de um banco com um único registro de mamografia por mulher para análises mais reais acerca do programa de rastreamento em Minas Gerais, uma vez que no SISMAMA constam registros de exames, o que gera duplicidade de informações de um mesmo indivíduo e pode superestimar as análises e resultados.

3 JUSTIFICATIVA

O câncer de mama é um grande desafio para a saúde pública, principalmente nos países e regiões de baixa e média renda, onde a infraestrutura de saúde deficiente e falta de organização dos serviços para prevenção, detecção e diagnóstico precoces contribuem para o aumento de casos identificados em estágio avançado e mortalidade, resultante de atrasos ou não acesso ao tratamento oportuno e adequado. (SUNG et al., 2021; WOODS; RACHET; COLEMAN, 2006).

Na maioria dos países, a mamografia é considerada como primeira escolha e como um dos métodos mais efetivos na detecção precoce, garantia de diagnóstico e tratamento do câncer de mama em tempo oportuno. Entretanto, devido à falta de recursos e problemas estruturais, nem todos os países de baixa a média renda conseguem ofertar rastreamento mamográfico populacional, optando em alguns casos pelo rastreamento oportunístico (WU; LEE, 2019).

Nesse tipo de rastreamento, não existe um acompanhamento sistemático e nem diretrizes bem definidas, o que resulta em ações inadequadas, como por exemplo: realização de exames em idades em que as evidências demonstram que não há benefícios, pequeno intervalo entre exames, diagnósticos de câncer de mama em estágio avançado e não alcance da meta ideal de cobertura de mulheres da faixa etária alvo, o que torna esse tipo de programa mais oneroso para o sistema de saúde e menos efetivo na redução da morbidade e mortalidade (INCA, 2015a).

O que se tem observado no Brasil é que apesar de existir diretrizes oficiais definidas e baseadas em evidências para o rastreamento mamográfico de mulheres de 50 a 69 anos com periodicidade bienal (INCA, 2015a), na prática as recomendações não são seguidas pelos gestores, profissionais de saúde e população. Há, inclusive, associações médicas que orientam o rastreamento a partir de 40 anos, com periodicidade anual (URBAN et al., 2012). No país, dos estudos realizados até 2019, nenhum tinha abordado acerca das recomendações oficiais de periodicidade para realização de mamografia, demonstrando se o intervalo entre exames está adequado ou não, e seu percentual entre as regiões. O primeiro estudo que apresentou resultados dentro dessa temática foi o de Rodrigues et al. (2019), que analisou uma microrregião de saúde do sudeste do país, constatando que em 2010 o sobrerastreamento por periodicidade atingiu 21% das mulheres e teve uma taxa de 150 mil mulheres/ano, o que demonstra que o programa de rastreamento precisa ser revisto.

No Brasil, para monitorar as ações do rastreamento do câncer de mama foi implantado em 2009 o SISMAMA - Sistema de Informação do Câncer de Mama, porém nesse sistema os registros são feitos por exame e não por mulher. Assim, desde 2013, o SISMAMA vem sendo substituído por um sistema que permite uma gestão mais detalhada das ações de controle do câncer de mama, com registro por mulher e não por exames, o SISCAN - Sistema de Controle do Câncer (BRASIL, 2020), no entanto devido as dificuldades de implantação, o registro de mamografias ainda é realizado tanto no SISMAMA quanto no SISCAN (INCA, 2019).

Este estudo se justifica pela necessidade de se conhecer e avaliar a verdadeira situação do rastreamento mamográfico no país, a partir da análise de um único exame por mulher. Como o SISCAN não está totalmente implantado, nesse estudo, utilizando o SISMAMA e a técnica de relacionamento probabilístico, foi possível chegar a um único registro por mulher, o que permitiu calcular a verdadeira cobertura de mamografia, verificar o sobrerastreamento por faixa etária e por periodicidade. O estado de Minas Gerais foi escolhido para o estudo porque apresenta características geográficas, estruturais e econômicas muito parecidas com o Brasil, com regiões mais pobres e menos desenvolvidas ao norte e nordeste e regiões mais ricas e desenvolvidas ao sul. Dessa maneira, entende-se que pesquisas realizadas no estado podem retratar importantes situações sociais, econômicas e de saúde que ocorrem em todo o país, contribuindo assim para tomada de decisões.

A utilização de uma metodologia de baixo custo, automatizada e de fácil e rápida execução, como nesse estudo, pode auxiliar gestores de saúde na avaliação das ações de rastreamento do câncer de mama, contribuindo para que estas sejam alinhadas às diretrizes de 2015 e com isso, proporcionem maior cobertura da população alvo e o alcance até 2030 da meta proposta no Plano de Ações Estratégica para o Enfrentamento das DCNT (70%), diminuição do sobrerastreamento por faixa etária e periodicidade, tratamento em tempo oportuno, redução dos casos diagnosticados em estágio avançados e consequentemente diminuição da mortalidade por câncer de mama.

4 OBJETIVOS

4.1 Objetivo Geral

Avaliar o rastreamento do câncer de mama em Minas Gerais e suas regiões de saúde a partir de dados do SISMAMA.

4.2 Objetivos Específicos

- Testar e avaliar a acurácia de um método para seleção de pares em relacionamento probabilístico de banco de dados, de forma a viabilizar a automatização da identificação de pares verdadeiros (Artigo 1);
- Avaliar a cobertura anual do rastreamento mamográfico em Minas Gerais, no período de 2010 a 2012 e fatores associados (Artigo 2);
- Estimar o sobrerastreamento mamográfico por faixa etária e periodicidade excessiva a partir de bases identificadas do SISMAMA, e sua associação com fatores sociodemográficos, em Minas Gerais, no período de 2010 a 2012 (Artigo 3).

5 METODOLOGIA

5.1 Tipo de Estudo

Trata-se de um estudo descritivo, de coorte não-concorrente de mulheres com rastreamento mamográfico com registro no SISMAMA, seguido por um estudo ecológico com indicadores sociodemográficos e assistenciais estimados por microrregiões de saúde, no período de 2010-2012.

5.2 População de Estudo e Fonte dos Dados

O estudo teve como cenário o estado de Minas Gerais, situado na região Sudeste do Brasil, dividido em 853 municípios e regionalizado em 13 macrorregiões e 77 microrregiões de saúde (figura 5). De acordo com o Censo de 2010, o estado possuía uma população de 19.597.330 habitantes, sendo 9.955.453 mulheres.

Foram utilizados dados identificados do SISMAMA, cedidos pela Secretaria de Saúde do Estado de Minas Gerais, referente ao período de 2009-2012.

As variáveis sobre os indicadores sociodemográficos e assistenciais das regiões de saúde utilizadas foram: população feminina total e na faixa etária de 50 a 69 anos, população feminina na faixa etária de 50 a 69 anos beneficiária de plano de saúde, população total, índice de desenvolvimento humano nas regiões de saúde (IDHr), renda, índice de vulnerabilidade na saúde (IVS), Gini, grau de urbanização, analfabetismo, gasto público em saúde per capita, cobertura estratégia saúde da família, razão de médicos e enfermeiros por 1.000 habitantes, razão de médicos radiologistas por 100.000 habitantes, número de mamógrafos habitantes por 1.000 habitantes e distância percorrida para realização de mamografia dentro e fora da microrregião de saúde. Tais variáveis foram selecionadas de acordo com a revisão de literatura e a disponibilidade em bases de acesso público. Assim, foram coletadas das páginas eletrônicas da Agência Nacional de Saúde Suplementar – ANS (<http://www.ans.gov.br/>), do Datasus (<http://www2.datasus.gov.br>), do Programa de Avaliação do Desempenho do Sistema de Saúde – PROADESS (<https://www.proadess.icict.fiocruz.br/>) e do Instituto Brasileiro de Geografia e Estatística - IBGE (<https://ww2.ibge.gov.br>). Informações não disponíveis foram solicitadas via sítio

eletrônico para acesso à informação do governo federal (<http://www.acessoinformacao.gov.br/>).

A distribuição dos indicadores sociodemográficos de cada microrregião encontram-se no apêndice I.



Figura 5 – Microrregiões e Macrorregiões de Saúde de Minas Gerais.

5.3 Construção da base de dados

Os arquivos mensais de exames de mamografia do período de 2009 a 2012, armazenados no SISMAMA foram agregados por ano, incluindo as variáveis descritas no quadro 4.

Quadro 4 – Variáveis selecionadas no SISMAMA.

Variável	Descrição	Variável	Descrição
C_CNES	Cadastro Nacional de Estabelecimentos de Saúde	C_CLI_DIAG	Diagnóstico da Mamografia anterior
C_ID_COMP	Competência	C_MAMO_RAS	Mamografia de Rastreamento
C_EXAME	Número do exame	C_RAD_NUM_	Número de filmes mama direita
C_ID_NOME	Nome da paciente	C_RAD_PEL_	Pele mama direita
C_ID_NOMEM	Nome da mãe	C_RAD_COMP	Composição mama direita
D_ID_DTNAS	Data de nascimento	C_NOD_SIM_	Nódulo mama direita
C_ID_ESCO	Escolaridade	C_NOD_TAM_	Tamanho do nódulo da mama direita
C_ID_ENDER	Endereço	XRAD_NUM_F	Número de filmes mama esquerda
C_ID_NUMER	Número	XRAD_PEL_	Pele mama esquerda
C_ID_COMPL	Complemento	XRAD_COMPO	Composição mama esquerda
C_ID_BAIRR	Bairro	NOD_SIM_01	Nódulo mama esquerda
C_ID_UF	Código do IBGE da unidade da federação - residência	NOD_TAM_01	Tamanho do nódulo mama esquerda
C_IBGE	Código do IBGE do município de residência	C_CON_DIA_	Diagnóstico mama direita
C_ID_SUS	Número do cartão SUS do paciente	XCON_DIA_C	Diagnóstico mama esquerda
C_US_UF	Código do UF da unidade de saúde - requisitante	C_CON_RECO	Recomendação mama direita
C_US_IBGE	Código do Município da unidade de saúde - requisitante	XCON_RECOM	Recomendação mama esquerda
C_US_CNES	Código do CNES da unidade de saúde - requisitante	D_RECEBE	Data do exame
C_US_NOME	Nome da unidade de saúde	D_EXAME	Data da solicitação do exame
C_ANM_NOD	Apresenta nódulo ou caroço na mama	D_LIBERA	Data da liberação do exame
XANM_NOD_M	Nódulo mama direita	C_ID_SEXO	Sexo
C_ANM_PARE	Apresenta risco elevado para o câncer de mama?	C_ID_RACAC	Raça
C_ANM_EXA_	Teve suas mamas examinadas por um profissional de saúde?	C_ETNIA	Etnia
C_ANM_MAMO	Se fez mamografia	C_NACIONAL	Nacionalidade
XANM_MAMO_	Ano da mamografia anterior		

Fonte: SISMAMA, 2009.

Dado que o SISMAMA é utilizado para pagamento dos prestadores de serviços, em determinado ano podem existir exames de anos anteriores. Desse modo, utilizando a variável data de solicitação do exame (D_EXAME) para as bases de exames alterados, referência e seguimento e data de realização do exame (D_RECEBE) para a base cobertura, foram selecionados e extraídos dos arquivos anuais, exames solicitados e/ou realizados de acordo com o ano, agregando-os ao arquivo ao qual se referiam. Cada base anual foi renomeada: base de exames alterados, base cobertura, base referência e base seguimento (figuras 6, 7, 8 e 9: letra a).

5.3.1 Limpeza das bases e padronização das variáveis

De cada base anual, foram extraídos apenas exames de mulheres maiores de 18 anos e residentes no estado de Minas Gerais (figura 6, 7, 8 e 9: letras b-d). Do banco 2009 (exames alterados) foram selecionados apenas exames classificados como BI-RADS® 3, 4, 5 e 6 (figura 6: letra l).

Com o objetivo de minimizar a ocorrência de erros durante o pareamento de diferentes bases, foi realizada a padronização dos registros. Os caracteres alfabéticos foram convertidos em maiúsculo, foram eliminadas as pontuações, caracteres especiais e espaços em branco. As variáveis numéricas e de data que não estavam no formato adequado, foram corrigidas. Os campos em branco das variáveis nome da mãe, data de nascimento e CNS, foram preenchidos com o código M09, M10, M11 e M12, de acordo com o ano ao qual o registro pertencia. Foram excluídos registros sem nome (figura 8: letra e).

Para viabilizar posteriores análises, foram criadas as seguintes variáveis: mês de solicitação do exame, identificação única de cada registro, idade, macrorregião e microrregiões de saúde. Todo o processo de construção, limpeza, padronização e criação de variáveis foi realizado no *R Studio* (<http://www.r-project.org/> versão 3.6.1) e os scripts encontram-se no apêndice II.

Para o relacionamento probabilístico no RecLink III (<http://reclink.sourceforge.net/>), no módulo padroniza, as variáveis nome e nome da mãe foram padronizadas, e o primeiro e último nomes foram transformados em código *soundex* apêndice III.

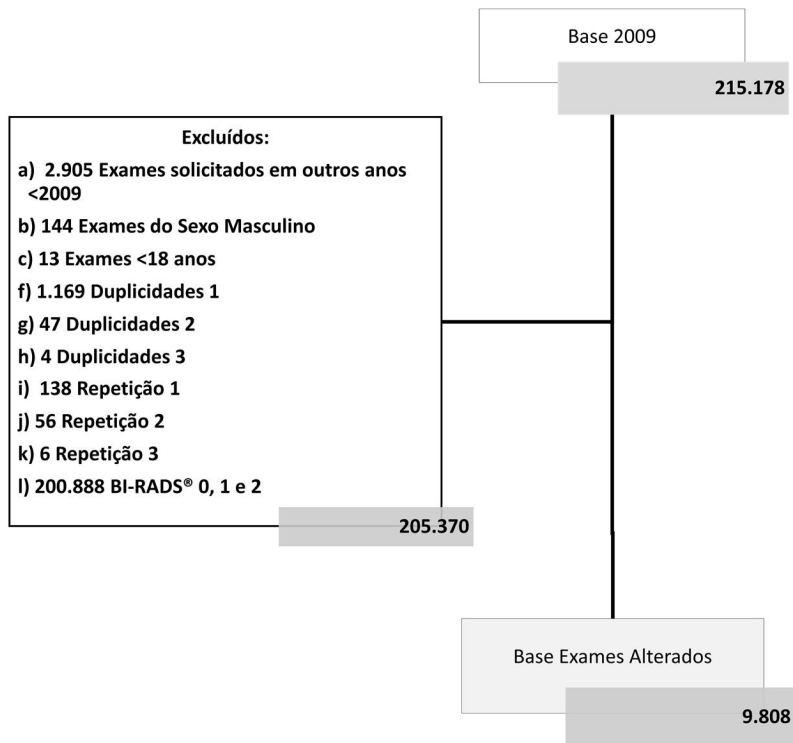


Figura 6 – Processo de limpeza da base exames alterados (ano 2009).

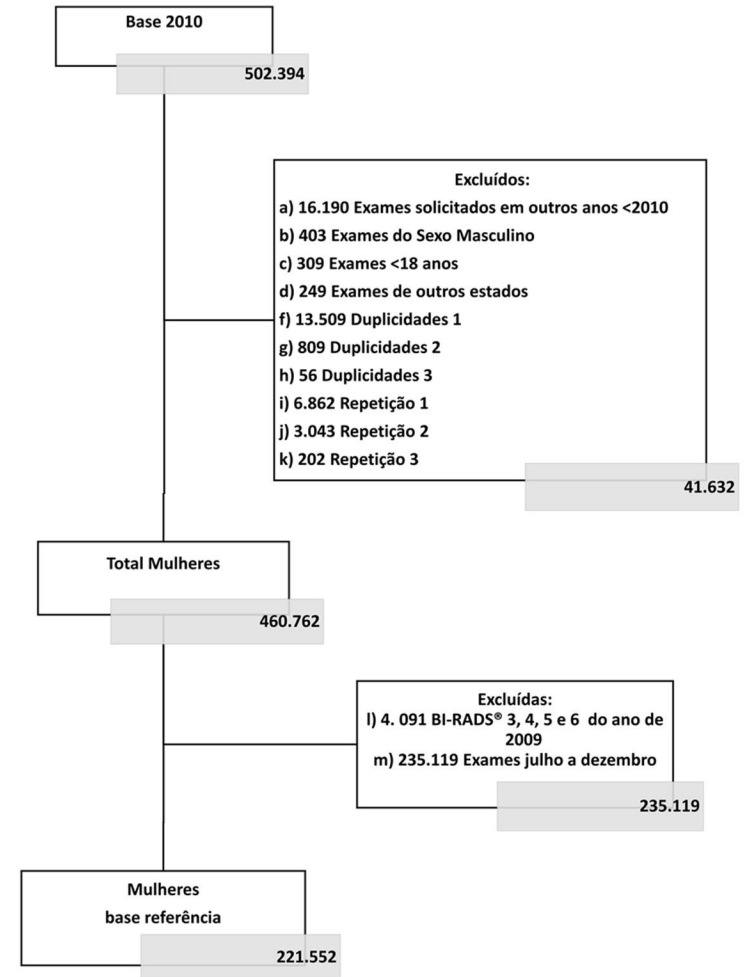


Figura 7 - Processo de limpeza da base referência (2010)

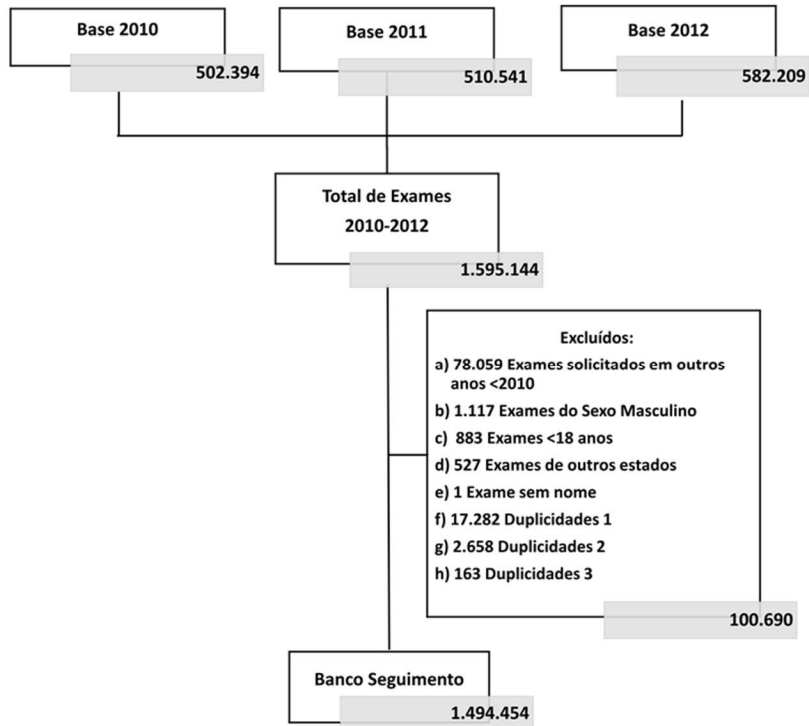


Figura 8 – Processo de limpeza da base seguimento (2010-2012).

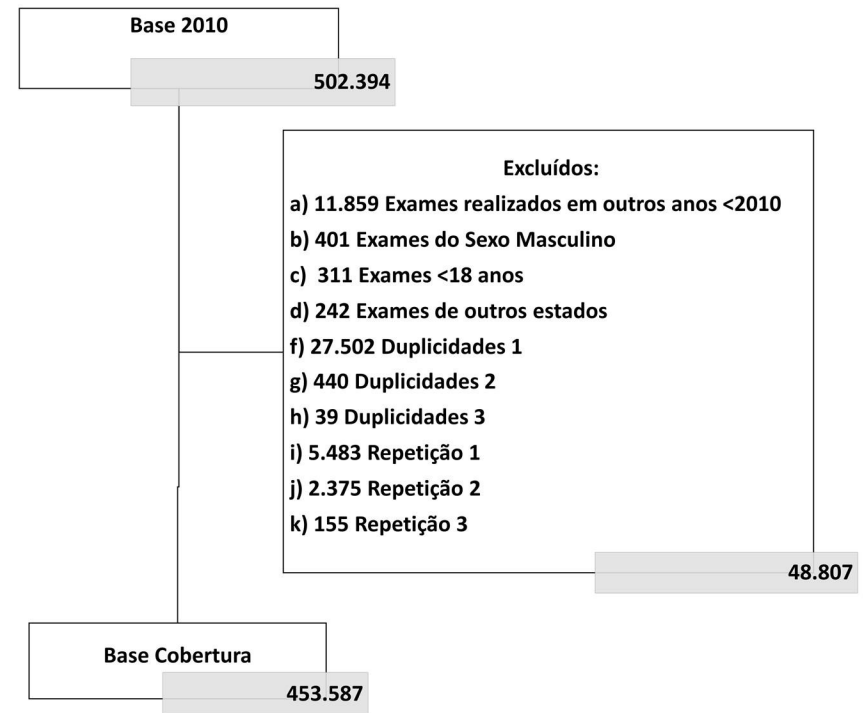


Figura 9 – Processo de limpeza da base cobertura (ano 2010).

5.3.2 Remoção de duplicidades e de exames repetidos

Para manter apenas um registro de cada indivíduo na base de exames alterados, referência e cobertura, foram removidas as duplicidades pelos métodos determinístico e probabilístico (figura 6, 7 e 9: letras f-h). No método determinístico, a ligação de registros foi feita com base na concordância exata de uma ou mais variáveis. No método probabilístico, a ligação de registros é baseada em uma probabilidade, calculada a partir da comparação das variáveis escolhidas e semelhança entre elas (CAPUANI et al., 2014).

Para retirada de duplicidades foram utilizadas as variáveis: nome, nome da mãe, data de nascimento, data de solicitação do exame (D_EXAME), cadastro nacional de estabelecimentos de saúde (CNES) e código do município. E para retirada de repetições (seguimento): nome, nome da mãe, data de nascimento e código do município.

Os relacionamentos probabilísticos foram realizados em duas etapas. Nos campos de bloqueio foram utilizadas as variáveis: códigos *soundex* do primeiro e último nome da mulher ou da mãe, ano de nascimento, mês de solicitação do exame e CNES requisitante; como campos de comparação: nome da mulher, nome da mãe, data de nascimento, data de solicitação do exame (D_EXAME) e código do município. Na retirada de repetição de exames (seguimento) foram utilizadas as variáveis: códigos *soundex* do primeiro e último nome da mulher ou da mãe, ano de nascimento; como campos de comparação: nome da mulher, nome da mãe, data de nascimento e código do município.

O relacionamento determinístico foi realizado no R Studio (duplicidade1 e repetição 1) e o relacionamento probabilístico no módulo Duplicidade do RecLink III (duplicidades 2 e 3; repetições 2 e 3), conforme recomendações do manual do programa (CAMARGO JR; COELI, 2006). No quadro 5 são apresentadas as etapas e a forma como cada variável foi combinada para retirada das duplicidades e repetições.

Quadro 5 – Combinação das Variáveis para retirada de duplicidades e repetições.

Etapa	Método	Variáveis
Duplicidade 1	Determinístico	NOME, MÃE, DATA DE NASCIMENTO, DATA DE SOLICITAÇÃO DO EXAME (D_EXAME), CNES REQUISITANTE E CÓDIGO DO MUNICÍPIO.
Duplicidade 2	Probabilístico	Blocagem: PNOME, UNOME, MÊS DE SOLICITAÇÃO DO EXAME, ANO DE NASCIMENTO E CNES REQUISITANTE; Comparação: NOMEPAD, MAEPAD, DATA DE NASCIMENTO, DATA DE SOLICITAÇÃO DO EXAME (D_EXAME) E CÓDIGO DO MUNICÍPIO.
Duplicidade 3	Probabilístico	Blocagem: PMAE, UMAE, MÊS DE SOLICITAÇÃO DO EXAME, ANO DE NASCIMENTO E CNES REQUISITANTE; Comparação: NOMEPAD, MAEPAD, DATA DE NASCIMENTO, DATA DE SOLICITAÇÃO DO EXAME (D_EXAME) E CÓDIGO DO MUNICÍPIO.
Repetição 1	Determinístico	NOME, MÃE, DATA DE NASCIMENTO E CÓDIGO DO MUNICÍPIO.
Repetição 2	Probabilístico	Blocagem: PNOME, UNOME, ANO DE NASCIMENTO; Comparação: NOMEPAD, MAEPAD, DATA DE NASCIMENTO E CÓDIGO DO MUNICÍPIO.
Repetição 3	Probabilístico	Blocagem: PMAE, UMAE, ANO DE NASCIMENTO. Comparação: NOMEPAD, MAEPAD, DATA DE NASCIMENTO E CÓDIGO DO MUNICÍPIO.

5.3.3 Exclusão de exames com resultados alterados

Com a finalidade de evitar a superestimação do sobrerastreamento mamográfico, foram excluídas base de referência as mulheres que tiveram resultado alterado no ano de 2009 (BI-RADS® 3, 4, 5 e 6), uma vez que o menor intervalo entre exames pode ser consequência do acompanhamento das anormalidades conforme as recomendações, como por exemplo resultado BI-RADS® 3 é recomendada a repetição em 6 meses a 1 ano. O banco de exames alterados foi então, relacionado com a base referência (figura 7: letra l).

O relacionamento probabilístico foi realizado em dezesseis passos, constituídos de diferentes variáveis usadas como chaves de blocagem e pareamento, adaptadas de um estudo que usou os dados do SISMAMA⁴ e segundo as recomendações do manual do RecLinkIII¹⁰ (Tabela 1).

Tabela 1 – Passos, Variáveis de bloqueio e pareamento, escores, pares formados e pares verdadeiros do relacionamento dos bancos do SISMAMA 2009 e 2010.

Passos	Variáveis de Bloqueio	Variáveis de pareamento	Escore	Pares formados	Pares verdadeiros
1	PN + UN + AN	N+DN+M+MUN	-13 a 21	2.108	1.076
2	PN + UN + AN	N+DN	-6 a 10	6.067	1.069
3	CNS	N + DN + M	-10 a 17	216	216
4	PN + UN + AN + PM + UM	N + DN + M	-10 a 17	932	923
5	PN + UN + AN	N + DN + M	-10 a 17	1.373	1.069
6	PN + AN + PM + UM	N + DN + M	-10 a 17	1.067	983
7	PN + AN + UM	N + DN + M	-10 a 17	1.188	1.022
8	PN + AN + PM	N + DN + M	-10 a 17	1.223	1.075
9	PN + AN	N + DN + M	-10 a 17	2.620	1.142
10	PN + UN + PM + UM	N + DN + M	-10 a 17	991	950
11	PN + UN + PM	N + DN + M	-10 a 17	1.181	1.042
12	PN + UN	N + DN + M	-10 a 17	2.143	1.106
13	PN + PM + UM	N + DN + M	-10 a 17	1.286	1.012
14	PN + UM	N + DN + M	-10 a 17	1.591	1.053
15	UN + NA	N + DN + M	-10 a 17	1.919	1.080
16	NA	N + DN + M	-10 a 17	12.581	1.145

AN: Ano de Nascimento; CNS: Cartão Nacional do SUS; DN: Data de Nascimento; M: Mãe; MUN: Município; N: Nome; PM: Primeiro Nome da Mãe; PN: Primeiro nome; UM: Último Nome da Mãe; UN: Último nome.

Após o relacionamento, os pares formados em cada passo foram inspecionados manualmente e classificados no módulo Combina do *RecLink III* como par e não par, segundo os critérios descritos por Ribeiro, 2019 e apresentados no quadro 6. Os pares que atendiam a, pelo menos, um dos critérios foram considerados verdadeiros.

Após o processo de classificação, os pares verdadeiros de cada passo foram reunidos em um arquivo único, denominado exames alterados. No programa *R Studio* os nomes que constavam na base exames alterados foram retirados da base de referência (figura 6).

Quadro 6 – 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

5.3.4 Construção da base referência, seguimento e cobertura

Com a finalidade de obter maior tempo de seguimento, após o relacionamento e exclusão dos exames alterados foram selecionadas para compor a base referência as mulheres que tiveram exames solicitados no período de janeiro a junho/2010 (figura 7: letra m).

Após o processo de limpeza, os exames dos anos 2010, 2011 e 2012 foram agregados em uma única base, constituindo assim base seguimento (figura 8). E a base cobertura foi criada com todas as mulheres que realizaram exame no 2010 (figura 9).

5.3.5 Acurácia de um método para seleção de pares em relacionamento probabilístico de banco de dados

Os arquivos dos pares formados nos dezesseis passos do relacionamento entre o banco de exames alterados e a base referência, classificados manualmente como par ou não, foram testados com a finalidade de avaliar um método de classificação automatizada de pares no relacionamento probabilístico. Esses arquivos foram considerados como padrão-ouro.

Do arquivo original de cada passo, ainda não classificados, foram extraídas três diferentes amostras de pares, compostas por 100, 200 e 300 pares, respectivamente, para posterior revisão manual e comparação com o padrão-ouro.

Para a definição do tamanho da amostra foi utilizada a função `power.roc.teste` do pacote `pROC` do programa R, que estimou que um tamanho amostral de cerca de 100 registros seria suficiente. Para fins de comparação foram testadas ainda amostras com 200 e 300 pares.

Para avaliar a acurácia do método em cada passo e nas amostras foram utilizadas a sensibilidade (*recall*) e valor preditivo positivo (*precision*) através da curva *precision-recall* e da medida F, que possibilitou a determinação do escore de maior poder discriminatório na identificação dos pares verdadeiros¹¹. Tal análise utilizou os pacotes `pROC` e `ROCR` do programa R.

Por fim, foi feita a combinação de todos os passos para encontrar o número total de pares verdadeiros. Procedeu-se então, à combinação dos 16 passos entre si para investigar quais passos em conjunto abrangem maior número de pares verdadeiros. O programa R foi

utilizado no cálculo das amostras e construção da curva *precision-recall*. Os resultados são no artigo 1.

5.4 Análise dos dados

5.4.1 Cálculo da cobertura do rastreamento

Para o cálculo da cobertura do rastreamento do câncer de mama por mulher examinada, em 2010, em Minas Gerais foi utilizado a base cobertura, dado que o referido banco após limpeza e exclusão de duplicidades e repetições tinha apenas o registro de uma única mulher (figura 6). A cobertura foi calculada considerando no numerador o número de mulheres examinadas na faixa etária de 50 a 69 anos e no denominador a metade da população feminina na faixa etária de 50 a 69 anos e que não são beneficiárias de plano de saúde, justificada pela recomendação de mamografia bianual para seguimento após exame normal. Para obtenção da população feminina não beneficiária de plano de saúde, subtraiu-se a população feminina beneficiária de planos de saúde na faixa etária de 50 a 69 anos, da população feminina nessa mesma faixa etária. As informações utilizadas para cálculo da cobertura da mamografia na população feminina total e alvo do programa de rastreamento (50 a 69 anos) são apresentadas na tabela 2.

Tabela 2 – População feminina de Minas Gerais e cobertura da mamografia, 2010.

População	Total	Faixa etária 50 a 69 anos
Feminina	9.955.453	1.730.927
Beneficiárias Planos de Saúde	2.513.589	443.145
Não Beneficiárias de Planos de Saúde	7.441.864	1.287.782
1/2 População Feminina Não Beneficiárias de Planos de Saúde	3.720.932	643.891

Fonte: DATASUS, 2019; ANS, 2019.

Utilizou-se para análise dos dados o programa R Studio. Os resultados são apresentados no artigo 2.

5.4.2 Avaliação do sobrerastreamento por faixa etária e periodicidade excessiva

Foi realizado o relacionamento probabilístico entre base referência e seguimento, por meio de três passos, utilizando as variáveis de blocagem e pareamento descritas na tabela 3. Com a finalidade de obter o maior número de pares verdadeiros possíveis, a cada passo foram mantidos todos os registros das bases referência e seguimento, mesmo aqueles que foram classificados como par no passo anterior.

Tabela 3 – Passos, Variáveis de blocagem e pareamento para relacionamento probabilístico das bases do SISMAMA.

Passos	Variáveis de blocagem	Variáveis de pareamento
1	PN + AN	N + DN + M
2	PN + UN	N + DN + M
3	AN	N + DN + M

AN: Ano de Nascimento; DN: Data de Nascimento; M: Mãe; N: Nome; PN: Primeiro nome; UN: Último nome.

De cada passo foi extraída uma amostra composta por 300 pares, para posterior revisão manual e definição de ponto de corte para classificação dos pares verdadeiros.

Para determinar o escore de maior poder discriminatório na identificação dos pares verdadeiros e avaliar sua acurácia foi utilizada a sensibilidade (*recall*) e valor preditivo positivo (*precision*) através da curva *precision-recall* e da medida F, conforme Duarte et al., 2019.

Os pares de todos os passos foram combinados para encontrar o número total de pares verdadeiros. Desses pares para análise do sobrerastreamento por faixa etária, ou seja, realização de exames em mulheres com idade <50 anos ou >=70 anos, foram selecionadas somente as mulheres com diagnóstico BI-RADS® 0, 1, 2, 3, 4 e 5 para ambas as mamas, que não apresentavam nódulo na anamnese, visto que essa descrição caracteriza mamografia diagnóstica; e de mulheres que foram identificadas como de baixo risco para câncer de mama (selecionadas as opções não e não sabe se tem risco elevado), dado que a recomendação para rastreamento de mulheres de alto risco é diferente. Foi calculado o percentual de exames por micro e macrorregiões de saúde nas seguintes faixas etárias menor que 30, 30 a 39, 40 a 49, maior ou igual a 70 anos e total (figura 6).

Posteriormente, para análise do rastreamento por periodicidade excessiva, ou seja, a repetição do exame em um período inferior a 18 meses, foram selecionadas no banco

anterior as mulheres com diagnóstico BI-RADS® 1 e 2 para ambas as mamas e na faixa etária de 50 a 69 anos.

A data do primeiro exame normal no ano de 2010 (variável D_RECEBE) foi tomada como data inicial para análise do seguimento. O desfecho foi sobrerastreamento mamográfico por periodicidade excessiva, definido como mamografia realizada <18 meses após resultado normal, uma vez que variações de ≤ 6 meses do preconizado foram consideradas pouco relevantes e possivelmente logísticas (RODRIGUES et al., 2019). Para as mulheres que não realizaram o exame subsequente até 30/06/2012 e para aquelas cujo intervalo foi superior a 18 meses ocorreu a censura.

A análise dos tempos entre o primeiro exame com resultado normal e o subsequente foi feita por meio do método de *Kaplan-Meier*. Em seguida, foi construído um gráfico das curvas de realização do exame subsequente (curvas de sobrerastreamento). A distribuição do tempo foi apresentada, em valor mínimo, primeiro quartil, mediana, média, terceiro quartil e valor máximo. Por fim, foi analisado o sobrerastreamento por macro e microrregiões de saúde. O programa empregado na análise estatística foi o R (<https://www.r-project.org/>). Os resultados são apresentados no artigo 3.

5.4.3 Associação da cobertura e do sobrerastreamento com variáveis sociodemográficas e assistenciais

No quadro 7 são descritas as variáveis sobre os indicadores sociodemográficos e assistenciais das regiões de saúde do ano de 2010 utilizadas no estudo ecológico, bem como a forma de cálculo e fonte de cada uma. Tais variáveis foram selecionadas de acordo com a revisão de literatura e a disponibilidade em bases de acesso público com o objetivo de se desenvolver um estudo exploratório sobre a associação entre a cobertura e sobrerastreamento com as características das microrregiões de saúde de Minas Gerais.

Foi utilizado também o Índice de Vulnerabilidade na Saúde e as variáveis para construção desse índice foram extraídas da página do Instituto Brasileiro de Geografia e Estatística. As informações que não estavam disponíveis foram solicitadas via sítio eletrônico para acesso à informação do governo federal¹.

¹Endereço da página de acesso à informação: <http://www.acessoainformacao.gov.br/>

Quadro 7– Cálculo dos Indicadores Socioeconômicos e de Saúde, 2010.

Variáveis	Descrição/Ano	Fonte/Endereço Eletrônico	Cálculo
Bloco Socioeconômico – distal			
IDH	Índice de Desenvolvimento Humano da Regiões de Saúde, 2010	Proadess 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.
Renda	Renda média domiciliar per capita, 2010	Proadess 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.
Gini	Índice de Gini, 2010	Atlas do desenvolvimento humano no Brasil http://www.atlasbrasil.org.br/2013/pt/consulta/	Para o cálculo do índice de Gini das microrregiões, foram agregados os índices municipais e realizada média simples.
Grau de Urbanização	Percentual da População Urbana, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/tabgi.exe?ibge/cnv/popmg.def	$(\text{População urbana} / \text{População total}) * 100$
Analfabetismo	Percentual da população com ≥ 15 anos analfabeta, 2010	Proadess https://www.proadess.icict.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.
População	Tamanho da População Total e Feminina, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popmg.def	Dado bruto, extraído por município e agregados por microrregiões de saúde.
Bloco Assistencial - proximal			
Médicos	Número de Médicos/ 1000 hab, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def	$(\text{Número de médicos/habitantes}) * 1000$

Enfermeiros	Número de Enfermeiros/ 1000 hab, 2010	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def	(Número de enfermeiros/habitantes) *1000
Cobertura Estratégia Saúde da Família	Percentual da população coberta pela Estratégia Saúde da Família, 2010	Proadess https://www.proadess.icict.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
Médicos Radiologistas	Número de Médicos Radiologistas/100.000 habitantes	DATASUS http://tabnet.datasus.gov.br/cgi/deftohtm.exe?cnes/cnv/prid02mg.def	(Médicos Radiologistas / Habitantes) *100.000
Gasto Público Total em Saúde por habitante (R\$)	Gasto em saúde por habitante, advindo de todas as fontes, quer sejam impostos, transferências SUS de outras esferas governamentais, operações de créditos, e outras, 2010	Proadess https://www.proadess.icict.fiocruz.br/index.php?pag=fic_r&cod=K51&tab=1	Numerador: gasto total em saúde (todas as fontes de financiamento). Denominador: população total residente.
Distância Percorrida para realização de mamografia na região de residência	Distância percorrida em Km para realização de procedimentos em municípios da mesma região de saúde, 2010	Proadess https://www.proadess.icict.fiocruz.br/index.php?pag=fic_r&cod=A26&tab=1	Numerador: soma do número de procedimentos realizados por residentes de uma RS na própria RS multiplicados pela distância entre as sedes dos municípios de residência e de atendimento. Denominador: soma de procedimentos realizados por residentes de uma RS na própria RS
Distância Percorrida para realização de mamografia fora da região de residência	Distância percorrida em Km para realização de procedimentos em outras regiões de saúde, 2010	Proadess https://www.proadess.icict.fiocruz.br/index.php?pag=fic_r&cod=A26&tab=1	Numerador: soma do número de procedimentos realizados por residentes de uma RS em outra RS multiplicada pelas distâncias entre as sedes dos municípios de residência e de atendimento. Denominador: soma de procedimentos realizados por residentes de uma RS em outra RS.
Mamógrafos Disponíveis ao SUS	Mamógrafos/100.000 habitantes	Proadess https://www.proadess.icict.fiocruz.br/index.php?pag=fic_r&cod=Z01&tab=1	Numerador: número de mamógrafos em uso X 100.000. Denominador: população total residente. O indicador considera três tipos do equipamento (mamógrafo com comando simples, mamógrafo com estereotaxia e mamógrafo computadorizado).

O Índice de Vulnerabilidade na Saúde (IVS) é um indicador composto pelas seguintes variáveis:

- **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 (PITCHON et al., 2013).

Após a coleta, essas variáveis foram padronizadas em uma escala de 0 a 1 para cada microrregião em estudo com a finalidade de se comparar e agregar cada uma delas para construção do IVS, conforme a seguinte fórmula:

(Valor bruto – Valor mínimo / Valor máximo – Valor mínimo), em que:

Valor Bruto: Valor referente à microrregião em análise; Valor Mínimo: Valor referente à microrregião que apresentou menor valor; Valor Máximo: Valor referente à microrregião que apresentou maior valor.

Além disto, cada variável tem um peso diferente, ou seja, pode ter maior ou menor importância quando comparada às demais na avaliação da vulnerabilidade em saúde. Dessa forma, elas foram ponderadas de acordo com a proposta dos idealizadores do IVS apresentada no quadro 8. Posteriormente, os valores foram somados para cada microrregião, formando assim, o IVS.

Quadro 8 – Pesos atribuídos aos indicadores selecionados, IVS 2012.

INDICADOR/DIMENSÃO	PESOS
Abastecimento inadequado de água	0,424
Esgotamento sanitário inadequado	0,375
Coleta inadequada de lixo	0,201
Moradores por domicílio	0,073
População analfabeta	0,283
Renda per capita de até ½ SM	0,288
Renda média dos responsáveis	0,173
Percentual de pop. Negra e indígena	0,185
Saneamento	0,396
Socioeconômica	0,604

Fonte: PITCHON et al., 2013.

Por fim, para demonstrar a extensão, localização e características da cobertura e do sobrerastreamento e indicadores socioeconômicos e de saúde nas diferentes microrregiões de Minas Gerais foram construídos mapas temáticos. A divisão dos indicadores em cinco classes foi realizada de acordo com 0,5 e 1,0 desvios padrões abaixo e acima da média e em escalas com tons mais claros para os menores valores/pior desempenho do indicador e mais escuros para os maiores valores/melhor desempenho do indicador.

Por meio da regressão linear, foi verificada a associação entre variáveis sociodemográficas e assistenciais, a cobertura e o sobrerastreamento por faixa etária e periodicidade excessiva. Na primeira etapa foi realizada a regressão linear simples. Em seguida, para cada modelo da regressão múltipla foram inseridas as variáveis explicativas de acordo com o grupo de exposição ao qual pertenciam, sendo seguido o critério do grupo distal para o proximal. Após o ajuste os resíduos de cada modelo foram investigados para detectar evidências de heterocedasticidade e/ou autocorrelação espacial (respectivamente com os testes de Breusch-Pagan e de Moran I). No modelo que apresentou heterocedasticidade foi inserida a variável população classificada segundo o porte populacional, a saber: pequeno até 100.000 habitantes; médio de 100.001 a 300.000; e grande >300.000. Os resultados são apresentados nos artigos 2 e 3.

5.5 Aspectos Éticos

Como as informações procedem de uma base de dados secundária, com variáveis identificadoras das pacientes, como nome, nome da mãe e endereço, este projeto foi submetido ao Comitê de Ética da Universidade Federal de Juiz de Fora, juntamente com o documento de solicitação de dispensa do Termo de Consentimento Livre e Esclarecido (TCLE) e foi aprovado sob o parecer nº 1.431.916. Esta proposta faz parte de um projeto maior intitulado “Desigualdades de acesso às ações de controle de câncer através dos Sistemas de Informações de Saúde (DAAC-SIS)”, que é resultado da parceria entre pesquisadores da UERJ, USP, UFBA, UFJF e *London School of Hygiene and Tropical Medicine* (LSHTM, do Reino Unido). Além disto, o projeto já foi aprovado no âmbito do Edital nº 07/2017, financiado pela Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), sob o título “Desigualdade no rastreamento do câncer de mama e do colo do útero no estado de Minas Gerais e suas microrregiões de saúde”.

Durante todo o trabalho foram adotadas medidas preventivas para garantir a segurança dos dados pessoais disponibilizados. Os dados foram arquivados em unidade de armazenamento externo (HD externo) protegido por senha e sempre que os bancos de dados foram manipulados, a internet do notebook foi desabilitada.

6 RESULTADOS

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

6.1 ARTIGO 1: AVALIAÇÃO DE MÉTODO PARA CLASSIFICAÇÃO AUTOMATIZADA DE PARES EM RELACIONAMENTOS PROBABILÍSTICOS DE BANCOS DE DADOS

Artigo publicado na Revista Cadernos Saúde Pública (DUARTE *et al.*, 2019 - <https://scielosp.org/pdf/csp/2019.v35n11/e00066419/pt>)

Resumo

Objetivo: Testar e avaliar a acurácia de um método para seleção de escore em relacionamento probabilístico de banco de dados, de forma a viabilizar a automatização da identificação de pares verdadeiros dispensando a etapa de inspeção manual. **Metodologia:** Estudo de acurácia utilizando dados do SISMAMA de Minas Gerais de 2009 e 2010. Após o processo de limpeza e padronização, foi realizado o relacionamento probabilístico dos bancos 2009 e 2010 utilizando 16 passos, sendo que, cada passo foi inspecionado manualmente para se obter um padrão-ouro. Posteriormente, selecionaram-se amostras que foram inspecionadas e avaliadas para calcular a acurácia do método de seleção dos pares verdadeiros. **Resultados:** Todos os passos e amostras com 200 e 300 pares apresentaram alta sensibilidade (*recall*) >0,97, alto valor preditivo positivo (*precision*) >0,95 e altas acurácia (> 0,97), medida F (>0,96) e área sob a curva *precision-recall* (>0,98). A amostra com 100 pares evidenciou altos valores para essas medidas, porém com escores mais baixos. Dos 16 passos avaliados, o uso de apenas três de forma combinada, foram suficientes para identificar 99,24% dos pares verdadeiros no banco total. **Conclusão:** O método proposto permite automatizar o relacionamento das bases de dados, mantendo a acurácia do método. Facilita a utilização de relacionamento probabilístico no âmbito dos serviços de saúde, especialmente para a vigilância e gestão em saúde.

Palavras-Chave: Sistema de Informação em Saúde, Relacionamento de bases de dados, Integração de Sistemas, Ponto de Corte, Acurácia.

Abstract

Objective: test and assess the accuracy of a scoring method in probabilistic data linkage in order to enable automatic identification of true matches, dispensing with the manual inspection stage. **Methods:** Accuracy study using data from the Breast Cancer Information System (SISMAMA) base in Minas Gerais State, Brazil, from 2009 and 2010. After cleaning and standardization, a 16-step probabilistic linkage of the 2009 and 2010 databases was performed, where each step was inspected manually to obtain a gold standard. Samples were then selected, inspected, and assessed to calculate the method's accuracy in selecting true matches. **Results:** All the steps and samples with 200 and 300 matches showed high sensitivity (recall) > 0.97 , high positive predictive value (precision) > 0.95 , high accuracy (> 0.97) and F measure (> 0.96), and high area under the curve precision-recall (> 0.98). The sample with 100 matches showed high values for these measures, but with low scores. Of the 16 steps assessed, the combined use of only three was sufficient to identify 99.24% of the true matches in the total database. **Conclusion:** The proposed method allows automatically linking databases, maintaining the method's accuracy. It facilitates the use of probabilistic linkage in health services, especially for health surveillance and management. **Keywords:** Health Information System; Systems Integration; Data Accuracy

INTRODUÇÃO

Os Sistemas de Informação em Saúde (SIS) existentes no Brasil são numerosos e diversificados e a maioria foi criada para atender as necessidades de gerenciamento financeiro e/ou clínico. Avanços no campo da Saúde Coletiva, nos conceitos de saúde e qualidade de vida da população exigem cada vez mais dos profissionais conhecimento acerca da realidade social e de saúde das pessoas com a finalidade de planejamento e priorização de suas ações. Dessa maneira, os SIS tornaram-se grandes aliados dos gestores e profissionais da saúde.

Cada setor da sociedade produz informações específicas para sua área de atuação. Para o setor saúde, a vinculação de dados de indivíduos de uma determinada população utilizando diferentes SIS é importante para se estabelecer e avaliar uma linha de cuidados, já que a saúde é resultante das condições sociais, políticas e econômicas do meio onde cada indivíduo está inserido.

O relacionamento de bases de dados é definido como estratégia que reúne duas ou mais informações, registradas em diferentes fontes e que se referem a um mesmo indivíduo. Vincular essas fontes nem sempre é uma tarefa fácil e rápida, uma vez que algumas bases não apresentam um identificador único, como o Cartão Nacional do Sistema Único de Saúde (CNS) ou Cadastro de Pessoas Físicas (CPF), que permitiriam a realização de um relacionamento determinístico. Além disso, há possibilidade de erros de registros, presença de homônimos, incompletude dos dados e mudanças de endereço. Dessa forma, uma solução é o relacionamento de bases por meio do método probabilístico, utilizando variáveis demográficas como nome, nome da mãe, data de nascimento e endereço, que pondera a vinculação dessas informações, resultando na construção de escores de classificação (COUTINHO e COELI, 2007).

Posteriormente é preciso estabelecer se as ligações estão ou não corretas. A classificação manual de pares usada nesse procedimento é considerada padrão-ouro. Porém, é uma etapa que demanda muito tempo, principalmente quando existe grande número de pares a serem inspecionados, situação que pode inviabilizar o processo (SILVEIRA e ARTMANN, 2009). Em estudo sobre relacionamento de dados referentes ao setor de alta complexidade em cardiologia foram despendidas duas horas para conferir todos os pares formados, gerando a recomendação de que nos casos em que exista grande número de pares após o relacionamento, seja realizada inspeção seletiva nas faixas de escore de maior interesse (MIGOWSKI et al., 2011). Outros estudos utilizaram escores para o limiar superior e inferior (TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2018; PERES et al., 2015; ROMERO et al., 2016), indicando para a inspeção os pares com escores intermediários. Camargo Jr e Coeli et al. (2000), sugeriram a definição do ponto de corte para o limiar inferior de forma arbitrária, determinando que registros com escore negativo deveriam ser considerados falsos pares e com escores positivos deveriam ser revistos. No entanto essas práticas tornam-se ineficazes quando grande número de pares atende ao critério de revisão. Uma terceira estratégia seria automatizar a categorização dos pares como verdadeiros ou falsos, dispensando a inspeção manual, entretanto, permaneceria a necessidade da definição de um valor de escore como ponto de corte (SILVEIRA e ARTMANN, 2009).

Diante da necessidade de otimização do processo e de uma elevada acurácia dos resultados, é preciso que para cada relacionamento de dados seja estabelecido um escore específico que identifique os pares verdadeiros. Mas como definir esse valor? Estudos sugerem que o escore e a validade das estratégias de relacionamento sejam determinados a

partir da comparação com um conjunto de dados referência cujo status de correspondência verdadeira seja conhecido, isto é, um padrão-ouro (COUTINHO e COELI, 2007; FONSECA et al., 2010; HARRON et al., 2017).

O padrão-ouro pode ser procedente de uma fonte de dados com identificadores completos ou de uma subamostra dos registros que foram revisados manualmente (HARRON et al., 2017). Revisão sistemática sobre a acurácia de relacionamento probabilístico de bases em saúde também sugere a geração de uma amostra e revisão manual dos pares dessa, porém não apresenta nenhum estudo onde esse método foi aplicado (SILVEIRA e ARTMANN, 2009). Outro estudo utilizou como padrão-ouro os bancos de pares verdadeiros, originados inicialmente no processo de relacionamento, que possuíam como identificador único de cada indivíduo, o CNS e/ou CPF (TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2014).

Para avaliar a qualidade desses métodos, são utilizados parâmetros como sensibilidade e especificidade e a curva ROC (*Receiver Operating Characteristics*) para ilustrar o desempenho e, menos frequentemente, para conjuntos de dados desequilibrados, a curva *precision-recall* e a medida F (SILVEIRA e ARTMANN, 2009; CAMARGO JR e COELI, 2000; HARRON et al., 2017; CAPUANI et al., 2014; SAITO e REHMSMEIER, 2015).

Valor preditivo positivo (*precision*) é definido como a proporção de pares formados que estão corretos. Sensibilidade (*recall*) é a proporção de todos os pares verdadeiros que foram vinculados corretamente (BOYD et al., 2016). Tais medidas podem ser estimadas adotando-se como padrão-ouro a revisão manual (CAMARGO JR e COELI, 2000; HARRON et al., 2017; CAPUANI et al., 2014).

Outro parâmetro utilizado para avaliar a qualidade do *linkage* é a medida F, uma média harmônica dos parâmetros *precision* e *recall*. Altos valores de *precision* e de *recall* levam a uma medida F elevada, que permite a definição de um escore de alto poder discriminatório na identificação de pares verdadeiros (FERRANTE e BOYD, 2012; BOYD; ENG; PAGE, 2013).

Esse estudo tem como objetivo testar e avaliar a acurácia de um método para seleção de escore em relacionamento probabilístico de banco de dados, a partir do qual seja possível identificar de forma automatizada todos os pares verdadeiros, dispensando a inspeção manual de todos os pares formados.

METODOLOGIA

Estudo de acurácia de um método automatizado para a classificação dos pares formados a partir do relacionamento probabilístico, realizado com dados do Sistema de Informação do Câncer de Mama (SISMAMA) de Minas Gerais. O banco de dados para o relacionamento foi construído a partir de exames alterados (BI-RADS® 4, 5 e 6) do ano de 2009 e este foi comparado com o ano de 2010. Foram selecionados apenas os exames de mulheres residentes em Minas Gerais com 18 anos ou mais.

A padronização das bases de dados consistiu em remoção de acentos, cedilhas, espaços duplos, caracteres especiais tais como interrogação, exclamação e outros. Em seguida foram excluídos os exames duplicados (mesmo nome, nome da mãe, data de nascimento e data de exame) e repetidos (mesmo nome, nome da mãe e data de nascimento) com a finalidade de manter apenas um registro de cada indivíduo.

O relacionamento probabilístico foi realizado em dezesseis passos e as chaves de blocagem e pareamento de cada um dos passos foram adaptadas de um estudo que usou os dados do SISMAMA (TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2018) e seguiu as recomendações do manual do RecLinkIII (CAMARGO JR e COELI, 2006) (Tabela 1).

Tabela 1 – Passos, Variáveis de blocagem e pareamento, escores, pares formados e pares verdadeiros do relacionamento dos bancos do SISMAMA 2009 e 2010.

Passos	Variáveis de blocagem	Variáveis de pareamento	Escore	Pares formados	Pares verdadeiros
1	PN + UN + AN	N+DN+M+MUN	-13 a 21	2.108	1.076
2	PN + UN + AN	N+DN	-6 a 10	6.067	1.069
3	CNS	N + DN + M	-10 a 17	216	216
4	PN + UN + AN + PM + UM	N + DN + M	-10 a 17	932	923
5	PN + UN + AN	N + DN + M	-10 a 17	1.373	1.069
6	PN + AN + PM + UM	N + DN + M	-10 a 17	1.067	983
7	PN + AN + UM	N + DN + M	-10 a 17	1.188	1.022
8	PN + AN + PM	N + DN + M	-10 a 17	1.223	1.075
9	PN + AN	N + DN + M	-10 a 17	2.620	1.142
10	PN + UN + PM + UM	N + DN + M	-10 a 17	991	950
11	PN + UN + PM	N + DN + M	-10 a 17	1.181	1.042
12	PN + UN	N + DN + M	-10 a 17	2.143	1.106
13	PN + PM + UM	N + DN + M	-10 a 17	1.286	1.012
14	PN + UM	N + DN + M	-10 a 17	1.591	1.053
15	UN + AN	N + DN + M	-10 a 17	1.919	1.080
16	AN	N + DN + M	-10 a 17	12.581	1.145

AN: Ano de Nascimento; CNS: Cartão Nacional do SUS; DN: Data de Nascimento; M: Mãe; MUN: Município; N: Nome; PM: Código Soudex do Primeiro Nome da Mãe; PN: Código Soudex do

Primeiro Nome; UM: Código Soundex do Último Nome da Mãe; UN: Código Soundex do Último nome.

Diferente de outros estudos (MIGOWSKI et al., 2011; TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2018) nos quais a cada passo foram mantidos apenas registros classificados como não par, no presente estudo todos os registros foram testados em todos os passos. Além disso, esses estudos utilizaram como padrão-ouro dados gerados no primeiro passo, cuja chave de bloqueio utilizava identificador único (CPF e/ou CNS) (TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2018) e revisão manual (MIGOWSKI et al., 2011). No presente estudo, cada passo foi inspecionado manualmente com a finalidade de se estabelecer um padrão-ouro de ponto de corte para os pares verdadeiros. Os critérios para definição de pares verdadeiros foram baseados em Girianelli e colaboradores (2008).

De cada passo foram extraídas três diferentes amostras aleatórias de pares, compostas por 100, 200 e 300 pares, respectivamente, para posterior revisão manual e comparação com o padrão-ouro. Para a definição do tamanho da amostra foi utilizada a função `power.roc.teste` do pacote `pROC` do programa R, que estimou que um tamanho amostral de cerca de 100 registros seria suficiente. Para fins de comparação foram testadas ainda amostras com 200 e 300 pares.

Para avaliar a acurácia do método em cada passo e nas amostras foram utilizadas a sensibilidade (*recall*) e valor preditivo positivo (*precision*) através da curva *precision-recall* e da medida F, que possibilitou a determinação do escore de maior poder discriminatório na identificação dos pares verdadeiros (SAITO e REHMSMEIER, 2015). Tal análise utilizou os pacotes `PRROC` e `ROCR` do programa R.

Por fim, foi feita a combinação de todos os passos para encontrar o número total de pares verdadeiros. Procedeu-se então, à combinação desses passos entre si para investigar quais passos em conjunto abrangem maior número de pares verdadeiros. O programa R foi utilizado na limpeza dos dados, remoção das duplicidades e repetições, cálculo das amostras e construção da curva *precision-recall* e o programa `RecLink III` (GIRIANELLI; THULER; SILVA, 2008) no relacionamento e também na remoção das duplicidades e repetições.

Este projeto foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora sob o número de parecer 1.431.916.

RESULTADOS

Inicialmente os bancos 2009 e 2010 tinham 165.115 e 490.535 registros, respectivamente. Do número total de observações encontradas na base referência (2009) 3.030 foram classificadas como BI-RADS® 4, 5 e 6. Após a execução do processo de limpeza permaneceram nos bancos 2.764 (2009) e 445.531(2010) registros.

No relacionamento foi constatado que o passo cuja variável de blocagem era ano de nascimento e as variáveis de pareamento eram nome, data de nascimento e nome da mãe (passo 16), formou maior número de pares verdadeiros enquanto o passo que utilizava o CNS como variável de blocagem e as mesmas várias de pareamento (passo 3), formou o menor número de pares verdadeiros (Tabela 1). Foi observado que o escore do relacionamento é influenciado pelo número de variáveis utilizadas no pareamento. Após a combinação de todos os passos chegou-se ao número total de 1.194 pares verdadeiros. Foi constatado que com apenas três passos é possível encontrar 99,24% dos pares verdadeiros e que a partir de quatro passos poucos pares verdadeiros são acrescentados, no caso desse estudo um a três pares a cada passo incluído (Tabela 2).

Tabela 2 – Combinação de pares do relacionamento dos bancos Sistema de Informação do Câncer de Mama (SISMAMA), Minas Gerais, Brasil, 2009 e 2010.

Passos	Número de Pares	% (Total=1.194)
16, 9	1.151	96,40
16, 9, 12	1.185	99,24
16, 9, 12, 8	1.185	99,24
16, 9, 12 e 1	1.188	99,50
16, 9, 12, 1 e 15	1.188	99,50
16, 9, 12, 1 e 2	1.188	99,50
16, 9, 12, 1 e 5	1.188	99,50
16, 9, 12, 1 e 14	1.190	99,67
16, 9, 12, 1, 14 e 11	1.192	99,83
16, 9, 12, 1, 14, 11 e 7	1.193	99,92
16, 9, 12, 1, 14, 11, 7 e 13	1.194	100,00

No que se refere ao ponto de corte para classificação de pares verdadeiros, a maioria dos passos possui escore acima de 6,5 e as amostras com 200 e 300 pares apresentaram valores bem próximos a estes, variando entre 6,5 e 6,8. Houve pouca diferença nos escores das amostras com 200 e 300 pares (Tabela 3).

Tabela 3 – Ponto de corte dos passos e amostras do Relacionamento dos bancos do Sistema de Informação do Câncer de Mama (SISMAMA), Minas Gerais, Brasil, 2009 e 2010.

Passo	Ponto de Corte			
	Todos os pares	A1	A2	A3
1	9.670	14,50	10,42	10,09
2	9.734	2,50	9,73	9,76
3	-	-	-	-
4	9.437	15,50	6,80	6,28
5	6.553	20,50	6,80	6,62
6	6.553	17,50	6,64	6,62
7	6.553	12,50	6,64	6,60
8	6.553	11,50	6,80	6,55
9	6.553	2,50	6,60	6,62
10	12.690	14,50	12,82	12,79
11	6.627	19,50	0,63	6,64
12	6.553	9,50	6,64	6,62
13	6.553	10,50	6,62	6,62
14	6.553	15,50	6,64	6,64
15	6.553	10,50	6,62	6,55
16	6.553	2,50	6,80	6,80

P: Passo; T: Total de pares (n=1.194); A1: Amostra de 100 pares; A2: Amostra de 200 pares; A3: Amostra de 300 pares.

Todos os pares formados no passo três, cuja variável de blocagem era CNS e de pareamento nome, data de nascimento e nome da mãe, eram verdadeiros e apenas três registros tinham baixos escores. A inexatidão encontrada em um dos pares estava no último nome e mês de nascimento; em outro o nome do pai foi registrado no local destinado para registro do nome da mãe; e o terceiro registro a letra M do nome e do nome da mãe foi trocada por N e o último sobrenome era diferente, contudo, as demais variáveis eram iguais: data de nascimento, endereço, código do município e CNS.

Todos os passos e amostras com 200 e 300 pares apresentaram alta sensibilidade (*recall*) >0,97, alto valor preditivo positivo (*precision*) >0,95 e altas acurácia (> 0,97), medida F (>0,96) e área sob a curva *precision-recall* (>0,98). A maioria das amostras de 100 pares evidenciou altos valores para essas medidas, porém com escores mais baixos. Somente o passo cuja variável de blocagem foi ano de nascimento e de comparação nome, data de nascimento e nome da mãe (passo 16) apresentou baixa sensibilidade (*recall*), medida F e área sob a curva *precision-recall* (Tabelas 4 e 5).

Tabela 4 – Recall e precision do ponto de corte dos passos e amostras do relacionamento dos bancos do Sistema de Informação do Câncer de Mama (SISMAMA), Minas Gerais Brasil, 2009 e 2010.

Passo	Recall				Precision			
	T	A1	A2	A3	T	A1	A2	A3
1	0,98	0,98	1,00	0,99	0,98	0,80	1,00	0,99
2	0,98	1,00	0,97	0,98	0,98	1,00	0,97	0,98
3	-	-	-	-	-	-	-	-
4	0,99	1,00	1,00	0,99	0,95	0,93	1,00	0,99
5	0,99	1,00	0,98	0,99	0,99	0,85	0,98	0,99
6	0,98	1,00	0,98	0,99	0,98	0,93	0,98	0,99
7	0,98	1,00	0,98	1,00	0,98	0,90	0,98	0,99
8	0,98	1,00	0,99	0,97	0,98	0,82	0,98	0,97
9	0,97	1,00	0,98	0,97	0,97	1,00	0,98	0,97
10	0,99	1,00	1,00	1,00	0,94	0,92	0,95	0,96
11	0,98	1,00	0,99	0,99	0,98	0,90	0,99	0,97
12	0,98	1,00	0,98	0,98	0,98	0,92	0,99	0,98
13	0,98	1,00	0,99	0,99	0,98	0,90	0,98	0,99
14	0,97	1,00	1,00	1,00	0,97	0,80	0,98	0,97
15	0,99	1,00	0,99	0,99	0,99	0,74	0,99	0,99
16	0,98	0,00	1,00	1,00	0,98	1,00	1,00	1,00

P: Passo; T: Total de pares (n=1.194); A1: Amostra de 100 pares; A2: Amostra de 200 pares; A3: Amostra de 300 pares.

Tabela 5 – Acurácia, medida F e área sob a curva *precision-recall* (AUPRC) do ponto de corte dos passos e amostras do relacionamento dos bancos do SISMAMA 2009 e 2010.

P	Acurácia				Medida F				AUPRC			
	T	A1	A2	A3	T	A1	A2	A3	T	A1	A2	A3
1	0,99	0,99	1,00	0,99	0,99	0,99	1,00	0,99	0,99	0,99	1,00	0,99
2	0,99	0,89	0,98	0,99	0,97	0,71	0,96	0,98	0,99	0,42	0,98	0,98
3	-	-	-	-	-	-	-	-	-	-	-	-
4	0,99	0,97	1,00	0,99	0,99	0,98	1,00	0,99	0,99	0,92	1,00	0,99
5	0,99	0,87	0,97	0,97	0,99	0,92	0,98	0,99	0,99	0,79	0,99	0,99
6	0,97	0,91	0,99	0,99	0,99	0,95	0,99	0,99	0,99	0,81	0,99	0,99
7	0,98	0,82	0,98	0,98	0,99	0,89	0,98	0,99	0,99	0,70	0,99	0,99
8	0,97	0,86	0,99	0,97	0,99	0,92	0,99	0,98	0,99	0,79	0,99	0,99
9	0,97	0,73	0,98	0,99	0,98	0,72	0,98	0,98	0,99	0,46	0,99	0,99
10	0,97	0,95	0,98	0,97	0,99	0,97	0,98	0,98	0,99	0,87	0,99	0,99
11	0,97	0,90	0,98	0,97	0,98	0,94	0,99	0,98	0,99	0,81	0,99	0,99
12	0,97	0,58	0,96	0,98	0,97	0,72	0,96	0,98	0,99	0,41	0,99	0,99
13	0,97	0,77	0,97	0,97	0,98	0,87	0,98	0,98	0,99	0,62	0,99	0,99
14	0,97	0,70	0,98	0,98	0,97	0,81	0,98	0,98	0,99	0,63	0,99	0,99
15	0,99	0,72	0,98	0,99	0,99	0,77	0,98	0,99	0,99	0,62	0,99	0,99
16	0,99	0,95	1,00	0,99	0,98	0,46	1,00	0,98	0,99	0,32	1,00	0,99

P: Passo; T: Total de pares (n=1.194); A1: Amostra de 100 pares; A2: Amostra de 200 pares; A3: Amostra de 300 pares.

DISCUSSÃO

O presente estudo demonstrou que é possível automatizar o relacionamento dos bancos de dados mantendo a acurácia do método, com base na inspeção manual de uma amostra retirada do próprio banco e posterior seleção de um escore de alto poder discriminatório na identificação de pares verdadeiros dessa amostra.

Alguns estudos utilizaram escores para o limiar superior e inferior (TOMAZELLI; GIRIANELLI; AZEVEDO e SILVA, 2014; PERES et al., 2015; ROMERO et al., 2016), indicando para a inspeção os pares com escores intermediários, porém isso pode não ser aplicável, quando existe um grande número de pares a serem revisados. Desse modo, recomendamos amostragem aleatória com no mínimo 200 registros, inspeção manual, cálculo da sensibilidade (*recall*), do valor preditivo positivo (*precision*), da medida F, construção da curva *precision-recall* dos escores dessa amostra e posterior aplicação desse ponto de corte para classificação automática dos pares do banco total.

Diversos estudos (COUTINHO e COELI, 2007; PERES et al., 2015; FONSECA et al., 2010) sobre acurácia do linkage encontraram medidas de qualidade semelhantes às do presente estudo tanto para o total de cada passo quanto para as amostras de 200 e 300 pares: sensibilidade >95%, valor preditivo positivo >95% e acurácia >97%.

Por meio da aplicação de três passos foi verificado que é possível encontrar quase a totalidade de pares verdadeiros e que a inclusão de mais passos agrega poucos pares. O estudo realizado por Girianelli e colaboradoras (2018) propõe a utilização de quinze passos, contudo dependendo do tamanho da base e da disponibilidade de tempo não é possível o uso de todas essas estratégias. As autoras referem ainda que com quatro passos é possível identificar a maioria dos pares verdadeiros e que ao utilizar a chave de blocagem primeiro nome, último nome, primeiro nome da mãe e último nome da mãe as autoras conseguiram recuperar pares que não tinham sido identificados nos passos anteriores. No presente estudo, o passo que usa as chaves de blocagem primeiro nome e ano de nascimento (passo 9) compartilha 97,2% de pares verdadeiros com o passo que usa as chaves primeiro nome, último nome, primeiro nome da mãe e último nome da mãe (passo 10).

O uso de maior número de variáveis na blocagem e pareamento, como por exemplo nos passos quatro, seis e dez, aumenta o número de ligações verdadeiras (>90%) e restringe as falsas (<8%). Em contrapartida, estratégias poucas restritas, como nos passos dois e

dezesseis, aumentam muito o número de pares formados e, conseqüentemente, eleva a quantidade de falsos pares.

A utilização de um identificador único como chave de bloqueio contribui para que a maior parte ou totalidade dos pares formados sejam verdadeiros, como verificado no presente estudo que com o CNS (passo 3) formou 216 pares, todos verdadeiros. Os escores baixos apresentados por pares desse passo podem ser explicados pela incompletude de informações. Em um relacionamento entre banco de doadores de sangue e o Sistema de Informação de Mortalidade (SIM) foram encontrados dezesseis pares verdadeiros em baixos escores, sendo observado que eles apresentavam dados incompletos, principalmente para a variável nome (CAPUANI et al., 2014). Esse mesmo estudo obteve sensibilidade de 94%.

Variáveis importantes na identificação do indivíduo, tais como: nome, nome da mãe e data de nascimento devem estar muito bem preenchidas nos bancos. Estudo sobre a definição de um ponto de corte para identificação de pares verdadeiros em registros de câncer usou três variáveis na bloqueio e duas no pareamento, encontrando nos relacionamentos realizados, especificamente para o câncer de mama, 47,5% e 78,8% pares falsos aproximadamente (PERES et al., 2015). Destaca-se que a qualidade da informação e sua completude são importantes fatores que influenciam o processo de formação de pares e que o escore é influenciado pelas variáveis selecionadas para o pareamento (SILVEIRA e ARTMANN, 2009). Elevado percentual de registros sem informação de nome da mãe no Sistema de Informação de Nascidos Vivos (SINASC) resultou em um relacionamento com baixa sensibilidade (DA MATTA COUTINHO et a., 2008).

Relacionamentos probabilísticos tem como objetivo obter o maior número de pares verdadeiros possíveis, mesmo que para isso sejam obtidos alguns falsos pares. Dessa forma, o interesse está principalmente na sensibilidade do método (HARRON et al., 2017). Na escolha do ponto de corte também pode ser considerada a máxima *precision* com a finalidade de minimizar a possibilidade de fazer correspondências falsas e, se necessário, aceitar algum nível de falhas para encontrar correspondências verdadeiras (CAPUANI et al., 2014).

Atualmente o grande volume de informações sociodemográficas, de saúde e de outras áreas, desperta a necessidade de agregação desses dados, para que os gestores desses setores tenham conhecimento da realidade e possam definir prioridades. Assim, o relacionamento de bases de dados ganha destaque e importância como ferramenta de integração dos diversos sistemas de informação, tornando-se essencial a otimização destes processos.

As limitações do estudo são inerentes ao uso de bases de dados secundários devido a sua qualidade variável, podendo conter incompletitudes e erros de digitação. Além disso cabe assinalar que a utilização da revisão manual como padrão-ouro pode produzir erros em cenários em que os revisores não estejam preparados adequadamente ou estejam submetidos a rotinas fatigantes, o que pode proporcionar classificações equivocadas entre pares verdadeiros e falsos.

Destaca-se que o processo de limpeza das bases de dados e retirada de duplicidades e repetições, que antecede a etapa de relacionamento, é fundamental para que sejam mantidas no banco apenas as variáveis de interesse, um registro de cada indivíduo e para a eliminação de caracteres especiais, facilitando o processo de ligação de pares.

Os resultados apresentados dão suporte aos gestores de saúde e àqueles que utilizam o relacionamento de bancos de dados, para que com base no seu próprio banco identifiquem, a partir de uma amostra selecionada aleatoriamente, o escore com melhor desempenho na identificação de pares verdadeiros e com mínimo possível de perdas, o que permite automatizar o relacionamento dos bancos mantendo a acurácia do método.

REFERÊNCIAS

BOYD, K.; ENG, K. H.; PAGE, C. D. Área sob a curva de recuperação de precisão: estimativas pontuais e intervalos de confiança. In: BLOCKEEL, H.; KERSTING, K.; NIJSSEN, S.; ŽELEZNÝ, F. Aprendizado de Máquina e Descoberta de Conhecimento em Bancos de Dados. ECML PKDD 2013. Notas de Aula em Ciência da Computação, vol 8190. Springer, Berlin, Heidelberg; Disponível em: https://link.springer.com/content/pdf/10.1007%2F978-3-642-40994-3_29.pdf. Acesso em jun 2019.

BOYD, J. H.; et al. A simple sampling method for estimating the accuracy of large-scale record linkage projects. **Methods Information in Medicine**, v. 55, n. 3, p. 276-283, 2016.

CAMARGO JR., K. R. DE; COELI, C. M. RecLink 3: nova versão do programa que implementa a técnica de associação probabilística de registros (probabilistic record linkage). **Cadernos de Saúde Coletiva**, v. 14, n. 2, p. 399-404, 2006.

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

CAPUANI, L. et al. 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.

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, 2007.

DA MATTA COUTINHO, R. G.; COELI, C. M.; FAERSTEIN, E.; CHOR, D. Sensitivity of probabilistic record linkage for reported birth identification: Pró-Saúde Study. **Revista de Saúde Pública**, v. 42, n. 6, p. 1097–1100, 2008.

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.

FONSECA, M. G. P. et al. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. **Cadernos de Saúde Pública**, v. 26, n. 7, p. 1431–1438, 2010.

GIRIANELLI, V. R.; THULER, L. C. S.; SILVA, G. A. E. Qualidade do sistema de informação do câncer do colo do útero no estado do Rio de Janeiro. **Revista de Saúde Pública**, v. 43, n. 4, p. 580–588, 2008.

HARRON, K. L.; et al. A guide to evaluating linkage quality for the analysis of linked data. **International Journal of Epidemiology**, v. 46, n.5, p. 1699–1710, 2017.

MIGOWSKI, A.; et al. Acurácia do relacionamento probabilístico na avaliação da alta complexidade em cardiologia. **Revista de Saúde Pública**, v. 45, n. 2, p. 269-275, 2011.

PERES, S. V. et al. 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. 428–436, 2015.

ROMERO, R. O. G. et al. Subnotificação de casos de tuberculose a partir da vigilância do óbito. **Revista Eletrônica de Enfermagem**, v. 18, 2016. Disponível em: <<http://dx.doi.org/10.5216/ree.v18.37249>>. Acesso em 04 abril 2018.

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.

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. 875-882, 2009.

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, n. 0, 2018.

6.2 ARTIGO 2: COBERTURA DO RASTREAMENTO MAMOGRÁFICO ANUAL EM MINAS GERAIS A PARTIR DE DADOS IDENTIFICADOS DO SISMAMA

Resumo

Objetivo: Estimar a cobertura do rastreamento mamográfico em Minas Gerais e avaliar fatores associados. **Método:** Estudo ecológico que analisou 77 microrregiões de saúde de Minas Gerais utilizando dados identificados da base de dados do SISMAMA e indicadores sociodemográficos e de saúde de 2010. Após limpeza e padronização dos dados, foram removidas duplicidades e repetições pelos métodos determinístico e probabilístico e calculada a cobertura do rastreio. A associação entre a cobertura mamográfica e as demais variáveis foi analisada por modelos de regressão linear. **Resultados:** A partir da análise de 453.587 mulheres rastreadas observou-se que as maiores coberturas ocorreram nas microrregiões Centro e Sul e as menores nas Norte e Nordeste. Foi encontrada associação positiva entre a cobertura e renda, índice de Gini, número de mamógrafos e sua utilização. **Considerações Finais:** Conhecer a cobertura mamográfica permite uma avaliação do programa de rastreamento e pode contribuir para a otimização do acesso nas áreas onde existem desigualdades socioeconômicas e de saúde.

Palavras-chave: Cobertura de Mamografia. Disparidades em Assistência à Saúde. Fatores Socioeconômicos.

Abstract

Objective: to estimate mammography screening coverage and evaluate associated factors in Minas Gerais. Method: ecological study analyzing 77 health micro-regions from Minas Gerais using data from SISMAMA database and sociodemographic and health indicators from 2010. Deterministic and probabilistic methods removed duplicates and repetitions and calculated screening coverage. Associations between mammography coverage and other variables were analyzed using linear regression model. Results: analysis of 453.587 screened women found high coverage in central and south micro-regions and low coverage in north and northeast. Positive associations were found between coverage and income, Gini Index, and number and use of mammograms. Final Considerations: Knowing mammography coverage allows assessing the screening program and may optimize access in areas with socioeconomic and health inequities.

Keywords: Mammography coverage. Health Care Disparities. Socioeconomic Factors.

INTRODUÇÃO

No sexo feminino, o câncer de mama é a neoplasia mais incidente em 159 países, enquanto figura como a principal causa de mortalidade em 110 países. No Brasil, para 2020 foram estimados 20.725 óbitos de mulheres por esta causa e 88.492 casos novos (SUNG et al., 2021).

A mamografia de rastreamento é considerada um dos métodos mais efetivos na detecção precoce, diagnóstico e tratamento em tempo oportuno. No Brasil, segundo as diretrizes do Ministério da Saúde (INCA, 2015), deve ser realizada a cada dois anos, em mulheres assintomáticas, na faixa etária de 50 a 69 anos. Contudo, o rastreamento é considerado oportunístico, ou seja, o exame só é solicitado quando a mulher procura o serviço de saúde ou em datas específicas, como por exemplo, o “outubro rosa”, o que resulta em poucos benefícios e torna o rastreio mais oneroso (SANTOS-SILVA, et al., 2014).

No Brasil, utiliza-se um *proxy* da cobertura de mamografia expresso pela razão entre o número de exames realizados (extraídos do SISMAMA e, algumas vezes, do SIA), e a metade da população feminina no período, que pode fornecer valores superestimados, gerando a falsa sensação de população coberta (INCA, 2021). Para verdadeira cobertura é ideal que no numerador seja utilizado o número de mulheres rastreadas e não de exames.

A detecção precoce do câncer de mama no Brasil, compreende ações de rastreamento, direcionado às mulheres assintomáticas e de diagnóstico precoce em mulheres sintomáticas. No entanto, apesar de implantado desde 2004, o rastreamento com mamografia não é efetivo e grande proporção de mulheres são diagnosticadas com câncer de mama em estágios avançados (SANTOS-SILVA, et al., 2014).

A quantidade de mamografias de rastreamento realizadas é inferior à capacidade de produção dos mamógrafos (TOMAZELLI e AZEVEDO e SILVA, 2017) e quase metade dos exames são realizadas em faixas etárias não recomendadas para o rastreamento pelo Ministério da Saúde (AZEVEDO e SILVA et al., 2014). No país, de 2010 a 2011, a cobertura de mulheres na faixa etária de 50 a 69 anos não atingiu a meta considerada ideal pela Organização Mundial de Saúde e diretrizes europeias (70%-75%), sendo observada maior cobertura nas regiões Sul e Sudeste e menor no Norte e Nordeste (NOGUEIRA et al., 2019).

Estudos apontam que nas áreas de menor nível socioeconômico e alta vulnerabilidade na saúde, as coberturas são menores, há menor disponibilidade de equipamentos e recursos humanos e as mulheres percorrem as maiores distâncias para

realização de mamografia (TOMAZELLI e AZEVEDO e SILVA, 2017; NOGUEIRA et al., 2019).

Minas Gerais é um estado brasileiro com grande extensão territorial e apresenta desigualdades regionais referentes aos aspectos demográficos e socioeconômicos (ALVES; MAGALHÃES; COELHO, 2017). Por retratar características geográficas, estruturais, econômicas e de saúde semelhantes às do país, as pesquisas realizadas no estado podem apontar importantes situações sociais, econômicas e de saúde que ocorrem no país, demonstrando as áreas que necessitam de maior atenção dos gestores. Nessa perspectiva, o objetivo deste estudo foi estimar, a partir de dados identificados do SISMAMA, a cobertura do rastreamento mamográfico em Minas Gerais por macro e microrregiões de saúde e avaliar fatores sociodemográficos e de saúde que podem contribuir para a variação regional da cobertura da mamografia de rastreamento.

MÉTODO

Estudo ecológico que utilizou dados identificados do Sistema de Informação do Câncer de Mama (SISMAMA), cedidos pela Secretaria de Saúde do Estado de Minas Gerais. O estudo teve como unidade de análise as 77 microrregiões de saúde de Minas Gerais, agrupadas em 13 macrorregiões (figura 1). Foram utilizados dados identificados do Sistema de Informação do Câncer de Mama (SISMAMA), cedidos pela Secretaria de Saúde do Estado de Minas Gerais, referente ao período de 2010-2012.

Utilizando a variável data de exame (D_RECEBE), foram selecionados e extraídos dos arquivos anuais de 2010 a 2012, os exames realizados no ano de 2010 com idade igual ou maior a 18 anos e residentes em Minas Gerais. Os dados foram padronizados, sendo removidos acentos, cedilhas, espaços duplos e caracteres especiais. Para construir a base referência com um único registro por mulher, foram removidas duplicidades e repetições de exames (seguimento) pelos métodos determinístico e probabilístico, sendo utilizados os Programas R (<http://www.r-project.org/>) e *Reclink* III (<http://reclink.sourceforge.net/>).

As variáveis empregadas no método determinístico foram nome, nome da mãe, data de nascimento, código do município, data de solicitação do exame (D_EXAME) e número do CNES da unidade requisitante, sendo as duas últimas variáveis suprimidas na remoção de repetições. O método probabilístico foi realizado em duas etapas utilizando as mesmas variáveis do método determinístico. Na segunda etapa a variável nome foi substituída pelo

nome da mãe, considerando alterações no sobrenome das mulheres devido mudanças do estado civil.



Figura 1 – Macrorregiões e Microrregiões de Saúde de Minas Gerais, 2010.

Com a base referência criada, foi estimada para cada microrregião a cobertura de mamografia para a faixa etária alvo dividida da seguinte forma: 50-59, 60-69 e 50-69 anos. Como as diretrizes definem que a mamografia deve ser bienal, para calcular a cobertura de mamografia, o número de mulheres examinadas foi dividido pela metade da população feminina não beneficiárias de plano de saúde, número obtido ao se subtrair as mulheres beneficiárias de planos de saúde da população feminina total de cada microrregião. Ressalta-se que o número de mulheres com plano de saúde é estimado, uma vez que não é possível identificar quais mulheres possuem plano de saúde. Essas informações foram extraídas da Agência Nacional de Saúde Suplementar – ANS (<http://www.ans.gov.br/>) e Datasus (<http://www2.datasus.gov.br>). O uso de metade da população é justificada pela recomendação de mamografia bianual para seguimento após exame normal (INCA, 2021).

As variáveis de exposição foram os seguintes indicadores sociodemográficos e assistenciais: índice de desenvolvimento humano nas regiões de saúde (IDHr), renda domiciliar per capita média, índice de vulnerabilidade na saúde (IVS), índice de Gini, grau de urbanização, analfabetismo, tamanho da população, razão de médicos, médicos radiologistas e enfermeiros por 1.000 habitantes, razão de mamógrafos por x habitantes e razão de mamografias por mamógrafos (utilização), distância média percorrida para realização de mamografia dentro e fora da microrregião de saúde (DPRM). Tais variáveis foram selecionadas de acordo com a revisão de literatura e a disponibilidade em bases de acesso público com a finalidade de se desenvolver um estudo exploratório sobre a associação entre a cobertura do rastreamento e características das microrregiões de saúde de Minas Gerais.

A capacidade utilizada do mamógrafo é uma referência utilizada pelo INCA para estimar a necessidade de equipamento para atendimento à demanda e aponta possíveis problemas na sua distribuição e utilização. O Instituto considera que em um turno de 8 horas diárias de trabalho é possível realizar 3 exames por hora, em 22 dias úteis e 12 meses por ano, perfazendo um total de 6.336 mamografias. Assim, o INCA define que durante um ano uma capacidade de utilização de 80% (5.069 exames) é ideal (INCA, 2015c).

Informações sobre os indicadores sociodemográficos e assistenciais foram coletadas das páginas eletrônicas do Datasus (<http://www2.datasus.gov.br>), do Programa de Avaliação do Desempenho do Sistema de Saúde – PROADESS (<https://www.proadess.icict.fiocruz.br/>) e do Instituto Brasileiro de Geografia e Estatística -

IBGE (<https://ww2.ibge.gov.br>). Informações não disponíveis foram solicitadas via sítio eletrônico para acesso à informação do governo federal (<http://www.acessoinformacao.gov.br/>).

A caracterização da distribuição das coberturas e dos indicadores foi realizada por meio da estatística descritiva e expressas em percentuais e por meio de mapas temáticos. Foi gerado também o mapa do índice de Moran Local (Lisa map), que permite a comparação de uma microrregião com seus vizinhos, identificando padrões espaciais localizados, por exemplo agrupamento de regiões com valores de atributos semelhantes (clusters de regiões) dos tipos alto-alto, baixo-baixo, alto-baixo, baixo alto e possíveis outliers.

Análises de regressão linear foram realizadas com a finalidade de identificar associações entre as coberturas estimadas e fatores sociodemográficos e de saúde. Na primeira etapa foi realizada a regressão linear simples.

Em seguida, para cada modelo da regressão múltipla foram inseridas as variáveis explicativas de acordo com o grupo de exposição ao qual pertenciam, sendo seguido o critério do grupo distal para o proximal.

Após o ajuste os resíduos de cada modelo foram investigados para detectar evidências de heterocedasticidade e/ou autocorrelação espacial (respectivamente com os testes de Breusch-Pagan e de Moran I). Um dos modelos apresentou heterocedasticidade e para corrigir esse fenômeno estatístico, foi inserida a variável população classificada segundo o porte populacional, a saber: pequeno até 100.000 habitantes; médio de 100.001 a 300.000; e grande >300.000.

Este estudo é parte do projeto de pesquisa “Desigualdades de acesso às ações de controle de câncer através dos Sistemas de Informações de Saúde (DAAC-SIS)” e foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora sob o parecer nº. 1.431.916.

RESULTADOS

Em 2010 foram registrados, nas bases do SISMAMA, 502.394 exames de mamografia no estado de Minas Gerais,. Após o processo de retirada das duplicidades e repetições chegou-se ao número de 453.587 mulheres examinadas. Dessas 51,6% pertenciam a população alvo (50-69 anos), enquanto que 6,4% tinham menos de 40 anos, 35,8% de 40 a 49 anos e 6,2% 70 anos ou mais. Pará de Minas, Diamantina e Minas

Novas/Turmalina/Capelinha, microrregiões do Oeste e Jequitinhonha apresentaram menor percentual de exames na faixa etária de 50 a 69 anos. E Manga, Resplendor e Coronel Fabriciano/Timóteo, microrregiões do Norte e Leste apresentaram os maiores percentuais.

A cobertura estimada para a população alvo do rastreamento mamográfico (50-69 anos) no estado foi de 36,3%. A faixa etária de 50-59 anos apresentou maior cobertura (40,0%) quando comparada à faixa etária de 60-69 anos (30,5%). Na tabela 1, são apresentadas a cobertura por mamografia e os indicadores socioeconômicos e de saúde referente às macrorregiões de saúde de Minas Gerais.

Tabela 1 – Caracterização da cobertura de mamografias em mulheres nas faixas etárias de 50-59, 60-69 e 50-69 anos e dos indicadores socioeconômicos e de saúde das 77 microrregiões de saúde de Minas Gerais, Brasil, 2010.

Variável (n=77)	N	M	DP	MIN	Q1	MED	Q3	MAX
Cobertura (%)								
50 a 69 anos	233.903	32,32	14,51	2,66	18,85	33,91	43,25	61,70
50 a 59 anos	158.246	35,76	15,31	3,06	23,27	37,88	46,61	67,62
60 a 69 anos	75.657	26,97	13,31	2,10	15,64	28,50	36,29	52,83
Indicadores Sociodemográficos								
IDHr		0,67	0,04	0,56	0,64	0,68	0,70	0,74
Renda (R\$)		561,80	178,47	238,50	427,90	590,30	671,40	1205,10
IVS		0,38	0,21	0,14	0,22	0,29	0,50	0,90
Índice de Gini		0,48	0,02	0,41	0,46	0,47	0,49	0,55
Grau de								
Urbanização (%)		78,84	13,43	42,46	70,40	83,28	89,49	99,66
Analfabetismo (%)		10,89	6,28	3,40	6,30	7,90	15,20	26,30
Indicadores Assistenciais								
Razão de Médicos*		0,95	0,45	0,32	0,62	0,87	1,14	2,48
Razão de Enfermeiros*		0,58	0,12	0,38	0,51	0,56	0,66	0,95
Médicos Radiologistas**		1,82	1,66	0,00	0,00	1,63	2,66	7,22
Número de Mamógrafos**		1,17	0,82	0,00	0,60	1,10	1,60	3,50
Utilização de Mamógrafos***		1954	1668,71	0	906	1505	2282	9762
DPRM dentro da região (Km)		28,49	21,46	-1,00	16,5	26,50	43,30	86,50
DPRM fora da região (Km)		135,50	110,91	15,30	64,60	98,80	178,40	522,70

Fonte: SISMAMA/Datasus/MS, Censo Demográfico 2010, PROADESS 2010.

*por 1.000 habitantes; **por 100.000 habitantes; *** n° mamografias/n° mamógrafos

N: Número de Mulheres; M: média; DP: desvio padrão; MIN: valor mínimo; Q1: primeiro quartil; MED: mediana; Q3: terceiro quartil; MAX: valor máximo; CV: coeficiente de variação (=M*DP/100).

Em todas as faixas etárias analisadas, foi observado nas macros Noroeste, Norte e Nordeste a formação de um grande cluster do tipo baixo-baixo, ou seja agrupamento de microrregiões com menor cobertura. Unaí, João Pinheiro e Montes Claros/Bocaiúva foram as únicas microrregiões que não participaram desse cluster. Nas macros Triângulo, Oeste, Sul, Centro e Leste do Sul formou-se um cluster menor do tipo alto-alto (figura 2).

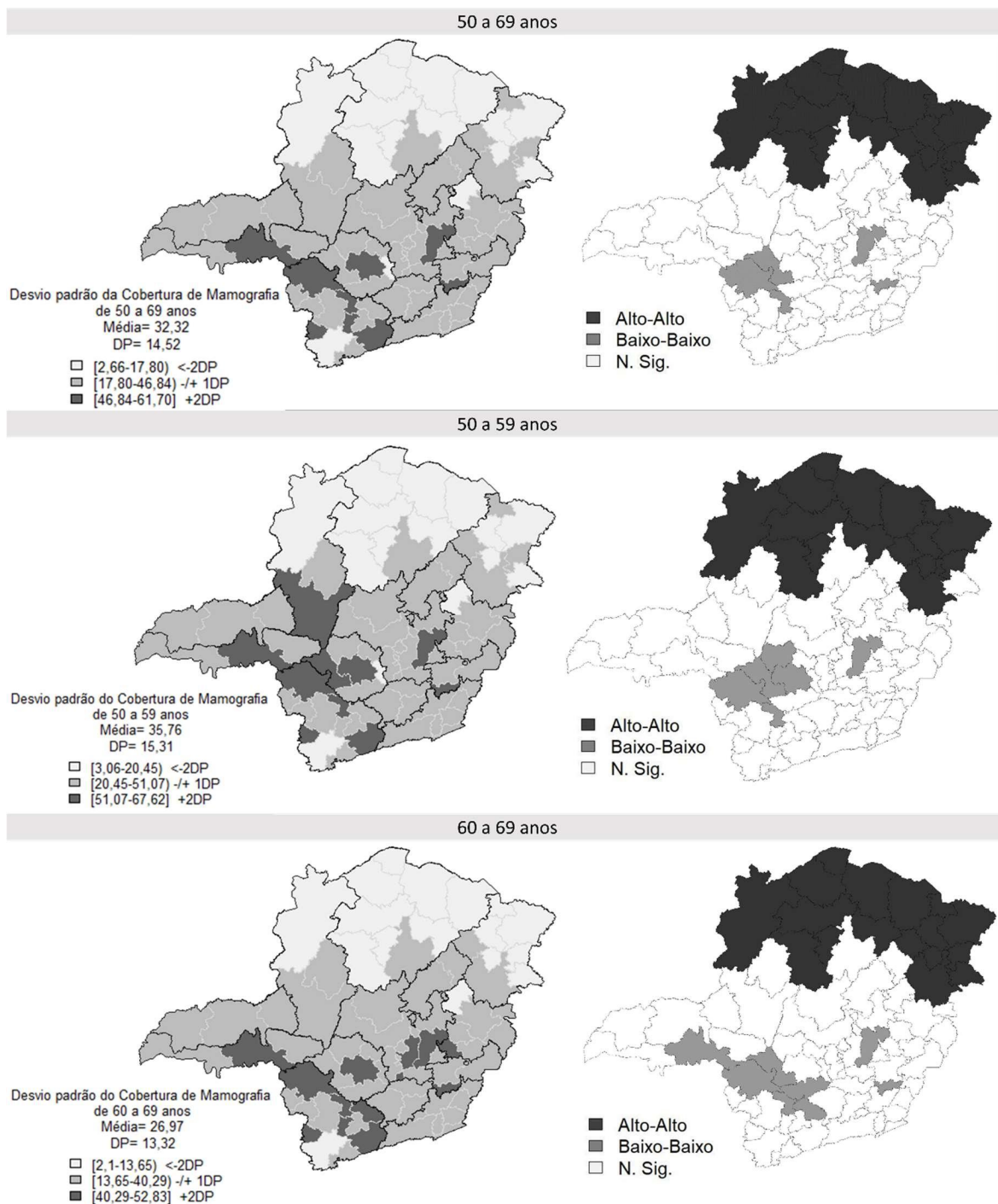


Figura 2 – Distribuição e Lisa Map das coberturas de mamografias nas microrregiões de saúde de Minas Gerais, 2010.

Todas as coberturas e indicadores socioeconômicos e assistenciais tiveram autocorrelação espacial positiva e significativa a 5% de probabilidade, exceto número de enfermeiros e distância percorrida para realização de mamografia fora da microrregião de saúde (DPRMF). Analfabetismo, IVS e IDH apresentaram forte autocorrelação. As coberturas, renda, grau de urbanização e índice de Gini, autocorrelação moderada.

Nos modelos simples de regressão, IVS, Gini e analfabetismo obtiveram coeficientes negativos e significativos com todas as coberturas, ou seja, quanto menor a vulnerabilidade na saúde, desigualdade de renda e analfabetismo, maior a cobertura de mamografia (tabela 2).

O modelo com melhor ajuste foi o que incluiu Gini e renda como variáveis independentes. Posteriormente, foram acrescentadas a esses modelos, cada uma das variáveis proximais, permanecendo significativas apenas o número de mamógrafos e sua utilização (tabela 2).

No modelo final da cobertura na faixa etária de 50-59 anos foi incluída a variável população categorizada para corrigir a heterocedasticidade dos resíduos. Essa variável não foi significativa, mas as variâncias apresentaram-se mais homogêneas e menos dispersas e a explicação das variáveis independentes sobre a cobertura nas faixas etárias de 50-59 anos, passou de 59,46% para 60,21% (tabela 2).

Após os ajustes, as variáveis que permaneceram associadas diretamente às coberturas foram renda, Gini, número de mamógrafos e utilização dos mamógrafos. Assim observou-se que para cada aumento de R\$ 1.000,00 na renda ocorreu um aumento de 31,71, 35,46 e 30,57 nas coberturas de 50-69, 50-59 e 60-69 anos, respectivamente. O índice de Gini se comportou de maneira inversa, demonstrando que conforme aumentaram as desigualdades, a cobertura diminuiu, ou seja, para cada 1% de acréscimo na desigualdade, houve uma redução de 1,26, 1,48 e 0,98 nas coberturas de 50-69, 50-59 e 60-69 anos, respectivamente (tabela 2).

As microrregiões das macros do Noroeste, Norte, Nordeste e Jequitinhonha apresentaram maior vulnerabilidade na saúde, menor renda e maior desigualdade de renda. (figura 3).

Tabela 2 – Modelos de Regressão Simples e Múltiplos entre as coberturas de mamografias e indicadores sociodemográficos e assistenciais das microrregiões de saúde de Minas Gerais, 2010.

Variável (n=77)	Modelos Simples			Modelos Múltiplos		
	50-69	50-59	60-69	50-69	50-59	60-69
IDH (%)	2,13 ($< 0,001$)	2,20 ($< 0,001$)	1,95 ($< 0,001$)			
Renda (x1.000 BRL)	54,06 ($< 0,001$)	55,81 ($< 0,001$)	50,04 ($< 0,001$)	31,71 ($< 0,001$)	35,46 ($< 0,001$)	30,57 ($< 0,001$)
IVS (%)	-0,48 ($< 0,001$)	-0,50 ($< 0,001$)	-0,44 ($< 0,001$)			
Gini (%)	-3,17 ($< 0,001$)	-3,36 ($< 0,001$)	2,82 ($< 0,001$)	-1,26 (0,013)	-1,48 (0,008)	-0,98 0,034
Grau de Urbanização (%)	0,65 ($< 0,001$)	0,67 ($< 0,001$)	0,59 ($< 0,001$)			
Analfabetismo (%)	-1,61 ($< 0,001$)	-1,68 ($< 0,001$)	-1,47 ($< 0,001$)			
População						
Médio porte	13,85 ($< 0,001$)	14,24 (0,001)	12,86 ($< 0,001$)		1,62 (0,614)	
Grande porte	18,46 ($< 0,001$)	19,36 ($< 0,001$)	16,50 ($< 0,001$)		5,32 (0,267)	
Médicos*	16,30 ($< 0,001$)	16,60 ($< 0,001$)	15,52 ($< 0,001$)			
Enfermeiros*	17,80 (0,212)	18,42 (0,220)	17,43 (0,182)			
Radiologistas**	3,14 (0,001)	3,26 (0,001)	2,88 (0,001)			
Nº Mamógrafos**	6,17 (0,002)	6,12 (0,004)	6,26 ($< 0,001$)	4,08 (0,007)	4,18 (0,012)	4,33 (0,002)
Cap. Util. Eq.***	3,64 (0,001)	3,95 ($< 0,001$)	3,07 ($< 0,001$)	2,71 ($< 0,001$)	3,34 ($< 0,001$)	2,31 ($< 0,001$)
DPRMR(Km)	0,100 (0,198)	0,11 (0,180)	0,09 (0,230)			
DPRMF (Km)	0,02 (0,261)	-0,015 (0,333)	-0,02 (0,174)			
R ² ajustado				0,58	0,57	0,58
R ²				0,61	0,60	0,60
Autocorrelação dos resíduos						
Moran Global I				$< -0,01$	-0,03	0,04
<i>p</i> -valor				0,797	0,985	0,347
Teste de Normalidade						
Jarque-Bera				1,11	1,57	0,12
<i>p</i> -valor				0,575	0,456	0,942
Teste de Heterocedasticidade						
Breusch-Pagan				8,96	8,77	6,10
<i>p</i> -valor				0,062	0,187	0,190

*Razão por 1.000 habitantes; **Razão por 100.000 habitantes; ***Capacidade Utilizada do Equipamento: A cada 1000 mamografias/mamógrafos; DPRMR: Distância Percorrida para Realização de Mamografia dentro da Região; DPRMF: Distância Percorrida para Realização de Mamografia fora da Região.

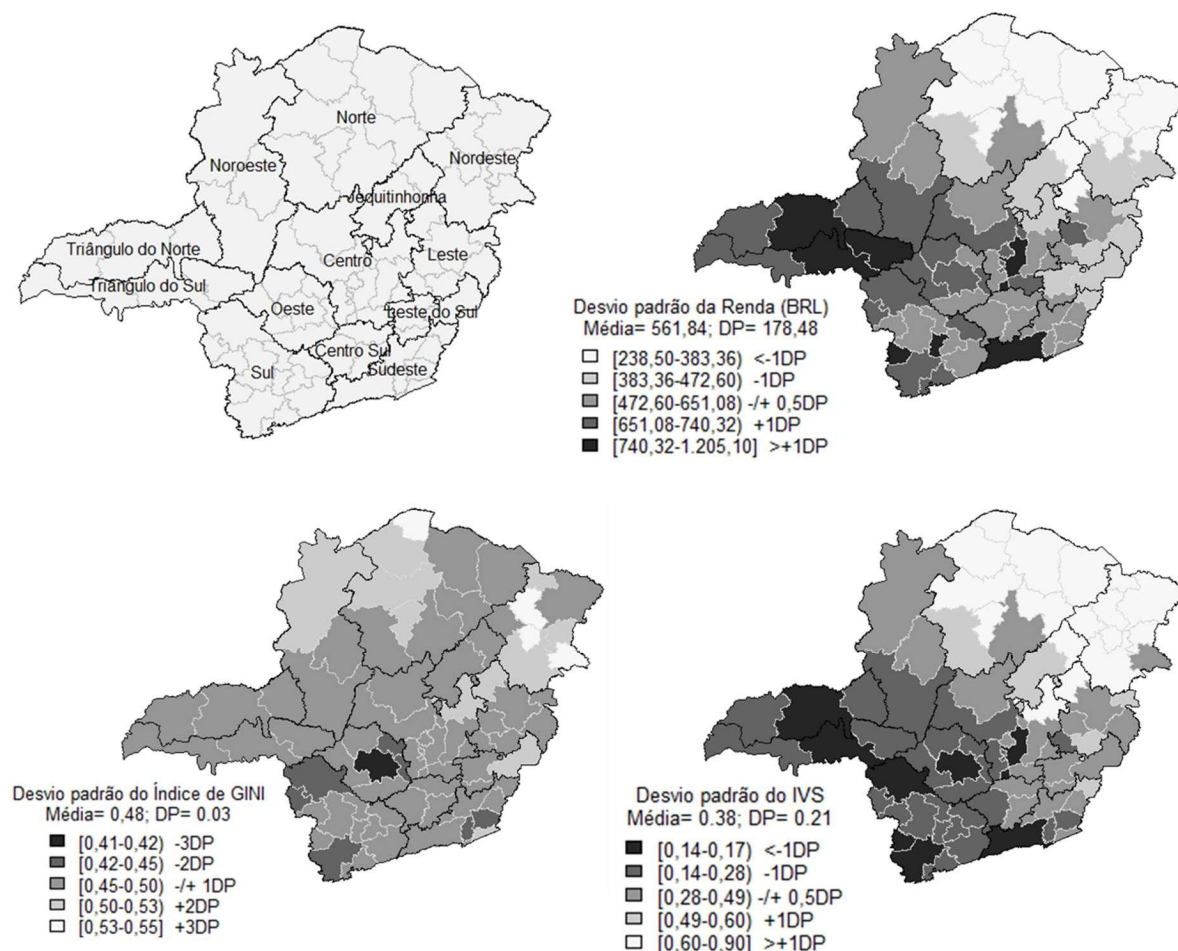


Figura 3 – Macrorregiões de saúde e indicadores socioeconômicos das microrregiões de saúde de Minas Gerais, 2010.

Dentre as 9 microrregiões que não possuem mamógrafos disponíveis para o SUS, 4 estão localizadas na macro Nordeste e 2 na macro Norte. A distância que as mulheres dessas microrregiões precisam percorrer para realizar mamografia em outra microrregião varia de 76,6 a 281,10 quilômetros. Nas regiões que possuem mamógrafos, as mulheres das micros de Patos de Minas, Francisco Sá, Janaúba/Monte Azul, Almenara, Curvelo, Araxá e Formiga percorrem distâncias superiores a 60 Km para realização do exame (figura 4).

No que se refere à utilização do equipamento, somente a microrregião de Ipatinga ultrapassa a sua capacidade em 44,45%. As demais microrregiões estão aquém da sua capacidade, com apenas 5 microrregiões utilizando mais de 70% da sua capacidade. Dentre as que utilizam menos de 25% o equipamento, 30% estão localizadas nas macros Norte, Nordeste e Noroeste (figura 4).

O número de enfermeiros e médicos é menor na maioria das microrregiões que utilizam menos de 25% da capacidade do mamógrafo, principalmente no Noroeste, Norte,

Jequitinhonha e Nordeste. No Centro, Oeste, Sudeste e Sul também existem microrregiões que possuem menor número de enfermeiros e utilização do mamógrafo (figura 4).

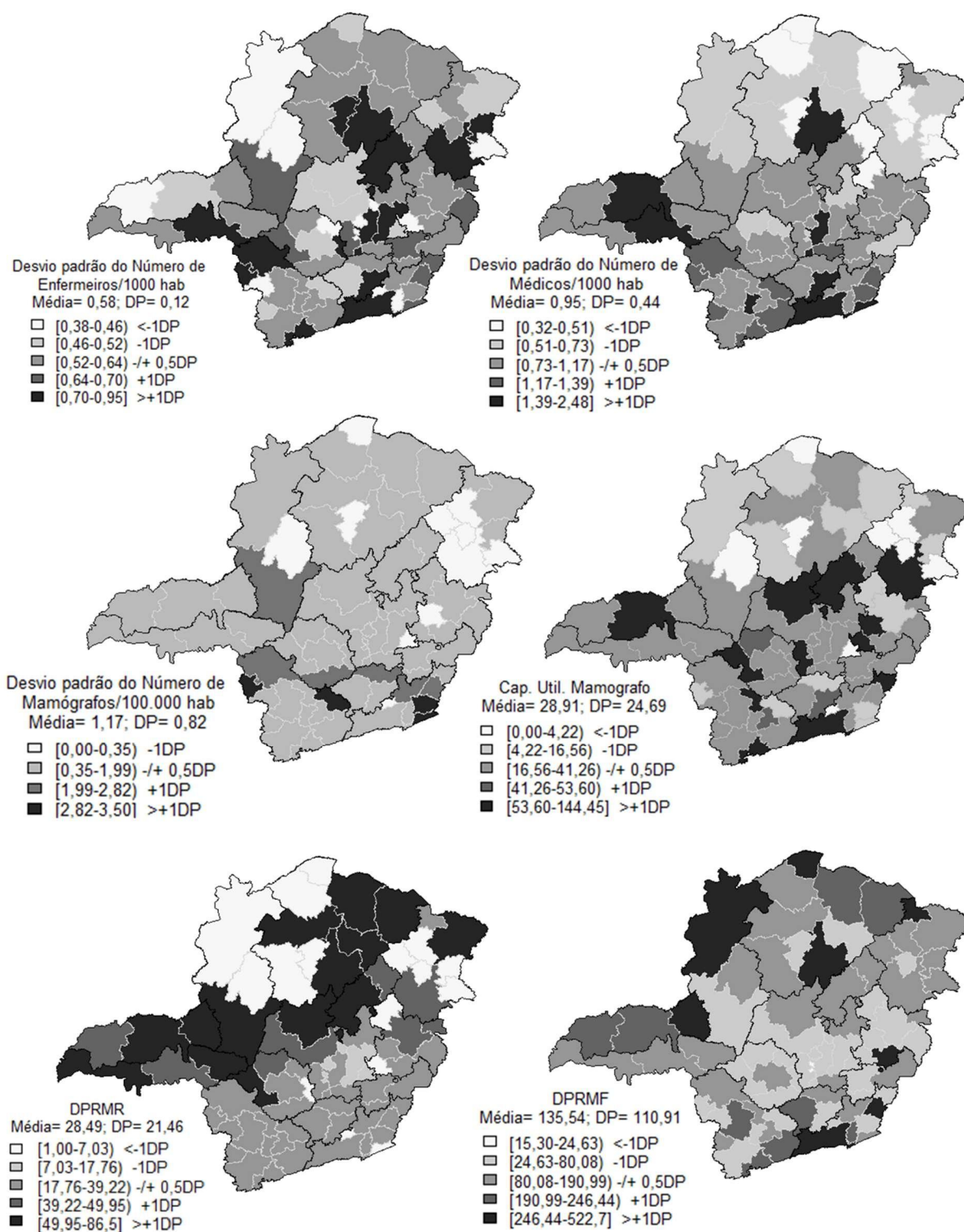


Figura 4 – Indicadores assistenciais das microrregiões de saúde de Minas Gerais, 2010.

DISCUSSÃO

Por meio do relacionamento probabilístico constatou-se que 90% (453.587) dos registros do SISMAMA eram referentes a uma única mulher, o que viabilizou o cálculo de uma cobertura de mamografia, com maior validade para o estado (36,3% na faixa etária de 50 a 69 anos) e suas microrregiões de saúde. A forma habitual de cálculo, que considera a razão entre número de exames realizados e população feminina na faixa etária preconizada, pode fornecer valores superestimados, principalmente nas áreas mais desenvolvidas e com maior oferta de exames. Estudo utilizando esse proxy encontrou razão de 0,14 em 2010 e 0,15 em 2011 para o estado, sendo as maiores razões encontradas em 2010 no Jequitinhonha, Triângulo Sul e Oeste (0,24, 0,19 e 0,18, respectivamente) e em 2011 no Sul, Oeste e Triângulo do Sul e Norte (0,18; 0,18 e 0,17 respectivamente) (CORREA et al., 2017). Esses valores demonstram a limitação desse proxy para traduzir a realidade do rastreamento mamográfico, pois nesse estudo, considerando o número de mulheres examinadas e não de exames, Jequitinhonha foi uma das macrorregiões que apresentou menor cobertura e Oeste, Centro e Triângulo do Sul as maiores.

Do número total de mulheres examinadas, pouco mais da metade (51,6%) pertenciam à faixa etária de 50-69 anos, recomendada para exames de rastreamento. Esse achado está de acordo com outros estudos. No Brasil, em 2010, a cobertura de mamografia na faixa etária de 40 a 49 anos foi ligeiramente superior (26,5%) à da faixa etária de 60-69 anos (25%)⁶. Nesse mesmo ano, estudo observou que no estado de São Paulo 42% das mulheres com idade inferior a 50 anos realizaram mamografia (FAYER et al., 2020). Ainda na região sudeste do país, outro estudo verificou que em 2010/2011, Minas Gerais realizou aproximadamente 50% das mamografias de rastreamento em mulheres com idade inferior a 50 anos, sendo o maior percentual encontrado na macrorregião Jequitinhonha (CORREA et al., 2017). Destaca-se que alto percentual de mamografias em idade não preconizada pode dificultar a ampliação do acesso para mulheres elegíveis ao rastreamento e diminuir sua efetividade.

Entre as microrregiões mineiras, Manga que em uma das macrorregiões com os piores indicadores socioeconômicos e de saúde (Norte), foi a que realizou a maior parte das mamografias na faixa etária de 50 a 69 anos (73,0%) e foi ainda a única a se aproximar da recomendação do Ministério da Saúde que indica que 50% dos exames de mamografia realizados devem contemplar as mulheres da faixa etária de 50 a 59 anos (BRASIL, 2015),

atingindo 48,6%. Destaca-se que Manga, tem um IVS de 0.88, não possui mamógrafo e que as mulheres se deslocam 281,1 Km para realizar o exame, e talvez esses aspectos expliquem uma seleção mais criteriosa e de acordo com as recomendações, das mulheres elegíveis para o rastreamento. Desta forma apesar dessa microrregião estar localizada na macrorregião Norte, uma das que apresenta as menores coberturas, grande parte dos seus exames foram direcionados para a faixa etária alvo do rastreamento, possibilitando atingir bons resultados.

O percentual mínimo de cobertura, recomendado pela Organização Mundial de Saúde, 70% a 75%, não foi atingido por nenhuma das microrregiões de saúde do estado. As menores coberturas foram encontradas em microrregiões do Noroeste, Norte, Nordeste e Leste. A baixa cobertura em microrregiões de Minas Gerais pode estar relacionada à falta de conhecimento da população em relação à doença e a escassez de recursos médico-hospitalares (ALVES; MAGALHÃES; COELHO, 2017). Mulheres residentes nas regiões Norte e Nordeste do Brasil, tem menor chance de obter um pedido médico de mamografia, e mesmo após receber a solicitação, algumas não conseguem realizar o exame devido as dificuldades que encontram para agendá-lo (AZEVEDO e SILVA et al., 2017).

As microrregiões do Norte, Noroeste e Nordeste compõem um grande cluster de regiões com cobertura abaixo da média, indicadores socioeconômicos ruins e condições assistenciais de saúde precárias, exceto Montes Claros, Unaí e João Pinheiro. O fato de estarem próximas às regiões mais ricas das quais são dependentes e partilham serviços com elas, pode explicar em parte o agrupamento de microrregiões de baixa cobertura com aquelas que apresentam melhor desempenho desse indicador, uma vez que algumas dessas microrregiões possuem municípios responsáveis por ofertar assistência ambulatorial e hospitalar para o tratamento do câncer, atendendo usuários de outras regiões (ALVES; MAGALHÃES; COELHO, 2017).

O acesso aos serviços de saúde é importante para desfechos favoráveis nos casos de doença, e a sua efetividade pode ser percebida nas facilidades e/ou dificuldades que os usuários encontram para utilizar os serviços e atender suas necessidades. No caso do câncer de mama, a dificuldade de acesso pode atrasar seu diagnóstico e tratamento contribuindo para um pior prognóstico, uma vez que a doença é detectada em estádios mais avançados, o que demanda serviços especializados de diagnóstico/tratamento, muitas vezes não disponíveis nas regiões menos desenvolvidas (GUERRA et al., 2015).

Fatores como ausência ou número insuficiente de mamógrafos na microrregião de residência da usuária, presentes nas regiões Norte e Nordeste, podem dificultar a marcação

de mamografia. Por outro lado, as macrorregiões Triângulo, Centro, Sul e Sudeste possuem a maior proporção de equipamentos. Situação semelhante foi observada em estudos realizados no Brasil, identificando menor número de mamógrafos disponíveis para o SUS no Norte e Nordeste e maior no Sul e Sudeste (TOMAZELLI e AZEVEDO e SILVA, 2017; NOGUEIRA et al., 2019).

Contudo, além do mamógrafo estar disponível e em condições de uso, é essencial que ele esteja próximo à região de residência da usuária, no máximo à 60 Km de distância (BRASIL, 2015). Sete microrregiões apresentaram distância superior a essa quilometragem e dessas quatro se localizam no Noroeste, Norte e Nordeste. Além disso, mulheres de microrregiões sem mamógrafos, podem chegar a se deslocar quase cinco vezes mais a distância máxima recomendada para realizar o exame.

A grande distância entre a residência da usuária até os serviços de saúde está associada às menores taxas de rastreamento, principalmente para aquelas que possuem baixa renda (ZHA et al., 2019). Dependendo da região, a locomoção por meio de transporte público ou particular tem que ser custeada pela própria pessoa, que traz ônus e pode inviabilizar sua adesão ao exame. No Canadá, mesmo existindo um programa de rastreio gratuito e universal, foi observado que a participação de grupos populacionais de baixo nível socioeconômico e residentes em áreas remotas nos programas de triagem é baixa (ZHA et al., 2014).

Além disso, fatores como péssimas condições das estradas, muitas sem pavimentação, alto índices de acidentes em alguns trechos, localidades cortadas por rios onde as travessias são possíveis somente por meio de balsas, dificultam ainda mais o acesso e diminuem a adesão às ações de rastreamento do câncer de mama (ALVES; MAGALHÃES; COELHO, 2017).

Nesse estudo foi observada escassez de profissionais aptos a solicitar e laudar mamografias, principalmente nas microrregiões ao Noroeste, Norte, Jequitinhonha e Nordeste. O menor número de enfermeiros e médicos nessas microrregiões talvez possa explicar em parte a menor utilização dos mamógrafos existentes, uma vez que sem o pedido as mulheres não conseguem realizar o exame.

A presença do enfermeiro nas unidades de saúde pode contribuir para uma maior adesão ao exame de mamografia e retorno bianual, conforme recomendado pelo programa de rastreamento, uma vez que esse profissional promove atividades de educação em saúde,

solicita exames conforme protocolos do serviço e esclarece possíveis dúvidas das mulheres acerca da saúde da mama e o processo de rastreamento (CHIARELLI, et al., 2010).

Apesar de algumas microrregiões de saúde de Minas Gerais não possuírem mamógrafos ou apresentarem menor número de equipamentos comparadas às demais, somente Ipatinga excede a sua capacidade de utilização. Se considerarmos esse atributo como critério para aumento da realização de exames e elevação da cobertura na faixa etária alvo, conclui-se que o número de equipamentos é suficiente para atender as demandas das microrregiões do estado, exceto em Ipatinga.

O valor tomado como referência para calcular a capacidade de utilização do equipamento é referente ao número total de exames que um mamógrafo tem condições de realizar (ALVES; MAGALHÃES; COELHO, 2017). No entanto, alguns fatores podem limitar sua utilização. Um deles está relacionado à manutenção, equipamentos podem apresentar defeitos, o que inviabiliza a realização de exames num período de dias ou meses. Outro fator está diretamente ligado aos contratos de prestação de serviços realizados entre instituições privadas e o SUS, essas instituições podem pactuar um número de exames abaixo da capacidade do equipamento (TOMAZELLI et al., 2017). Esses elementos individualmente ou em conjunto, provocam redução no número de mamografias disponíveis para as mulheres que tem acesso somente por meio do serviço público.

Minas Gerais possui um território extenso e apresenta diferenças geográficas, socioeconômicas, políticas e culturais marcantes entre suas microrregiões, impactando no setor saúde. Nesse estudo foi observado que as regiões centrais e sul têm maior cobertura de mamografia, disponibilidade de serviços, equipamentos/recursos humanos, maior renda e menor desigualdade de renda e vulnerabilidade na saúde, ao contrário das regiões do Norte, Nordeste e Noroeste.

Os principais fatores sociodemográficos relacionados à não realização ou baixa cobertura de mamografia são: menor renda, desigualdade de renda e desigualdades regionais que interferem na alocação e oferta de serviços e recursos humanos. Esses achados são corroborados por outros autores. Na África do Sul, estudo encontrou associação entre rastreamento do câncer de mama e ensino superior, maior renda e acesso a um plano de saúde (PELTZER e PHASWANA-MAFYA, 2014). Outro estudo, observou no Brasil, maior cobertura de mamografia nas microrregiões com maior desenvolvimento socioeconômico, menor desigualdade de renda e melhor acesso geral aos cuidados de saúde. As microrregiões

menos desenvolvidas apresentaram menor número de mamógrafos e de recursos humanos (NOGUEIRA et al., 2019).

A desigualdade de renda e vulnerabilidade na saúde é bem delimitada no território mineiro, podendo ser observada nos mapas temáticos, principalmente ao Norte, Nordeste, Jequitinhonha e Leste. Por meio do Índice de Gini, observa-se que mesmo em locais desenvolvidos e com maior renda e menor vulnerabilidade na saúde, a riqueza tem distribuição desigual, o que prejudica o acesso aos serviços de saúde e conseqüentemente impacta na cobertura do rastreamento, sendo esse fato percebido nas baixas coberturas de microrregiões do Triângulo, Centro e Sul, consideradas as mais ricas do estado. Estudo realizado em Uberaba, município pertencente à microrregião do Triângulo Sul de Minas Gerais, encontrou associação entre maior renda e rastreamento do câncer de mama (BURANELLO, et al., 2018).

A vulnerabilidade na saúde associada à natureza oportunística do rastreamento do câncer de mama existente no país, é outro fator que prejudica ainda mais o acesso das mulheres aos serviços prevenção/detecção precoce. Assim, usuárias residentes em regiões menos desenvolvidas e com recursos escassos podem deixar de procurar atendimento ou buscá-lo apenas quando a sua situação de saúde já está agravada.

Grupos mais favorecidos têm maior probabilidade de participar no rastreio. A explicação parcial para esse fato é que essas pessoas se preocupam mais com a sua saúde e quando decidem realizar uma avaliação, optam pelo serviço que no momento oferecem acesso mais fácil (CHOI et al., 2018). Por outro lado, mulheres com menor renda e sem cobertura adequada de um sistema de saúde tem menos acesso à mamografia, maior probabilidade de apresentar câncer de mama em estágio avançado e menor chance de cura e sobrevida (RIBA et al., 2019).

A introdução de um programa de rastreamento organizado consiste em um desafio para políticas públicas de controle do câncer de mama em Minas e em todo país e pode contribuir para redução da desigualdade de acesso, uma vez que esse tipo de programa convida todas as mulheres de uma determinada faixa etária a participar, independentemente de sua renda, raça, escolaridade e outros fatores socioeconômicos. O planejamento das ações de rastreamento e organização dos serviços permite alocação dos equipamentos de acordo com a necessidade de cada região, contratação de recursos humanos e otimização na sua utilização, com conseqüente melhoria da qualidade da assistência e elevação da cobertura na

população alvo das ações (ZHA et al., 2019; BURANELLO et al., 2018), o que se coaduna com os princípios e diretrizes do Sistema Único de Saúde (SUS).

Além disso, a política brasileira de controle do câncer de mama precisa priorizar com urgência o diagnóstico de mulheres sintomáticas, uma vez que no país, a prevalência de casos de câncer de mama em estadiamento avançado (III e IV) é alta e não diminuiu desde a implantação do programa de rastreamento, em 2004.

O SISMAMA vem sendo substituído pelo Sistema de Informação do Câncer (SISCAN), implantado a partir de 2013, que possibilita a identificação da pessoa examinada ao invés de exames e permite a construção do histórico da mulher e da sua situação de seguimento dos cânceres femininos. Devido as dificuldades de implantação como utilização de sistema próprio por parte de alguns serviços e ausência de ferramenta de *webservice* para migração de dados para o Siscan, principalmente na região Sudeste, o registro de mamografias ainda é realizado tanto no Sismama quanto no Siscan, assim a identificação e acompanhamento longitudinal das usuárias examinadas pode ser realizado no SISMAMA por meio de técnicas de relacionamento probabilístico e classificação automatizada dos registros vinculados (DUARTE et al., 2019).

As limitações desse estudo derivam da utilização de bases de dados secundárias que podem apresentar incompletude e erros de preenchimento e dos relacionamentos probabilísticos que utilizam classificação automatizada dos registros vinculados que não são totalmente perfeitos e alguns pares falsos podem ser classificados como verdadeiros. No entanto, medidas que avaliam a qualidade dessa técnica, principalmente nos casos de dados desequilibrados, como *precision*, *recall* e medida F, apresentaram valores superiores a 98%, nesse estudo. Por se tratar de um estudo ecológico, o tipo de associação encontrada entre fator e desfecho reflete o que ocorre a nível da população, não podendo ser realizadas inferências para o nível individual. Além disso, devido ao delineamento transversal utilizado, não se pode estabelecer uma relação temporal de causa e efeito, mesmo no nível ecológico.

CONSIDERAÇÕES FINAIS

Ainda que o SISMAMA ofereça apenas informações sobre exames realizados, é possível, por meio de um relacionamento probabilístico, identificar o número de mulheres

atendidas, calcular a cobertura do rastreamento do câncer de mama e realizar o seguimento das mulheres rastreadas e registradas nesse sistema.

Considerando que todos os mamógrafos em Minas Gerais estejam em condições de uso, conclui-se que a baixa cobertura não pode ser atribuída à falta do equipamento, sendo a capacidade instalada suficiente para atendimento das demandas do estado. Fatores como má distribuição dos mamógrafos e ausência e/ou menor número de profissionais aptos a solicitar mamografias em algumas microrregiões, prejudicam o acesso ao exame.

Observou-se que fatores socioeconômicos e de saúde influenciam a cobertura de mamografia, sendo verificado que mulheres residentes em microrregiões com menor renda, maior desigualdade de riqueza e maior vulnerabilidade na saúde apresentam menor cobertura de mamografia, mesmo nas microrregiões que compõem as macrorregiões mais ricas.

Por meio do estudo, foi possível compreender, que em Minas Gerais o acesso da população feminina às ações de rastreamento do câncer de mama é afetado pelas desigualdades sociais e de saúde, o que representa um grande desafio para os gestores de saúde. Destaca-se que a implantação no estado de um rastreamento organizado, com convocação das mulheres num período determinado e a busca ativa das ausentes, pode ampliar ainda mais o acesso e cobertura, contribuindo para aumentar a efetividade do programa de rastreamento.

REFERÊNCIAS

ALVES, M. O.; MAGALHÃES, S. C. M.; COELHO, B. A. A regionalização da saúde e a assistência aos usuários com câncer de mama. **Saúde e Sociedade**, v. 26, n. 1, p. 141-154, 2017.

AZEVEDO E SILVA, G. et al. Acesso à detecção precoce do câncer de mama no Sistema Único de Saúde: uma análise a partir dos dados do Sistema de Informações em Saúde. **Cadernos de Saúde Pública**, v. 30, n. 7, p. 1537-1550, 2014.

AZEVEDO e SILVA, G.; et al. Detecção precoce do câncer de mama no Brasil: dados da Pesquisa Nacional de Saúde, 2013. **Revista de Saúde Pública**, v. 51, n. 1, p. 1s-9s, 2017.

BRASIL. Ministério da Saúde. Critérios e Parâmetros para o Planejamento e Programação de Ações e Serviços de Saúde no âmbito do Sistema Único de Saúde. Brasília, Ministério da Saúde, 2015.

BURANELLO, M. C.; et al. Prática de exames de rastreio para câncer de mama e fatores associados – Inquérito de Saúde da Mulher em Uberaba MG, Brasil, 2014. **Ciência e Saúde Coletiva**, v. 23, n. 8, p. 2661-2670, 2018.

CHIARELLI, A. M.; et al. Influence of Nurses on Compliance with Breast Screening Recommendations in an Organized Breast Screening Program. **Cancer Epidemiology, Biomarkers & Prevention**, v. 19, n. 3, p. 697-706, 2010.

CHOI, E. et al. Socioeconomic Inequalities in Cervical and Breast Cancer Screening among Women in Korea , 2005 – 2015. **Yonsei Medical Journal**, v. 59, n. 9, p. 1026–1033, 2018.

CORREA, C. S. L.; PEREIRA, L. C.; LEITE, I. C.G.; FAYER, V. A.; GUERRA, M. R.; BUSTAMANTE-TEIXEIRA, M. T. Rastreamento do câncer de mama em Minas Gerais: avaliação a partir de dados dos sistemas de informações do Sistema Único de Saúde. **Epidemiologia e Serviços de Saúde**, v. 26, n. 3, p. 481-492, 2017.

DUARTE, D. A. P.; et al. Avaliação de método para classificação automatizada de pares em relacionamentos probabilísticos de bancos de dados. **Cadernos de Saúde Pública**, v. 35, n. 11, p. e00066419, 2019.

FAYER, V. A.; et al. Controle do câncer de mama no estado de São Paulo: uma avaliação do rastreamento mamográfico. **Cadernos de Saúde Coletiva**, v. 28, n. 1, p. 140-152, 2020. GUERRA, M. R.; et al. Sobrevida por câncer de mama e iniquidade em saúde. **Cadernos de Saúde Pública**, v. 31, n. 8, p. 1673-1684, 2015.

INCA. Instituto Nacional do Câncer. Diretrizes para a detecção precoce do câncer de mama no Brasil. Rio de Janeiro: INCA, 2015. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//diretrizes_deteccao_precoce_cancer_mama_brasil.pdf>. Acesso em jun 2019.

INCA. Instituto Nacional do Câncer. Revisão do Parâmetro para Cálculo da Capacidade de Produção do Mamógrafo Simples - 2015. Rio de Janeiro: INCA, 2015c. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document/nota-revisao-capacidade-dos-mamografos-2015_0.pdf>. Acesso em jun 2021.

INCA. Instituto Nacional do Câncer. Parâmetros técnicos para rastreamento do câncer de mama. Rio de Janeiro: INCA, 2021. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document/parametrostecrastreamentocamama_2021_1.pdf>. Acesso em mai 2021.

NOGUEIRA, M. C.; et al. Inequities in access to mammographic screening in Brazil. **Cadernos de Saúde Pública**, v. 35, n. 6, e0009981, 2019.

PELTZER, K.; PHASWANA-MAFYA, N. Breast and Cervical Cancer Screening and Associated Factors among Older Adult Women in South Africa. **Asian Pacific Journal of Cancer Prevention**, v. 15, n. 6, p. 2473-2476, 2014.

RIBA, L. A. et al. Association between socioeconomic factors and outcomes in breast cancer. **Breast Journal**, v. 25, n. 3, p. 488–492, 2019.

SANTOS-SILVA, I. et al. Ethnoracial and social trends in breast cancer staging at diagnosis in Brazil, 2001-14: a case only analysis. **Lancet Glob Health**, v. 7, p. e784-e797, 2019.

SUNG, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA Cancer Journal Clinicians**, Nova York, v. 71, n. 3, p. 209-249, 2021.

TOMAZELLI, J. G.; AZEVEDO E SILVA, G. Rastreamento do câncer de mama no Brasil: uma avaliação da oferta e utilização da rede assistencial do Sistema Único de Saúde no período 2010-2012. **Epidemiologia e Serviços de Saúde**, v. 26, n.4, p. 713-724, 2017.

ZHA, N. et al. Beyond Universal Health Care: Barriers to Breast Cancer Screening Participation in. **Journal of the American College of Radiology**, v. 16, n. 4, p. 570–579, 2019.

6.3. ARTIGO 3: SOBRERRASTREAMENTO MAMOGRÁFICO POR FAIXA ETÁRIA E PERIODICIDADE E FATORES ASSOCIADOS

Resumo

Objetivos: estimar o sobrerastreamento mamográfico por periodicidade excessiva e por faixa etária a partir do seguimento de bases identificadas do SISMAMA, em Minas Gerais e analisar possíveis fatores sociodemográficos e assistenciais associados. **Métodos:** estudo descritivo, de coorte não-concorrente seguido por um estudo ecológico com indicadores sociodemográficos e assistenciais estimados por microrregiões de saúde, no período de 2010-2012. Após limpeza e padronização dos dados, foram removidas duplicidades e repetições pelos métodos determinístico e probabilístico. Foi estimado o percentual de sobrerastreamento por faixa etária (realização de exames em mulheres com idade <50 anos ou ≥ 70 anos) e periodicidade (repetição do exame em um período inferior a 18 meses). A associação entre o sobrerastreamento e demais variáveis foi analisada por modelos de regressão linear. **Resultados:** o sobrerastreamento por faixa etária atingiu 47,1% das mulheres e por periodicidade 27,8%. Maior renda, número de mamógrafos e menor desigualdade de renda se associaram ao sobrerastreamento por periodicidade. Essas características são encontradas nas macrorregiões Centro, Sul, Sudeste, Triângulo do Norte e Oeste. **Conclusão:** as recomendações de faixa etária e periodicidade não têm sido adotadas pelas microrregiões de saúde de Minas Gerais. O sobrerastreamento esteve associado às desigualdades regionais existentes no estado e sua prevalência é maior nas regiões de melhor nível socioeconômico e que dispõem de mais recursos humanos e equipamentos. As ações de rastreamento no estado precisam de melhorias e adequações e estas devem focar na capacitação dos profissionais, sobretudo acerca das evidências que embasaram a elaboração das diretrizes vigentes no Brasil e diversos outros países, conscientização da população em geral, distribuição de recursos de acordo com a necessidade e característica de cada região, organização e otimização dos recursos, principalmente nas regiões em que estes são insuficientes ou que não utilizam toda a capacidade instalada.

Palavras-Chave: Mamografia, Rastreamento, Periodicidade, Faixa Etária, Fatores Socioeconômicos

Abstract

Objectives: to estimate mammographic over-screening in Minas Gerais by excessive periodicity and age using databases identified in SISMAMA and analyze associated sociodemographic and health care factors. **Methods:** descriptive non-concurrent cohort followed by an ecological study using sociodemographic and assistance indicators estimated by health micro-regions from 2010 to 2012. Deterministic and probabilistic methods removed duplicates and repetitions. Percentage of over-screening was calculated by age (exams performed in women aging < 50 or ≥ 70 years) and periodicity (repetition of the exam in a period lower than 18 months). Linear regression model verified associations between over-screening and other variables. **Results:** over-screening by age and periodicity reached 47.1% and 27.8% of women. Over-screening by periodicity was associated with high income, number of mammography, and low-income inequality. These characteristics were found in central, south, southeast, north triangle, and west macro-regions. **Conclusion:** health micro-regions of Minas Gerais did not adopt age and periodicity recommendations. Over-screening was associated with regional inequalities and high prevalence in regions with better socioeconomic levels, human resources, and equipment. Screening actions require improvements and adequacies focusing on professional training, especially evidence supporting current guidelines in Brazil and other countries; awareness of general population; resource distribution according to characteristics and needs of each region; and organized and optimized resources, particularly in regions with insufficient resources or not using total capacity.

Key words: Mammography, Screening, Periodicity, Age group, Socioeconomic factors

INTRODUÇÃO

O câncer de mama é a neoplasia mais comumente diagnosticada no sexo feminino e é considerada a principal causa de morte de mulheres por câncer no mundo (SUNG et al., 2020).

A mamografia de rastreamento é um dos métodos indicados e mais efetivos na detecção precoce do câncer de mama, sendo o seu principal benefício a redução da mortalidade, resultante da diminuição do diagnóstico da doença em estágios avançados da doença e oferta tratamento adequado em tempo oportuno (LOBERG et al., 2015; SANTOS-SILVA et al., 2019). No Brasil aproximadamente 40% dos casos de câncer de mama são

diagnosticados em estágios tardios e, apesar da introdução do programa de rastreamento em 2004, essa proporção tem se mantido constante (SANTOS-SILVA et al., 2019).

As diretrizes nacionais recomendam o rastreamento mamográfico bienal para mulheres na faixa etária de 50 a 69 anos, porém o rastreamento no Brasil é considerado oportunístico, ou seja, não há convocação da população alvo em intervalos pré-estabelecidos (INCA, 2015). Estudos observaram que no país a cobertura de mulheres na faixa etária alvo não atinge a meta de 70-75% considerada ideal pela Organização Mundial de Saúde (NOGUEIRA et al., 2019). Além disto, quase metade das mamografias são realizadas em mulheres com menos de 50 anos, com periodicidade anual (AZEVEDO e SILVA et al., 2014; CORREA et al., 2017), tendo sido observada uma taxa de sobrerastreamento por periodicidade de 150 mil mulheres/ano, o que demonstra que as recomendações do Ministério da Saúde não estão sendo seguidas (RODRIGUES et al., 2019).

Além dos benefícios, o rastreamento pode ocasionar danos, como exposição à radiação ionizante, resultados falso-positivos, sobretreamento e impactos psicológicos resultantes do sobrediagnóstico e intervenções desnecessárias, e estes podem ser ainda maiores se o intervalo entre exames for pequeno e se o rastreio for aplicado em grupos onde estudos demonstram que os malefícios superam os benefícios, como por exemplo mulheres jovens (INCA, 2015).

Desse modo, torna-se relevante conhecer detalhes acerca do rastreamento do câncer de mama no Brasil e avaliar se as ações executadas seguem as diretrizes e recomendações propostas pelo programa, como por exemplo faixa etária e periodicidade. O objetivo deste estudo foi estimar o sobrerastreamento mamográfico por faixa etária e periodicidade excessiva segundo as diretrizes e recomendações do Ministério da Saúde, a partir do seguimento de bases identificadas do SISMAMA, em Minas Gerais e analisar as desigualdades regionais e fatores associados. Isto se faz necessário, visto que os achados possibilitarão a identificação das mulheres sobrerastreadas e conhecimento acerca das características das regiões de residência dessas mulheres para que assim seja definido se o foco de intervenção para resolução de possíveis problemas no rastreamento do câncer de mama seria a mulher, o profissional de saúde ou o sistema de saúde.

METODOLOGIA

Trata-se de um estudo de coorte não-concorrente seguido por um estudo ecológico com indicadores sociodemográficos e assistenciais estimados por microrregiões de saúde, no período de 2010-2012.

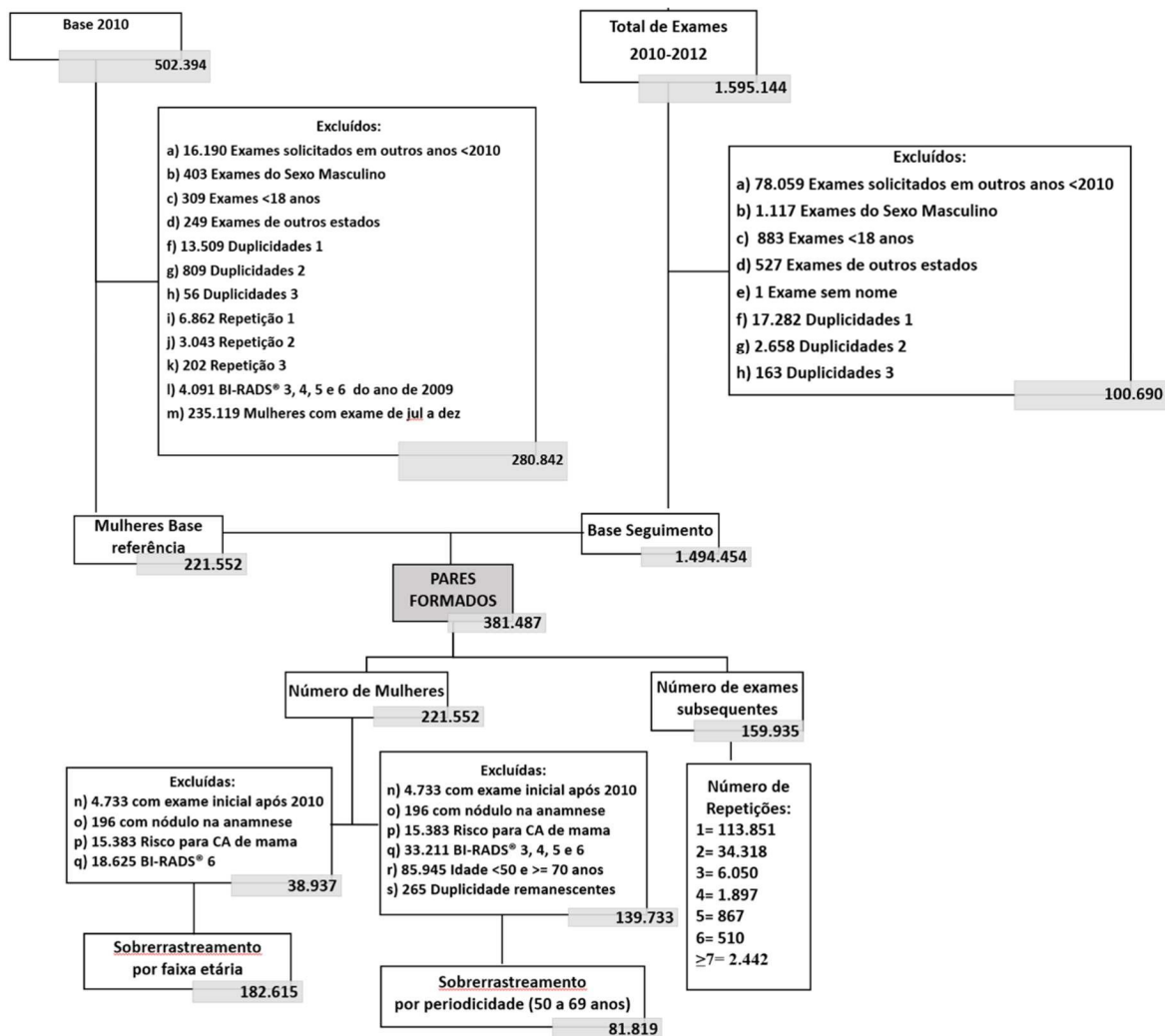
Foi feita uma coorte retrospectiva com dados secundários identificados do SISMAMA para estimar o sobrerastreamento por faixa etária e periodicidade. Os dados utilizados no estudo foram cedidos pela Secretaria de Saúde do Estado de Minas Gerais. Utilizando a variável data de solicitação do exame (D_EXAME), foram selecionados e extraídos dos arquivos anuais do SISMAMA exames de mulheres com idade maior ou igual a 18 anos, residentes em Minas Gerais e solicitados no ano de 2010 para construção da base referência e nos anos de 2010-2012 para construção da base seguimento, sendo excluídos os registros em que o campo da variável nome estava em branco (figura 1: a-e).

Foram removidas duplicidades e para se obter um único registro por mulher, foram excluídas da base referência as repetições de exames. Os métodos determinístico e probabilístico foram aplicados nesse processo e foram utilizados os Programas R (<http://www.r-project.org/> versão 3.6.1) e *Reclink III* (<http://reclink.sourceforge.net/>) (figura 1: f-k).

Foram excluídas da base referência mulheres que tiveram resultado alterado no ano de 2009 (BI-RADS® 3, 4, 5 e 6), uma vez que o menor intervalo entre exames pode ser consequência do acompanhamento das anormalidades conforme as recomendações, como por exemplo resultado BI-RADS® 3 é recomendada a repetição em 6 meses a 1 ano. Esse processo foi realizado por meio do relacionamento probabilístico em dezesseis passos (DUARTE et al., 2019) (figura 1: l). Por fim, visando obter maior tempo de seguimento, o banco de referência incluiu apenas os exames solicitados entre os meses de janeiro/junho de 2010, totalizando 221.552 registros (figura 1:m). A limpeza da base cobertura de mamografia e o cálculo da cobertura foram realizados conforme descrito no artigo 2 da presente tese.

O relacionamento probabilístico entre as bases referência e seguimento foi realizado no programa *Reclink III* seguindo 3 passos que utilizaram como variáveis de bloqueio ano de nascimento, primeiro nome/ano de nascimento e primeiro nome/último nome e como variáveis de pareamento nome, data de nascimento e nome da mãe (DUARTE et al., 2019). Com a finalidade de obter o maior número possível de pares verdadeiros, a cada

passo foram mantidos todos os registros das bases referência e seguimento, mesmo aqueles classificados como par no passo anterior.



Duplicidade 1- Método Determinístico utilizando as variáveis: N+M+DN+DS+CNES+CODMUN. **Duplicidade 2:** Método Probabilístico utilizando na blocagem as variáveis: PN+UN+MS+AN+CNES; e na comparação as variáveis: NP+MP+DN+DS+CODMUN. **Duplicidade 3:** Método Probabilístico utilizando na blocagem as variáveis: PM+UM+MS+AN+CNES; e na comparação as variáveis: NP+MP+DN+DS+CODMUN. **Repetição 1-** Método Determinístico utilizando as variáveis: N+M+DN+CODMUN. **Repetição 2:** Método Probabilístico utilizando na blocagem as variáveis: PN+UN+NA; e na comparação as variáveis: NP+MP+DN+ CODMUN. **Repetição 3:** Método Probabilístico utilizando na blocagem as variáveis: PM+UM+AN; e na comparação as variáveis: NP+MP+DN+ CODMUN.

N: NOME, M: MÃE, DN: DATA DE NASCIMENTO, DS: DATA DE SOLICITAÇÃO DO EXAME (D_EXAME), CNES: CNES REQUISITANTE, CODMUN: CÓDIGO DO MUNICÍPIO, PN: PNOOME, UN: UNOME, MS: MÊS DE SOLICITAÇÃO DO EXAME, AN: ANO DE NASCIMENTO, NP: NOMEPAD, MP: MAEPAD, PM: PMAE, UM: UMAE

Figura 1 - Processo de limpeza das bases referência e seguimento para relacionamento probabilístico e do banco de sobrerrestreamento.

De cada passo foi extraída uma amostra composta por 300 pares, para posterior revisão manual e definição de ponto de corte para classificação dos pares verdadeiros. Para

determinar o escore de maior poder discriminatório na identificação dos pares verdadeiros e avaliar sua acurácia foi utilizada a sensibilidade (*recall*) e valor preditivo positivo (*precision*) através da curva *precision-recall* e da medida F (DUARTE et al., 2019). Após a identificação dos pares verdadeiros, foi criado um banco com os pares organizados de acordo com as datas do primeiro exame e exames subsequentes de cada mulher. Posteriormente foram excluídas as mulheres que realizaram o primeiro exame após o ano de 2010 (figura 1:n).

Para análise do sobrerastreamento por faixa etária foram excluídas as mulheres que no primeiro exame em 2010 tinham biópsia prévia com malignidade comprovada, ou seja, diagnóstico BI-RADS® 6, com nódulo na anamnese, visto que essa descrição caracteriza mamografia diagnóstica; e daquelas identificadas como de alto risco para câncer de mama, dado que a recomendação para rastreamento de mulheres de alto risco é diferente (figura 1: o-q). Foi calculado o percentual de exames por micro e macrorregiões de saúde nas seguintes faixas etárias menor que 30, 30 a 39, 40 a 49, maior ou igual a 70 anos e total.

Em seguida, para análise do sobrerastreamento por periodicidade foram selecionadas as mulheres faixa etária de 50 a 69 e que no primeiro exame em 2010 apresentaram diagnóstico BI-RADS® 1 e 2 para ambas as mamas, ou seja, exame negativo ou com achados benignos. Foram removidas duplicidades remanescentes (figura 1: n-s). A data do primeiro exame normal no ano de 2010 (variável D_RECEBE) foi tomada como data inicial para análise do seguimento. O desfecho foi sobrerastreamento mamográfico por periodicidade excessiva, definido como mamografia realizada <18 meses após resultado normal, uma vez que variações de ≤ 6 meses do preconizado foram consideradas pouco relevantes e possivelmente logísticas (RODRIGUES et al., 2019). Para as mulheres que não realizaram o exame subsequente até 30/06/2012 e para aquelas cujo intervalo foi superior a 18 meses ocorreu a censura.

A análise dos tempos entre o primeiro exame e o subsequente foi feita por meio do método de *Kaplan-Meier*. Em seguida, foi construído um gráfico das curvas de realização do exame subsequente (curvas de sobrerastreamento). Por fim, foi analisado o sobrerastreamento por macro e microrregiões de saúde. O programa empregado nas análises estatísticas foi o R (<http://www.r-project.org/> versão 3.6.1).

Para o estudo ecológico foram utilizadas as seguintes variáveis de exposição: índice de desenvolvimento humano nas regiões de saúde (IDHr), renda domiciliar per capita média, índice de vulnerabilidade na saúde (IVS), índice de Gini, grau de urbanização, analfabetismo, tamanho da população, razão de médicos, médicos radiologistas e enfermeiros por 1.000

habitantes, razão de mamógrafos por x habitantes e razão de mamografias por mamógrafos (utilização), distância média percorrida para realização de mamografia dentro e fora da microrregião de saúde (DPRM). Tais variáveis foram selecionadas de acordo com a revisão de literatura e a disponibilidade em bases de acesso público com o objetivo de se desenvolver um estudo exploratório sobre a associação entre o sobrerastreamento e características das microrregiões de saúde de Minas Gerais.

As variáveis sobre os indicadores sociodemográficos e assistenciais das regiões de saúde foram coletadas das páginas eletrônicas do Datasus (<http://www2.datasus.gov.br>), do Programa de Avaliação do Desempenho do Sistema de Saúde – PROADESS (<https://www.proadess.icict.fiocruz.br/>) e do Instituto Brasileiro de Geografia e Estatística - IBGE (<https://ww2.ibge.gov.br>). Informações não disponíveis foram solicitadas via sítio eletrônico para acesso à informação do governo federal (<http://www.acessoinformacao.gov.br/>).

Foram elaborados mapas temáticos do sobrerastreamento, cobertura e indicadores sociodemográficos e assistenciais. A divisão dos indicadores em cinco classes foi realizada de acordo com 0,5 e 1,0 desvios padrões abaixo e acima da média e em escalas com tons mais claros para os menores valores/pior desempenho do indicador e mais escuros para os maiores valores/melhor desempenho do indicador.

O agrupamento de regiões com valores de atributos semelhantes foi avaliado pelo índice I de Moran local e apresentados no LisaMap que permite a identificação de clusters espaciais de regiões com valores muito altos, muito baixos ou não significativos.

Análises de regressão linear foram realizadas com a finalidade de identificar associações entre o sobrerastreamento por faixa etária e periodicidade e fatores sociodemográficos e assistenciais. Na primeira etapa foi realizada a regressão linear simples. Em seguida, para cada modelo da regressão múltipla foram inseridas as variáveis explicativas de acordo com o grupo de exposição ao qual pertenciam, sendo seguido o critério do grupo distal para o proximal. Após o ajuste os resíduos de cada modelo foram investigados para detectar evidências de heterocedasticidade e/ou autocorrelação espacial (respectivamente com os testes de Breusch-Pagan e de Moran I).

Este estudo é parte de um projeto de pesquisa mais amplo, denominado “Desigualdades de acesso às ações de controle de câncer através dos Sistemas de Informações de Saúde (DAAC-SIS)” e foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Juiz de Fora sob o parecer nº. 1.431.916.

RESULTADOS

Em Minas Gerais no período de janeiro a junho/2010, 182.615 mulheres realizaram mamografia de rastreamento, sendo 96.670 (52,9%) da faixa etária alvo (50 a 69 anos). O sobrerastreamento total por faixa etária atingiu 85.945 mulheres (47,1%) sendo maior nas mulheres de 40 a 49 anos, em todas as macrorregiões. As macrorregiões Jequitinhonha, Oeste e Leste do Sul foram as que apresentaram os maiores percentuais de sobrerastreamento total por faixa etária (56,6; 51,5 e 50,4%, respectivamente) e Leste, Norte e Centro, os menores (36,6; 44,9 e 45,5%) (tabela 1).

A figura 2 mostra a distribuição do sobrerastreamento por faixa etária nas microrregiões de saúde do estado. No LisaMap foram observados clusters com baixos e altos percentuais de sobrerastreamento em regiões pouco desenvolvidas e com alta vulnerabilidade na saúde, Norte e Jequitinhonha, respectivamente; e clusters com alto percentual de sobrerastreamento em regiões desenvolvidas e menos vulneráveis, Leste e Centro.

Tabela 1 – Sobrerrastreamento por faixa etária* no estado de Minas Gerais e macrorregiões, janeiro a junho de 2010.

Macrorregião	Total de Mulheres	Total de Mulheres rastreadas	<30		30-39		40-49		≥70		Sobrerrastreio Total	
	N	N	N	%	N	%	N	%	N	%	N	%
Jequitinhonha	89.632	2896	15	0,5	337	11,6	1079	37,3	209	7,2	1640	56,6
Oeste	418.247	15930	26	0,2	1359	8,5	5869	36,8	949	6,0	8203	51,5
Leste do Sul	230.401	8256	51	0,6	481	5,8	3130	37,9	500	6,1	4162	50,4
Triângulo do Sul	248.596	8089	12	0,1	506	6,3	2921	36,1	613	7,6	4052	50,1
Sudeste	578.319	17800	44	0,2	1025	5,8	6378	35,8	1214	6,8	8661	48,7
Nordeste	289.842	4397	15	0,3	227	5,2	1611	36,6	271	6,2	2124	48,3
Noroeste	217.392	6013	17	0,3	222	3,7	2347	39,0	314	5,2	2900	48,2
Triângulo do Norte	423.868	12943	24	0,2	544	4,2	4714	36,4	880	6,8	6162	47,6
Sul	919.607	27493	46	0,2	1229	4,5	10121	36,8	1521	5,5	12917	47,0
Centro Sul	268.248	7306	20	0,3	356	4,9	2623	35,9	403	5,5	3402	46,6
Centro	2.247.754	56481	105	0,2	2294	4,1	19495	34,5	3826	6,8	25720	45,5
Norte	504.321	5831	22	0,4	216	3,7	2028	34,8	353	6,1	2619	44,9
Leste	507.755	9180	22	0,2	394	4,3	2508	27,3	459	5,0	3383	36,9
Minas Gerais	6.943.982	182615	419	0,2	9190	5,0	64824	35,5	11512	6,3	85945	47,1

* Sobrerrastreamento por faixa etária= n e % de mulheres fora da faixa etária alvo para rastreamento (<50 anos e ≥70 anos)/ total de mulheres rastreadas.

SOBRERRASTREAMENTO POR FAIXA ETÁRIA, POR PERIODICIDADE E COBERTURA

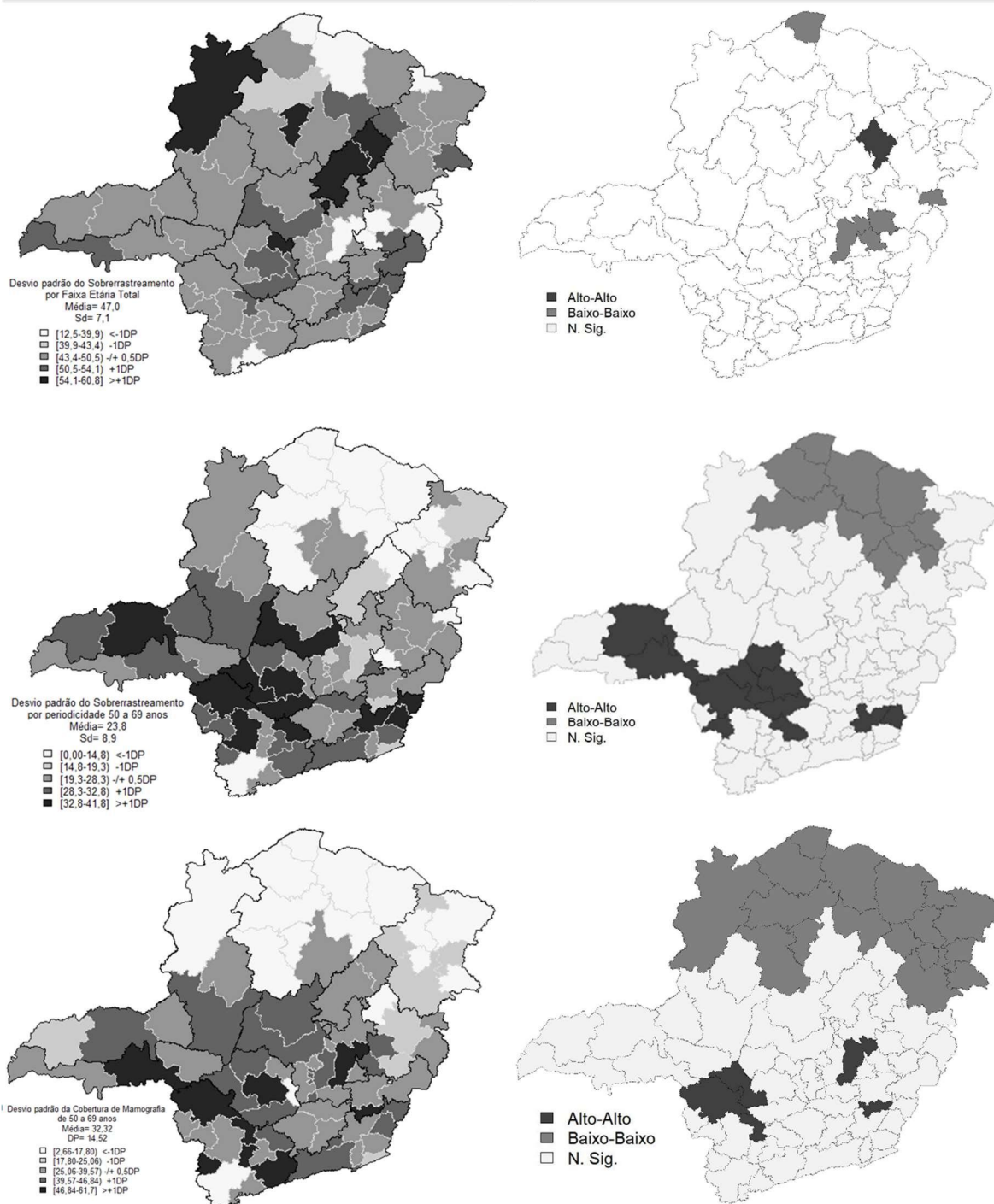


Figura 2 – Mapas temáticos sobrerrastreamento por por faixa etária, periodicidade e cobertura de mamografia (esquerda), índice de Moran local - LisaMap (direita), microrregiões de saúde de Minas Gerais, 2010.

Na análise da coorte do sobrerrastreamento por periodicidade foram consideradas 81.819 mulheres da faixa etária de 50 a 69 anos. Dessas, 22.584 repetiram a mamografia no

período de 18 meses, sendo o percentual de sobrerastreamento por periodicidade de 27,6% (IC95%: 27,3%-27,9%) (tabela 2).

Tabela 2 – Sobrerastreamento por periodicidade nas macrorregiões de saúde de Minas Gerais, 2010.

Macrorregiões	Sobrerastreamento por Periodicidade (%)	IC95%
Jequitinhonha	14,5	(12,4 - 16,5)
Norte	17,4	(15,9 - 18,8)
Leste	17,5	(16,4 - 18,6)
Nordeste	18,2	(16,4 - 19,9)
Leste do Sul	21,2	(19,9 - 22,5)
Centro Sul	26,0	(24,5 - 27,5)
Centro	26,6	(26,0 - 27,1)
Triângulo do Sul	30,0	(28,5 - 31,5)
Sul	30,3	(29,5 - 31,1)
Sudeste	31,2	(30,2 - 32,2)
Noroeste	31,5	(29,8 - 33,3)
Oeste	33,8	(32,5 - 35,0)
Triângulo do Norte	35,1	(33,9 - 36,3)
Minas Gerais	27,6	(27,3 - 27,9)

Log-rank: *p*-valor <0,001

Os maiores percentuais de sobrerastreamento por periodicidade foram observados no Triângulo do Norte, Oeste e Noroeste e os menores no Norte, Nordeste e Jequitinhonha (tabela 2). Um grande cluster de regiões com baixos percentuais de sobrerastreamento se formou nas regiões menos desenvolvidas, Norte, Nordeste e Jequitinhonha e outro de regiões com altos percentuais se formou nas regiões de melhor nível de desenvolvimento, Triângulo Norte e Sul, Oeste, Sul e Sudeste (figura 2).

A tabela 3 apresenta a distribuição do tempo entre a data do primeiro exame realizado em 2010 (exame de referência) e a data da primeira repetição (seguimento).

Tabela 3 – Distribuição do tempo em meses entre a data do primeiro exame realizado em 2010 e a data da primeira repetição.

	Valor Mínimo	Q1	Mediana	Média	Q3	Valor Máximo
Meses	1,0	17,0	28,0	24,5	31,0	36,0

Legenda: Q1: primeiro quartil; Q3: terceiro quartil.

Observa-se que a média de tempo entre o primeiro exame realizado em 2010 e a repetição posterior foi de 24,5 meses. Após o primeiro exame a probabilidade de uma mulher repetir a mamografia com 6 meses é de 1,6%, enquanto em um ano a probabilidade aumenta

para 8,5%. O gráfico da curva de sobrerastreamento apresenta elevação mais acentuada após um ano (365 dias) do primeiro exame (figura 3).

Quanto à cobertura, foram examinadas em 2010, 453.587 mulheres, sendo 233,903 (51,6%) da faixa etária de 50 a 69 anos. No estado, a cobertura de mamografia nessa faixa etária foi de 36,33%. Nas macrorregiões de saúde mais desenvolvidas e com menor vulnerabilidade na saúde foi observada maior cobertura, Oeste, Triângulo do Sul e Sudeste (44,96; 42,12 e 40,22% respectivamente). E menor cobertura nas macrorregiões menos desenvolvidas e com maior vulnerabilidade na saúde, Norte, Nordeste e Leste (18,03; 16,25 e 29,55%, respectivamente) (figura 2).

Sobrerastreamento por Periodicidade

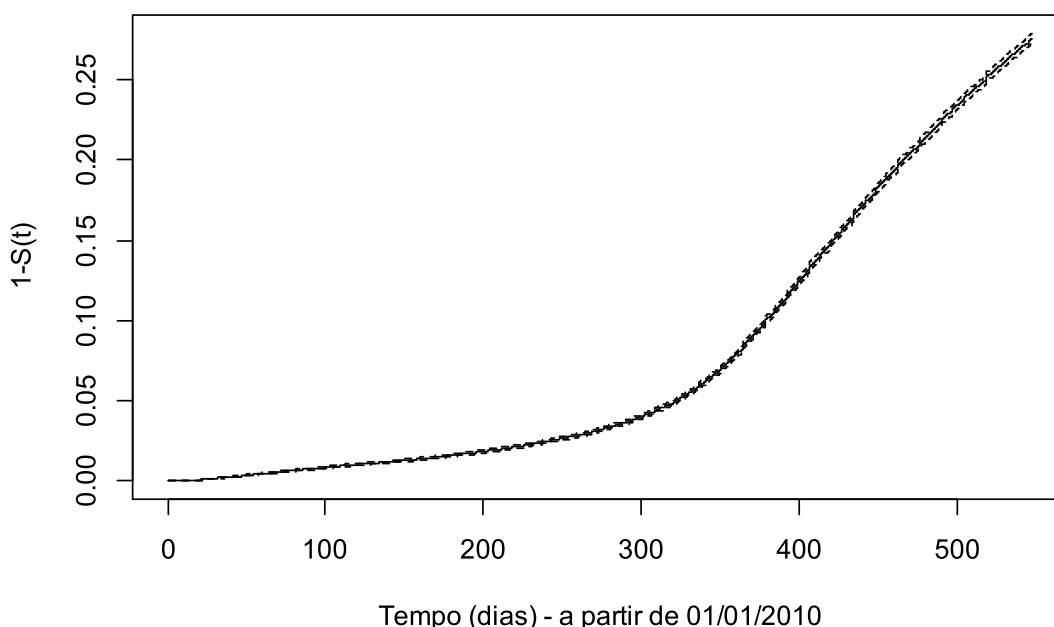


Figura 3 – Curva do sobrerastreamento por periodicidade (faixa etária de 50 a 69 anos).

Um grande conglomerado de microrregiões com baixa cobertura de mamografia de mulheres da faixa etária alvo que se formou nas regiões com piores indicadores socioeconômicos e assistenciais, Noroeste, Norte e Nordeste. Outro conglomerado de regiões de melhor nível socioeconômico e maior disponibilidade de recursos se formou nas regiões Oeste, Sul, Centro e Leste do Sul (figura 2).

Na análise de regressão nenhuma variável teve associação com o sobrerastreamento por faixa etária. Com o sobrerastreamento por periodicidade, todas as variáveis distais e proximais, exceto número de enfermeiros e distâncias percorridas para

realização de mamografia tiveram associação (tabela 3). O modelo múltiplo que apresentou melhor ajuste para explicar o sobrerastreamento por periodicidade foi o que incluiu as variáveis renda, índice de Gini e número de mamógrafos. Observou-se que para cada acréscimo de R\$ 1.000,00 na renda e de um mamógrafo (por x habitantes) para realização de exames, ocorreu um aumento de 19,39% e 3,15%, respectivamente, no sobrerastreamento por periodicidade. O índice de Gini se comportou de maneira inversa, demonstrando que conforme aumentaram as desigualdades, o sobrerastreamento diminuiu, ou seja, para cada 1% de acréscimo na desigualdade, houve uma redução 0,70% no sobrerastreamento por periodicidade (tabela 3).

Tabela 4– Modelos de Regressão Simples e Múltiplo entre o sobrerastreamento e indicadores sociodemográficos e assistenciais das microrregiões de saúde de Minas Gerais, 2010.

Variável (n=77)	Sobrerastreamento por faixa etária		Sobrerastreamento por Periodicidade
	Modelo Simples Coeficiente (valor de p)	Modelo Simples Coeficiente (valor de p)	Modelo Múltiplo Coeficiente (valor de p)
IDH (%)	0,19 (0,343)	1,27 (0,001)	
Renda (x1.000 BRL)	0,79 (0,864)	29,41 (< 0,001)	19,39 (<0,001)
IVS (%)	-0,01 (0,805)	-0,28 (< 0,001)	
Gini (%)	-20,17 (0,513)	-168,32 (< 0,001)	-0,70 (0,046)
Grau de Urbanização (%)	0,01 (0,818)	0,36 (< 0,001)	
Analfabetismo (%)	-0,11 (0,405)	-0,86 (< 0,001)	
População			
Médio porte	0,34 (0,871)	-2,72 (<0,267)	
Pequeno porte	-2,34 (0,364)	-10,03 (0,001)	
Médicos*	-0,546 (0,768)	8,41 (< 0,001)	
Enfermeiros*	2,10 (0,765)	11,38 (0,195)	
Radiologistas**	0,13 (0,793)	1,73 (0,004)	
Nº Mamógrafos**	0,66 (0,507)	5,07 (<0,001)	3,15 (0,002)
Cap. Util. Eq.***	1,36 (0,972)	0,05 (0,213)	
DPRMR(Km)	0,01 (0,954)	0,07 (0,158)	
DPRMF (Km)/1000	1,103 (0,882)	-0,99 (0,916)	
R ² ajustado			44,52
Autocorrelação dos resíduos			
Moran Global I			1,25
<i>p-valor</i>			(0,212)
Teste de Normalidade			
Jarque-Bera			2,71
<i>p-valor</i>			(0,258)
Teste de Heterocedasticidade			
Breusch-Pagan			2,96
<i>p-valor</i>			(0,398)

*Razão por 1.000 habitantes; **Razão por 100.000 habitantes; ***Capacidade Utilizada do Equipamento: A cada 1000 mamografias/mamógrafos; DPRMR: Distância Percorrida para Realização de Mamografia dentro da Região; DPRMF: Distância Percorrida para Realização de Mamografia fora da Região.

A partir dos mapas temáticos observa-se que as microrregiões com maiores taxas de sobrerastreamento tanto por faixa etária quanto por periodicidade são as que apresentaram maior cobertura de mamografia, menor desigualdade de renda, menor vulnerabilidade na saúde e maior número de médicos (figura 2).

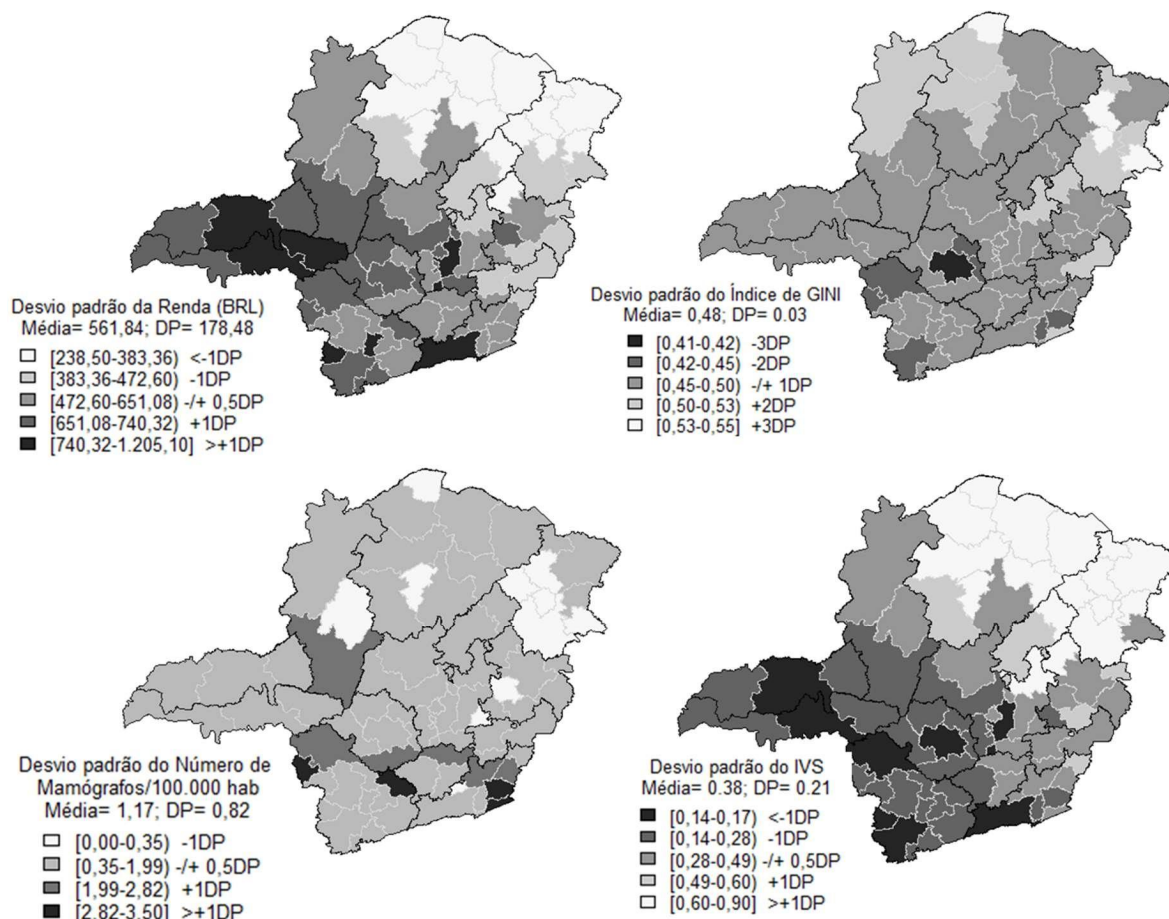


Figura 4 – Mapas temáticos dos indicadores sociodemográficos e assistenciais, microrregiões de saúde de Minas Gerais, 2010.

DISCUSSÃO

Em Minas Gerais, foi observado que a cobertura de mamografia na faixa etária alvo, 50 a 69 anos é baixa (36,33%), que o sobrerastreamento por faixa etária é elevado (47,1%) e se aproxima do percentual de mulheres examinadas na faixa etária de 50 a 69 anos (52,9%). O sobrerastreamento por periodicidade ocorre em 27,8% das mulheres rastreadas com elevação mais acentuada após um ano do primeiro exame. Maior renda, menor desigualdade de renda e maior número de mamógrafos se associaram ao sobrerastreamento por periodicidade,

sendo essas características apresentadas pelas macrorregiões Oeste Sul, Sudeste, Triângulo do Norte e do Sul.

O monitoramento das mulheres rastreadas é um dos benefícios chave de um programa de rastreamento, pois permite avaliar se as ações executadas seguem as diretrizes propostas e caso contrário, sejam adequadas visando maior qualidade e abrangência da população alvo (MARMARÁ; MARMARÁ; HUBBARD, 2018). No Brasil, conforme as recomendações da OMS e de países que possuem programas de rastreamento organizados, as diretrizes do programa de rastreamento do câncer de mama recomendam que a mamografia seja realizada em mulheres da faixa etária de 50 a 69 anos com periodicidade bienal (INCA, 2015). No entanto, nesse estudo, foi observado que quase metade das mamografias de rastreamento foram realizadas em mulheres fora da faixa etária alvo. Esse dado é corroborado por outro estudo também realizado em Minas Gerais que analisou número de exames e verificou que de 2010-2011, a maioria das macrorregiões de saúde apresentou percentual de aproximadamente 50% de mamografia de rastreamento em idades não preconizadas (CORRÊA et al., 2017).

O rastreamento com mamografia não é recomendado para mulheres com idade superior a 70 anos, pois estudos demonstram que o equilíbrio entre os possíveis danos e benefícios ainda são incertos, dessa maneira, não existem evidências fortes que apoiem essa prática. Quanto ao rastreio de mulheres jovens, estudos demonstraram balanço desfavorável entre seus riscos e benefícios devido à densidade mamária, fator que diminui a sensibilidade do exame e pode contribuir com maior número de resultados falso-positivos e sobretratamento, principalmente se o exame for realizado anualmente (INCA, 2015; MIGOWSKI et al., 2018a).

O sobrerastreamento por periodicidade nesse estudo foi calculado para as mulheres da faixa etária de 50 a 69 anos, mas vale destacar que no banco de mulheres examinadas existem registros de mulheres com idade inferior a 50 anos e maior ou igual a 70 anos que repetiram o exame com menos de 18 meses (dados não apresentados). A repetição de exames em intervalo inferior ao preconizado além de não oferecer maior benefício, aumenta a exposição à radiação ionizante e pode contribuir com o sobrediagnóstico, ou seja, detecção de um tumor que não terá evolução para ser confirmado clinicamente durante a vida da paciente, consequentemente essa mulher será submetida a tratamentos desnecessários (MORRIS et al., 2015). Mamografias com resultados falso-positivos são mais comuns em exames anuais do que bienais. O rastreamento bienal comparado ao anual pode reduzir em

19% a chance de uma mulher receber um resultado falso-positivo e de 1-4% do risco de biópsia desnecessária (NELSON et al., 2009; INCA, 2015; SIU, 2016).

O menor intervalo entre mamografias de rastreamento pode refletir falhas na coordenação do cuidado da mulher na ESF ou na organização das redes de atenção à saúde que prejudicam o acesso à mamografia, uma vez que os exames que poderiam ser direcionados à ampliação da cobertura são empregados em mulheres fora da faixa etária alvo e em período menor do que os preconizados pelo MS (RODRIGUES et al, 2019; RAMOS et al, 2018).

O papel de uma diretriz é orientar gestores, profissionais de saúde e usuários acerca da melhor prática a partir das melhores evidências científicas disponíveis. No Brasil, as primeiras diretrizes do programa de rastreamento do câncer, do ano de 2004, eram baseadas na opinião de especialistas. A nova diretriz de 2015 inovou nesse processo ao reforçar e validar as recomendações existentes a partir de evidências que discutiam sobre faixa etária, periodicidade e balanço entre riscos e benefícios de cada intervenção e ao integrar às recomendações o rastreamento de mulheres sintomáticas, demonstrando a importância da sua aplicação (MIGOWSKI et al., 2018a).

No estado, assim como nas macros e microrregiões, foi observado que a cobertura de mamografia não se aproxima da meta considerada ideal pela Organização Mundial de Saúde (70%-75%), e as microrregiões com maiores coberturas situam-se nas regiões mais desenvolvidas e com maiores taxas de sobrerastreamento. Destaca-se que em Minas Gerais, aproximadamente 25% das mulheres possuem plano de saúde e aquelas que residem nas regiões com maior renda tem maior probabilidade de possuir plano de saúde, assim é possível que as taxas de sobrerastreamento sejam ainda maiores nessas regiões.

Em contrapartida, microrregiões com menores cobertura e menor sobrerastreamento situam-se nas macrorregiões que apresentam maior vulnerabilidade na saúde e desigualdade de renda. Cabe lembrar que por possuir grande extensão territorial e diversidade regional. Minas Gerais é considerada uma representação da estrutura regional brasileira, com regiões menos desenvolvidas ao Norte/Nordeste e mais desenvolvidas ao Sul/Sudeste. Segundo a Pesquisa Nacional de Saúde, no Brasil, em 2013, mais de 50% das mulheres residentes no Norte e com idade entre 50 e 69 anos nunca realizaram mamografia de rastreamento (THEME-FILHA et al., 2016) e no país, a razão de cobertura de mamografia na faixa etária alvo do rastreamento é baixa, sendo as menores coberturas observadas nas

regiões economicamente mais pobres, Norte e Nordeste e as mais elevadas naquelas mais ricas, Sudeste e Sul (AZEVEDO e SILVA et al., 2014).

As regiões de saúde de Minas Gerais menos desenvolvidas possuem menor número de médicos e enfermeiros, profissionais responsáveis pela solicitação de mamografia. Estudo observou que no Brasil, mulheres residentes nas regiões Norte e Nordeste, com menor escolaridade e que não consultaram com médico no último ano, tendem a não realizar o exame de mamografia num período de até 2 anos (BARBOSA et al., 2019). A baixa disponibilidade de recursos humanos e equipamentos limitam o acesso ao rastreamento (NOGUEIRA et al., 2019), de outro modo, o maior acesso aos recursos de saúde não só promove o uso adequado da mamografia de rastreamento, mas também está associado positivamente ao uso excessivo (TAN; KUO; GOODWIN, 2014).

Estudo realizado em um município de Minas Gerais sobre conhecimento, atitudes e práticas dos profissionais médicos e enfermeiros da ESF nas ações de prevenção e detecção precoce do câncer de mama, verificou que 75,6% tinham conhecimento adequado sobre a faixa etária e periodicidade recomendada para o rastreamento mamográfico (FERREIRA, 2020). No entanto, outros estudos (MELO et al., 2017; MORAES et al., 2016; FONSECA et al., 2016; JÁCOME e BARBOSA, 2011), relatam que mais da metade dos profissionais indicam mamografia a partir dos 40 anos e em São Paulo, região sudeste, 72,9% dos profissionais recomendam a periodicidade anual (MELO et al., 2017).

As diferentes opiniões das sociedades de especialidades médicas acerca da periodicidade e faixa etária de início do rastreamento, o desconhecimento e a aceitação por parte dos profissionais de saúde das diretrizes e recomendações do Ministério da Saúde, dificultam a adequada implementação e adesão ao rastreamento mamográfico (SANTOS; RAMOS; MIGOWSKI, 2019). Ademais, profissionais de saúde em geral podem ser influenciados por recomendações que são mais difundidas ou legitimadas por sociedade médicas ou líderes de opinião, uma vez que alguns podem não possuir formação para identificar potenciais vieses e discernir sobre a validade de determinada orientação (MIGOWSKI et al., 2018c). Dessa forma, regiões com maior número de profissionais que não seguem as recomendações nacionais tem maior probabilidade de realizar o sobrerastreamento.

Por outro lado, regiões menos desenvolvidas como Norte e Nordeste, enfrentam o desafio da alta rotatividade de profissionais na atenção básica, fator que compromete o rastreamento porque o intervalo até a contratação de um profissional pode demorar e esse

ainda pode precisar de capacitação acerca da organização do serviço e recomendações para o rastreamento (SANTOS; RAMOS; MIGOWSKI, 2019).

Em Minas Gerais foi observada maior disponibilidade de mamógrafos nas áreas mais desenvolvidas e com menor desigualdade de renda e vulnerabilidade na saúde. Destaca-se que essas macrorregiões não utilizam toda a sua capacidade de realização de exames e ainda realizam sobrerastreamento, principalmente o Sul e Oeste. No Brasil, estudos analisaram o quantitativo de mamógrafos e constataram que embora tenha ocorrido no período de 2008 a 2013 uma expansão do número de equipamentos financiados pelo SUS, as regiões Norte e Nordeste, menos desenvolvidas, ainda apresentam menor número quando comparadas à região Sul, mais desenvolvida; e que apesar da oferta de mamografia ser suficiente, a cobertura de mulheres na faixa etária preconizada continua baixa (VIACAVA et al., 2019; TOMAZELLI e AZEVEDO e SILVA, 2017).

Microrregiões com menor disponibilidade de recursos humanos e equipamentos, maior desigualdade de renda, maior vulnerabilidade na saúde e menor cobertura de mamografia apresentaram sobrerastreamento por faixa etária e periodicidade superiores a 50% e 20%, respectivamente. Este fato chama atenção porque além de possuírem barreiras relacionadas ao sistema de saúde que prejudicam o rastreamento, como desigualdade de acesso, número insuficiente de profissionais aptos a solicitar e laudar mamografias e distribuição desigual de mamógrafos no território, essas regiões ainda realizam exames em faixas etárias não preconizadas e em intervalo inferior ao que é orientado pelas diretrizes de rastreamento do câncer de mama.

Este estudo foi desenvolvido a partir de bases secundárias de dados, assim apresenta limitações relacionadas à possíveis erros de preenchimento e completude das informações. Apesar disso, a avaliação do sobrerastreamento não foi prejudicada, visto que a técnica de relacionamento probabilístico e classificação automatizada dos registros (DUARTE et al., 2019) possibilitou a identificação de um único exame por mulher e as medidas de avaliação da qualidade dos dados vinculados apresentaram nesse estudo valores superiores a 98%, o que demonstra que as chances de classificação de um par falso como verdadeiro foram minimizadas.

No Brasil, a política de controle do câncer de mama publicada no documento de consenso, em 2004 (INCA), teve seu foco no rastreamento mamográfico para as mulheres de 50 a 69 anos e incluía a realização de exame clínico das mamas, anualmente, para as mulheres com 40 anos ou mais. Ainda assim, tais esforços não resultaram na diminuição da

alta prevalência de mulheres diagnosticadas em estadiamento avançado (III e IV), que foi de aproximadamente 40%, segundo o estudo de Santos-Silva e colaboradores (2018). Tal estudo confirmou ainda a grave situação da desigualdade no país, pois tal prevalência foi significativamente maior nas mulheres de cor preta ou parda com baixa ou nenhuma escolaridade (48,8%) e menor para as mulheres brancas com nível universitário (29,4%) (SANTOS-SILVA, 2018). As atuais diretrizes brasileiras para a detecção precoce do câncer de mama, em vigência a partir de 2015, baseiam-se em duas estratégias: rastreamento e diagnóstico precoce. A estratégia para o diagnóstico precoce das mulheres sintomáticas se desenvolve em um tripé que contempla: população capaz de identificar os sinais e sintomas suspeitos de câncer de mama; profissionais de saúde capacitados para identificar e reconhecer os casos suspeitos; e sistema e serviços de saúde capazes de fornecer a confirmação do diagnóstico e garantir assistência integral em toda a linha de cuidado. Já o rastreamento utiliza a mamografia em população-alvo de mulheres sadias, com o intuito de identificar doenças em sua fase pré-clínica (assintomática).

Por meio desse estudo, foi possível verificar que a recomendação de periodicidade e faixa etária para realização de mamografia não tem sido adotada pelas microrregiões de saúde de Minas Gerais e que o sobrerastreamento resultante está associado de forma direta à cobertura e reflete as desigualdades regionais existentes no estado. Torna-se urgente a necessidade de reforçar a implementação das diretrizes vigentes no país, ou seja, garantir o diagnóstico precoce para as mulheres sintomáticas e a implantação de um rastreamento mamográfico organizado, que permita o acesso ao diagnóstico e tratamento, em tempo oportuno. Deve-se focar na capacitação dos profissionais, sobretudo acerca das evidências que embasaram a elaboração das diretrizes adotadas no Brasil e por países com programas de rastreamento organizado; na sensibilização da população para a detecção precoce; na alocação de recursos de acordo com a necessidade e característica de cada região e na organização e otimização das ações e recursos, com vistas a reduzir a desigualdade de acesso e ampliar a cobertura, principalmente nas regiões mais carentes. Somente desta forma poderemos assegurar a detecção precoce da doença e garantir que as pacientes se beneficiem das possibilidades terapêuticas disponíveis, trazendo impactos na qualidade de vida e redução da mortalidade por câncer de mama.

REFERÊNCIAS

AZEVEDO E SILVA, G. et al. Acesso à detecção precoce do câncer de mama no Sistema Único de Saúde: uma análise a partir dos dados do Sistema de Informações em Saúde. **Cadernos de Saúde Pública**, v. 30, n. 7, p. 1537–1550, 2014.

BARBOSA, Y. C.; et al. Fatores associados à não realização de mamografia: Pesquisa Nacional de Saúde, 2013. **Revista Brasileira de Epidemiologia**, v. 22, e 190069, 2019.

CORREA, C. S. L.; PEREIRA, L. C.; LEITE, I. C.G.; FAYER, V. A.; GUERRA, M. R.; BUSTAMANTE-TEIXEIRA, M. T. Rastreamento do câncer de mama em Minas Gerais: avaliação a partir de dados dos sistemas de informações do Sistema Único de Saúde. **Epidemiologia e Serviços de Saúde**, v. 26, n. 3, p. 481-492, 2017.

DUARTE, D. A. P.; et al. Avaliação de método para classificação automatizada de pares em relacionamentos probabilísticos de bancos de dados. **Cadernos de Saúde Pública**, v. 35, n. 11, p. e00066419, 2019.

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 (MG): Universidade Federal de Juiz de Fora; 2020.

FONSECA, D. C. O.; SILVA, F. T.; SANTOS, N. M. G.; DAVIM, R. M. B. Ações na prevenção do exame de câncer de mama na consulta do enfermeiro. **Revista de Enfermagem da UFPE**, v. 10, n. 12, p. 4563-4571, 2016.

INCA. Instituto Nacional do Câncer. Diretrizes para a detecção precoce do câncer de mama no Brasil. Rio de Janeiro: INCA, 2015a. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//diretrizes_deteccao_precoce_cancer_mama_brasil.pdf>. Acesso em jun 2019.

JÁCOME, E. M.; BARBOSA, I. L. Breast Cancer Detection: Knowledge, Attitude and Practices of Doctors and Nurses from the Family Health Strategy of Mossoró, RN, Brazil. **Revista Brasileira de Cancerologia**, v. 57, n. 2, p. 189-198, 2011.

LOBERG, M.; et al. Benefits and harms of mammography screening. **Breast Cancer Research**, v. 17, n. 63, p. 1-12, 2015.

MARMARÁ, D.; MARMARÁ, V.; HUBBARD, G.; A national cross-sectional study of adherence to timely mammography use in Malta. **BMC Cancer**, v. 18, n. 346, p. 1-12, 2018.

MELO, F. B. B.; et al. Actions of nurses in early detection of breast cancer. **Revista Brasileira de Enfermagem**, v. 70, n. 6, p. 1119-1128, 2017.

MIGOWSKI A.; et al. Diretrizes para detecção precoce do câncer de mama no Brasil. II – Novas recomendações nacionais, principais evidências e controvérsias. **Cadernos de Saúde Pública**, v. 34, n. 6, p. e00074817, 2018.

MIGOWSKI A.; et al. Diretrizes para detecção precoce do câncer de mama no Brasil. III – Desafios à implementação. **Cadernos de Saúde Pública**, v. 34, n. 6, p. e00046317, 2018c

MORAES, D. C.; et al. Opportunistic screening actions for breast cancer performed by nurses working in primary health care. **Revista da Escola de Enfermagem da USP**, v.50, n. 1, p. 14–21, 2016.

MORRIS, E.; FEIG, A. S.; DREXLER, M.; LEHMAN, C. Implications of Overdiagnosis: Impacto on Screening Mammography Practices. *Population Health Management*, v. 18, s. 1, s1-s11, 2015.

NELSON, H. D.; et al. Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. *Annals of Internal Medicine*, v. 164, n. 4, p. 257-268.

NOGUEIRA, M. C.; et al. Inequities in access to mammographic screening in Brazil. **Cadernos de Saúde Pública**, v. 35, n. 6, e0009981, 2019.

PEREIRA, D. A. Avaliação do Rastreamento do Câncer de Mama em Minas Gerais a partir de Dados do Sismama. [tese]. Juiz de Fora (MG): Universidade Federal de Juiz de Fora; 2021.

RAMOS, A. C. V.; et al. Estratégia Saúde da Família, saúde suplementar e desigualdade no acesso à mamografia no Brasil. **Revista Panamericana Salud Publica**, v. 42, p. 1-9, 2018.

RODRIGUES, T. B. Sobrerrastreio mamográfico: avaliação a partir de bases identificadas do Sistema de Informação do Câncer de Mama (SISMAMA). **Cadernos de Saúde Pública**, v. 35, n. 1, e00049718, 2019.

SANTOS, R. O. M.; RAMOS, D. N.; MIGOWSKI, A. Barreiras na implementação das diretrizes de detecção precoce dos cânceres de mama e colo do útero no Brasil. **Physis: Revista de Saúde Coletiva**, v. 29, n. 4, p. e290402, 2019.

SANTOS-SILVA, I. et al. Ethnoracial and social trends in breast cancer staging at diagnosis in Brazil, 2001-14: a case only analysis. **Lancet Glob Health**, v. 7, p. e784-e797, 2019.

SIU, A. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. **Annals of Internal Medicine**, v. 164, n. 4, p. 279-296, 2016.

SUNG, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA Cancer Journal Clinicians**, Nova York, v. 71, n. 3, p. 209-249, 2021.

TAN, A.; KUO, Y. F.; GOODWIN, J. S. Potential Overuse of Screening Mammography and Its Association with Access to Primary Care. **Med Care**, v. 52, n. 6, p. 490-495, 2014.

THEME FILHA, M. M. et al. 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. 1–8, 2016.

TOMAZELLI, J. G.; AZEVEDO E SILVA, G. Rastreamento do câncer de mama no Brasil: uma avaliação da oferta e utilização da rede assistencial do Sistema Único de Saúde no período 2010-2012. **Epidemiologia e Serviços de Saúde**, v. 26, n.4, p. 713-724, 2017.

VIACAVA, F.; et al. Desigualdades regionais e sociais em saúde segundo inquéritos domiciliares (Brasil, 1998-2013). **Ciência e Saúde Coletiva**, v. 24, n. 7, p. 2745-2760, 2019.

WENDER, R. C. et al. A blueprint for cancer screening and early detection: Advancing screening's contribution to cancer control. **CA: A Cancer Journal for Clinicians**, v. 69, n. 1, p. 50–79, 2019.

CONSIDERAÇÕES FINAIS

A cobertura de mamografia de mulheres na faixa etária alvo, em Minas Gerais, é baixa e fatores socioeconômicos e assistenciais, tais como renda, desigualdade de riqueza e saúde, se associam com a cobertura de mamografia, ou seja, mulheres residentes em microrregiões com menor renda, maior desigualdade de riqueza e maior vulnerabilidade na saúde apresentam menor cobertura. Além disso, a desigualdade de renda presente nas macrorregiões mais ricas contribui para que em determinadas microrregiões a cobertura seja menor quando comparadas às demais. Considerando que todos os mamógrafos existentes em Minas Gerais estão com a sua manutenção em dia e em condições adequadas de funcionamento, conclui-se que a baixa cobertura não pode ser atribuída à falta do equipamento e que o número é suficiente para atendimento das demandas do estado, porém fatores como má distribuição dos mamógrafos, sobrerastreamento por faixa etária e/ou periodicidade e ausência e/ou menor número de profissionais aptos a solicitar mamografias em algumas microrregiões, prejudicam o acesso ao exame.

A recomendação de periodicidade e faixa etária para realização de mamografia não tem sido adotada pelas microrregiões de saúde de Minas Gerais e o sobrerastreamento está associado às desigualdades regionais existentes no estado, ou seja, a prevalência deste é maior nas regiões de melhor nível socioeconômico, maior número de recursos humanos e equipamentos, menor vulnerabilidade na saúde e maior cobertura de mamografia, assim nos locais onde existe maior disponibilidade de recursos, o sobrerastreamento ocorre com mais frequência, principalmente por periodicidade. Além disso, chama atenção o fato de que microrregiões com limitação de recursos humanos e equipamentos, maior vulnerabilidade na saúde, menor renda e cobertura de mamografia apresentam percentuais consideráveis de sobrerastreamento tanto por faixa etária quanto por periodicidade.

A automatização do relacionamento dos bancos de dados mantendo a acurácia do método, com base na inspeção manual de uma amostra retirada do próprio banco e posterior seleção de um escore de alto poder discriminatório na identificação de pares verdadeiros dessa amostra, diminuiu o tempo dispendido na etapa de identificação dos pares formados no relacionamento, otimizando o processo. Relacionamentos probabilísticos e que utilizam classificação automatizada dos registros vinculados não são totalmente perfeitos e alguns pares falsos podem ser classificados como verdadeiros, porém, existem medidas que avaliam a qualidade dessa técnica, principalmente nos casos de dados desequilibrados, sendo elas

precision, *recall* e medida F. No caso desse estudo, essas medidas apresentaram valores superiores a 98%. Assim, o método desenvolvido deu suporte para criação dos bancos com único registro de mamografia por mulher que foram utilizados na análise da cobertura e do sobrerastreamento.

Ainda que o SISMAMA ofereça apenas informações sobre exames realizados, é possível, por meio de um relacionamento probabilístico, identificar o número de mulheres atendidas, calcular a cobertura do rastreamento do câncer de mama e realizar o seguimento das mulheres rastreadas e registradas nesse sistema.

Os dados desse estudo, corroboram as características oportunísticas do rastreamento do câncer de mama em Minas Gerais: baixa cobertura, realização de exames de rastreamento em faixas etárias não preconizadas e em intervalo inferior ao recomendado e não utilização de toda a capacidade instalada do mamógrafo, demonstrando a necessidade de adequações. O sobrerastreamento onera o sistema de saúde, traz prejuízos para a mulher devido ao sobrediagnóstico e tratamento e compromete o alcance do resultado esperado em programas de rastreamento do câncer de mama, redução do número de casos diagnosticados em estágios avançados e menor mortalidade por esta causa.

É urgente a necessidade de adequações das ações do programa e estas devem focar no planejamento e implantação do rastreamento organizado para mulheres assintomáticas e facilitação do acesso para mulheres sintomáticas, melhor alocação dos mamógrafos entre as microrregiões do estado visando atender as regiões em que estes inexistem ou são insuficientes, distribuição de recursos humanos de acordo com a necessidade e característica de cada região, organização e otimização das ações e recursos, principalmente nas microrregiões com baixa cobertura de mamografia e que não utilizam toda a sua capacidade de realização de exames e capacitação dos profissionais e orientação da população quanto as diretrizes do rastreamento, demonstrando que as recomendações são embasadas por evidências e que também são adotadas por outros países. Essas mudanças e adequações podem contribuir para uma melhor aplicação dos recursos do SUS e para que as mulheres da faixa etária alvo tenham acesso à mamografia de acordo com periodicidade adequada e mulheres sintomáticas tenham acesso ao diagnóstico o mais precocemente possível.

Por fim, espera-se que as lacunas e fragilidades encontradas nesse estudo contribuam para que ações do programa de rastreamento no país sejam avaliadas e se necessário passem por adequações. E que este estudo sirva de base e direcionamento para estudos em regiões que precisam levantar mais detalhes sobre a execução de seus programas

com a finalidade de planejar, melhorar e ampliar as ações futuras, conforme as recomendações do Ministério da Saúde.

REFERÊNCIAS

ACR. American College of Radiology. ACR BI-RADS® Follow-up and Outcome Monitoring. In: **ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System**. Reston, VA, American College of Radiology; 2013.

ACOG. American College of Obstetricians and Gynecologists. Breast Cancer Risk Assessment and Screening in Average-Risk Women. ACOG Practice Bulletin, n. 179, 2017. Disponível em: <<https://www.acog.org/Clinical-Guidance-and-Publications/Practice-Bulletins/Committee-on-Practice-Bulletins-Gynecology/Breast-Cancer-Risk-Assessment-and-Screening-in-Average-Risk-Women>>. Acesso em abr. 2019.

ADAMS, E. K. et al. Enhancing screening and early detection among women transitioning to Medicare from the NBCCEDP in Georgia. **HHS Public Access**, v. 26, n. 5, p. 795–803, 2015.

ADIB, S. M. et al. Guidelines for breast cancer screening in Lebanon. **Lebanese Medical Journal**, v. 57, n. 2, p. 72–74, 2009.

ALBRECHT, C. A. M. et al. Mortalidade por câncer de mama em hospital de referência em oncologia , Vitória , ES Breast cancer mortality among. **Revista Brasileira de Epidemiologia**, v. 16, n. 3, p. 582–591, 2013.

ALTOBELLI, E.; LATTANZI, A. Breast cancer in European Union : An update of screening programmes as of March 2014 (Review). **International Journal of Oncology**, v. 2014, n. March, p. 1785–1792, 2014.

AMORIM, V. M. S. L. et al. Fatores associados a não realização da mamografia e do exame clínico das mamas: um estudo de base populacional em Campinas, São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 24, n. 11, p. 2623-2632, 2008.

ANDERS, W. Costa Rica Invests Over \$15M in Breast Cancer Prevention Program. **The Costa Rica Star**. 2017. Disponível em: <<https://news.co.cr/costa-rica-invests-over-15m-in-breast-cancer-prevention-program/65701/>>. Acesso em abr. 2019.

ANS. Agência Nacional de Saúde Suplementar. Dados do Setor. Beneficiários de planos privados de saúde. Disponível em: <<http://www.ans.gov.br/perfil-do-setor/dados-e-indicadores-do-setor>>. Acesso em mai 2019.

ANOTHASINTAWEE, T.; Risk Factors of Breast Cancer: A Systematic Review and Meta-Analysis. **Asia-Pacific Journal of Public Health**, v. 25. n. 5, 368-387, 2013.

ARIAS-ORTIZ, N. E.; VRIES, E. DE. Original articles Health inequities and cancer survival in Manizales , Colombia : a population-based study. **Colombia Médica**, v. 49, n. 6, p. 63–72, 2018.

ASCUNCE, N. et al. original article Cancer screening in Spain. **Annals of Oncology**, v. 21, n. Supplement 3, p. 43–51, 2010.

AUGUSTYNOWICZ, A. et al. Breast cancer and cervical cancer prevention programmes carried out by local government units in Poland in 2009 – 2014. **Oncotarget**, v. 9, n. 31, p. 21943–21949, 2018.

AUTIER, P.; BONIOL, M.; KOEHLIN, A. Effectiveness of and overdiagnosis from mammography screening in the Netherlands: population based study. **BMJ**, v. 359, n. j5224, p. 1–20, 2017.

AZEVEDO E SILVA, G. et al. Acesso à detecção precoce do câncer de mama no Sistema Único de Saúde : uma análise a partir dos dados do Sistema de Informações em Saúde. **Cadernos de Saúde Pública**, v. 30, n. 7, p. 1537–1550, 2014.

AZEVEDO e SILVA, G.; et al. 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.

BATTESINI, M.; ANDRADE, C. L. T.; SETA, M. H. DE. Financiamento federal da Vigilância Sanitária no Brasil de 2005 a 2012 : análise da distribuição dos recursos. **Ciência & Saúde Coletiva**, v. 22, n. 10, p. 3295–3306, 2017.

BEAU, A. et al. Benefit-to-harm ratio of the Danish breast cancer screening programme. **International Journal of Cancer**, v. 141, n. 3, p. 512–518, 2017.

BECKMANN, K. R. et al. diagnosis of breast cancer due to organised population-based mammography screening. **International Journal of Cancer**, v. 136, n. 6, p. 1411–1421, 2015.

BELLANGER, M. et al. Are Global Breast Cancer Incidence and Mortality Patterns Related to Country- Specific Economic Development and Prevention Strategies ? **Journal of Global Oncology**, v. 4, p. 1–16, 2018.

BIESHEUVEL, C.; WEIGEL, S.; HEINDEL, W. Breast Care Mammography Screening : Evidence , History and Current Practice in Germany and Other European Countries. **Breast Care**, v. 6, p. 104–109, 2011.

BILLER-ADORNO; JUNI, P. The New engla nd journal. **Perspective**, v. 370, n. 21, p. 1965–1967, 2014.

BOLIVIA. Ministerio de Salud y Deportes. Dirección General de Servicios de Salud. Unidad de Redes de Servicios de Salud y Calidad. Guía de tamizaje de cáncer de cuello uterino y mama./Ministerio de Salud y Deportes. La Paz : Sistemas Gráficos Color, 2013.

BONCZ, I. et al. The effect of an organized, nationwide breast cancer screening programme on non-organized mammography activities. **Journal of Medical Screening**, v. 15, n. 1, p. 14–17, 2008.

_____. Secretaria de Vigilância em Saúde. Guia de vigilância epidemiológica / Ministério da Saúde, Secretaria de Vigilância em Saúde. 6. ed. Brasília: Ministério da Saúde, 2005.

_____. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Controle dos cânceres do colo do útero e da mama / Secretaria de Atenção à Saúde, Departamento de Atenção Básica. Brasília: Ministério da Saúde, 2006a.

_____. Secretaria Executiva. Departamento de Apoio à Descentralização. Coordenação Geral de Apoio à Gestão Descentralizada. Diretrizes operacionais dos Pactos pela Vida, em Defesa do SUS e de Gestão. V.1. Brasília: Ministério da Saúde, 2006b.

_____. Portaria nº 779 de 31 de dezembro de 2008. Institui o Sistema de Informação do Câncer de Mama (SISMAMA). Diário Oficial da União, Brasília, DF, p. 38.

_____. Portaria nº 215 de 25 de junho de 2009. Prorroga o prazo para início da utilização exclusiva do SISMAMA para faturamento dos exames. Diário Oficial da União, Brasília, DF, p. 68.

_____. Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Rastreamento / Ministério da Saúde, Secretaria de Atenção à Saúde, Departamento de Atenção Básica. – Brasília : Ministério da Saúde, 2010.

_____. 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). Diário Oficial da União, Brasília, DF, p. 57.

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

_____. Ministério da Saúde. Secretaria de Vigilância em Saúde. Plano de ações estratégicas para o enfrentamento das doenças crônicas e agravos não transmissíveis no Brasil 2021-2030. Brasília: Editora do Ministério da Saúde, 2020. Disponível em: <<https://antigo.saude.gov.br/images/pdf/2020/October/01/Plano-DANT-vers--o-Consulta-p-blica.pdf>>. Acesso mai 2021.

BRAY, F.; et al. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. **CA: A Cancer Journal for Clinicians**, v. 68, p. 394-424, 2018.

BRINTON, L. et al. Factors Contributing to Delays in Diagnosis of Breast Cancers in Ghana, West Africa. **Breast Cancer Res Treat**, v. 162, n. 1, p. 105–114, 2017.

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

CAMARGO JR., K. R. DE; COELI, C. M. RecLink 3: nova versão do programa que implementa a técnica de associação probabilística de registros (probabilistic record linkage). **Cadernos de Saúde Coletiva**, v. 14, n. 2, p. 399-404, 2006.

CAPUANI, L. et al. 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.

CARIOLI, G. et al. Trends and predictions to 2020 in breast cancer mortality : Americas and Australasia. **The Breast**, v. 37, p. 163–169, 2018.

CARNEY, P.; NEILL, C. O. Income inequality in uptake of voluntary versus organised breast cancer screening : evidence from the British Household Panel Survey. **BMC Pub**, v. 18, n. 252, p. 1–7, 2018.

CHOI, E. et al. Socioeconomic Inequalities in Cervical and Breast Cancer Screening among Women in Korea , 2005 – 2015. **Yonsei Medical Journal**, v. 59, n. 9, p. 1026–1033, 2018.

CHONGTHAWONSATID, S. Inequity of healthcare utilization on mammography examination and Pap smear screening in Thailand : Analysis of a population-based household survey. **PLoS ONE**, v. 12, n. 3, p. 1–12, 2017.

CNDSS. Comissão Nacional sobre Determinantes Sociais da Saúde. As causas sociais das iniquidades em saúde no Brasil. Rio de Janeiro: Editora Fiocruz, 2008. 215p.

CORREA, C. S. L.; PEREIRA, L. C.; LEITE, I. C.G.; FAYER, V. A.; GUERRA, M. R.; BUSTAMANTE-TEIXEIRA, M. T. Rastreamento do câncer de mama em Minas Gerais: avaliação a partir de dados dos sistemas de informações do Sistema Único de Saúde. **Epidemiologia e Serviços de Saúde**, v. 26, n. 3, p. 481-492, 2017.

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.

CTFPHC. Canadian Task Force on Preventive Health Care. Breast cancer screening: protocol for an evidence report to inform an update of the Canadian Task Force on Preventive Health Care 2011 Guidelines. 2017. Disponível em: <<https://canadiantaskforce.ca/wp-content/uploads/2017/03/BCU-protocol-14Mar2017-FINAL.pdf>>. Acesso em abr. 2019.

DA MATTA COUTINHO, R. G.; COELI, C. M.; FAERSTEIN, E.; CHOR, D. Sensitivity of probabilistic record linkage for reported birth identification: Pró-Saúde Study. **Revista de Saúde Pública**, v. 42, n. 6, p. 1097–1100, 2008.

DATASUS. Departamento de Informática do SUS. Informações de Saúde. Demográficas e Socioeconômicas. 2019. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?area=0206>>. Acesso em mai 2019.

DI SIBIO, A. J. Efectividad del tamizaje mamográfico en la reducción de la mortalidad por cáncer de mama / Alejandro Javier Di Sibio. - 1a ed. - Ciudad Autónoma de Buenos Aires : Instituto Nacional del Cáncer, 2018.

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. **Revista Pesquisa e Planejamento Econômico**, Rio de Janeiro, v. 39, n. 1, p. 121-158, 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 (MG): Universidade Federal de Juiz de Fora; 2016.

DUARTE, D. A. P.; NOGUEIRA, M. C.; MAGALHÃES, M. C.; BUSTAMANTE-TEIXEIRA, M. T. Iniquidade social e câncer em mulheres: análise da mortalidade por câncer de mama. **Cadernos de Saúde Coletiva**, v. 28, n. 4, p. 465-476, 2020.

EQUADOR. Ministerio de Salud Pública. Información general cáncer de mama. Disponível em: < <https://www.salud.gob.ec/cancer-de-mama/>>. Acesso em abr. 2019.

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 (MG): Universidade Federal de Juiz de Fora; 2020.

FERREIRA, N. A. S. et al. Treatment delays among women with breast cancer in a low socio-economic status region in Brazil. **BMC Women's Health**, v. 17, n. 13, p. 4–11, 2017.

FIGUEIREDO, F. W. DOS S.; et al. The role of health policy in the burden of breast cancer in Brazil.” **BMC Women's Health**, v. 17 n. 121, p. 1-6, 2017.

FIGUEIREDO, F. W. DOS S.; ADAMI, F. Effects of the high-inequality of income on the breast cancer mortality in Brazil. **Scientific Reports**, v. 9, n. 1, p. 8–13, 2019.

FONSECA, M. G. P. et al. Accuracy of a probabilistic record linkage strategy applied to identify deaths among cases reported to the Brazilian AIDS surveillance database. **Cadernos de Saúde Pública**, v. 26, n. 7, p. 1431–1438, 2010.

FONSECA, D. C. O.; SILVA, F. T.; SANTOS, N. M. G.; DAVIM, R. M. B. Ações na prevenção do exame de câncer de mama na consulta do enfermeiro. **Revista de Enfermagem da UFPE**, v. 10, n. 12, p. 4563-4571, 2016.

FRANCO, J. L. F. Sistemas de Informação. In: Bourget, Ir. Monique Marie. (Org.). PROGRAMA SAUDE DA FAMILIA: GUIA PARA O PLANEJAMENTO LOCAL. 1ªed.São Paulo: Editora Martinari, 2005, v. 1, p. 001-104.

GHOSH, K.; WAHNER-ROEDLER, D.; BRANDT, K. Screening mammography starting at age 40: Still relevant. **Cleveland Clinic Journal of Medicine**, v. 82, n. 5, p. 276–279, 2015.

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. 459–467, 2014.

GIRIANELLI, V. R.; THULER, L. C. S.; SILVA, G. A. E. Qualidade do sistema de informação do câncer do colo do útero no estado do Rio de Janeiro. **Revista de Saúde Pública**, v. 43, n. 4, p. 580–588, 2008.

GLOBOCAN. Global Cancer Observatory. Cancer Today. Mapas. 2018. Disponível em: <<https://gco.iarc.fr/today/online-analysis-map>>. Acesso em abr 2019.

GLOBOCAN. Global Cancer Observatory. Cancer Today. Mapas. 2020. Disponível em: <<https://gco.iarc.fr/today/online-analysis-map>>. Acesso em mai 2021.

GONZÁLEZ, R. S. Cancer screening: global debates and Cuban experience. **MEDICC Review**, v. 16, n. 3–4, p. 73–77, 2014.

GØTZSCHE, P. C.; JORGENSEN, K. J. **Screening for breast cancer with mammography** *Base de Dados Cochrane de Revisões Sistemáticas*, 2013.

GUERRA, M. R. et al. Inequalities in the burden of female breastcancer in Brazil, 1990–2017. **Population Health Metrics**, v. 18, n. suppl 1, p. 8, 2020.

HADJISAVVAS, A. et al. An investigation of breast cancer risk factors in Cyprus: A case control study. **BMC Cancer**, v. 10, n. 447, p. 1–10, 2010.

_____. SISMAMA: Informação para o avanço das ações de controle do câncer de mama no Brasil. Rio de Janeiro: INCA, 2010. Disponível em: <<http://bvsmms.saude.gov.br/bvs/publicacoes/inca/Sismama.pdf>>. Acesso em abr 2019.

_____. Controle do Câncer de Mama: Documento de Consenso. Rio de Janeiro: INCA, 2004. Disponível em: <<http://bvsmms.saude.gov.br/bvs/publicacoes/Consensointegra.pdf>>. Acesso em jun 2019.

_____. 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 / Instituto Nacional de Câncer. Coordenação Geral de Ações Estratégicas. Divisão de Apoio à Rede de Atenção Oncológica. Rio de Janeiro: INCA, 2011.

_____. Sistema de informação do câncer: manual preliminar para apoio à implantação. Rio de Janeiro: INCA, 2013.

_____. Diretrizes para a detecção precoce do câncer de mama no Brasil. Rio de Janeiro: INCA, 2015a. Disponível em: <https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//diretrizes_deteccao_precoce_cancer_mama_brasil.pdf>. Acesso em jun 2019.

_____. Monitoramento das ações de controle do câncer de mama: Linha de cuidado e rede de atenção ao câncer de mama. Boletim ano 6, nº 3. Rio de Janeiro: INCA, 2015b.

_____. Tipos de Câncer: Câncer de Mama. Rio de Janeiro: INCA, 2019a.
Disponível em: < <https://www.inca.gov.br/tipos-de-cancer/cancer-de-mama>>. Acesso em abr 2019.

_____. Atualização em mamografia para técnicos em radiologia. 2. ed. rev. atual. Rio de Janeiro: INCA, 2019b.

_____. Informativo Detecção Precoce: Monitoramento das ações de controle dos cânceres do colo do útero e da mama. Boletim ano 10, n. 2, 2019. Rio de Janeiro: INCA, 2019c.

_____. Parâmetros técnicos para rastreamento do câncer de mama. Rio de Janeiro: INCA, 2021. Disponível em:
<https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document/parametrostecrastramentocamama_2021_1.pdf>. Acesso em mai 2021.

INUMARU, L. E.; SILVEIRA, E. A.; NAVES, M. M. V. Fatores de risco e de proteção para câncer de mama: uma revisão sistemática. **Cadernos de Saúde Pública**, v. 27, n. 7, p. 1259-1270, 2011.

IOL. Institute of Oncology Ljubljana. Registro e call center DORA. Disponível em: < <https://www.onko-i.si/eng/sectors/epidemiology-and-cancer-registry/registry-dora/>>. Acesso em abr 2019.

JÁCOME, E. M.; BARBOSA, I. L. Breast Cancer Detection: Knowledge, Attitude and Practices of Doctors and Nurses from the Family Health Strategy of Mossoró, RN, Brazil. **Revista Brasileira de Cancerologia**, v. 57, n. 2, p. 189-198, 2011.

KASAHARA, Y. et al. Harms of screening mammography for breast cancer in Japanese women. **Breast Cancer**, v. 20, n. 4, p. 310–315, 2012.

KELAVA, I. et al. Breast and gynecological cancers in Croatia, 1988-2008. **Croatian Medical Journal**, v. 53, n. 2, p. 100–108, 2012.

KISIANGANI, J. et al. Determinants of breast cancer early detection for cues to expanded control and care: the lived experiences among women from Western Kenya. **BMC women's health**, v. 18, n. 1, p. 81, 2018.

KOÇ, H.; O'DONNELL, O.; VAN OURTI, T. What Explains Education Disparities in Screening Mammography in the United States? A Comparison with The Netherlands. **International Journal of Environmental Research and Public Health**, v. 15, n. 9, p. 1961, 2018.

KONG, Q.; MONDSCHHEIN, S.; PEREIRA, A. Effectiveness of breast cancer screening policies in countries with medium-low incidence rates. **Revista de Saude Publica**, v. 52, n. 7, p. 1–9, 2018.

KWEON, S. et al. Difference of stage at cancer diagnosis by socioeconomic status for four target cancers of the National Cancer Screening Program in Korea : Results from the Gwangju and Jeonnam cancer registries. **Journal of Epidemiology**, v. 27, n. 7, p. 299–304, 2017.

LEE, S. Y.; LEE, E. E. Cancer screening in Koreans: A focus group approach. **BMC Public Health**, v. 18, n. 254, p. 1–12, 2018.

LIPSCHITZ, S. Screening mammography with special reference to guidelines in South Africa. **South African Journal of Radiology**, v. 22, n. 2, p. a1370, 2017.

LUNDQVIST, A. et al. Socioeconomic inequalities in breast cancer incidence and mortality in Europe - A systematic review and meta-analysis. **European Journal of Public Health**, v. 26, n. 5, p. 804–813, 2016.

MAHMUD, A.; ALJUNID, S. M. The uptake of mammogram screening in Malaysia and its associated factors: A systematic review. **Medical Journal of Malaysia**, v. 73, n. 4, p. 202–211, 2018.

MAJEK, O. et al. Breast cancer screening in the Czech Republic : time trends in performance indicators during the first seven years of the organised programme. **BMC Public Health**, v. 11, n. 1, p. 288, 2011.

MANDRIK, O. et al. Systematic reviews as a ‘lens of evidence’: Determinants of benefits and harms of breast cancer screening. **International Journal of Cancer**, p. 1–14, 2019.

MARMARÁ, D.; MARMARÁ, V.; HUBBARD, G. A national cross-sectional study of adherence to timely mammography use in Malta. **BMC Cancer**, v. 18, n. 346, p. 1-12, 2018.

MARMOT, M. G. et al. The benefits and harms of breast cancer screening : an independent review. **British journal of cancer**, v. 108, n. 11, p. 2205–2240, 2013.

MARROCOS. Ministère De La Santé. Guide de détection précoce des cancers du sein et du col de l’utérus, 2011.

MCNAMARA, C.; BAYAKLY, A. R.; WARD, K. C. Georgia Cancer Data Report, 2016. Georgia Department of Public Health, Georgia Comprehensive Cancer Registry, December 2016.

MELO, F. B. B.; et al. Actions of nurses in early detection of breast cancer. **Revista Brasileira de Enfermagem**, v. 70, n. 6, p. 1119–1128, 2017.

MÉXICO. Norma Oficial Mexicana NOM-041-SSA2-2011. Para la prevención, diagnóstico, tratamiento, control y vigilancia epidemiológica del cáncer de mama. Diario Oficial. 9 jun 2011. Segunda Sección.

- MHCI. Minister for Health and Children Ireland. Breast Check. The National Breast Screening Programme. Programme Report 2016-2017. 2016. Disponível: <https://www.breastcheck.ie/sites/default/files/bcheck/documents/bc-pr-pm-10_rev01_breastcheck_programme_report_2016_-_2017.pdf>. Acesso abr 2019.
- MIGOWSKI, A.; et al. Acurácia do relacionamento probabilístico na avaliação da alta complexidade em cardiologia. **Revista de Saúde Pública**, v. 45, n. 2, p. 269-275, 2011.
- MIGOWSKI, A.; et al. Diretrizes para detecção precoce do câncer de mama no Brasil. I - Métodos de elaboração. **Cadernos de Saúde Pública**, v. 34, n. 6, p. e00116317, 2018a.
- MIGOWSKI A.; et al. Diretrizes para detecção precoce do câncer de mama no Brasil. II – Novas recomendações nacionais, principais evidências e controvérsias. **Cadernos de Saúde Pública**, v. 34, n. 6, p. e00074817, 2018b.
- MIGOWSKI A.; et al. Diretrizes para detecção precoce do câncer de mama no Brasil. III – Desafios à implementação. **Cadernos de Saúde Pública**, v. 34, n. 6, p. e00046317, 2018c.
- MILES, A. et al. A perspective from countries using organized screening programs. **Cancer**, v. 101, n. S5, p. 1201–1213, 2004.
- MILLER, A. B. et al. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. **BMJ**, v. 348, n. g366, p. 1–16, 2014.
- MITTRA, I. et al. A cluster randomized, controlled trial of breast and cervix cancer screening in Mumbai, India: Methodology and interim results after three rounds of screening. **International Journal of Cancer**, v. 126, n. 4, p. 976–984, 2009.
- MOLASSIOTIS, A.; et al. Organized breast cancer screening not only reduces mortality from breast cancer but also significantly decreases disability-adjusted life years: analysis of the Global Burden of Disease Study and screening programme availability in 130 countries. **Esmo Open Cancer Horizons**, v. 6, n. 3, p. 1-8, 2021.
- MONTICCIOLO, D. L.; et al. Breast Cancer Screening for Average-Risk Women: Recommendations From the ACR Commission on Breast Imaging. **Journal of the American College of Radiology**, v. 14, n. 9, p. 1137-1143, 2017.
- MORAES, D. C.; et al. Opportunistic screening actions for breast cancer performed by nurses working in primary health care. **Revista da Escola de Enfermagem da USP**, v.50, n. 1, p. 14–21, 2016.
- MUKEM, S. et al. Breast cancer screening among women in Thailand: Analyses of population-based household surveys. **Journal of the Medical Association of Thailand**, v. 97, n. 11, p. 1106–1118, 2014.
- MURILLO, R. et al. Increased breast cancer screening and downstaging in Colombian women: A randomized trial of opportunistic breast-screening. **International Journal of Cancer**, v. 138, n. 3, p. 705–713, 2015.

NICARÁGUA. Ministerio de Salud. Jornada de Mamografias em El Hospital de Masaya.2014. Disponível em: <<http://www.minsa.gob.ni/index.php/103-noticias-2014/708-jornada-de-mamografias-en-el-hospital-de-masaya>>. Acesso abr. 2019.

NOGUEIRA, M. C.; et al. Disparidade racial na sobrevivência em 10 anos para o câncer de mama: uma análise de mediação usando abordagem de respostas potenciais. **Cadernos de Saúde Pública**, v. 34, n. 9, e00211717, 2018.

NOGUEIRA, M. C.; et al. Inequities in access to mammographic screening in Brazil. **Cadernos de Saúde Pública**, v. 35, n. 6, e0009981, 2019.

OECD. Organisation for Economic Co-operation and Development. OECD Health Statistics 2018 Definitions, Sources and Methods. Breast cancer screening (mammography), survey data and programme data. Disponível: <<https://stats.oecd.org/FileView2.aspx?IDFile=eb5acd7d-2445-401a-b624-62fcdad85091>>. Acesso em abr. 2019.

OEFFINGER, K. C.; et al., Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society. **JAMA**, v. 314, n. 15, p. 1599–1614, 2015.

OLVER, I.; RODER, D. History, development and future of cancer screening in Australia. **Public Health Research & Practice**, v. 27, n. 3, p. 1–7, 2017.

OZMEN, V. et al. Cost-Effectiveness of Breast Cancer Screening in Turkey, a Developing Country: Results from Bahçeşehir Mammography Screening Project. **European Journal of Breast Health**, v. 13, n. 3, p. 117–122, 2017.

PANAMÁ. Ministerie de Salud. Normas Técnicas-Administrativas y Protocolo de Atención del Programa de Salud Integral de la Mujer. Programa de Salud Integral de la Mujer, 2015.

PARAGUAY. Ministerio de Salud Publica Y Bienestar Social. Manual Nacional de Normas y Procedimientos para la Prevención y el Control del Cáncer de Mama em el Paraguay, 2016.

PERES, S. V. et al. 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. 428–436, 2015.

PERU. Ministerio de Salud. Resolución Ministerial n° 442-2017/MINSA. Aproba el Documento Técnico: "Plan Nacional para la Prevención y Control de Cáncer de Mama em el Perú (2017-2021)", 2017.

PITCHON, A.; et al. Minas Gerais. Índice de Vulnerabilidade Social (IVS) 2012. Belo Horizonte, 2013. 15 p.

RENNA JUNIOR, N. L.; AZEVEDO E SILVA, G. Late-Stage Diagnosis of Breast Cancer in Brazil : Analysis of Data from Hospital-Based Cancer Registries (2000 – 2012). **Revista Brasileira de Ginecologia e Obstetrícia**, v. 40, n. 3, p. 127–136, 2018.

RIBA, L. A. et al. Association between socioeconomic factors and outcomes in breast cancer. **Breast Journal**, v. 25, n. 3, p. 488–492, 2019.

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.

ROCHA-BRISCHILIARI, S. C.; et al. Spatial distribution of breast cancer mortality: Socioeconomic disparities and access to treatment in the state of Parana, Brazil. **PlosOne**, v. 13, n. 10, p. e0205253, 2018.

RODRIGUES, T. B. Sobrerrastreio mamográfico: avaliação a partir de bases identificadas do Sistema de Informação do Câncer de Mama (SISMAMA). **Cadernos de Saúde Pública**, v. 35, n. 1, e00049718, 2019.

ROMERO, R. O. G. et al. Subnotificação de casos de tuberculose a partir da vigilância do óbito. **Revista Eletrônica de Enfermagem**, v. 18, 2016. Disponível em: <<http://dx.doi.org/10.5216/ree.v18.37249>>. Acesso em 04 abril 2018.

SANTOS, R. O. M.; RAMOS, D. N.; MIGOWSKI, A. Barreiras na implementação das diretrizes de detecção precoce dos cânceres de mama e colo do útero no Brasil. **Physis: Revista de Saúde Coletiva**, v. 29, n. 4, p. e290402, 2019.

SANTOS, J. S.; et al. Prevenção e controle do câncer de mama na Atenção Primária à Saúde. **Brazilian Journal of Surgery and Clinical Research**, v. 28, n. 1, p. 87-95, 2019.

SANTOS-SILVA, I. et al. Ethnoracial and social trends in breast cancer staging at diagnosis in Brazil, 2001-14: a case only analysis. **Lancet Glob Health**, v. 7, p. e784-e797, 2019.

SANKARANARAYANAN, R. Screening for Cancer in Low- and Middle-Income Countries. **Annals of Global Health**, v. 80, n. 5, p. 412–417, 2014.

SARKEALA, T. et al. Breast carcinoma detection modes and death in a female population in relation to population-based mammography screening. **SpringerPlus**, v. 3, n. 1, p. 1–9, 2014.

SARKEALA, T.; HEINÄVAARA, S.; ANTTILA, A. Breast cancer mortality with varying invitational policies in organised mammography. **British Journal of Cancer**, v. 98, n. 3, p. 641–645, 2008.

SELIG, L. et al. Óbitos atribuídos à tuberculose no Estado do Rio de Janeiro TT - Deaths attributed to tuberculosis in the State of Rio de Janeiro. **Jornal Brasileiro de Pneumologia**, v. 30, n. 4, p. 327–334, 2004.

- SHAH, T.; GURAYA, S. Breast cancer screening programs: Review of merits, demerits, and recent recommendations practiced across the world. **Journal of Microscopy and Ultrastructure**, v. 5, n. 2, p. 59, 2017.
- SHANNOUN, F. et al. Organisational aspects of mammography screening in digital settings: First experiences of Luxembourg. **Radiation Protection Dosimetry**, v. 129, n. 1–3, p. 195–198, 2008.
- SIEGEL, R. L. et al. An assessment of progress in cancer control. **CA: A Cancer Journal for Clinicians**, v. 68, n. 5, p. 329–339, 2018.
- SILVA, P. A.; RIUL, S. S. Câncer de mama: fatores de risco e detecção precoce. **Revista Brasileira de Enfermagem**, v. 64, n.6, p. 1016-1021, 2011.
- SILVEIRA, D. P. DA; 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. 875–882, 2009.
- SIMBRICH, A. et al. Trends in advanced breast cancer incidence rates after implementation of a mammography screening program in a German population. **Cancer Epidemiology**, v. 44, p. 44–51, 2016.
- SIU, A. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. **Annals of Internal Medicine**, v. 164, n. 4, p. 279-296, 2016.
- SONG, Q.K.; et al. Breast Cancer Challenges and Screening in China: Lessons From Current Registry Data and Population Screening Studies. **The Oncologist**, v. 20, p. 773-779, 2015.
- SRIPAIBOONKIJ, N. et al. Breast Cancer Detection Rate , Incidence , Prevalence and Interval Cancer-related Mammography Screening Times among Thai Women. **Asian Pacific Journal of Cancer Prevention**, v. 17, p. 4137–4141, 2016.
- STEPONAVICIENE, L. et al. Trends in Breast Cancer Incidence and Stage Distribution Before and During the Introduction of the Mammography Screening Program in Lithuania. **Cancer Control**, v. 26, n. 1, p. 1–7, 2019.
- SUIÇA. Associação do Rastreamento do Câncer na Suíça. Departamento Federal de Saúde. Liga Suíça Contra o Câncer. Folhetos e folhetos para a detecção precoce do câncer de mama. 2014. Disponível em: < <https://www.swisscancerscreening.ch/krebs-frueherkennung/brust/broschueren-und-flyer>>. Acesso em jul. 2019.
- SUNG, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. **CA Cancer Journal Clinicians**, Nova York, v. 71, n. 3, p. 209-249, 2021.
- SUONIO, E.; DEAN, P. B. **Estonian breast cancer screening programme pre-conditions and current status report of a pre-audit assessment on 28-29 December**

2010 and 9-11 May 2011. [s.l: s.n.]. Disponível em:

<<http://rahvatervis.ut.ee/handle/1/6344>>. Acesso em mai 2019.

TEIXEIRA, L. A. S.; NETO, L. A. A. Câncer de mama no Brasil: medicina e saúde pública no século XX. **Saúde e Sociedade**, v. 29, n. 3, p. e180753, 2020a.

TEIXEIRA, L. A. S.; NETO, L. A. A. Still Controversial: Early Detection and Screening for Breast Cancer in Brazil, 1950–2010s. **Medical History**, v. 64, n. 1, p. 52-70, 2020b.

THEME FILHA, M. M. et al. 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. 1–8, 2016.

TIENSOLI, S. D.; FELISBINO-MENDES, M. S.; VELASQUEZ-MELENDEZ, G. Iniquidade em saúde, comportamentos não saudáveis e cobertura de mamografia no Brasil. **Revista Brasileira de Enfermagem**, v. 73, s. 5. p. e20200011, 2020.

TOMAZELLI, J. G.; AZEVEDO E SILVA, G. Rastreamento do câncer de mama no Brasil: uma avaliação da oferta e utilização da rede assistencial do Sistema Único de Saúde no período 2010-2012. **Epidemiologia e Serviços de Saúde**, v. 26, n.4, p. 713-724, 2017.

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, n. 0, 2018.

TORRE, L. A.; ISLAMI, F.; SIEGEL, R.L.; WARD, E. M.; JEMAL, A. Global Cancer in Women: Burden and Trends. **Cancer Epidemiology, Biomarkers & Prevention**, v. 26, n. 4, p. 444-457, 2017.

TRAVASSOS, C.; CASTRO M. S. M. Determinantes e desigualdades sociais no acesso e na utilização de serviços de saúde. In: GIOVANELLA, L.; et al. Políticas e Sistema de Saúde no Brasil. Rio de Janeiro: Editora Fiocruz; 2008. p. 215-243.

UJHELYI, M. et al. Barriers to organized mammography screening program in Hungary: a questionnaire-based study of 3 313 women. **European Journal of Surgical Oncology**, v. 38, p. 1727–1734, 2018.

URBAN, L. A. B. D.; SONDERMANN, V. R. Recomendações do Colégio Brasileiro de Radiologia e Diagnóstico por Imagem, da Sociedade Brasileira de Mastologia e da Federação Brasileira das Associações de Ginecologia e Obstetrícia para rastreamento do câncer de mama por métodos de imagem. **Radiologia Brasileira**, v. 45, n. 6, p. 334-339, 2012.

URUGUAY. Ministerio de Salud Pública. Guía de Prática Clínica de Detección Temprana del Cáncer de Mama: Tamizaje y diagnóstico precoz. 2ª ed. Uruguay, 2015.

VAHABI, M. et al. Breast cancer screening disparities among urban immigrants : a population-based study in Ontario , Canada. **BMC Public Health**, p. 1–12, 2015.

VAN LUIJT, P. A.; HEIJNSDIJK, E. A. M.; DE KONING, H. J. Cost-effectiveness of the Norwegian breast cancer screening program. **International Journal of Cancer**, v. 140, n. 4, p. 833–840, 2017.

VERLEYE, L.; et al. Dépistage Du Cancer Du Sein: Comment Identifier Les Femmes Exposées À Un Risque Accru – Quelles Techniques D’imagerie Utiliser ? Good Clinical Practice (GCP). Bruxelles: Centre fédéral d’expertise des soins de santé (KCE). 2011.KCE Reports 172B. D/2011/10.273/91.

VIEIRA-DA-SILVA, L. M.; FILHO, N. A.; Equidade em saúde: uma análise crítica de conceitos. *Cadernos de Saúde Pública*, São Paulo, v. 25, s. 2, p. 217-226, 2009.

VIEIRA, S. C. Câncer de mama: Consenso da Sociedade Brasileira de Mastologia Regional Piauí. Editora: EDUFPI, 2017. Disponível em: <<https://www.sbmastologia.com.br/medicos/wp-content/uploads/2018/03/Câncer-de-Mama-Consenso-da-SBM-Regional-Piauí-2017.pdf>>. Acesso em jun 2021.

VUTUC, C.; WALDHOER, T.; HAIDINGER, G. Breast cancer trends: Opportunistic screening in Austria versus controlled screening in Finland and Sweden. **European Journal of Cancer Prevention**, v. 15, n. 4, p. 343–346, 2006.

WEE, L. E. et al. Socioeconomic factors affecting colorectal, breast and cervical cancer screening in an Asian urban low-income setting at baseline and post-intervention. **Preventive Medicine**, v. 55, n. 1, p. 61–67, 2012.

WEIGEL, S.; et al. Reduction of Advanced Breast Cancer Stages at Subsequent Participation in Mammography Screening. **Fortschr Röntgenstr**, v. 188, p. 33-37, 2016.

WENDER, R. C. et al. A blueprint for cancer screening and early detection: Advancing screening’s contribution to cancer control. **CA: A Cancer Journal for Clinicians**, v. 69, n. 1, p. 50–79, 2019.

WHITEHEAD, M. The concepts and principles of equity and health. *International Journal of Health Services*, Estados Unidos, v. 22. n. 3, p. 429-45, 1992.

WILLEMS, B.; BRACKE, P. The impact of regional screening policies on the diffusion of cancer screening participation in Belgium: Time trends in educational inequalities in Flanders and Wallonia. **BMC Health Services Research**, v. 18, n. 1, p. 1–13, 2018.

WOODS, L. M.; RACHET, B.; COLEMAN, M. P. Origins of socio-economic inequalities in cancer survival: A review. **Annals of Oncology**, v. 17, n. 1, p. 5-19, 2006.

WHO. WORLD HEALTH ORGANIZATION. Early Detection. Geneva, 2007. (Cancer control: knowledge into action: WHO guide for effective programmes). Disponível em: <http://apps.who.int/iris/bitstream/handle/10665/43743/9241547338_eng.pdf>. Acesso em abr 2019.

WHO. WORLD HEALTH ORGANIZATION. **Who Position Paper on Mammography Screening**. 1ª Edição. Geneva: World Health Organization, 2014. Disponível em: <https://www.who.int/cancer/publications/mammography_screening/en/>. Acesso em jun 2021.

WU, T.; LEE, J. Promoting Breast Cancer Awareness and Screening Practices for Early Detection in Low-Resource Settings. **European Journal of Breast Health**, v. 15, n. 1, p. 18–25, 2018.

YANG, L. et al. Quality assurance target for community-based breast cancer screening in China: A model simulation. **BMC Cancer**, v. 18, n. 1, p. 1–9, 2018.

YOU, H. et al. Why hasn't this woman been screened for breast and cervical cancer? – Evidence from a Chinese population-based study. **Public Health**, v. 168, p. 83–91, 2019.

ZAKHAROVA, N. et al. The introduction of a breast cancer screening programme in a region with a population at medium risk for developing breast cancer: Khanty-Mansiysky autonomous Okrug - Ugra (Russian Federation). **e Cancer Medical Science**, v. 5, n. 1, p. 1–10, 2011.

ZHA, N. et al. Beyond Universal Health Care : Barriers to Breast Cancer Screening Participation in. **Journal of the American College of Radiology**, v. 16, n. 4, p. 570–579, 2019.

ZIDAR, M. N. et al. Non-attendance of mammographic screening: The roles of age and municipality in a population-based Swedish sample. **International Journal for Equity in Health**, v. 14, n. 157, p. 1–11, 2015.

APÊNDICES

**APÊNDICE I – DISTRIBUIÇÃO DOS INDICADORES SOCIOECONOMICOS DAS
MACROS E MICRORREGIÕES DE SAÚDE.**

CÓDIGO DA MACRO/ MICRO	MACRO/MICRO	IDH	RENDA (R\$)	IVS	GRAUURB (%)	ANALF* (%)	GINI
3103	CENTRO						
31008	B.Horizonte/N.Lima/Caeté	0,71	1205,10	0,14	98,96	3,4	0,49
31009	Betim	0,69	589,70	0,27	94,09	5,9	0,47
31015	Contagem	0,73	707,70	0,20	99,66	4,0	0,45
31018	Curvelo	0,67	509,30	0,38	85,16	9,8	0,48
31025	Guanhães	0,62	392,50	0,63	61,95	17,0	0,51
31028	Itabira	0,67	587,50	0,34	83,31	7,5	0,46
31035	João Monlevade	0,70	601,50	0,28	90,35	5,6	0,46
31047	Ouro Preto	0,74	675,60	0,29	89,63	4,7	0,49
31067	Sete Lagoas	0,69	654,80	0,25	88,83	6,1	0,46
31076	Vespasiano	0,73	652,40	0,28	90,85	5,5	0,47
3102	CENTRO SUL						
31007	Barbacena	0,66	615,20	0,32	77,46	7,7	0,47
31014	Cons.Lafaiete/Congonhas	0,66	638,30	0,28	81,98	5,6	0,46
31063	São João del Rei	0,69	590,60	0,22	85,61	5,9	0,48
3104	JEQUITINHONHA						
31019	Diamantina	0,62	402,90	0,56	68,42	14,7	0,49
31043	M.Novas/Turmalim/Capelin	0,64	352,20	0,69	54,27	19,7	0,47
3106	LESTE						
31013	Caratinga	0,62	452,40	0,49	69,53	12,9	0,47
31017	Cel.Fabriciano/Timóteo	0,68	645,60	0,27	94,78	6,3	0,45
31024	Governador Valadares	0,63	532,90	0,34	85,64	10,9	0,49
31027	Ipatinga	0,66	664,50	0,30	90,83	7,9	0,46
31042	Mantena	0,64	400,70	0,50	72,09	18,5	0,49
31058	Resplendor	0,64	441,40	0,46	72,98	15,2	0,49
31060	S.Maria Suaçuí/S.João Ev	0,61	330,60	0,75	50,48	20,1	0,50
3110	LESTE DO SUL						
31041	Manhuaçu	0,65	452,90	0,46	62,93	13,1	0,50
31056	Ponte Nova	0,64	463,20	0,44	70,4	12,0	0,46
31077	Viçosa	0,64	646,50	0,38	74,98	9,0	0,49
3111	NORDESTE						
31001	Águas Formosas	0,61	313,20	0,73	68,67	26,3	0,51
31004	Almenara	0,6	321,20	0,67	72,3	25,9	0,49
31005	Araçuaí	0,63	327,70	0,74	53,08	18,7	0,49
31030	Itaobim	0,6	301,10	0,75	57,85	23,1	0,53
31046	Nanuque	0,67	455,10	0,46	81,55	17,4	0,53
31048	Padre Paraíso	0,56	238,50	0,90	42,46	23,8	0,53
31053	Pedra Azul	0,6	293,40	0,67	78,01	23,0	0,52
31068	T.Otoni/Malacac/Itambac.	0,61	411,70	0,63	62,75	19,9	0,52

3109	NOROESTE						
31036	João Pinheiro	0,68	505,00	0,40	81,54	9,4	0,48
31051	Patos de Minas	0,71	706,20	0,19	83,88	7,0	0,49
31074	Unaí	0,67	540,50	0,44	76,96	10,0	0,50
3108	NORTE						
31011	Brasília MG/S.Francisco	0,62	279,80	0,75	53,25	19,0	0,50
31016	Coração de Jesus	0,64	292,90	0,82	58,97	17,3	0,50
31022	Francisco Sá	0,61	277,10	0,74	57,46	19,4	0,48
31033	Janaúba Monte Azul	0,63	338,80	0,67	62,67	19,5	0,48
31034	Januária	0,61	308,90	0,76	59,46	18,6	0,52
31040	Manga	0,59	281,60	0,88	55,52	24,1	0,55
31044	Montes Claros Bocaiúva	0,67	569,70	0,33	89,51	7,7	0,46
31054	Pirapora	0,64	427,90	0,52	88,24	10,9	0,48
31059	Salinas Taiobeiras	0,6	349,10	0,72	55,23	23,0	0,49
3105	OESTE						
31010	Bom Despacho	0,71	688,80	0,23	90,86	7,8	0,45
31020	Divinópolis/S.Antº.Monte	0,71	728,00	0,15	91,75	5,1	0,41
31021	Formiga	0,71	673,10	0,20	85,78	6,6	0,45
31031	Itaúna	0,69	734,10	0,20	86,92	5,0	0,44
31049	Pará de Minas	0,70	628,60	0,20	90,17	4,7	0,43
31061	S.Antôn.Amparo/Cpo.Belo	0,68	562,40	0,25	83,28	8,5	0,45
3107	SUDESTE						
31002	Além Paraíba	0,70	597,50	0,31	88,06	7,1	0,50
31012	Carangola	0,64	450,40	0,50	61,1	12,6	0,49
31037	J.Fora/L.Duarte/B.Jardim	0,67	906,50	0,14	94,03	4,6	0,47
31039	Leopoldina Cataguases	0,70	591,30	0,23	90,17	7,2	0,44
31045	Muriaé	0,68	610,80	0,29	82,33	9,4	0,47
31062	Santos Dumont	0,68	530,30	0,31	87,46	7,7	0,45
31064	S.João Nepomuceno/Bicas	0,69	573,80	0,22	92,04	7,1	0,44
31071	Ubá	0,67	572,90	0,29	79,88	8,5	0,45
3101	SUL						
31003	Alfenas Machado	0,70	617,40	0,22	78,89	7,7	0,45
31026	Guaxupé	0,70	646,40	0,23	77,65	7,7	0,46
31029	Itajubá	0,69	686,50	0,23	73,22	6,4	0,47
31038	Lavras	0,72	704,50	0,21	88,71	5,6	0,48
31050	Passos Piumhi	0,71	687,40	0,16	84,46	6,7	0,44
31055	Poços de Caldas	0,71	840,10	0,15	89,23	4,8	0,45
31057	Pouso Alegre	0,70	661,10	0,16	77,17	7,5	0,43
31065	São Lourenço	0,69	590,30	0,21	78,54	7,7	0,49
31066	São Sebastião do Paraíso	0,70	652,90	0,21	85,04	8,5	0,43
31069	Três Corações	0,69	644,40	0,27	86,12	8,0	0,47
31070	Três Pontas	0,70	587,70	0,27	82,6	7,8	0,45
31075	Varginha	0,71	768,10	0,19	91,44	5,9	0,47
3113	TRIÂNGULO DO NORTE						
31032	Ituiutaba	0,71	714,30	0,24	89,49	9,8	0,46

31052	Patrocínio Monte Carmelo	0,71	671,40	0,24	84,63	7,0	0,48
31073	Uberlândia Araguari	0,72	883,80	0,14	94,71	4,4	0,47
3112	TRIÂNGULO DO SUL						
31006	Araxá	0,72	767,80	0,18	90,64	5,4	0,49
31023	Frutal Iturama	0,71	675,20	0,23	84,98	8,3	0,48
31072	Uberaba	0,70	868,40	0,14	95,03	4,7	0,47

*Analfabetismo.

APÊNDICE II - SCRIPTS EXECUTADOS NO RSTUDIO

1. Limpeza da base referência

```

# Programa de Pós-graduação em Saúde Coletiva - Doutorado
# Doutoranda: Daniela de Almeida Pereira

# ETAPA 1 - Preparar o banco
# Indicar o caminho onde as bases de dados estão

#### Passo 1
setwd("G:/SISMAMA/MAMOGRAFIA")

# Importar a biblioteca necessária com as funções
library(foreign)

## Banco 2010
# Ler os bancos de dados de cada mês em dbf

M10jan <- read.dbf("M10jan.dbf")
M10fev <- read.dbf("M10fev.dbf")
M10mar <- read.dbf("M10mar.dbf")
M10abr <- read.dbf("M10abr.dbf")
M10mai <- read.dbf("M10mai.dbf")
M10jun <- read.dbf("M10jun.dbf")
M10jul <- read.dbf("M10jul.dbf")
M10ago <- read.dbf("M10ago.dbf")
M10set <- read.dbf("M10set.dbf")
M10out <- read.dbf("M10out.dbf")
M10nov <- read.dbf("M10nov.dbf")
M10dez <- read.dbf("M10dez.dbf")

#### Passo 2
# Juntar os bancos de dados de cada mês de 2010
banco2010<-
rbind(M10jan,M10fev,M10mar,M10abr,M10mai,M10jun,M10jul,M10ago,M10set,M10out,M10no
v,M10dez)

#### Passo 3
# Selecionar somente as variáveis de interesse
banco2010 <-
subset(banco2010,select=c(C_CNES,C_ID_COMP,C_EXAME,C_ID_NOME,C_ID_NOMEM,D_
ID_DTNAS,C_ID_ESCO,C_ID_ENDER,C_ID_NUMER,C_ID_COMPL,C_ID_BAIRR,C_ID_U
F,C_IBGE,C_ID_SUS,C_US_UF,C_US_IBGE,C_US_CNES,C_US_NOME,C_ANM_NOD,C_A
NM_NOD_,XANM_NOD_M,C_ANM_PARE,C_ANM_EXA_,C_ANM_MAMO,XANM_MAM
O_,C_CLI_DIAG,C_MAMO_RAS,C_RAD_NUM_,C_RAD_PELE,C_RAD_COMP,C_NOD_SI
M_,C_NOD_TAM_,XRAD_NUM_F,XRAD_PELE_,XRAD_COMPO,NOD_SIM_01,NOD_TA
M_01,C_CON_DIA_,XCON_DIA_C,C_CON_RECO,XCON_RECOM,D_RECEBE,D_EXAME,
D_LIBERA,C_ID_SEXO,C_ID_RACAC,C_ETNIA,C_NACIONAL))

# Salvar e abrir o banco final 2010

write.dbf(banco2010, file="banco2010")
banco2010 <- read.dbf("banco2010.dbf")

```

```

#### Passo 4
# Nomes das variáveis
names(banco2010)

# Estrutura do banco (tipo de cada variável)
str(banco2010)

# Transformando as variáveis com data no formato data (mm/dd/aaaa)
library(zoo)
banco2010$D_EXAME2 <- as.Date(banco2010$D_EXAME, "%d/%m/%Y")
banco2010$D_ID_DTNAS2 <- as.Date(banco2010$D_ID_DTNAS, "%d/%m/%Y")
``

#### Passo 5
# Extraíndo o dia, mês ou ano de uma data

library(lubridate)
banco2010$ANO <- year(banco2010$D_EXAME2)

# Ver variáveis deste banco
names(banco2010)

# Ver variável ano no banco, usando NA para células vazias
table(banco2010$ANO,useNA = "ifany")

#### Passo 6
# Selecionando observações que tenham a data entre dois valores
novob2010<-banco2010[banco2010$D_EXAME2>"2009-12-31" &
banco2010$D_EXAME2<"2011-01-01",]

# Ver variável ano no banco novo, usando NA para células vazias
table(novob2010$ANO,useNA = "ifany")

#### Passo 7
# Salvar o banco
write.dbf(novob2010, file="novob2010")
novob2010<-read.dbf("novob2010.dbf")

Observação: Os passos de 1 a 7 foram repetidos com os bancos 2011 e 2012 e foram extraídos os exames solicitados no ano de 2010.

# Salvar o banco
write.dbf (novob2011, file="novob2011")
novob2011<-read.dbf("novob2011.dbf")

# Salvar o banco
write.dbf (novob2012, file="novob2012")
novob2012<-read.dbf("novob2012.dbf")

#### Passo 8
# JUNTANDO OS BANCOS, RENOMEAR VARIÁVEIS
# Ler os bancos de dados de interesse que estão em dbf

Banco2010 <- read.dbf("novob2010.dbf")

```



```

Banco2011 <- read.dbf("novob2011.dbf")
Banco2012 <- read.dbf("novob2012.dbf")

# Juntar os bancos de dados
Banco2010total<-rbind(Banco2010, Banco2011, Banco2012)
Banco2010total<-read.dbf("Banco2010total.dbf")

# Salvar o banco
write.dbf(Banco2010total, file="Banco2010total")

#### Passo 9
# Renomear variáveis
names(Banco2010total) <- c("CNES", "COMPET","N_EXAME", "NOME", "MAE", "DTNASC",
"ESCOL", "ENDERECO", "NUMERO","COMPLEM", "BAIRRO", "UF", "IBGE","CNSUS",
"US_UF", "US_IBGE", "CNES_REQ", "US_NOME","NODULO", "NOD_D", "NOD_E",
"RISCO","EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG",
"MAMOG_DIAG_ANT","RASTREAM","NUM_FIL_MD", "PELE_MD", "COMP_MD",
"NOD_MD","NOD_TAM_MD","NUM_FIL_ME", "PELE_ME", "COMP_ME",
"NOD_ME","NOD_TAM_ME","DIAG_MD", "DIAG_ME", "RECOM_MD","RECOM_ME",
"DT_EXAME", "DT_SOLICIT", "DT_LIBERA","SEXO","RACA","ETNIA","NACIONAL",
"D_EXAME2", "D_ID_DTNAS2", "ANO")

# Salvar o banco total
write.dbf(Banco2010total, file="Banco2010total")

#### Passo 10
# Restringir a base de dados ao sexo feminino
Banco2010totalN <- subset(Banco2010total,(SEXO=="2"))
Banco2010total <- read.dbf("Banco2010total.dbf")
table(Banco2010totalN$SEXO)
table(Banco2010total$SEXO)

# Salvar o banco só com sexo feminino
write.dbf(Banco2010totalN, file="Banco2010totalN")
Banco2010totalN<-read.dbf("Banco2010totalN.dbf")

#### Passo 11
# Extrair a variável ano de nascimento da data de nascimento e criar a variável idade
Banco2010totalN$DTNASC<- as.Date(Banco2010totalN$DTNASC, "%d/%m/%Y")

library(lubridate)
Banco2010totalN$ANO_NASC <- year(Banco2010totalN$DTNASC)
Banco2010totalN$IDADE<- 2010 - (Banco2010totalN$ANO_NASC)

# Verificar idades incorretas
library(descr)
freq(Banco2010totalN$ANO_NASC)

# Fazer uma cópia do banco Banco2010totalN
db0<-Banco2010totalN
write.dbf(db0, file="db0")

#### Passo 12

```

```
# Excluir os exames com idade inferior a 18 anos. Justificativa: Usar dados somente da população adulta.
```

```
db<- db0[db0$ANO_NASC>"1900" & db0$ANO_NASC<"1993",]
dbex<- db0[db0$ANO_NASC>"1992" & db0$ANO_NASC<"2058",]
```

```
freq(db$ANO_NASC)
freq(dbex$ANO_NASC)
```

```
freq(db$IDADE)
freq(dbex$IDADE)
```

```
write.dbf(db, file="db")
write.dbf(dbex, file = "dbex")
db<-read.dbf("db.dbf")
```

```
#### Passo 13
```

```
# Criar variável macro do IBGE com 6 dígitos
# Observação: A variável do município no IBGE tem 7 dígitos então tem que extrair os 6 primeiros para fazer o merge
library(stringr)
db$CODMUN6<- substr(db$IBGE,1,6)
```

```
# Verificar se é a coluna correta
head(db$CODMUN6)
```

```
# Carregar o banco no formato data table
library(data.table)
dbdt <- data.table(db)
write.dbf(dbdt, "dbdt.dbf")
```

```
#### Passo 14
```

```
# Ler arquivo csv, considerando a primeira linha como cabeçalho
##csv: separador de variáveis é a vírgula. Confunde no Brasil se o separador de decimal é virgula. Para usá-lo o excel tem que considerar o separador como ponto.
###csv2: separador de variáveis é o ponto e vírgula
library(csv)
tab_mun_cir_mac <- read.csv2("tab_mun_cir_mac.csv", header = T)
tab_mun_cir_mac$CODMUN6 <- tab_mun_cir_mac$MUN_COD
```

```
# Carregar o banco no formato data table
tab_mun_cir_mac_dt <- data.table(tab_mun_cir_mac)
```

```
# Ver tipo de variável
str(dbdt$CODMUN6)
str(tab_mun_cir_mac_dt)
```

```
# Transformar variável codmun6 em integer
dbdt$CODMUN6<- as.integer(dbdt$CODMUN6)
```

```
# Juntar os bancos - variável chave: codmun6
db1 <- merge(dbdt,tab_mun_cir_mac_dt,by="CODMUN6",all.x=TRUE)
```

```
#### Passo 15
```

```
# Ver os 60 primeiros registros das colunas 55-60 e preencher os campos com NA com M10
```

```
db1[1:60,c(1,55:60)]
```

```
# Preencher campos vazios de município,
```

```
db1$MUNICIPIO<-as.character(db1$MUNICIPIO)
```

```
db1$CIR<-as.character(db1$CIR)
```

```
db1$MACRO<-as.character(db1$MACRO)
```

```
db1$MUN_COD[is.na(db1$MUN_COD)]<-"M10"
```

```
db1$MUNICIPIO[is.na(db1$MUNICIPIO)]<- "M10"
```

```
db1$CIR_COD[is.na(db1$CIR_COD)]<-"M10"
```

```
db1$CIR[is.na(db1$CIR)]<-"M10"
```

```
db1$MACRO_COD[is.na(db1$MACRO_COD)]<-"M10"
```

```
db1$MACRO[is.na(db1$MACRO)]<-"M10"
```

```
db1[1:60,c(1,55:60)]
```

```
# Salvar o banco em dbf
```

```
write.dbf(db1,"db1.dbf")
```

```
db1<-read.dbf("db1.dbf")
```

```
#### Passo 16
```

```
# Criar variável estado extraíndo os 2 primeiros dígitos
```

```
library(stringr)
```

```
db1$ESTADO <- substr(db1$CODMUN6,1,2)
```

```
table(db1$ESTADO)
```

```
#### Passo 17
```

```
# Criar novo banco apenas com exames realizados em MG
```

```
db2 <- db1[db1$ESTADO=="31",]
```

```
write.dbf(db2,"db2.dbf")
```

```
# Verificar se ficou algum campo vazio em macro e microrregião
```

```
table(db2$MACRO,useNA = "ifany")
```

```
table(db2$CIR, useNA="ifany")
```

```
table(db2$ESTADO)
```

```
#### Passo 18
```

```
# Transformar as variáveis NOME, MAE, MUNICIPIO, CNSUS, DTNASC, DTEXAME em caracter
```

```
db2$NOME<-as.character(db2$NOME)
```

```
db2$MAE<-as.character(db2$MAE)
```

```
db2$MUNICIPIO<-as.character(db2$MUNICIPIO)
```

```
db2$DTNASC<-as.character(db2$DTNASC)
```

```
db2$CNSUS<-as.character(db2$CNSUS)
```

```
db2$DT_EXAME<-as.character(db2$DT_EXAME)
```

```
write.dbf(db2, file = "db2")
```

```
db2<-read.dbf("db2.dbf")
```

```
#### Passo 19
```

```
# Transformar todos os campos em maiusculo
```

```

cl <- sapply(db2, class)
wc <- which((cl=="factor")|(cl=="character"))
for(j in wc){db2[,j] <- toupper(db2[,j])}
#### Passo 20
#Retirar expressões regulares das variáveis nome e nome da mãe
aux<- db2$NOME
aux[-grep("[a-zA-Z]",aux)] <- ""
db2$NOME <- aux

aux <- db2$MAE
aux[-grep("[a-zA-Z]",aux)] <- ""
db2$MAE <- aux
``

#### Passo 21
# Substituicoes de acentos e cedilha no banco Banco2010totalN

for (j in 1:ncol(db2)) {
  db2 [,j] <- gsub("Ã", "A", db2 [,j])
}

```

Observação: O passo 21 foi realizado em todas as vogais combinadas com os acentos ~^^ e com o Ç

```

# Retirar números do campo NOME E MAE

db2$NOME<-gsub("[(0-9)]", "", db2$NOME)
db2$MAE<-gsub("[(0-9)]", "", db2$MAE)

# Retirar traço do NOME E MAE
db2$NOME<-gsub("[(-)]", "", db2$NOME)
db2$MAE<-gsub("[(-)]", "", db2$MAE)

# Substituir Mª por MARIA em NOME E MAE
db2$NOME<-gsub("[(ª)]", "ARIA", db2$NOME)
db2$MAE<-gsub("[(ª)]", "ARIA", db2$MAE)

# Retirar simbolos de NOME E MAE
db2$NOME<-gsub("[(.)]", "", db2$NOME)
db2$MAE<-gsub("[(.)]", "", db2$MAE)

```

Observação: Esse processo foi realizado para todos os símbolos (? ” ; ’ . : / \ | *

```

# Salvar o Banco e abrir para verificar informações indesejadas
write.dbf(db2, file = "db2")
db2<-read.dbf("db2.dbf")

# Transformar as variáveis NOME, MAE, MUNICIPIO, CNSUS, DTNASC, DTEXAME em
caracter
db2$NOME<-as.character(db2$NOME)
db2$MAE<-as.character(db2$MAE)

# Substituir os campos com informacoes indesejadas das variaveis nome, mae e data de nascimento
por campo em branco - se tiver aparecendo interrogacao substitua a sintaxe deste item pelo conteudo
do arquivo item9.docx

```

```

##MÃE
db2$MAE[which(db2$MAE=="A")] <- ""
db2$MAE[which(db2$MAE=="AAAAA")] <- ""
db2$MAE[which(db2$MAE=="AO CADASTRADA")] <- ""
db2$MAE[which(db2$MAE=="A MESMA")] <- ""
db2$MAE[which(db2$MAE=="AA")] <- ""
db2$MAE[which(db2$MAE=="AAA")] <- ""
db2$MAE[which(db2$MAE=="CAMP")] <- ""
db2$MAE[which(db2$MAE==" CAMPO NAO PREENCHIDO")] <- ""
db2$MAE[which(db2$MAE=="D")] <- ""
db2$MAE[which(db2$MAE=="DES")] <- ""
db2$MAE[which(db2$MAE=="DESC")] <- ""
db2$MAE[which(db2$MAE=="DESCONHECE")] <- ""
db2$MAE[which(db2$MAE=="DESCONHECIDA")] <- ""
db2$MAE[which(db2$MAE=="DESCONHECIDA")] <- ""
db2$MAE[which(db2$MAE=="DESCONHECIDO")] <- ""
db2$MAE[which(db2$MAE=="DOIS")] <- ""
db2$MAE[which(db2$MAE=="E")] <- ""
db2$MAE[which(db2$MAE=="F")] <- ""
db2$MAE[which(db2$MAE=="GE")] <- ""
db2$MAE[which(db2$MAE=="H")] <- ""
db2$MAE[which(db2$MAE=="I")] <- ""
db2$MAE[which(db2$MAE=="IGIN")] <- ""
db2$MAE[which(db2$MAE=="IGN")] <- ""
db2$MAE[which(db2$MAE=="IGNORADA")] <- ""
db2$MAE[which(db2$MAE=="IGNORADO")] <- ""
db2$MAE[which(db2$MAE=="IGNORADOS")] <- ""
db2$MAE[which(db2$MAE=="III")] <- ""
db2$MAE[which(db2$MAE=="ILEGIVEL")] <- ""
db2$MAE[which(db2$MAE=="INLEGIVEL")] <- ""
db2$MAE[which(db2$MAE=="JJJ")] <- ""
db2$MAE[which(db2$MAE=="JK")] <- ""
db2$MAE[which(db2$MAE=="LEI")] <- ""
db2$MAE[which(db2$MAE=="LLLLL")] <- ""
db2$MAE[which(db2$MAE=="LLLLLLLLLL")] <- ""
db2$MAE[which(db2$MAE=="LLLLLLLLLLLLL")] <- ""
db2$MAE[which(db2$MAE=="M,AE")] <- ""
db2$MAE[which(db2$MAE=="MAE")] <- ""
db2$MAE[which(db2$MAE=="MAE DESCONHECIDA")] <- ""
db2$MAE[which(db2$MAE=="MAE NAO IDENTIFICADA")] <- ""
db2$MAE[which(db2$MAE=="N")] <- ""
db2$MAE[which(db2$MAE=="N DECLARADO")] <- ""
db2$MAE[which(db2$MAE=="NAO")] <- ""
db2$MAE[which(db2$MAE=="N I")] <- ""
db2$MAE[which(db2$MAE=="N INF")] <- ""
db2$MAE[which(db2$MAE=="N INFORMADO")] <- ""
db2$MAE[which(db2$MAE=="N T")] <- ""
db2$MAE[which(db2$MAE=="N/C")] <- ""
db2$MAE[which(db2$MAE=="NAO")] <- ""
db2$MAE[which(db2$MAE=="NÃO")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM")] <- ""
db2$MAE[which(db2$MAE=="NAO CADASTRADO")] <- ""
db2$MAE[which(db2$MAE=="NAO CADASTRAR")] <- ""

```

db2\$MAE[which(db2\$MAE=="NAO CADASTRDA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADATRADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADATRASDA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTOR")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRAMOS")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRAOU")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTROU")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADADASTRADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRAR")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRDA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADASTRO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADATRADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADATRASDA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADSTRADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CADSTRADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO COLOCOU")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NA FICHA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NA FOLHA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO CADASTRO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NA REQUISICAO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO DOCUMENTO DE IDENTIDADE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO DOCUMENTO DO PACIENTE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO PEDIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO PEDIDO MEDICO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NOME DA MAE NO PEDIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NOME DA MAE PEDIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NOME NO RG")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA O NOME DA MAE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTATADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONTA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTATADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO COSNTA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO COSTA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONHECE A MAE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA FILIACAO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO DOCUMENTO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO CONSTA NO RG")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ENFORMANDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESTAR PREENCHIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO DECLADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO DECLARO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO DECLARADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO DECLARADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO DISPONIVEL")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ENCONTRADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ENFORMADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESPACIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESPCIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESPECIFICAD")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESPECIFICAD")] <- ""

db2\$MAE[which(db2\$MAE=="NAO ESPECIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESPECIFICAD")] <- ""
db2\$MAE[which(db2\$MAE=="NAO FOI INFORMADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO FORNECIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO GONSTA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADO NO PEDIDO DO PACIENTE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INDENTIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INF DORMADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFOMARDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO E REGISTRADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO E REGISTRADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ENTENDI")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ENTENDI O NOME DA MAE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ENFORMANDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO ESTAR PREENCHIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO HA")] <- ""
db2\$MAE[which(db2\$MAE=="NAI")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFACADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDETIFICADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADA NA FICHA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INDENTIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDETIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICA OO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INDICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADA NA FICHA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO IDENTIFICA OO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INDICADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORAMDA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORAMDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADAC")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADDA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADO NA FICHA DO PACIENTE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADO NO PEDIDO DO PACIENTE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMADO PELO PACIENTE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMANDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMAOD")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMNADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORMOU")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INFORNADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO INOFRMADA")] <- ""
db2\$MAE[which(db2\$MAE=="NAO POSSUI")] <- ""
db2\$MAE[which(db2\$MAE=="NAO POSSUI NO PEDIDO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO POSSUI NOME DA MAE")] <- ""
db2\$MAE[which(db2\$MAE=="NAO RELATADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO SOLICITADO")] <- ""
db2\$MAE[which(db2\$MAE=="NAO TEM")] <- ""

```

db2$MAE[which(db2$MAE=="NAO TEM NOME DA MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM O NOME DA MAE")] <- ""
db2$MAE[which(db2$MAE=="NAOCONSTA")] <- ""
db2$MAE[which(db2$MAE=="NAOINFORMADO")] <- ""
db2$MAE[which(db2$MAE=="NAOMINFORMADO")] <- ""
db2$MAE[which(db2$MAE=="NC")] <- ""
db2$MAE[which(db2$MAE=="ND")] <- ""
db2$MAE[which(db2$MAE=="NNN")] <- ""
db2$MAE[which(db2$MAE=="NNNNNNNNNNNNNNNNNN")] <- ""
db2$MAE[which(db2$MAE=="SN")] <- ""
db2$MAE[which(db2$MAE=="NOA CONSTA")] <- ""
db2$MAE[which(db2$MAE=="NOA INFORMADO")] <- ""
db2$MAE[which(db2$MAE=="NT")] <- ""
db2$MAE[which(db2$MAE=="O")] <- ""
db2$MAE[which(db2$MAE=="PORTO")] <- ""
db2$MAE[which(db2$MAE=="S")] <- ""
db2$MAE[which(db2$MAE=="S NOME")] <- ""
db2$MAE[which(db2$MAE=="SD")] <- ""
db2$MAE[which(db2$MAE=="SEM INFORMACAO")] <- ""
db2$MAE[which(db2$MAE=="SEM INFORMCAO")] <- ""
db2$MAE[which(db2$MAE=="SEM MAE")] <- ""
db2$MAE[which(db2$MAE=="SEM NADA")] <- ""
db2$MAE[which(db2$MAE=="SEM NOME")] <- ""
db2$MAE[which(db2$MAE=="SEM NOME DA MAE")] <- ""
db2$MAE[which(db2$MAE=="SEM REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="SM")] <- ""
db2$MAE[which(db2$MAE=="SN")] <- ""
db2$MAE[which(db2$MAE=="VVV")] <- ""
db2$MAE[which(db2$MAE=="1")] <- ""
db2$MAE[which(db2$MAE=="2")] <- ""
db2$MAE[which(db2$MAE=="3")] <- ""
db2$MAE[which(db2$MAE=="4")] <- ""
db2$MAE[which(db2$MAE=="5")] <- ""
db2$MAE[which(db2$MAE=="6")] <- ""
db2$MAE[which(db2$MAE=="NAO COLOCARAM O NOME DA MAE NA FICHA")] <- ""
db2$MAE[which(db2$MAE=="NAO CONSTA NA ID")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMADO NA RG")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMADO.")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMADO?")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMADO")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMOU")] <- ""
db2$MAE[which(db2$MAE=="NAO INF")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMA")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMADA")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMADO")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMOU")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORMOU NO QUESTIONARIO")] <- ""
db2$MAE[which(db2$MAE=="NAO INFRMADO")] <- ""
db2$MAE[which(db2$MAE=="NAO LEGIVEL")] <- ""
db2$MAE[which(db2$MAE=="NAO POSSUI")] <- ""
db2$MAE[which(db2$MAE=="NAO MENCIONADO")] <- ""
db2$MAE[which(db2$MAE=="NAO OSHIRO")] <- ""
db2$MAE[which(db2$MAE=="NAP")] <- ""
db2$MAE[which(db2$MAE=="NAPO")] <- ""

```



```

db2$MAE[which(db2$MAE=="NAO PEGAMOS")] <- ""
db2$MAE[which(db2$MAE=="NAO RECONHECIDA")] <- ""
db2$MAE[which(db2$MAE=="NAO SABE")] <- ""
db2$MAE[which(db2$MAE=="NAO SEI")] <- ""
db2$MAE[which(db2$MAE=="NA IDENTIFICADO")] <- ""
db2$MAE[which(db2$MAE=="NAO DISPONIVEL")] <- ""
db2$MAE[which(db2$MAE=="NAO HA REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="NAO INFOMRADO")] <- ""
db2$MAE[which(db2$MAE=="NAO INFORAMADO")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE NO REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE REGISTRADA")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE REEISTRADA")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM MAE REGISTRADA SEM MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM REGISTRO DE MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM NO RG")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM NOME")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM REGISTRO DE MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM TEM REGISTRO DE MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM REGITRO DE MAE")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM RESGITRO DE MAO")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM NOME")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM NO CADASTRO")] <- ""
db2$MAE[which(db2$MAE=="NAO TEM NO CADSTRO")] <- ""
db2$MAE[which(db2$MAE=="NAO TENHO")] <- ""
db2$MAE[which(db2$MAE=="NAO VEIO")] <- ""
db2$MAE[which(db2$MAE=="ORFAO")] <- ""
db2$MAE[which(db2$MAE=="ORFAOI")] <- ""
db2$MAE[which(db2$MAE=="NAO TENHO O NOME DA MAE")] <- ""
db2$MAE[which(db2$MAE=="S")] <- ""
db2$MAE[which(db2$MAE=="SM")] <- ""
db2$MAE[which(db2$MAE=="SN")] <- ""
db2$MAE[which(db2$MAE=="S N")] <- ""
db2$MAE[which(db2$MAE=="S REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="S/ NOME")] <- ""
db2$MAE[which(db2$MAE=="S/IDENTIFICACAO")] <- ""
db2$MAE[which(db2$MAE=="S/ INFORMACAO")] <- ""
db2$MAE[which(db2$MAE=="SEM CADASTRO")] <- ""
db2$MAE[which(db2$MAE=="SEM IDENTIFICACAO")] <- ""
db2$MAE[which(db2$MAE=="SEM IDENTIFICACAO NA IDENT")] <- ""
db2$MAE[which(db2$MAE=="SEM INFORMACAO")] <- ""
db2$MAE[which(db2$MAE=="SEM NOME")] <- ""
db2$MAE[which(db2$MAE=="SE MAE")] <- ""
db2$MAE[which(db2$MAE=="SEM MAE")] <- ""
db2$MAE[which(db2$MAE=="SEM DADOS")] <- ""
db2$MAE[which(db2$MAE=="SEM REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="S REGISTRO")] <- ""
db2$MAE[which(db2$MAE=="SDEM NOME")] <- ""

```



```

db2$MAE[which(db2$MAE=="MMMMMMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
M")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
MMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
MMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
MMMMMMMMMMMMM")] <- ""
db2$MAE[which(db2$MAE=="MNA")] <- ""
db2$MAE[which(db2$MAE=="MNFDKGDKGDSNIG")] <- ""
db2$MAE[which(db2$MAE=="MNJKJHKML,JK")] <- ""
db2$MAE[which(db2$MAE=="MNPA")] <- ""
db2$MAE[which(db2$MAE=="MDJDJFHDJ")] <- ""
db2$MAE[which(db2$MAE=="MDSFHFAJHJI")] <- ""
db2$MAE[which(db2$MAE=="MDSJHUHFUAHFG")] <- ""
db2$MAE[which(db2$MAE=="ZXV TPUOB")] <- ""
db2$MAE[which(db2$MAE=="ZXXX")] <- ""
db2$MAE[which(db2$MAE=="ZXZCZ")] <- ""
db2$MAE[which(db2$MAE=="ZXZCZ")] <- ""
db2$MAE[which(db2$MAE=="O.")] <- ""
db2$MAE[which(db2$MAE=="O")] <- ""
db2$MAE[which(db2$MAE=="NAP")] <- ""
db2$MAE[which(db2$MAE=="NPA")] <- ""
db2$MAE[which(db2$MAE=="NOA")] <- ""
db2$MAE[which(db2$MAE=="NA")] <- ""
db2$MAE[which(db2$MAE=="MAOA")] <- ""
db2$MAE[which(db2$MAE=="MNPA")] <- ""
db2$MAE[which(db2$MAE=="PAIOLE,A")] <- ""
db2$MAE[which(db2$MAE=="HHH")] <- ""
db2$MAE[which(db2$MAE=="X")] <- ""
db2$MAE[which(db2$MAE=="AA")] <- ""
db2$MAE[which(db2$MAE=="D")] <- ""
db2$MAE[which(db2$MAE=="N C")] <- ""
db2$MAE[which(db2$MAE=="AAA")] <- ""
db2$MAE[which(db2$MAE=="AAAA")] <- ""
db2$MAE[which(db2$MAE=="ARTIGO 61 LEI6015/73")] <- ""
db2$MAE[which(db2$MAE=="ARTIGO LEI")] <- ""
db2$MAE[which(db2$MAE=="ARTIGO SESENTA E UM, LEI SEIS MIL E QUINZE")] <- ""
db2$MAE[which(db2$MAE=="B")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""
db2$MAE[which(db2$MAE=="~")] <- ""
db2$MAE[which(db2$MAE=="D")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""
db2$MAE[which(db2$MAE=="[]")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""
db2$MAE[which(db2$MAE=="")] <- ""

```

```
db2$MAE[which(db2$MAE=="N/C")]<-""
db2$MAE[which(db2$MAE=="/")]<-""
```

```
db2$NOME[which(db2$NOME=="")] <- ""
db2$NOME[which(db2$NOME==";")] <- ""
db2$NOME[which(db2$NOME=="2")] <- ""
db2$NOME[which(db2$NOME=="III")] <- ""
```

Passo 22

```
# Acrescentar M10 nos códigos vazios das variáveis NOME, MAE, ENDEREÇO, MUNICIPIO,
CNS, DTNASC e DTEXAME
```

```
db2$NOME<-as.character(db2$NOME)
db2$MAE<-as.character(db2$MAE)
db2$ENDERECO<-as.character(db2$ENDERECO)
db2$MUNICIPIO<-as.character(db2$MUNICIPIO)
db2$DTNASC<-as.character(db2$DTNASC)
db2$CNSUS<-as.character(db2$CNSUS)
db2$DT_EXAME<-as.character(db2$DT_EXAME)
```

```
db2$MAE[is.na(db2$MAE)]<-"M10"
db2$NOME[is.na(db2$NOME)]<-"M10"
db2$ENDERECO[is.na(db2$ENDERECO)]<-"M10"
db2$MUNICIPIO[is.na(db2$MUNICIPIO)]<-"M10"
db2$DTNASC[is.na(db2$DTNASC)]<-"M10"
db2$CNSUS[is.na(db2$CNSUS)]<-"M10"
db2$DT_EXAME[is.na(db2$DT_EXAME)]<-"M10"
```

Passo 23

```
# Manter apenas um espaço entre as palavras
```

```
install.packages(stringr)
library(stringr)
db2$NOME<- str_replace_all(string = db2$NOME, pattern = " ", repl=" ")
db2$MAE<- str_replace_all(string = db2$MAE, pattern = " ", repl=" ")
```

```
# Salvar e abrir o arquivo
```

```
write.dbf(db2,"db2.dbf")
db2<- read.dbf("db2.dbf")
```

```
# Verificar se tem campos vazios
```

```
summary(db2)
```

Passo 24

```
#Criar variável identificadora
```

```
db2<-read.dbf("db2.dbf")
db2$FONTE<-"M10"
db2$REGISTRO<-1:485243
db2$FONTE<-paste(db2$FONTE, sep=" ", db2$REGISTRO)
write.dbf(db2,"db2.dbf")
```

Passo 25

```

#Retirar duplicidade de exames usando as variáveis NOME, MAE, DTNASC, D_EXAME2,
CNES_REQ E MUN_COD – DUPLICIDADE 1
db2<- read.dbf("db2.dbf")
names(db2)
aux<- which(duplicated(db2[,c("NOME","MAE","DTNASC","D_EXAME2", "CNES_REQ",
"MUN_COD")]))
db2dup <- db2[aux,]
if(length(aux)>0){db3 <- db2[-aux,]}else{db3 <- db2}

write.dbf(db2dup, file="db2dup.dbf")
write.dbf(db3, file="db3.dbf")

# Criar a variável mês de exame
library(zoo)
db3$MES_EXAME<- as.Date(db3$D_EXAME2,"%Y/%m/%d")

library(lubridate)
db3$MES_EXAME<- month(db3$D_EXAME2)

write.dbf(db3, file="db3.dbf")
db3<-read.dbf("db3.dbf")

###OBS.: LEVAR O BANCO PARA O RECLINK PARA RETIRAR A DUPLICIDADE DE
EXAMES (DUPLICIDADE 2)

### Passo 26
# Voltar com banco do RecLink
# Fazer o relacionamento entre db3PAD e dbMESTRE
db3MESTRE<-read.dbf("db3MESTRE.dbf")
db3PAD<-read.dbf("db3PAD.dbf")

names(db3MESTRE)
names(db3PAD)

# Retirar duplicidades do banco original (db3PAD menos db3Mestre)
aux<-which(duplicated(db3PAD[,c("UNICO")]))
db3dup<-db3PAD[aux,]
if(length(aux)>0){db3MESTRE<-db3PAD[-aux,]}else{db3MESTRE<-db3PAD}

names(db3dup)

write.dbf(db3dup, "db3dup.dbf")

#Retirar duplicidades do banco original do R (db3) - Utilizar a variável FONTE
db3<-read.dbf("db3.dbf")
names(db3)

#Fazer o merge entre db3MESTRE e db3 para ficar com todas as variáveis
db4<-merge(db3MESTRE,db3,by="FONTE",all.x=TRUE)
write.dbf(db4,"db4.dbf")
names(db4)

#Excluir variáveis .x criadas no merge

```

```

excluir<-c("DTNASC.x", "D_EXAME2.x", "ANO_NASC.x", "CNSUS.x", "MUN_COD.x",
"MES_EXAME.x","CNES.x","CNES_REQ.x")

db4 <- db4[!(names(db4)%in% excluir)]

names(db4)

#Renomear as variáveis .y
names(db4)<- c("FONTE", "NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE",
"UMAE", "UNICO", "CODMUN6", "CNES", "COMPET", "N_EXAME", "NOME", "MAE",
"DTNASC", "ESCOL", "ENDERECO", "NUMERO", "COMPLEM", "BAIRRO", "UF", "IBGE",
"CNSUS", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME", "NODULO", "NOD_D",
"NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG", "MAMOG_DIAG",
"RASTREAM", "NUM_FIL_MD", "PELE_MD", "COMP_MD", "NOD_MD",
"NOD_TAM_MD", "NUM_FIL_ME", "PELE_ME", "COMP_ME", "NOD_ME",
"NOD_TAM_ME", "DIAG_MD", "DIAG_ME", "RECOM_MD", "RECOM_ME",
"DT_EXAME", "DT_SOLICIT", "DT_LIBERA", "SEXO", "RACA", "ETNIA", "NACIONAL",
"D_EXAME2", "D_ID_DTNAS", "ANO", "ANO_NASC", "IDADE", "MUN_COD",
"MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD", "MACRO", "ESTADO", "REGISTRO",
"MES_EXAME")
write.dbf(db4,"db4.dbf")

#Retirar duplicidade RecLink – Duplicidade 3

# Voltar com banco do RecLink
# Fazer o relacionamento entre db4PAD e db4MESTRE
db4MESTRE<-read.dbf("db4MESTRE.dbf")
db4PAD<-read.dbf("db4PAD.dbf")

names(db4MESTRE)
names(db4PAD)

# Retirar duplicidades do banco original (db4PAD menos db4Mestre)
aux<-which(duplicated(db4PAD[,c("UNICO")]))
db4dup<-db4PAD[aux,]
if(length(aux)>0){db4MESTRE<-db4PAD[-aux,]}else{db4MESTRE<-db4PAD}

names(db4dup)

write.dbf(db4dup, "db4dup.dbf")

#Retirar duplicidades do banco original do R (db4) - Utilizar a variável FONTE
db4<-read.dbf("db4.dbf")
names(db4)

#Fazer o merge entre db4MESTRE e db4 para ficar com todas as variáveis
db5<-merge(db4MESTRE,db4,by="FONTE",all.x=TRUE)
write.dbf(db5,"db5.dbf")
names(db5)

#Excluir variáveis .x criadas no merge
excluir<-c("NOMEPAD.x", "PNOME.x","UNOME.x", "MAEPAD.x", "PMAE.x", "UMAE.x",
"DTNASC.x", "D_EXAME2.x", "ANO_NASC.x", "CNSUS.x", "MUN_COD.x",
"MES_EXAME.x","CNES.x", "CNES_REQ.x", "UNICO.x")

```

```
db5 <- db5[,!(names(db5)%in% excluir)]
```

```
names(db5)
```

```
#Renomear as variáveis .y
```

```
names(db5)<- c("FONTE", "NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE",
"UMAE", "UNICO", "CODMUN6", "CNES", "COMPET", "N_EXAME", "NOME", "MAE",
"DTNASC", "ESCOL", "ENDERECO", "NUMERO", "COMPLEM", "BAIRRO", "UF", "IBGE",
"CNSUS", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME", "NODULO", "NOD_D",
"NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG", "MAMOG_DIAG",
"RASTREAM", "NUM_FIL_MD", "PELE_MD", "COMP_MD", "NOD_MD",
"NOD_TAM_MD", "NUM_FIL_ME", "PELE_ME", "COMP_ME", "NOD_ME",
"NOD_TAM_ME", "DIAG_MD", "DIAG_ME", "RECOM_MD", "RECOM_ME",
"DT_EXAME", "DT_SOLICIT", "DT_LIBERA", "SEXO", "RACA", "ETNIA", "NACIONAL",
"D_EXAME2", "D_ID_DTNAS", "ANO", "ANO_NASC", "IDADE", "MUN_COD",
"MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD", "MACRO", "ESTADO", "REGISTRO",
"MES_EXAME")
write.dbf(db5,"db5.dbf")
```

```
###Passo 27
```

```
#Retirar duplicidade de nomes (repetição) – Repetição 1
```

```
db5<- read.dbf("db5.dbf")
```

```
aux<- which(duplicated(db5[,c("NOME", "MAE", "DTNASC", "MUN_COD")]))
```

```
db5dup<- db5[aux,]
```

```
if(length(aux)>0){db6 <- db5[-aux,]}else{db6 <- db5}
```

```
write.dbf(db5dup, file="db5dup.dbf")
```

```
write.dbf(db6, file="db6.dbf")
```

```
db5dup<- read.dbf("db5dup.dbf")
```

```
db6<-read.dbf("db6.dbf")
```

```
names(db6)
```

```
#OBS.: Levar o banco db6 para o RecLink e retirar a duplicidade de exames (repetição) no RecLink
-Repetição 2
```

```
### Passo 28
```

```
# Voltar com banco do RecLink
```

```
# Fazer o relacionamento entre db5PAD e db5MESTRE
```

```
db6MESTRE<-read.dbf("db6MESTRE.dbf")
```

```
db6PAD<-read.dbf("db6PAD.dbf")
```

```
names(db6MESTRE)
```

```
names(db6PAD)
```

```
# Retirar duplicidades do banco original (db6PAD menos db6Mestre)
```

```
aux<-which(duplicated(db6PAD[,c("UNICO")]))
```

```
db6dup<-db6PAD[aux,]
```

```
if(length(aux)>0){db6MESTRE<-db6PAD[-aux,]}else{db6MESTRE<-db6PAD}
```

```

names(db6dup)

write.dbf(db6dup, "db6dup.dbf")

#Retirar duplicidades do banco original do R (db5) - Utilizar a variável FONTE
db6<-read.dbf("db6.dbf")
names(db6)

#Fazer o merge entre db5MESTRE e db5 para ficar com todas as variáveis
db7<-merge(db6MESTRE,db6,by="FONTE",all.x=TRUE)
write.dbf(db7,"db7.dbf")
names(db7)

#Excluir variáveis .x criadas no merge
excluir<-c("NOMEPAD.x", "PNOME.x", "UNOME.x", "MAEPAD.x", "PMAE.x", "UMAE.x",
"DTNASC.x", "D_EXAME2.x", "ANO_NASC.x","CNSUS.x", "MUN_COD.x",
"MES_EXAME.x","CNES.x", "CNES_REQ.x", "UNICO.x")
db7 <- db7[,!(names(db7)%in%excluir)]

names(db7)

#Renomear as variáveis .y
names(db7) <- c("FONTE", "NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE",
"UMAE", "UNICO", "CODMUN6", "CNES", "COMPET","N_EXAME", "NOME", "MAE",
"DTNASC", "ESCOL", "ENDERECO", "NUMERO","COMPLEM", "BAIRRO", "UF", "IBGE",
"CNSUS", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME", "NODULO", "NOD_D",
"NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG", "MAMOG_DIAG",
"RASTREAM", "NUM_FIL_MD","PELE_MD", "COMP_MD", "NOD_MD", "NOD_TAM_MD",
"NUM_FIL_ME", "PELE_ME", "COMP_ME", "NOD_ME", "NOD_TAM_ME", "DIAG_MD",
"DIAG_ME", "RECOM_MD", "RECOM_ME", "DT_EXAME", "DT_SOLICIT", "DT_LIBERA",
"SEXO", "RACA", "ETNIA", "NACIONAL", "D_EXAME2" , "D_ID_DTNAS", "ANO",
"ANO_NASC", "IDADE", "MUN_COD", "MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD",
"MACRO", "ESTADO", "REGISTRO","MES_EXAME")
write.dbf(db7,"db7.dbf")

# Levar db7 para o RecLink para retirada de repetição – Repetição 3
# Voltar com banco do RecLink
# Fazer o relacionamento entre db7PAD e db7MESTRE
db7MESTRE<-read.dbf("db7MESTRE.dbf")
db7PAD<-read.dbf("db7PAD.dbf")

names(db7MESTRE)
names(db7PAD)

# Retirar duplicidades do banco original (db7PAD menos db7Mestre)
aux<-which(duplicated(db7PAD[,c("UNICO")]))
db7dup<-db7PAD[aux,]
if(length(aux)>0){db7MESTRE<-db7PAD[-aux,]}else{db7MESTRE<-db7PAD}

names(db7dup)

write.dbf(db7dup, "db7dup.dbf")

#Retirar duplicidades do banco original do R (db7) - Utilizar a variável FONTE

```



```

db7<-read.dbf("db7.dbf")
names(db7)

#Fazer o merge entre db7MESTRE e db7 para ficar com todas as variáveis
db8<-merge(db7MESTRE,db7,by="FONTE",all.x=TRUE)
write.dbf(db8,"db8.dbf")
names(db8)

#Excluir variáveis .x criadas no merge
excluir<-c("NOMEPAD.x", "PNOME.x", "UNOME.x", "MAEPAD.x", "PMAE.x", "UMAE.x",
"DTNASC.x", "D_EXAME2.x", "ANO_NASC.x","CNSUS.x", "MUN_COD.x",
"MES_EXAME.x","CNES.x", "CNES_REQ.x", "UNICO.x")
db8 <- db8[,!(names(db8)%in%excluir)]

names(db8)

#Renomear as variáveis .y
names(db8) <- c("FONTE", "NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE",
"UMAE", "UNICO", "CODMUN6", "CNES", "COMPET","N_EXAME", "NOME", "MAE",
"DTNASC", "ESCOL", "ENDERECO", "NUMERO","COMPLEM", "BAIRRO", "UF", "IBGE",
"CNSUS", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME", "NODULO", "NOD_D",
"NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG", "MAMOG_DIAG",
"RASTREAM", "NUM_FIL_MD","PELE_MD", "COMP_MD", "NOD_MD", "NOD_TAM_MD",
"NUM_FIL_ME", "PELE_ME", "COMP_ME", "NOD_ME", "NOD_TAM_ME", "DIAG_MD",
"DIAG_ME", "RECOM_MD", "RECOM_ME", "DT_EXAME", "DT_SOLICIT", "DT_LIBERA",
"SEXO", "RACA", "ETNIA", "NACIONAL", "D_EXAME2" , "D_ID_DTNAS", "ANO",
"ANO_NASC", "IDADE", "MUN_COD", "MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD",
"MACRO", "ESTADO", "REGISTRO","MES_EXAME")
write.dbf(db8,"db8.dbf")
db8<-read.dbf("db8.dbf")

# Fazer uma cópia do arquivo, salvar e abrir o arquivo
write.dbf(db8, file="Mamog2010.dbf")
Mamog2010<-read.dbf("Mamog2010.dbf")
``

#Mamog2010 será relacionado com Mamog2009 no RecLink para retirada de exames alterados.

```

2. Limpeza do Banco de Exames Alterados

```

# Script para limpeza do banco 2009 - retirada de exames alterados
##### PREPARAR O BANCO DE 2009
### Repetir os passos 1, 2, 3, 4, 5, 6, 9, 10, 11,
## Banco 2009
Observação no passo 6 foi utilizado o período de 31/12/2008 a 01/01/2010.

### Passo 7
#Passo 1 a 3 já prontos
library(foreign)
banco2010 <- read.dbf("banco2010.dbf")
banco2011 <- read.dbf("banco2011.dbf")
banco2012 <- read.dbf("banco2012.dbf")

```

Observação: dos bancos 2010, 2011 e 2012 foram extraídos exames solicitados no período de 31/12/2008 a 01/01/2010.

```
# Passo 8
# Juntar os bancos de dados
db2009<-rbind(novob2009,b2010,b2011,b2012)
```

```
# Salvar o banco
write.dbf (db2009, file="db2009")
```

Observação: etapas 9, 10 e 11 iguais ao da base referência

```
# Extrair os exames com Bi_rads 3, 4, 5 e 6 (códigos 04, 05, 06 e 07)
0 - Default
01 - categoria 0 (inconclusivo, avalia-se adicional por imagem ou exame anterior)
02 - categoria 1 (sem achados)
03 - categoria 2 (achados benignos)
04 - categoria 3 (achado provavelmente benigno - repetir exame em 6 meses)
05 - categoria 4 (achados suspeitos de malignidade - bi-ipsia)
06 - categoria 5 (achados altamente sugestivos de malignidade)
07 - categoria 6 (bi-ipsia pr-ativa com malignidade comprovada)
```

```
table(Banco2009total$DIAG_MD,useNA = "ifany")
table(Banco2009total$DIAG_ME,useNA = "ifany")
```

```
B2009<-
subset(Banco2009total,(DIAG_MD=="4")|(DIAG_MD=="5")|(DIAG_MD=="6")|(DIAG_MD=="
7")|(DIAG_ME=="4")|(DIAG_ME=="5")|(DIAG_ME=="6")|(DIAG_ME=="7"))
```

```
table (B2009$DIAG_MD, useNA="ifany")
table (B2009$DIAG_ME, useNA="ifany")
```

```
write.dbf(B2009,"B2009.dbf")
B2009<-read.dbf("B2009.dbf")
```

Observação: etapas de 12 a 28 iguais ao da base referência.

```
# Fazer uma cópia do arquivo, salvar e abrir o arquivo
write.dbf(BD5, file="Mamog2009.dbf")
Mamog2009<-read.dbf("Mamog2009.dbf")
```

```
#### ETAPA 2 - IR PARA O RECKLINK PARA FAZER RELACIONAMENTO
PROBABILÍSTICO
```

```
##Observação: Foi feito relacionamento entre o banco 2009 e 2010 por meio de 16 passos para
verificar as melhores variáveis e escore.
```

3. Análise dos pares formados após relacionamento probabilístico e dos melhores passos para demais relacionamentos.

```
Mamog2010PAD<-read.dbf("Mamog2010PAD.dbf")
Mamog2009PAD<-read.dbf("Mamog2009PAD.dbf")
```

```
#### ETAPA 3 - VOLTAR AO R - com cada passo
library(foreign)
```

```

setwd("G:/SISMAMA/Passos")
passo1 <- read.dbf("Passo1.dbf")
passo2 <- read.dbf("Passo2.dbf")
passo3 <- read.dbf("Passo3.dbf")
passo4 <- read.dbf("Passo4.dbf")
passo5 <- read.dbf("Passo5.dbf")
passo6 <- read.dbf("Passo6.dbf")
passo7 <- read.dbf("Passo7.dbf")
passo8 <- read.dbf("Passo8.dbf")
passo9 <- read.dbf("Passo9.dbf")
passo10 <- read.dbf("Passo10.dbf")
passo11 <- read.dbf("Passo11.dbf")
passo12 <- read.dbf("Passo12.dbf")
passo13 <- read.dbf("Passo13.dbf")
passo14 <- read.dbf("Passo14.dbf")
passo15 <- read.dbf("Passo15.dbf")
passo16 <- read.dbf("Passo16.dbf")

```

O comando abaixo foi repetido com todos os passos, apresento passo 1 como exemplo.

```
names(passo1)
```

```
# criando a variável pares
```

```
passo1$pares <- as.factor(paste(passo1$REFREC,sep=" ",passo1$COMPREC))
```

```
# selecionando apenas os pares verdadeiros
```

```
passo1v <- passo1[passo1$MATCH=="X",]
```

```
#Pares falsos negativos
```

```
table(passo1v$SCORE, passo1v$MATCH=="X")
```

```
# selecionando apenas os pares falsos
```

```
passo1f <- passo1[passo1$MATCH=="!",]
```

```
#Pares falsos positivos
```

```
table(passo1f$SCORE, passo1f$MATCH=="!")
```

```
predictions_passo1 <- predict(passo1, teste)
```

```
matriz_regressao <- confusionMatrix(predictions_passo1, teste$match)
```

```
matriz_regressao
```

```
# comparando os bancos
```

```
library(arsenal)
```

```
#Passo 1
```

```
compare(passo1v,passo2v,by="pares")[1]
```

```
compare(passo1v,passo3v,by="pares")[1]
```

```
compare(passo1v,passo4v,by="pares")[1]
```

```
compare(passo1v,passo5v,by="pares")[1]
```

```
compare(passo1v,passo6v,by="pares")[1]
```

```
compare(passo1v,passo7v,by="pares")[1]
```

```
compare(passo1v,passo8v,by="pares")[1]
```

```
compare(passo1v,passo9v,by="pares")[1]
```

```
compare(passo1v,passo10v,by="pares")[1]
```

```
compare(passo1v,passo11v,by="pares")[1]
```

```
compare(passo1v,passo12v,by="pares")[1]
compare(passo1v,passo13v,by="pares")[1]
compare(passo1v,passo14v,by="pares")[1]
compare(passo1v,passo15v,by="pares")[1]
compare(passo1v,passo16v,by="pares")[1]
```

Quadro gerado após a comparação de todos os passos

N	ntotal	ntotal v	passo1	passo2	passo3	passo4	passo5	passo6	passo7	passo8	passo9	passo10	passo11	passo12	passo13	passo14	passo15	passo16
passo1	2108	1076	-	1069	202	923	1069	922	956	1012	1073	923	1009	1072	922	957	1071	1070
passo2	6067	1069	1069	-	201	918	1064	918	951	1007	1067	918	1004	1066	918	951	1066	1066
passo3	216	216	202	201	-	190	201	199	203	205	212	192	199	205	201	205	203	212
passo4	932	923	923	918	190	-	922	922	922	923	923	923	921	923	922	923	922	921
passo5	1373	1069	1069	1064	201	922	-	922	956	1009	1069	922	1007	1069	922	956	1069	1069
passo6	1067	983	922	918	199	922	922	-	983	983	983	922	920	922	983	983	922	980
passo7	1188	1022	956	951	203	922	956	983	-	983	1021	922	920	956	983	1021	956	1018
passo8	1223	1078	1012	1007	205	923	1009	983	983	-	1075	923	1009	1011	983	984	1010	1071
passo9	2620	1142	1073	1067	212	923	1069	983	1021	1075	-	923	1009	1072	983	1022	1071	1136
passo10	991	950	923	918	192	923	922	922	922	923	923	-	948	950	949	950	922	921
passo11	1181	1042	1009	1004	199	921	1007	920	920	1009	1009	948	-	1040	947	948	1008	1007
passo12	2143	1106	1072	1066	205	923	1069	922	956	1011	1072	950	1040	-	949	986	1071	1069
passo13	1286	1012	922	918	201	922	922	983	983	983	983	949	947	949	-	1011	922	980
passo14	1591	1053	957	951	205	923	956	983	1021	984	1022	950	948	986	1011	-	956	1018
passo15	1919	1082	1071	1066	203	922	1069	922	956	1010	1071	922	1008	1071	922	956	-	1078
passo16	12581	1145	1070	1065	212	921	1068	980	1018	1071	1136	921	1007	1069	980	1018	1078	-

Observação: Em amarelo, seguindo a linha a combinação que tinha maior número de pares.

Exemplo: Passo 1 com passo 9.

juntando todos os bancos (apenas as variáveis pares e SCORE)

```
passo1s <- passo1v[,c(5,3)]
names(passo1s) <- c("pares","score1")
```

```
passo2s <- passo2v[,c(5,3)]
names(passo2s) <- c("pares","score2")
```

Observação: comando repetido até o passo 16.

```
passos <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="pares",all=T),
```

```
list(passo1s,passo2s,passo3s,passo4s,passo5s,passo6s,passo7s,passo8s,passo9s,passo10s,passo11s,
passo12s,passo13s,passo14s,passo15s,passo16s))
write.csv(passos,"passos.csv",row.names = F)
```

```
passos<-read.csv("passos.csv")
```

```
summary(passos)
```

testar os passos que juntos englobam a maioria dos passos

#Teste 1 - 16 e 9

```
teste1 <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="pares",all=T),
list(passo16s,passo9s))
write.csv(teste1,"teste1.csv",row.names = F)
```

```
teste1<-read.csv("teste1.csv")
```

Os passos foram agregados na seguinte ordem: 16, 9, 12, 1, 8, 2, 5, 14, 11, 7 e 13. Foi verificado se cada passo mantinha o mesmo número de pares ou agregava novos pares. Foram realizados 11 testes, conforme quadro abaixo:

#ETAPA 4 - Verificar a acurácia por meio de uma amostra

#verificando o tamanho da amostra

Teste	Passos	Pares	Percentual (1.194)	Observação
1	16, 9	1.151	96.4%	Faltam 43 pares para o total (1.194). O Passo 16 tem 1145 pares. O Passo 9 acrescentou mais 6 pares.
2	16, 9, 12	1.185	99.24%	Acrescentou mais 34 pares.
3	16, 9, 12, 8	1.185	99.24%	Não acrescentou nada. Retirar o 8
4	16, 9, 12 e 1	1.188	99.50%	Acrescentou mais 3 pares
5	16, 9, 12, 1 e 15	1.188	99.50%	Não acrescentou nada. Retirar o 15
6	16, 9, 12, 1 e 2	1.188	99.50%	Não acrescentou nada. Retirar o 2.
7	16, 9, 12, 1 e 5	1.188	99.50%	Não acrescentou nada. Retirar o 5.
8	16, 9, 12, 1 e 14	1.190	99.67%	Acrescentou 2 pares.
9	16, 9, 12, 1, 14 e 11	1.192	99.83%	Acrescentou 2 pares.
10	16, 9, 12, 1, 14, 11 e 7	1.193	99.92%	Acrescentou 1 par.
11	16, 9, 12, 1, 14, 11, 7 e 13	1.194	100.00%	Acrescenta 1 par.
Obs.: 1) Os passos que abrangem todos os pares são: 16, 9, 12, 1, 14, 11, 7 e 13.				

library(pROC)

```
power.roc.test(auc=0.80, sig.level=0.01, power=0.95, kappa=1)
power.roc.test(auc=0.90, sig.level=0.01, power=0.95, kappa=1)
power.roc.test(auc=0.80, sig.level=0.01, power=0.95, kappa=1.5)
power.roc.test(auc=0.90, sig.level=0.01, power=0.95, kappa=1.5)
```

```
power.roc.test(auc=0.80, sig.level=0.01, power=0.99, kappa=1)
power.roc.test(auc=0.90, sig.level=0.01, power=0.99, kappa=1)
power.roc.test(auc=0.80, sig.level=0.01, power=0.99, kappa=1.5)
power.roc.test(auc=0.90, sig.level=0.01, power=0.99, kappa=1.5)
```

Resultado:

sig.level					
0.01					
				poder=0.95	poder=0.99
kappa	1	auc	0.80	57	76
			0.9	27	34
	1.5	auc	0.80	60	80
			0.9	29	37

Observação: Uma amostra de tamanho 100 é suficiente. Foram realizados testes com tamanhos de amostra 200 e 300.

```
#Caminho do arquivo
library(foreign)
setwd("G:/SISMAMA/Amostra")
```

```
#Passo1
```

```

# selecionando uma amostra
library(foreign)
passo1 <- read.dbf("passo1.dbf")
library(dplyr)
passo1a <- sample_n(passo1, 100)
write.dbf(passo1a,"passo1a.dbf")

#OBSERVAÇÃO: Abrir o passo completo no RecLink e marcar todos como pares, nomear o passo
com ver

#ler o banco
passo1a<-read.dbf("passo1a.dbf")
passo1ver<-read.dbf("passo1ver.dbf")

#criar variáveis pares
passo1a$pares <- as.factor(paste(passo1a$REFREC,sep=" ",passo1a$COMPREC))
passo1ver$pares <- as.factor(paste(passo1ver$REFREC,sep=" ",passo1ver$COMPREC))

names(passo1a)
names(passo1ver)

#Fazer o merge
A1 <- merge(passo1a,passo1ver,by="pares",all.x=TRUE)
write.dbf(A1, "A1.dbf")

names(A1)

#Retirar variáveis excedentes
excluir<-c("COMPREC.y", "REFREC.y", "SCORE.y", "C_PNOME", "C_UNOME", "C_PMAE",
"C_UMAE", "C_DT_EXAME", "C_ANO_NASC", "R_PNOME", "R_UNOME", "R_PMAE",
"R_UMAE", "R_DT_EXAME", "R_ANO_NASC")

A1 <- A1[!(names(A1)%in% excluir)]
write.dbf(A1, "A1.dbf")

# Conferir o passo A1 no Excel
# Renomear variáveis
A1<-read.csv2("A1.csv")
names(A1)
names(A1)<-c("pares", "COMPREC", "REFREC", "SCORE","MATCH", "C_NOMEPAD",
"R_NOMEPAD", "C_MAEPAD", "R_MAEPAD", "C_DTNASC", "R_DTNASC",
"C_ENDERECO", "R_ENDERECO", "C_CNSUS", "R_CNSUS", "C_MUN_COD",
"R_MUN_COD", "C_FONTE", "R_FONTE")
write.dbf(A1,"A1.dbf")

#Converter variáveis SCORE número
A1$SCORE<-as.numeric(A1$SCORE)
write.dbf(A1,"A1.dbf")

# avaliar a acurácia da amostra
library(pROC)
bd1 <- read.dbf("A1.dbf")

```

```
roc(bd1$MATCH,bd1$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thr
es="best",legacy.axes=T)
```

```
roc1 <-
roc(bd1$MATCH,bd1$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thr
es="best",legacy.axes=T)
names(roc1)
```

```
acuracia <- as.data.frame(cbind(roc1$thresholds,roc1$sensitivities,roc1$specificities))
names(acuracia) <- c("escore","sensibilidade","especificidade")
write.csv(acuracia,"acuracia1.csv",row.names = F)
```

Amostra 200

```
passo1 <- read.dbf("passo1.dbf")
library(dplyr)
passo1b <- sample_n(passo1, 200)
write.dbf(passo1b,"passo1b.dbf")
```

```
bd1 <- read.dbf("passo1b.dbf")
library(pROC)
```

```
roc(bd1$MATCH,bd1$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thr
es="best",legacy.axes=T)
```

```
roc1 <-
roc(bd1$MATCH,bd1$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thr
es="best",legacy.axes=T)
names(roc1)
```

```
acuracia <- as.data.frame(cbind(roc1$thresholds,roc1$sensitivities,roc1$specificities))
names(acuracia) <- c("escore","sensibilidade","especificidade")
write.csv(acuracia,"acuracia1b.csv",row.names = F)
```

Amostra 300

```
passo1 <- read.dbf("passo1.dbf")
library(dplyr)
passo1c <- sample_n(passo1, 300)
write.dbf(passo1c,"passo1c.dbf")
```

```
bd1 <- read.dbf("passo1c.dbf")
library(pROC)
```

```
roc(bd1$MATCH,bd1$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thr
es="best",legacy.axes=T)
```

```
roc1 <-
roc(bd1$MATCH,bd1$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thr
es="best",legacy.axes=T)
names(roc1)
```

```
acuracia <- as.data.frame(cbind(roc1$thresholds,roc1$sensitivities,roc1$specificities))
names(acuracia) <- c("escore","sensibilidade","especificidade")
write.csv(acuracia,"acuracia1c.csv",row.names = F)
```

```

#ETAPA 5 - Curva ROC
# lendo os bancos
library(foreign)
passo1 <- read.dbf("passo1.dbf")

# escolhendo o passo para analisar a acurácia
db <- passo1

# curva ROC e AUC
library(pROC)
roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"))
roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"),ci=T)
roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T)
roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T)

roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thres
="best")

roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thres
="best",legacy.axes=T)

roc1 <-
roc(db$MATCH,db$SCORE,plot=T,levels=c("!", "X"),ci=T,auc.polygon=T,print.auc=T,print.thres
="best",legacy.axes=T)
names(roc1)
acuracia <- as.data.frame(cbind(roc1$thresholds,roc1$sensitivities,roc1$specificities))
names(acuracia) <- c("escore","sensibilidade","especificidade")
write.csv(acuracia,"acuracia.csv",row.names = F)

```

Observação: Comandos repetidos para todos os passos.

```

#ETAPA 6
#No RecLink executar o combina de cada um dos bancos que abrangem todos os pares verdadeiros
(16, 9, 12, 15, 5, 1, 8, 13, 6 e 10) e depois no R juntar todos esses bancos
### PASSO 26

```

```

#Juntar todos os passos verdadeiros num sã³ banco

```

```

#Caminho do arquivo
library(foreign)
setwd("G:/SISMAMA/Passos")

```

```

passo16ver<-read.dbf("passo16ver.dbf")
Mamog2009PAD16<-read.dbf("Mamog2009PAD16.dbf")
Mamog2010PAD16<-read.dbf("Mamog2010PAD16.dbf")

```

```

# criando a variável pares
passo16ver$pares <- as.factor(paste(passo16ver$REFREC,sep=" ",passo16ver$COMPREC))

```

Observação: Comando repetido para os passos 16, 9, 12, 15, 5, 1, 8, 13, 6 e 10

```

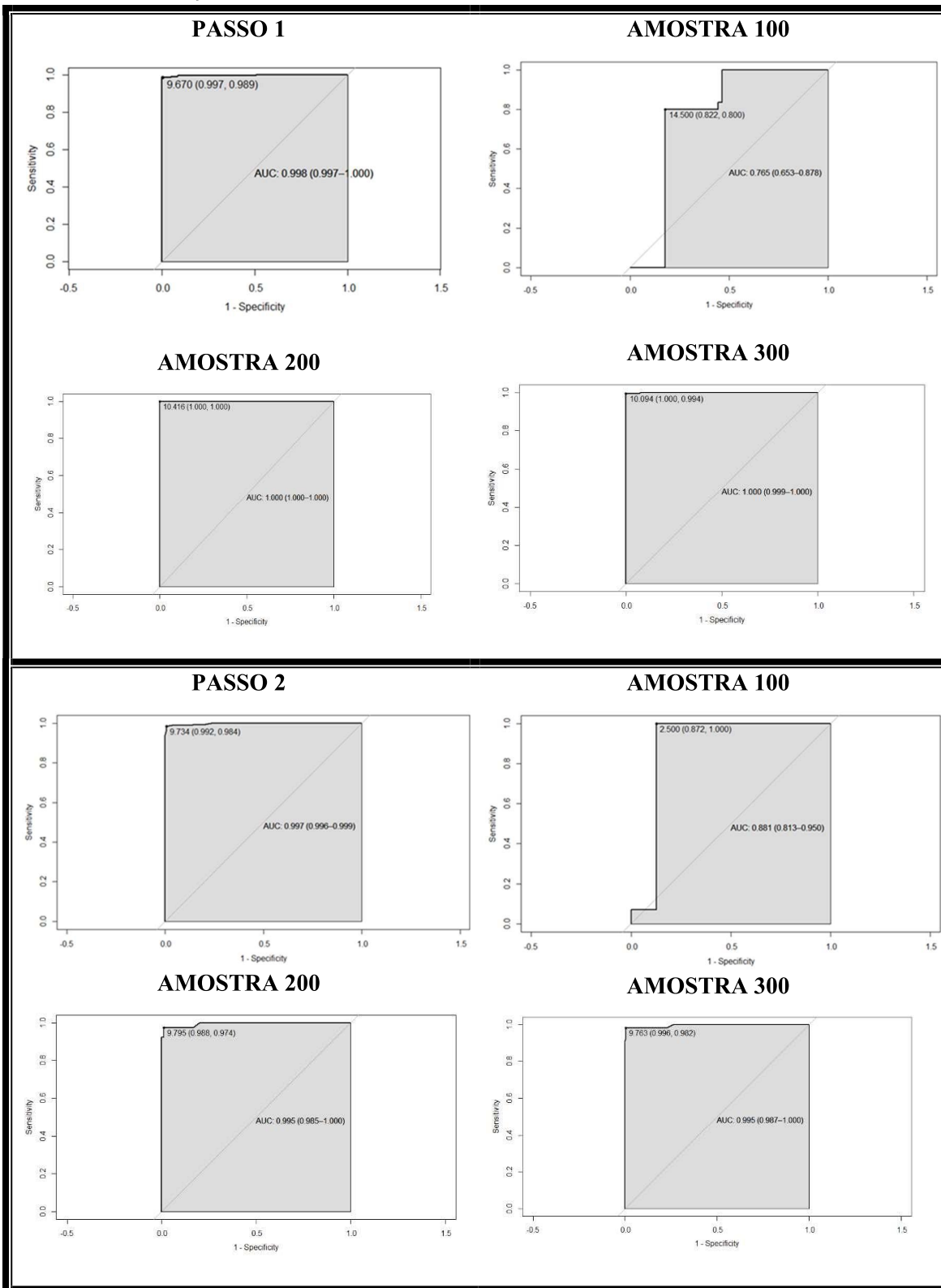
#Juntar o banco
Passosver <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="pares",all=T),

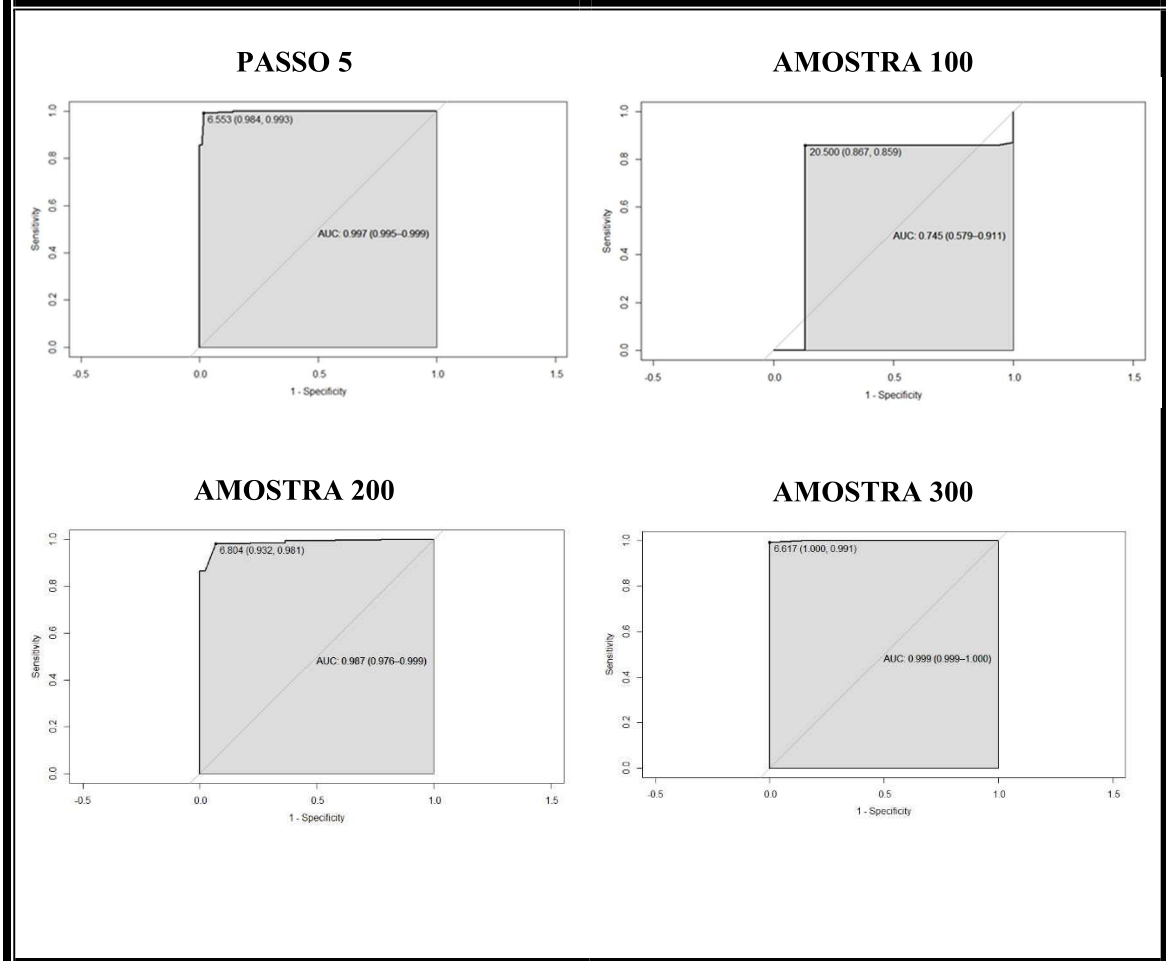
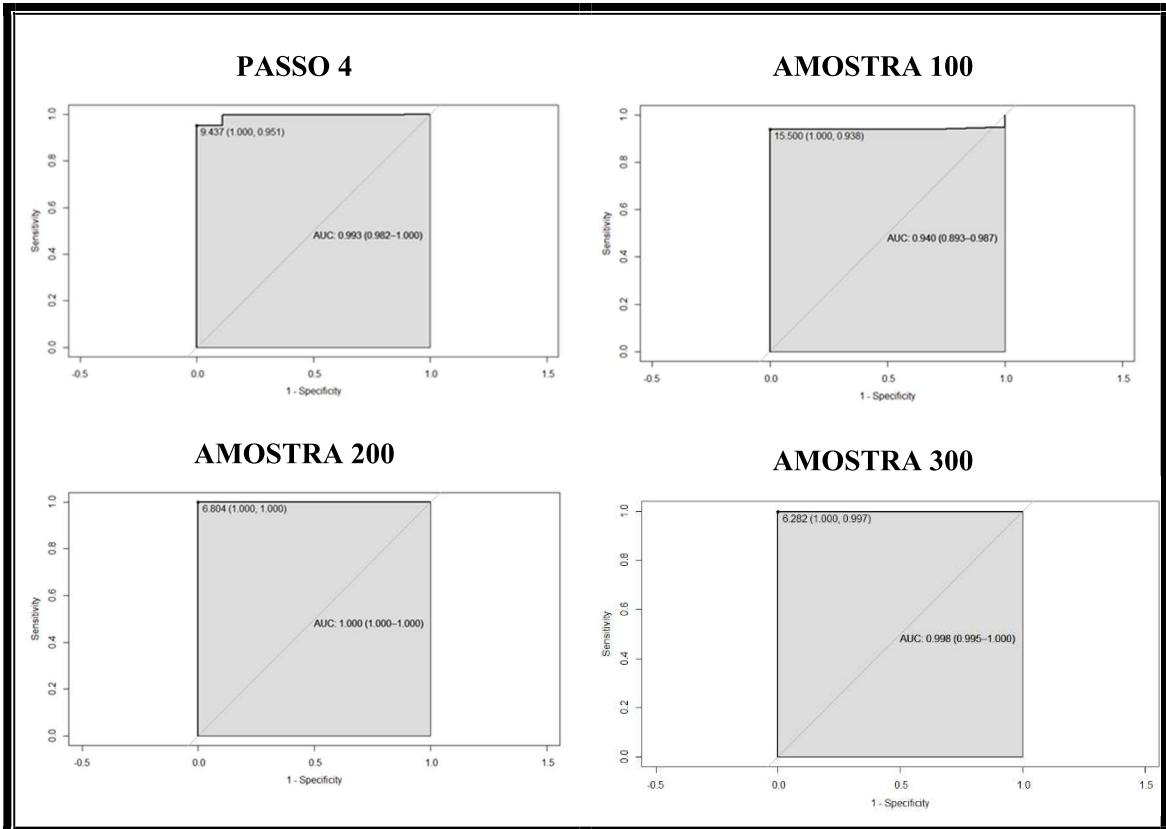
```

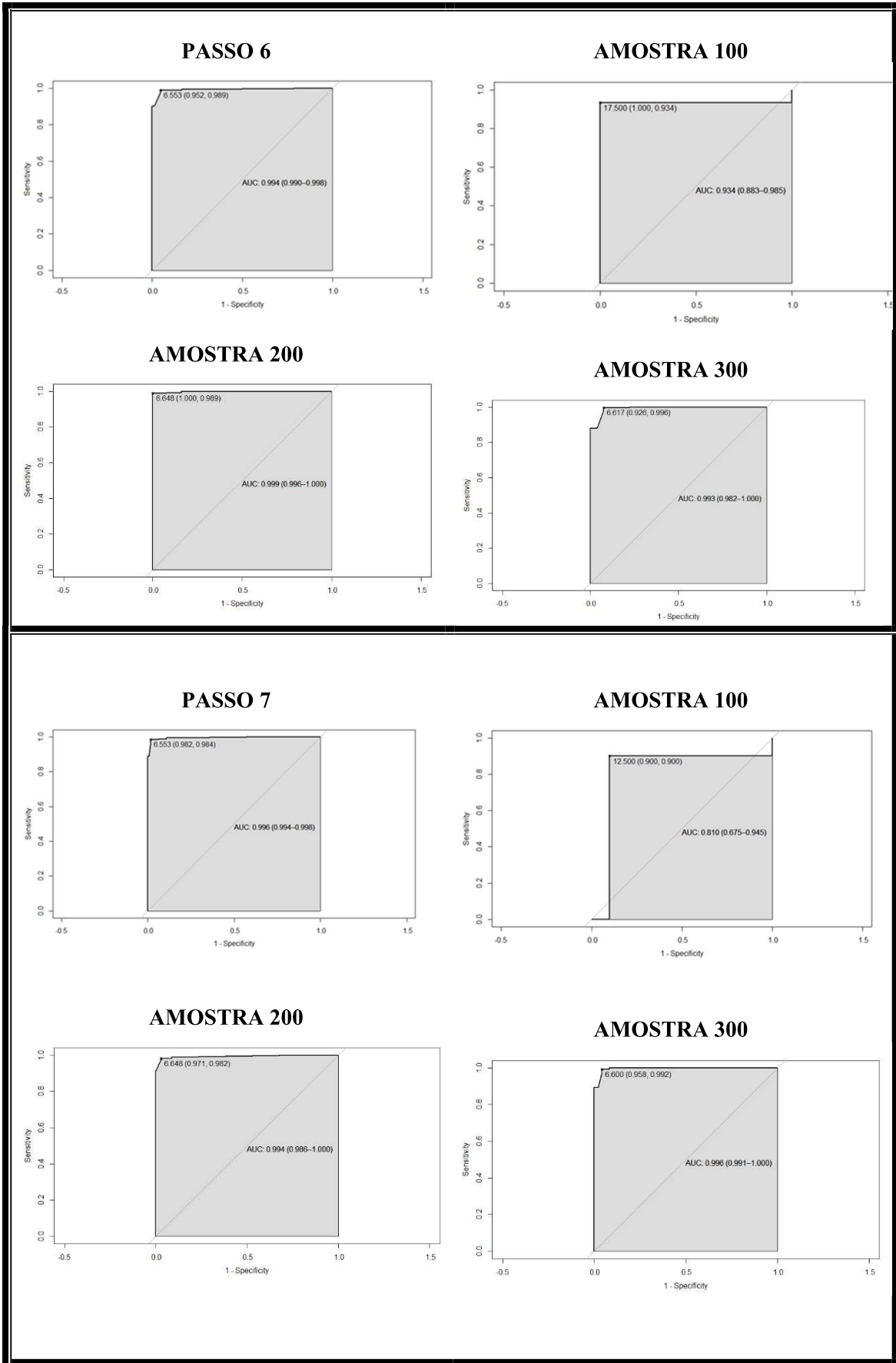


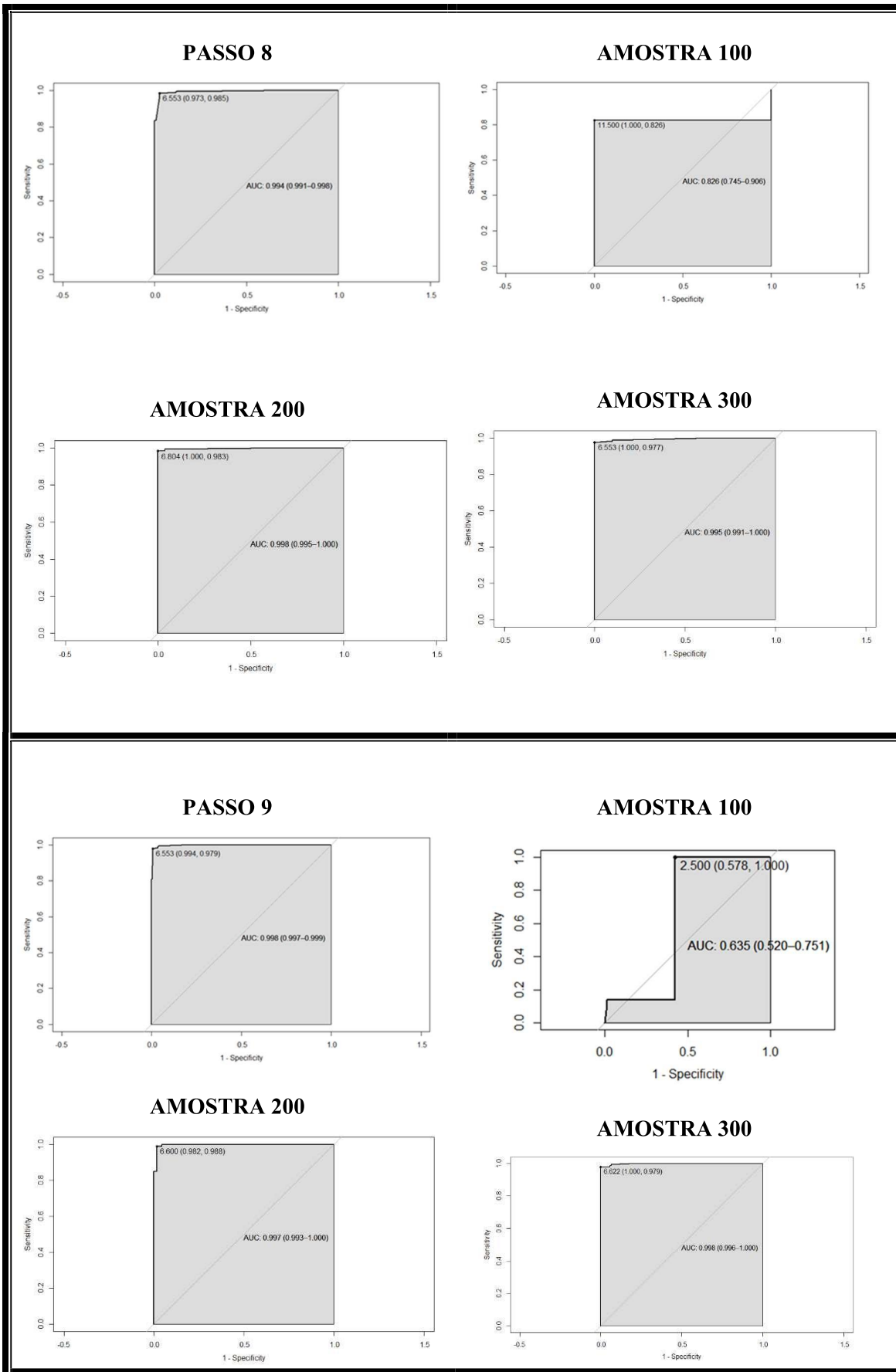
```
list(passo16ver,passo9ver, passo12ver,passo1ver,passo14ver, passo11ver, passo7ver, passo13ver))
write.csv(Passosver,"Passosver.csv",row.names = F)
Passosver<-read.csv("Passosver.csv")
```

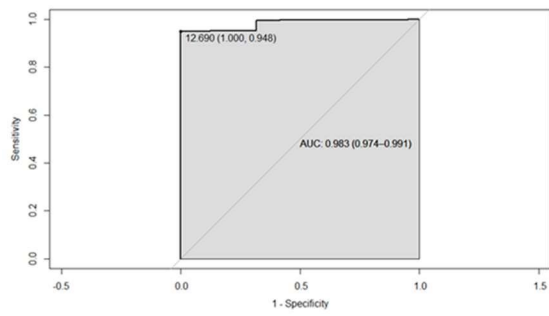
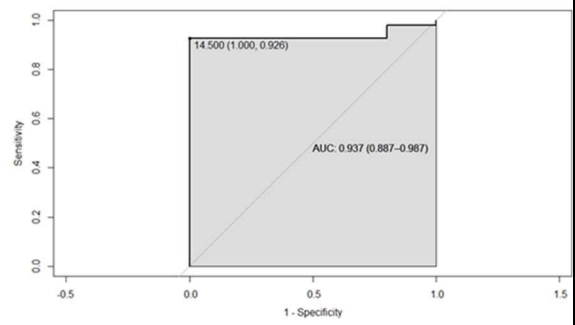
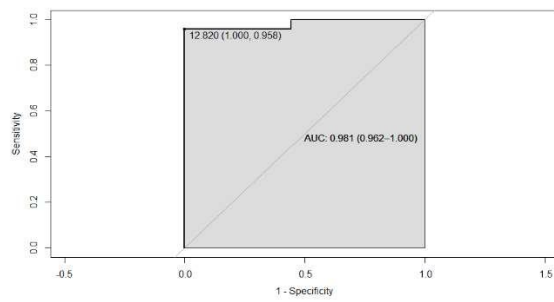
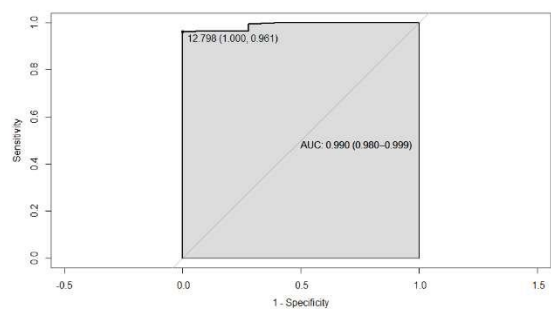
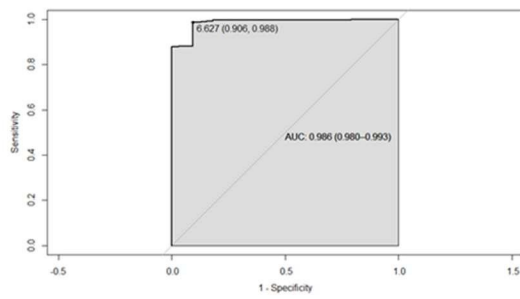
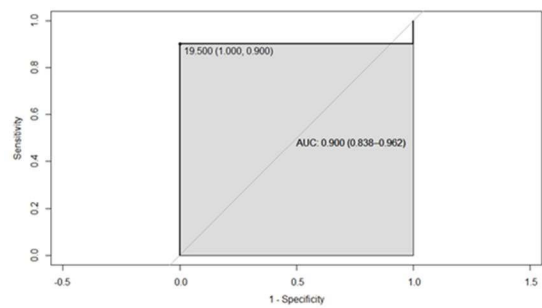
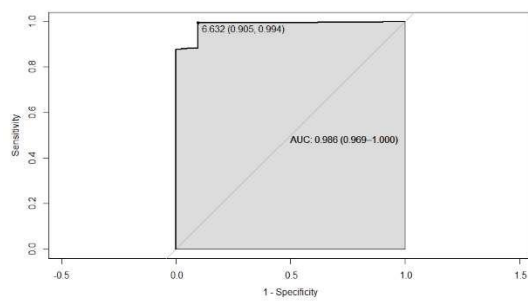
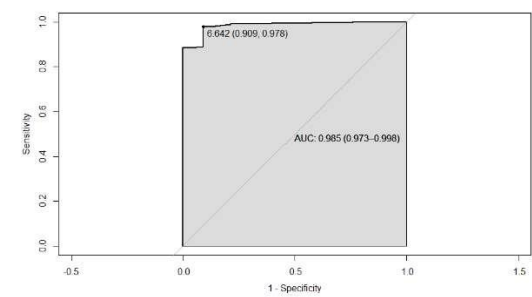
COMPARAÇÃO DO SCORE E CURVA ROC DE CADA PASSO E AMOSTRA



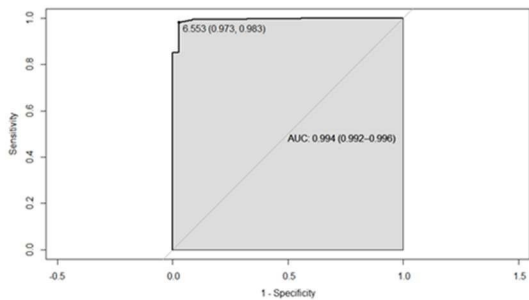




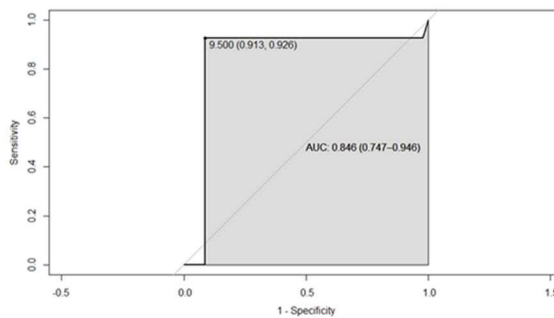


PASSO 10**AMOSTRA 100****AMOSTRA 200****AMOSTRA 300****PASSO 11****AMOSTRA 100****AMOSTRA 200****AMOSTRA 300**

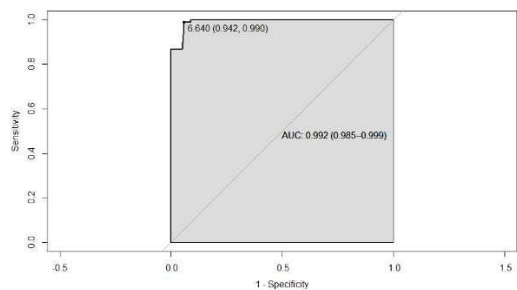
PASSO 12



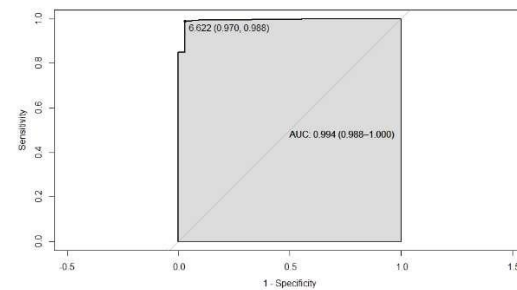
AMOSTRA 100



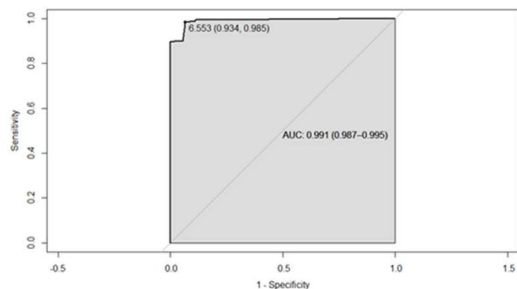
AMOSTRA 200



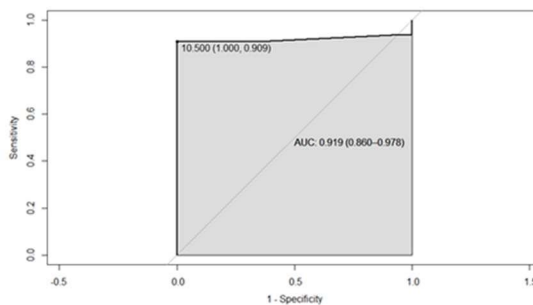
AMOSTRA 300



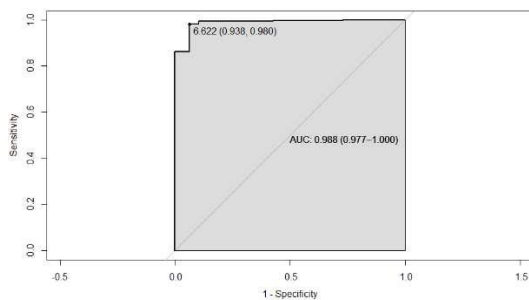
PASSO 13



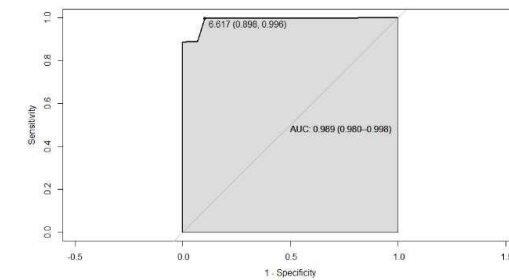
AMOSTRA 100



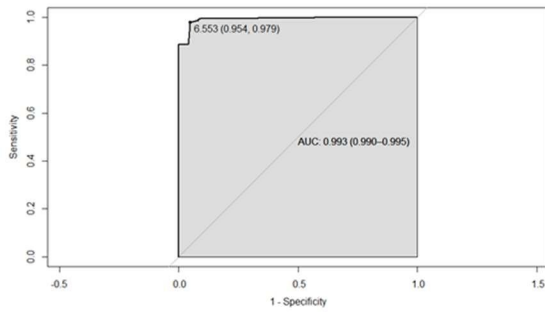
AMOSTRA 200



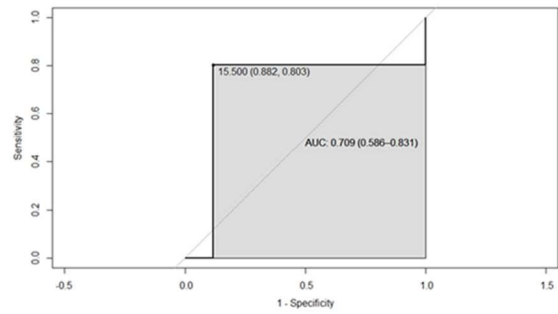
AMOSTRA 300



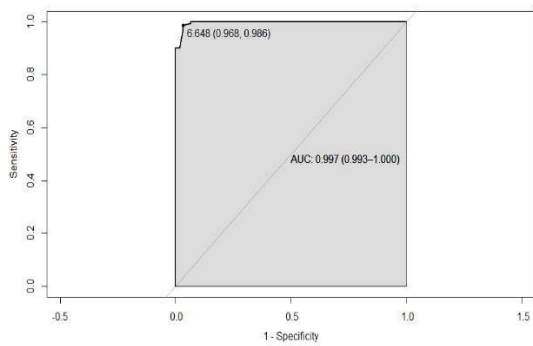
PASSO 14



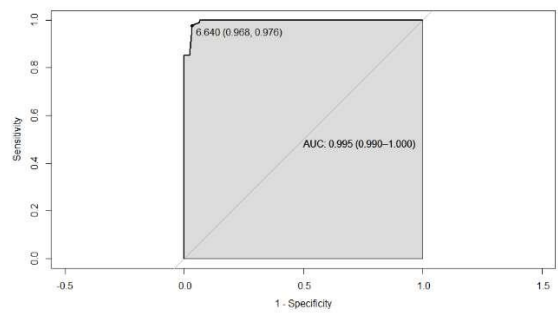
AMOSTRA 100



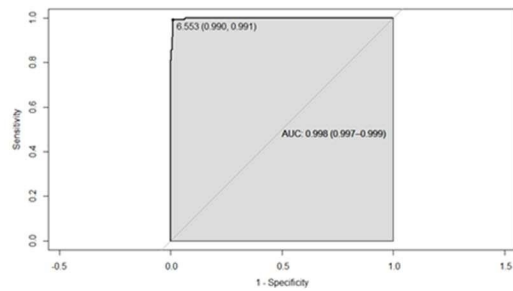
AMOSTRA 200



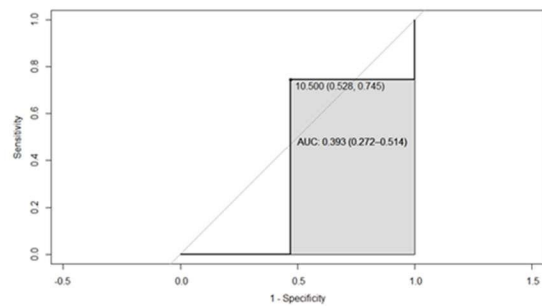
AMOSTRA 300



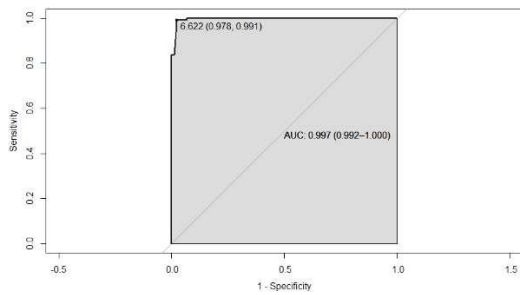
PASSO 15



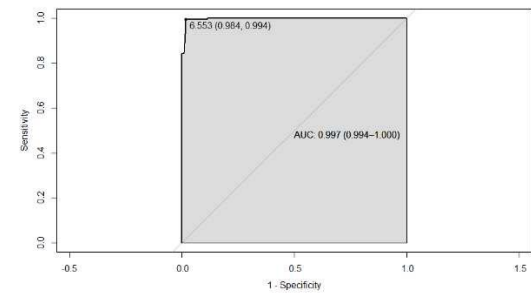
AMOSTRA 100

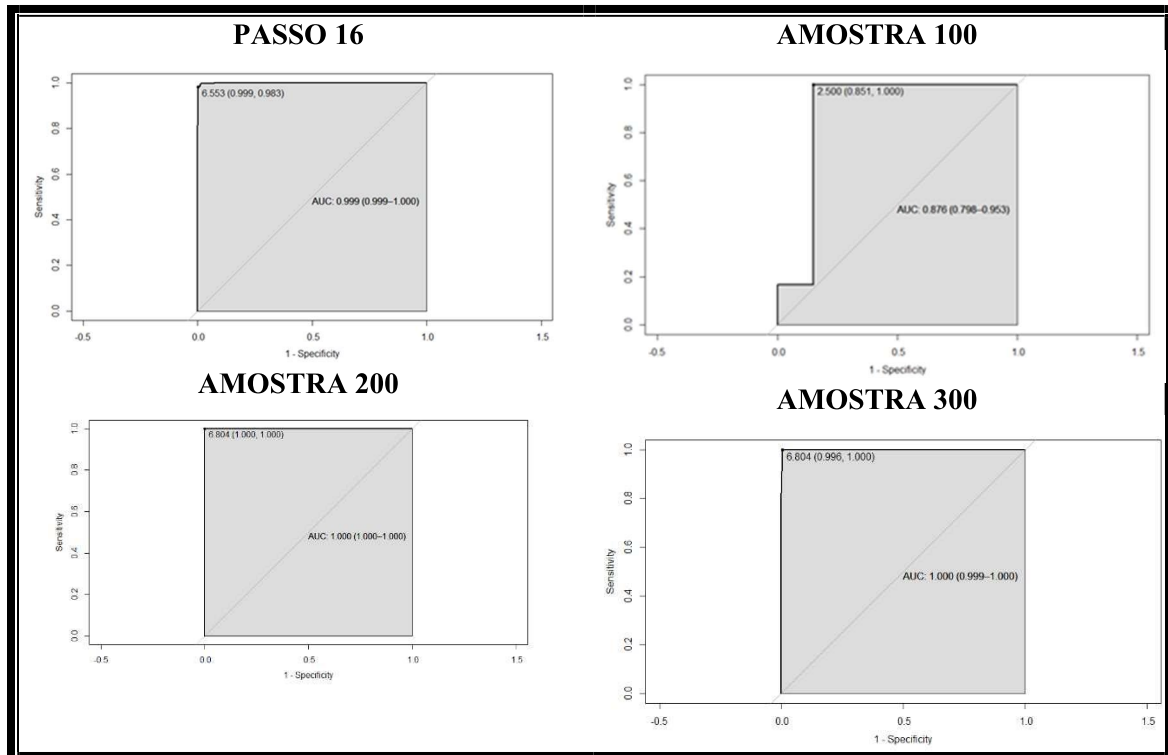


AMOSTRA 200



AMOSTRA 300





4. Retirada dos exames alterados da base referência

```
# Remover do banco Mamog2010 os pares de cada passo:
```

```
# Remover nomes do Passo16
```

```
v16<-read.dbf("passo16ver.dbf")
```

```
#Deixar os bancos com o mesmo número de variáveis
```

```
# Selecionar no banco Passosver as variáveis do banco de referência names(Mamog2010PAD)
```

```
names(Mamog2010PAD)
```

```
names(v16)
```

```
v16<-subset(v16,select=c(R_NOMEPAD,R_PNOME,R_UNOME, R_MAEPAD,R_PMAE,
R_UMAE, R_DTNASC, R_DT_EXAME, R_ANO_NASC, R_ENDERECO, R_CNSUS,
R_MUN_COD, R_FONTE))
```

```
#Renomear o banco
```

```
names(v16)<-c("NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE", "UMAE",
"DTNASC", "DT_EXAME", "ANO_NASC", "ENDERECO", "CNSUS", "MUN_COD",
"FONTE")
```

```
#Criar variável SUB
```

```
Mamog2010PAD$SUB<-"1"
```

```
v16$SUB<-"2"
```

```
#Fazer merge e excluir registros com variável SUB=2
```

```
bt<-merge(Mamog2010PAD,v16,by="FONTE",all.x=TRUE)
```

```
write.dbf(bt, "bt.dbf")
```

```
bt<-subset(bt,select=c(NOMEPAD.x,PNOME.x,UNOME.x, MAEPAD.x, PMAE.x, UMAE.x,
DTNASC.x, DT_EXAME.x, ANO_NASC.x, ENDERECO.x, CNSUS.x, MUN_COD.x, FONTE,
SUB.y))
```



```
str(bt)
```

```
bt$SUB.y<-as.integer(bt$SUB.y)
summary(bt$SUB.y)
```

```
bt$SUB.y[is.na(bt$SUB.y)]<-"1"
```

```
bt16<- subset(bt,(SUB.y=="1"))
```

```
#Renomear variáveis
names(bt16)<-c("NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE", "UMAE",
"DTNASC", "DT_EXAME", "ANO_NASC", "ENDERECO", "CNSUS", "MUN_COD",
"FONTE", "SUB")
write.dbf(bt16, "bt16.dbf")
```

Observação: Comandos repetidos para os passos 9, 12, 15, 5, 1, 8, 13, 6 e 10

```
#Salvar o banco 2010 final como db7
write.dbf(bt13, "db7.dbf")
db7<-read.dbf("db7.dbf")
```

```
# Fazer o merge entre Mamog2010 e db7 para ficar com banco completo - todas variáveis e
excluir variáveis excedentes
Mamog2010<-read.dbf("Mamog2010.dbf")
db9<-merge(db7, Mamog2010,by="FONTE",all.x=TRUE)
write.dbf(db9,"db9.dbf")
db9<-read.dbf(db9, "db9.dbf")
```

```
names(db9)
```

```
db9<-subset(db9,select=c("FONTE", "NOMEPAD.y", "PNOME.y", "UNOME.y", "MAEPAD.y",
"PMAE.y", "UMAE.y", "UNICO", "CODMUN6", "CNES", "COMPET", "N_EXAME",
"NOME", "MAE", "DTNASC.y", "ESCOL", "ENDERECO.y", "NUMERO", "COMPLEM",
"BAIRRO", "UF", "IBGE", "CNSUS.y", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME",
"NODULO", "NOD_D", "NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT",
"ANO_MAMOG", "MAMOG_DIAG", "RASTREAM", "NUM_FIL_MD", "PELE_MD",
"COMP_MD", "NOD_MD", "NOD_TAM_MD", "NUM_FIL_ME", "PELE_ME", "COMP_ME",
"NOD_ME", "NOD_TAM_ME", "DIAG_MD", "DIAG_ME", "RECOM_MD", "RECOM_ME",
"DT_EXAME.y", "DT_SOLICIT", "DT_LIBERA", "SEXO", "RACA", "ETNIA", "NACIONAL",
"D_EXAME2", "D_ID_DTNAS", "ANO", "ANO_NASC.y", "IDADE", "MUN_COD.y",
"MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD", "MACRO", "ESTADO", "REGISTRO",
"MES_EXAME"))
```

```
#Renomear variáveis
names(db9)<-c("FONTE", "NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE",
"UMAE", "UNICO", "CODMUN6", "CNES", "COMPET", "N_EXAME", "NOME", "MAE",
"DTNASC", "ESCOL", "ENDERECO", "NUMERO", "COMPLEM", "BAIRRO", "UF", "IBGE",
"CNSUS", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME", "NODULO", "NOD_D",
"NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG", "MAMOG_DIAG",
"RASTREAM", "NUM_FIL_MD", "PELE_MD", "COMP_MD", "NOD_MD",
"NOD_TAM_MD", "NUM_FIL_ME", "PELE_ME", "COMP_ME", "NOD_ME",
"NOD_TAM_ME", "DIAG_MD", "DIAG_ME", "RECOM_MD", "RECOM_ME",
```

```
"DT_EXAME", "DT_SOLICIT", "DT_LIBERA", "SEXO", "RACA", "ETNIA", "NACIONAL",
"D_EXAME2", "D_ID_DTNAS", "ANO", "ANO_NASC", "IDADE", "MUN_COD",
"MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD", "MACRO", "ESTADO", "REGISTRO",
"MES_EXAME")
```

```
write.dbf(db9,"db9.dbf")
db9<-read.dbf("db9.dbf")
```

```
# Fazer o merge entre Mamog2010 e db9 para ficar com banco completo - todas variáveis e excluir
variáveis excedentes
```

```
Mamog2010<-read.dbf("Mamog2010.dbf")
db10<-merge(db9, Mamog2010,by="FONTE",all.x=TRUE)
write.dbf(db10,"db10.dbf")
db10<-read.dbf(db10, "db10.dbf")
```

```
names(db10)
```

```
db11<-subset(db10,select=c("FONTE", "NOMEPAD.y", "PNOME.y", "UNOME.y",
"MAEPAD.y", "PMAE.y", "UMAE.y", "UNICO", "CODMUN6", "CNES.y", "COMPET",
"N_EXAME", "NOME", "MAE", "DTNASC.y", "ESCOL", "ENDERECO", "NUMERO",
"COMPLEM", "BAIRRO", "UF", "IBGE", "CNSUS.y", "US_UF", "US_IBGE", "CNES_REQ.y",
"US_NOME", "NODULO", "NOD_D", "NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT",
"ANO_MAMOG", "MAMOG_DIAG", "RASTREAM", "NUM_FIL_MD", "PELE_MD",
"COMP_MD", "NOD_MD", "NOD_TAM_MD", "NUM_FIL_ME", "PELE_ME", "COMP_ME",
"NOD_ME", "NOD_TAM_ME", "DIAG_MD", "DIAG_ME", "RECOM_MD", "RECOM_ME",
"DT_EXAME.y", "DT_SOLICIT", "DT_LIBERA", "SEXO", "RACA", "ETNIA", "NACIONAL",
"D_EXAME2", "D_ID_DTNAS", "ANO", "ANO_NASC.y", "IDADE", "MUN_COD.y",
"MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD", "MACRO", "ESTADO", "REGISTRO",
"MES_EXAME.y"))
```

```
names (db11)
```

```
#Renomear variáveis
```

```
names(db11)<-c("FONTE", "NOMEPAD", "PNOME", "UNOME", "MAEPAD", "PMAE",
"UMAE", "UNICO", "CODMUN6", "CNES", "COMPET", "N_EXAME", "NOME", "MAE",
"DTNASC", "ESCOL", "ENDERECO", "NUMERO", "COMPLEM", "BAIRRO", "UF", "IBGE",
"CNSUS", "US_UF", "US_IBGE", "CNES_REQ", "US_NOME", "NODULO", "NOD_D",
"NOD_E", "RISCO", "EXA_MAMA", "MAMOG_ANT", "ANO_MAMOG", "MAMOG_DIAG",
"RASTREAM", "NUM_FIL_MD", "PELE_MD", "COMP_MD", "NOD_MD",
"NOD_TAM_MD", "NUM_FIL_ME", "PELE_ME", "COMP_ME", "NOD_ME",
"NOD_TAM_ME", "DIAG_MD", "DIAG_ME", "RECOM_MD", "RECOM_ME",
"DT_EXAME", "DT_SOLICIT", "DT_LIBERA", "SEXO", "RACA", "ETNIA", "NACIONAL",
"D_EXAME2", "D_ID_DTNAS", "ANO", "ANO_NASC", "IDADE", "MUN_COD",
"MUNICIPIO", "CIR_COD", "CIR", "MACRO_COD", "MACRO", "ESTADO", "REGISTRO",
"MES_EXAME")
```

```
write.dbf(db11,"db11.dbf")
db11<-read.dbf(db11, "db11.dbf")
```

Observação: db11 é a base referência.

5. Limpeza da base seguimento

Repetir os passos de 1 a 26 da limpeza da base referência para limpar os bancos dos anos 2011 e 2012.

```
# Fazer uma cópia do arquivo, salvar e abrir o arquivo
write.dbf(t2011_8, file="Mamog2011.dbf")
Mamog2011<-read.dbf("Mamog2011.dbf")

write.dbf(t2012_8, file="Mamog2012.dbf")
Mamog2012<-read.dbf("Mamog2012.dbf")

# Ler os bancos de exames solicitados em cada ano
Mamog2010_10 <- read.dbf("db5.dbf")
Mamog2011<- read.dbf("Mamog2011.dbf")
Mamog2012 <- read.dbf("Mamog2012.dbf")

#Juntar as bases seguimento
Mamogseg<-rbind(Mamog2010_10, Mamog2011, Mamog2012)
write.dbf(Mamogseg,file="Mamogseg.dbf")
```

6. Relacionamento probabilístico entre base referência e seguimento

```
#Ler os bancos
Mamogseg<- read.dbf("Mamogseg.dbf")
Mamogref<- read.dbf("db11.dbf")

#Dividir o banco referÃancia (6 primeiros meses do ano)
table(Mamogref$MES_EXAME)
# Selecionando observaÃµes que tenham a data entre dois valores
Mamogref2<- subset(Mamogref,(MES_EXAME=="1")|(MES_EXAME=="2")|(MES_EXAME=="3")|(MES_EXAME=="4")|(MES_EXAME=="5")|(MES_EXAME=="6"))
Mamogref3<- subset(Mamogref,(MES_EXAME=="7")|(MES_EXAME=="8")|(MES_EXAME=="9")|(MES_EXAME=="10")|(MES_EXAME=="11")|(MES_EXAME=="12"))

table(Mamogref2$MES_EXAME)
table(Mamogref3$MES_EXAME)

write.dbf(Mamogref2,file="Mamogref2.dbf")
write.dbf(Mamogref3,file="Mamogref3.dbf")
Mamogref2<- read.dbf("Mamogref2.dbf")
```

Observação: Os bancos Mamogref2 e Mamogseg foram relacionados no RecLink por meio de 3 passos: 16, 9 e 12

7. Determinar ponto de corte após relacionamento entre base referência e seguimento

```
# Determinar ponto de corte
#Voltar com os bancos gerados nos passos 16, 9 e 12 para o R e calcular uma amostra de 300 pares e fazer conferÃncia manual para determinar ponto de corte.
```

```

#### Passo 1
{r}
setwd("G:/doutorado/SISMAMA/relacionamento_seguinto")
``

# Importar a biblioteca necess ria com as funcoes
{r}
library(foreign)

passo16 <- read.dbf("Passo16.dbf")

# selecionando uma amostra
library(dplyr)
passo16c <- sample_n(passo16, 300)
write.dbf(passo16c,"passo16c.dbf")

# Obs.: voltar no Reclink e marcar todos os pares como verdadeiros, executar e criar arquivo com
nomes.

# Voltar com o banco para o R
#Fazer merge dos bancos para conferir no excel

#PASSO 16
passo16c<-read.dbf('passo16c.dbf')
passo16ver<-read.dbf('passo16ver.dbf')

#criar vari vel pares
passo16c$pares <- as.factor(paste(passo16c$REFREC,sep=" ",passo16c$COMPREC))
passo16ver$pares <- as.factor(paste(passo16ver$REFREC,sep=" ",passo16ver$COMPREC))

names(passo16c)
names(passo16ver)

#Fazer o merge
A16 <- merge(passo16c,passo16ver,by="pares",all.x=TRUE)
write.dbf(A16, "A16.dbf")

names(A16)

#Retirar vari veis excedentes
excluir<-c("COMPREC.y", "REFREC.y", "SCORE.y")
A16 <- A16[!(names(A16)%in% excluir)]
write.dbf(A16, "A16.dbf")

# Conferir o passo A16 no Excel
#Observa  o marcar em match X para par e ! para n o par

# Voltar com o passo conferido do Excel
A16 <- read.csv2("A16.csv")
names(A16)

# Renomear vari veis

```

```
names(A16)<-c("pares", "COMPREC", "REFREC", "SCORE","MATCH", "C_NOMEPAD",
"C_PNOME", "C_UNOME", "C_MAEPAD","C_PMAE", "C_UMAE", "C_DTNASC",
"C_D_EXAME2", "C_ANO_NASC", "C_CNSUS", "C_MUN_COD", "C_FONTE",
"C_MES_EXAM", "C_CNES", "C_CNES_REQ", "R_NOMEPAD", "R_PNOME",
"R_UNOME", "R_MAEPAD", "R_PMAE", "R_UMAE", "R_DTNASC", "R_D_EXAME2",
"R_ANO_NASC", "R_CNSUS", "R_MUN_COD", "R_FONTE", "R_MES_EXAM", "R_CNES",
"R_CNES_REQ")
write.dbf(A16,"A16.dbf")
```

```
str(A16)
```

```
#Converter variÃ¡vel SCORE nÃºmero
A16$SCORE<-as.character(A16$SCORE)
A16$SCORE <- as.numeric(A16$SCORE)
```

```
write.dbf(A16,"A16.dbf")
```

```
str(A16)
```

```
#Curva Precision-Recall e AcurÃ¡cia
```

```
# lendo os bancos
setwd("G:/SISMAMA/relacionamento_seguinto")
library(foreign)
A16<- read.dbf("A16.dbf")
```

```
# ajustando o banco (separando os escores dos pares verdadeiros e falsos em bancos diferentes)
A16$par<-0
A16$par[A16$MATCH=="X"]<-1
```

```
write.dbf(A16,"A16.dbf")
```

```
# fazendo a curva precision-recall
library(PRROC)
pr <- pr.curve(scores.class0 = A16[A16$MATCH=="X",4], scores.class1 =
A16[A16$MATCH=="!",4], 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)]
print(best.cutoff)
```

```
#AcurÃ¡cia do ponto de corte
library(ROCR)
pred <- prediction(predictions = A16$SCORE, labels = A16$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)

library(pROC)
curva.roc <- roc(A16$par,A16$SCORE,plot=T,levels=c("0","1"),ci=T,auc.polygon=T,
  print.auc=T,print.thres="best",legacy.axes=T)
coords(curva.roc,x="best",transpose=F,ret=c("threshold","precision","recall","accuracy"))
ci.coords(curva.roc,x="best",transpose=F,ret=c("threshold","precision","recall","accuracy"))
```

Observação: Os comandos foram repetidos para os passos 9 e 12. Depois o ponto de corte foi aplicado em cada passo no RecLink.

```
# ETAPA 3 - VOLTAR AO R - com cada passo
##VERIFICAR QUANTOS REGISTROS CADA PASSO RETIRA DA BASE REFERÊNCIA

# lendo os bancos
setwd("G:/doutorado/SISMAMA/relacionamento_seguinto")
library(foreign)

#PASSO 16
passo16<- read.dbf("passo16.dbf")
passo16_pc <- read.dbf("passo16_pc.dbf")

library(descr)
freq(passo16$SCORE>7)
summary(passo16_pc$SCORE)

# criando a variável pares
passo16_pc$pares <- as.factor(paste(passo16_pc$REFREC,sep=" ",passo16_pc$COMPREC))
write.dbf(passo16_pc,"passo16_pc.dbf")
```

Observação: comando repetido com os passos 9 e 12.

```
passos <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="pares",all=T),
  list(passo16s,passo9s,passo12s))

write.dbf(passos,"passos.dbf")
passos<-read.dbf("passos.dbf")

summary(passos)

#comparar os bancos
library(arsenal)
passo9_pc <- read.dbf("passo9_pc.dbf")
passo16_pc <- read.dbf("passo16_pc.dbf")
compare(passo16_pc,passo9_pc,by="pares")[1]

compare(passo16_pc,passo12_pc,by="pares")[1]
```

8. Preparo do banco para análise do sobrerastreamento

```
#Identificar as datas das repetições dos exames e combinar os bancos em um novo banco
```

```

Mamogref2<-read.dbf("Mamogref2.dbf")
passo16_pc <- read.dbf("passo16_pc.dbf")
passo9_pc <- read.dbf("passo9_pc.dbf")
passo12_pc <- read.dbf("passo12_pc.dbf")

names (Mamogref2)
names (passo16_pc)
names(passo12_pc)

#Fazer merge dos bancos passos com base seguimento para pegar variáveis importantes (usar
C_FONTE)
Mamogseg<-read.dbf("Mamogseg.dbf")
names(Mamogseg)
names(Mamogseg)[1]<-"C_FONTE"

Passo16<-merge(passo16_pc,Mamogseg, by="C_FONTE", all.x=TRUE)
names(Passo16)

c16 <-
Passo16[,c("C_FONTE","R_FONTE","DIAG_MD","DIAG_ME","RECOM_MD","RECOM_ME"
,"DT_EXAME","DT_SOLICIT","DT_LIBERA", "IDADE", "MUN_COD", "CIR", "MACRO")]
names(c16)[1]<-"C_FONTE"
names(c16)
write.dbf(c16,"c16.dbf")

Passo12<-merge(passo12_pc,Mamogseg, by="C_FONTE", all.x=TRUE)
names(Passo12)

c12 <-
Passo12[,c("C_FONTE","R_FONTE","DIAG_MD","DIAG_ME","RECOM_MD","RECOM_ME"
,"DT_EXAME","DT_SOLICIT","DT_LIBERA", "IDADE", "MUN_COD", "CIR", "MACRO")]
names(c12)[1]<-"C_FONTE"
names(c12)
write.dbf(c12,"c12.dbf")

names(Mamogref2)

ref<-
Mamogref2[,c("FONTE","DIAG_MD","DIAG_ME","RECOM_MD","RECOM_ME","DT_EXA
ME","DT_SOLICIT","DT_LIBERA", "IDADE", "MUN_COD", "CIR", "MACRO")]

ref$R_FONTE<-paste(ref$FONTE)

names(ref)
names(c16)
names(c12)

names(ref)[1]<-"C_FONTE"
ref <-
ref[,c("C_FONTE","R_FONTE","DIAG_MD","DIAG_ME","RECOM_MD","RECOM_ME","DT
_EXAME","DT_SOLICIT","DT_LIBERA", "IDADE", "MUN_COD", "CIR", "MACRO")]
names(ref)

write.dbf(ref,file="ref.dbf")

```

```
write.dbf(c16,file="c16.dbf")
write.dbf(c12,file="c12.dbf")
```

Observação: não realizado com passo 9 porque ele não agregou nenhum par diferente.

```
#empilhar os bancos de dados e ordenar pela data
db14 <- rbind(ref,c16,c12)
db14$DT_EXAME <- as.Date(db14$DT_EXAME, "%d/%m/%Y")
db14 <- db14[order(db14$DT_EXAME),]
write.dbf(db14,file="db14.dbf")
db14<-read.dbf("db14.dbf")

# identificar e excluir duplicidades
aux <- which(duplicated(db14))
db14dup <- db14[aux,]
if(length(aux)>0){db15 <- db14[-aux,]}else{db15 <- db14}
rm(aux)
write.dbf(db15,file="db15.dbf")
db15<-read.dbf("db15.dbf")

# frequencias de repetições do exame
rep <- as.data.frame(table(db15$R_FONTE))
table(rep$Freq)

# identificar as primeiras datas de exame
aux <- which(duplicated(db15[,c("R_FONTE")]))
dbdupa <- db15[aux,]
if(length(aux)>0){data0 <- db15[-aux,]}else{data0 <- db15}
names(data0)<-
c("C_FONTE","R_FONTE","DIAG_MD0","DIAG_ME0","RECOM_MD0","RECOM_ME0","D
T_EXAME0","DT_SOLICIT0","DT_LIBERA0","IDADE","MUN_COD","CIR","MACRO")
rm(aux)

# identificar as primeiras repetições do exame
aux <- which(duplicated(dbdupa[,c("R_FONTE")]))
dbdupb <- dbdupa[aux,]
if(length(aux)>0){data1 <- dbdupa[-aux,]}else{data1 <- dbdupa}
names(data1)<-
c("C_FONTE","R_FONTE","DIAG_MD1","DIAG_ME1","RECOM_MD1","RECOM_ME1","D
T_EXAME1","DT_SOLICIT1","DT_LIBERA1","IDADE","MUN_COD","CIR","MACRO")
rm(aux)

# identificar as segundas repetições do exame
aux <- which(duplicated(dbdupb[,c("R_FONTE")]))
dbdupc <- dbdupb[aux,]
if(length(aux)>0){data2 <- dbdupb[-aux,]}else{data2 <- dbdupb}
names(data2)<-
c("C_FONTE","R_FONTE","DIAG_MD2","DIAG_ME2","RECOM_MD2","RECOM_ME2","D
T_EXAME2","DT_SOLICIT2","DT_LIBERA2","IDADE","MUN_COD","CIR","MACRO")
rm(aux)

# identificar as terceiras repetições do exame
aux <- which(duplicated(dbdupc[,c("R_FONTE")]))
dbdupd <- dbdupc[aux,]
```



```

if(length(aux)>0){data3 <- dbdupc[-aux,]}else{data3 <- dbdupc}
names(data3)<-
c("C_FONTE","R_FONTE","DIAG_MD3","DIAG_ME3","RECOM_MD3","RECOM_ME3","D
T_EXAME3","DT_SOLICIT3","DT_LIBERA3", "IDADE", "MUN_COD", "CIR", "MACRO")
rm(aux)

# identificar as quartas repetições do exame
aux <- which(duplicated(dbdupd[,c("R_FONTE")]))
dbdupe <- dbdupd[aux,]
if(length(aux)>0){data4 <- dbdupd[-aux,]}else{data4 <- dbdupd}
names(data4)<-
c("C_FONTE","R_FONTE","DIAG_MD4","DIAG_ME4","RECOM_MD4","RECOM_ME4","D
T_EXAME4","DT_SOLICIT4","DT_LIBERA4", "IDADE", "MUN_COD", "CIR", "MACRO")
rm(aux)

# identificar as quintas repetições do exame
aux <- which(duplicated(dbdupe[,c("R_FONTE")]))
dbdupf <- dbdupe[aux,]
if(length(aux)>0){data5 <- dbdupe[-aux,]}else{data5 <- dbdupe}
names(data5)<-
c("C_FONTE","R_FONTE","DIAG_MD5","DIAG_ME5","RECOM_MD5","RECOM_ME5","D
T_EXAME5","DT_SOLICIT5","DT_LIBERA5", "IDADE", "MUN_COD", "CIR", "MACRO")
rm(aux)

# identificar as sextas repetições do exame
aux <- which(duplicated(dbdupf[,c("R_FONTE")]))
dbdupg <- dbdupf[aux,]
if(length(aux)>0){data6 <- dbdupf[-aux,]}else{data6 <- dbdupf}
names(data6)<-
c("C_FONTE","R_FONTE","DIAG_MD6","DIAG_ME6","RECOM_MD6","RECOM_ME6","D
T_EXAME6","DT_SOLICIT6","DT_LIBERA6", "IDADE", "MUN_COD", "CIR", "MACRO")
rm(aux)

# combinar todas as datas de cada mulher
db16 <- Reduce(function(dtf1,dtf2) merge(dtf1,dtf2,by="R_FONTE",all.x=T),
               list(data0,data1,data2,data3,data4,data5,data6))

write.dbf(db16, "db16.dbf")
write.dbf(ref, "ref.dbf")
write.dbf(rep, "rep.dbf")

```

Relacionamento probabilístico da base referência com a base seguimento

Passo	PF	PV	PC	Recall	Precision	F	ACC	PCR
16	3.832.071	368.847	7,06	0,96	0,93	0,94	0,99	0,99
9	836.114	363.557	7,06	1,00	0,99	1,00	1,00	1,00
12	762.225	368.529	7,06	1,00	0,97	0,99	0,99	1,00
Total de pares combinados		381.487						
1º Exame	Repetições							Total
	1	2	3	4	5	6	≥7	
221.552	113.851	34.318	6.050	1.897	867	510	2.442	381.487

PF: Pares formados; **PV:** Pares Verdadeiros; **PC:** Ponto de Corte; **F:** Medida F; **ACC:** Acurácia; **PCR:** Curva Precision Recall.

```

db16<-read.dbf("db16.dbf")

names(db16)

db17<-db16[,c("R_FONTE","DT_EXAME0","DT_EXAME1", "DT_EXAME2",
"DT_EXAME3", "DT_EXAME4", "DT_EXAME5",
"DT_EXAME6","DT_SOLICIT0","DT_SOLICIT1", "DT_SOLICIT2", "DT_SOLICIT3",
"DT_SOLICIT4", "DT_SOLICIT5", "DT_SOLICIT6", "DT_LIBERA0","DT_LIBERA1",
"DT_LIBERA2", "DT_LIBERA3", "DT_LIBERA4", "DT_LIBERA5", "DT_LIBERA6",
"DIAG_MD0","DIAG_MD1", "DIAG_MD2", "DIAG_MD3", "DIAG_MD4", "DIAG_MD5",
"DIAG_MD6", "DIAG_ME0","DIAG_ME1", "DIAG_ME2", "DIAG_ME3", "DIAG_ME4",
"DIAG_ME5", "DIAG_ME6","RECOM_MD0","RECOM_MD1", "RECOM_MD2",
"RECOM_MD3", "RECOM_MD4", "RECOM_MD5",
"RECOM_MD6","RECOM_ME0","RECOM_ME1", "RECOM_ME2", "RECOM_ME3",
"RECOM_ME4", "RECOM_ME5", "RECOM_ME6")]
names(db17)
write.dbf(db17, "db17.dbf")

db17<-read.dbf("db17.dbf")
ref<-read.dbf("ref.dbf")
names(ref)

ref2<-ref[,c("R_FONTE","IDADE", "MUN_COD", "CIR", "MACRO")]

db17<-merge(db17,ref2, by="R_FONTE", all.x=TRUE)
summary(db17$MACRO)
write.dbf(db17, "db17.dbf")

```

9. Análise do sobrerastreamento por faixa etária

```

# ler os dados
setwd("G:/doutorado/SISMAMA/sobrevida_br_nod_risco/sobrerastreamento
final18_exame2010")

# Importar a biblioteca necessã  Ária com as funcoes
library(foreign)

Mamogref2<-read.dbf("Mamogref2.dbf")
db17<-read.dbf("db17.dbf")

names(Mamogref2)
names(db17)

#Excluir exames depois 2010
dbfetotal<-read.dbf("db17.dbf")

#criar a variável ano de exame
library(lubridate)
dbfetotal$ANO_EXAME<-year(dbfetotal$DT_EXAME0)
freq(dbfetotal$ANO_EXAME)
dbfetotal<- dbfetotal[dbfetotal$DT_EXAME0>"2009-12-31" & dbfetotal$DT_EXAME0<"2011-
01-01",]

```

```

freq(dbfetotal$ANO_EXAME)

##Excluir exames Bi_rads 6, códigos 07

#Legenda
#0 - Default
#01 - categoria 0 (inconclusivo, avaliação adicional por imagem ou exame anterior)
#02 - categoria 1 (sem achados)
#03 - categoria 2 (achados benignos)
#04 - categoria 3 (achado provavelmente benigno - repetir exame em 6 meses)
#05 - categoria 4 (achados suspeitos de malignidade - biópsia)
#06 - categoria 5 (achados altamente sugestivos de malignidade)
#07 - categoria 6 (biópsia prévia com malignidade comprovada)

# retirar notaÃ§Ã£o cientÃfica
options(scipen=999)

freq(dbfetotal$DIAG_MD0)
dbfetotal<-subset(dbfetotal,
(DIAG_MD0=="0")|(DIAG_MD0=="1")|(DIAG_MD0=="2")|(DIAG_MD0=="3")|(DIAG_MD0=
=="4")|(DIAG_MD0=="5")|(DIAG_MD0=="6"))
freq(dbfetotal$DIAG_MD0)
freq(dbfetotal$DIAG_ME0)
dbfetotal<-subset(dbfetotal,
(DIAG_ME0=="0")|(DIAG_ME0=="1")|(DIAG_ME0=="2")|(DIAG_ME0=="3")|(DIAG_ME0==
"4")|(DIAG_ME0=="5")|(DIAG_ME0=="6"))
freq(dbfetotal$DIAG_ME0)
table(dbfetotal$DIAG_MD0, dbfetotal$DIAG_ME0)

# Selecionar mulheres sem nódulo na mama
#"NODULO": Apresenta Nódulo ou Carço na Mama ( 0 - Def ; 3 - Não)
library(descr)
freq(dbfetotal$NODULO_x)
dbfetotal<-subset(dbfetotal, (NODULO_x=="3"))
freq(dbfetotal$NODULO_x)

# Selecionar mulheres sem risco
#"RISCO": mulheres com história familiar, de pelo menos, um parente de primeiro grau com
diagnóstico de: - câncer de mama antes dos 50 anos de idade; - câncer de mama bilateral ou câncer
de ovário; - mulheres com história familiar de câncer de mama masculino; - mulheres com
diagnóstico histopatológico de lesão mamária (1 - Não; 2 - Não Sabe; 3 - Sim)
freq(dbfetotal$RISCO)
dbfetotal<-subset(dbfetotal, (RISCO=="1")|(RISCO=="2"))
freq(dbfetotal$RISCO)

# criar variáveis faixa etaria (uma com 3 categorias e outra com duas categorias)
dbfetotal$IDADE <- as.character(dbfetotal$IDADE)
dbfetotal$IDADE <- as.numeric(dbfetotal$IDADE)
summary(dbfetotal$IDADE)

dbfetotal$faixaetaria<- cut(dbfetotal$IDADE,breaks=c(0,49,69,110),labels=c("<50","50-
69",">=70"))
summary(dbfetotal$faixaetaria)

```

```

freq(dbfetotal$faixaetaria)

dbfetotal$faixaetaria3<- cut(dbfetotal$IDADE,breaks=c(0,29,39,49,69,110),labels=c("<30","30-39","40-49","50-69",">=70"))
library(descr)
freq(dbfetotal$faixaetaria3)
write.dbf(dbfetotal, "sobrerrastreamentofaixaetariacorreto.dbf")

db30<-read.dbf("sobrerrastreamentofaixaetariacorreto.dbf")
freq(db30$MACRO)
freq(db30$IDADE)
library(descr)
names(db30)
freq(db30$faixaetari.1)

table(db30$MACRO,db30$faixaetari.1)
db31<-as.data.frame(table(db30$MACRO, db30$faixaetari.1))
write.csv2(db31,"sobrerrastreamentofe_macrocorreto.csv")

#Olhar o sobrerrastreamento por faixa etária por microrregião
freq(db30$CIR)
freq(db30$IDADE)
library(descr)
names(db30)
freq(db30$faixaetari.1)

table(db30$CIR,db30$faixaetari.1)
db31<-as.data.frame(table(db30$CIR, db30$faixaetari.1))
write.csv2(db31,"sobrerrastreamentofe_microcorreto.csv")

#Salvar o número de exames em cada micro
db32<-as.data.frame(table(db30$CIR))
write.csv2(db32,"mulheresfe_microcorreto.csv")

```

10. Análise do Sobrerrastreamento por periodicidade

```

#Separar banco para periodicidade
#Extrair faixa etária 50 a 69 anos
db18<-read.dbf("sobrerrastreamentofaixaetariacorreto.dbf")
db18<-subset(db18, faixaetari=="50-69")

#Bi-rads
#Legenda
#0 - Default
#01 - categoria 0 (inconclusivo, avaliação adicional por imagem ou exame anterior)
#02 - categoria 1 (sem achados)
#03 - categoria 2 (achados benignos)
#04 - categoria 3 (achado provavelmente benigno - repetir exame em 6 meses)
#05 - categoria 4 (achados suspeitos de malignidade - biópsia)
#06 - categoria 5 (achados altamente sugestivos de malignidade)
#07 - categoria 6 (biópsia prévia com malignidade comprovada)

```

```

# retirar notaÃ§Ã£o cientÃfica
options(scipen=999)

# Selecionar Bi-rads do primeiro exame juntando o diagnÃstico da mama direita e esquerda
db18$BR_GERAL<-paste(db18$DIAG_MD0, sep=" ", db18$DIAG_ME0)
freq(db18$BR_GERAL)

db18a<-subset(db18, (BR_GERAL=="2 2")|(BR_GERAL=="2 3")|(BR_GERAL=="3
2")|(BR_GERAL=="3 3"))
freq(db18a$DIAG_MD0)
freq(db18a$DIAG_ME0)
table(db18a$DIAG_MD0, db18a$DIAG_ME0)
write.dbf(db18a,"sobrerrastreamentoperiodicidade.dbf")

db18<-read.dbf("sobrerrastreamentoperiodicidade.dbf")
names(db18)

names(db18)[65]<-"ANO_EXAME"
names(db18)[66]<-"faixaetaria"
names(db18)[67]<-"faixaetaria2"

db18<-subset(db18, select=c(R_FONTE, DT_EXAME0, DT_EXAME1, IDADE, MUN_COD,
CIR, MACRO, DIAG_MD0, DIAG_ME0, BR_GERAL, NODULO, RISCO, ANO_EXAME,
faixaetaria, faixaetaria2))
write.dbf(db18, file="db18b.dbf")
names(db18)

# Se precisar transformar as variaveis com data (aaaa/mm/dd) no formato data
db18$DT_EXAME0<- as.Date(db18$DT_EXAME0, "%Y/%m/%d")
db18$DT_EXAME1 <- as.Date(db18$DT_EXAME1, "%Y/%m/%d")

# criar data final para sobrevida em 18 meses
db18$datafinal <- db18$DT_EXAME1
db18$datafinal[is.na(db18$DT_EXAME1)] <- as.Date("2012-12-31")
summary(db18$datafinal)

# criar as variaveis tempo e status
str(db18$datafinal)
str(db18$DT_EXAME0)
db18$tempodias <- as.numeric(db18$datafinal - db18$DT_EXAME0)
db18$tempomes <- round(db18$tempodias/30.4)
db18$status <- ifelse(db18$tempodias > 547,0, 1)

# limitar o tempo a 18 meses
db18$tempo <- db18$tempodias
db18$tempo[db18$tempo> 548 ] <- 548
write.dbf(db18, "db18b.dbf")
db18<-read.dbf("db18b.dbf")
names(db18)

names(db18)
names(db18)[14] <- "faixaetaria"
names(db18)[15] <- "faixaetaria2"

```

```

table(db18$tempo, db18$status)
table(db18$tempomes, db18$status)
library(descr)
freq(db18$status)
freq(db18$tempomes)
table(db18$tempomes, db18$faixaetaria)
table(db18$status, db18$faixaetaria)

write.dbf(db18, "db18c.dbf")
db18<-read.dbf("db18c.dbf")
db18<-read.dbf("db18c.dbf")

## ANALISE NAO PARAMETRICA - KAPLAN-MEIER

# pacote necessario
library(survival)
options(max.print=10000,scipen=999)

# estimar o modelo
KM <- survfit(Surv(tempo,status) ~ 1, data=db18)
print(KM, print.rmean=TRUE)
par(mar=c(4,4,4,4),mfrow=c(1,1))
plot(KM, xlab = "Tempo (dias)", ylab = "S(t)", main = "Sobrerastreamento por Periodicidade",
mark.time = F)
plot(KM, xlab = "Tempo (dias)", ylab = "1-S(t)", main = "Sobrerastreamento por Periodicidade",
mark.time = F, fun="event")
summary(KM)

# tabela completa, com todos os tempos e taxas de sobrevida

# mostrar a taxa de sobrevida no tempo final (18 meses), com IC95%
round(min(KM[["surv"]])*100,2)
round(min(KM[["lower"]])*100,2) # limite inferior
round(min(KM[["upper"]])*100,2) # limite superior

# mostrar o tempo mediano de sobrevida (com IC95%), se houver
quantile(KM, probs = 0.5)[1]
quantile(KM, probs = 0.5)[2] # limite inferior
quantile(KM, probs = 0.5)[3] # limite superior

freq(db18$status)

# mostrar o tempo médio restrito (18 meses) de sobrevida e seu erro-padrão
survival:::survmean(KM, rmean=548) [[1]]["*rmean"]
survival:::survmean(KM, rmean=548) [[1]]["*se(rmean)"]

# salvar grafico mais informativo como imagem
library("survminer")
tiff("curvakm.tiff",width=10, height=10,units="in",res=72)
par(mar=c(2,2,2,2),mfrow=c(1,1))
ggsurvplot(KM,risk.table="nrisk_cumevents",conf.int=T,surv.median.line="v")
dev.off()
ggsurvplot(KM,risk.table="nrisk_cumevents",conf.int=T,surv.median.line="v", fun="event")

```

```
## LOOP PARA ESTIMAR SOBREVIDA PARA CADA MACRO
```

```
# criar um vetor com os nomes das macros
```

```
macros <- as.character(unique(db18$MACRO))
```

```
# criar uma lista com bancos separados para cada macro
```

```
lista <- lapply(macros,function(x) db18[db18$MACRO==x,])
```

```
names(lista) <- macros
```

```
n <- length(macros)
```

```
# criar um banco de dados para receber os resultados (taxa de sobrevida e tempos medios restritos)
```

```
sobrevida_macro <-
```

```
data.frame(macros,sob=rep(NA,n),sobLI=rep(NA,n),sobLS=rep(NA,n),rmst=rep(NA,n),rmst_se=rep(NA,n))
```

```
# loop para cada região
```

```
for(i in seq(lista)) {
```

```
  sub <- lista[[i]]
```

```
  KM <- survfit(Surv(tempo,status) ~ 1, data=sub)
```

```
  sobrevida_macro$sob[i] <- round(min(KM[["surv"]])*100,2)
```

```
  sobrevida_macro$sobLI[i] <- round(min(KM[["lower"]])*100,2)
```

```
  sobrevida_macro$sobLS[i] <- round(min(KM[["upper"]])*100,2)
```

```
  sobrevida_macro$rmst[i] <- round(survival::survmean(KM, rmean=548) [[1]]["*rmean"],2)
```

```
  sobrevida_macro$rmst_se[i] <- round(survival::survmean(KM, rmean=548)
```

```
  [[1]]["*se(rmean)",2)
```

```
}
```

```
# salvar os resultados como arquivo externo
```

```
write.csv2(sobrevida_macro,"sobrevida_macro2.csv",row.names=F)
```

```
#criar um banco de soberrastreamento
```

```
sobrevida_macro$soberrastreamento<- 100-sobrevida_macro$sob
```

```
sobrevida_macro$soberrastreamentoLI<- 100-sobrevida_macro$sobLI
```

```
sobrevida_macro$soberrastreamentoS<- 100-sobrevida_macro$sobLS
```

```
write.csv2(sobrevida_macro,"sobrevida_rastreamento_macro2.csv",row.names=F)
```

```
# remover os arquivos temporarios
```

```
rm(macros,lista,n,i,sub,KM,sobrevida_macro)
```

```
## LOOP PARA ESTIMAR SOBREVIDA PARA CADA MICRO
```

```
# criar um vetor com os nomes das micros
```

```
micros <- as.character(unique(db18$CIR))
```

```
# criar uma lista com bancos separados para cada micro
```

```
lista <- lapply(micros,function(x) db18[db18$CIR==x,])
```

```
names(lista) <- micros
```

```
n <- length(micros)
```

```
# criar um banco de dados para receber os resultados (taxa de sobrevida e tempos medios restritos)
```

```
sobrevida_micro <-
```

```
data.frame(micros,sob=rep(NA,n),sobLI=rep(NA,n),sobLS=rep(NA,n),rmst=rep(NA,n),rmst_se=rep(NA,n))
```

```

# loop para cada região
for(i in seq(lista)) {
  sub <- lista[[i]]
  KM <- survfit(Surv(tempo,status) ~ 1, data=sub)
  sobrevida_micro$sob[i] <- round(min(KM[["surv"]])*100,2)
  sobrevida_micro$sobLI[i] <- round(min(KM[["lower"]])*100,2)
  sobrevida_micro$sobLS[i] <- round(min(KM[["upper"]])*100,2)
  sobrevida_micro$rmst[i] <- round(survival::survmean(KM, rmean=548) [[1]]["*rmean"],2)
  sobrevida_micro$rmst_se[i] <- round(survival::survmean(KM, rmean=548)
[[1]]["*se(rmean)",2)
}

# salvar os resultados como arquivo externo
write.csv2(sobrevida_micro,"sobrevida_micro.csv",row.names=F)

#criar um banco de sobrerastreamento
sobrevida_micro$sobrerastreamento<- 100-sobrevida_micro$sob
sobrevida_micro$sobrerastreamentoLI<- 100-sobrevida_micro$sobLI
sobrevida_micro$sobrerastreamentoS<- 100-sobrevida_micro$sobLS
write.csv2(sobrevida_micro,"sobrevida_rastreamento_micro2.csv",row.names=F)

# remover os arquivos temporarios
rm(micros,n,lista,i,sub,KM,sobrevida_micro)

```

11. Limpeza do banco para cálculo da cobertura

Repetir os passos de 1 a 28 da limpeza da base referência para limpar os bancos dos anos 2010, 2011 e 2012. No passo 4 selecionar a variável D_RECEBE que corresponde a data em que o exame foi realizado. E no passo 6 selecionar o período de 31/12/2009 a 01/01/2011.

Observação: O processo foi repetido com os bancos 2011 e 2012 e foram extraídos os exames realizados no ano de 2010.

```

write.dbf(db08,"db08.dbf")
db08<-read.dbf("db08.dbf")

table(db08$IDADE)

str(db08$IDADE)

db08$IDADE<-as.character(db08$IDADE)
db08$IDADE<-as.numeric(db08$IDADE)

db08$F_ETARIA[18<=db08$IDADE&db08$IDADE<40]<- '<39 anos'
db08$F_ETARIA[40<=db08$IDADE&db08$IDADE<50]<- '40 a 49 anos'
db08$F_ETARIA[50<=db08$IDADE&db08$IDADE<60]<- '50 a 59 anos'
db08$F_ETARIA[60<=db08$IDADE&db08$IDADE<70]<- '60 a 69 anos'
db08$F_ETARIA[70<=db08$IDADE&db08$IDADE<110]<- '>70 anos'
table(db08$F_ETARIA)

table(db08$IDADE)
library (descr)
freq(db08$IDADE)

```



```

table(db08$F_ETARIA)
freq(db08$F_ETARIA)

table(db08$IDADE==50)

names(db08)

library(descr)
table(db08$MACRO)
table(db08$MACRO, db08$F_ETARIA=='50 a 59 anos')
table(db08$MACRO, db08$F_ETARIA=='60 a 69 anos')

#Macro
exames_macro <- as.data.frame(table(db08$MACRO_COD, db08$MACRO))
names(exames_macro)<-c("MACRO_COD", "MACRO", "N°EXAMES")
exames_macro<-exames_macro[exames_macro$N°EXAMES>0,]
write.csv(exames_macro,"exames_macro.csv")

table(db08$F_ETARIA)
exames_macro1<- as.data.frame(table(db08$MACRO_COD, db08$MACRO,
db08$F_ETARIA=='50 a 59 anos'))
names(exames_macro1)<-c("MACRO_COD", "MACRO", "CONDIÇÃO", "N_EXAMES")
exames_macro1<-exames_macro1[exames_macro1$N_EXAMES>0,]
exames_macro1<- subset(exames_macro1,(CONDIÇÃO=="TRUE"))
write.csv(exames_macro1,"exames_macro1.CSV")

exames_macro2<- as.data.frame(table(db08$MACRO_COD, db08$MACRO,
db08$F_ETARIA=='60 a 69 anos'))
names(exames_macro2)<-c("MACRO_COD", "MACRO", "CONDIÇÃO", "N_EXAMES")
exames_macro2<-exames_macro2[exames_macro2$N_EXAMES>0,]
exames_macro2<- subset(exames_macro2,(CONDIÇÃO=="TRUE"))
write.csv(exames_macro2,"exames_macro2.CSV")

#Micro
exames_CIR <- as.data.frame(table(db08$CIR_COD, db08$CIR))
names(exames_CIR)<-c("CIR_COD", "CIR", "N°EXAMES")
exames_CIR<-exames_CIR[exames_CIR$N°EXAMES>0,]
write.csv(exames_CIR,"exames_CIR.csv")

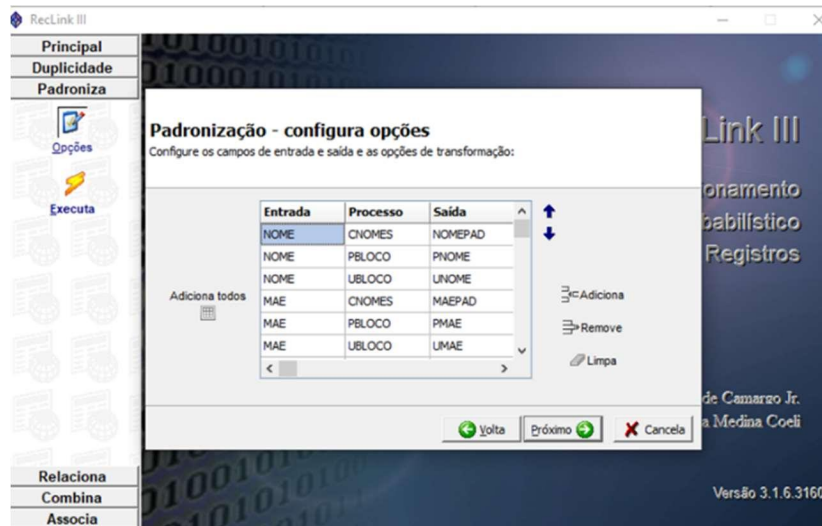
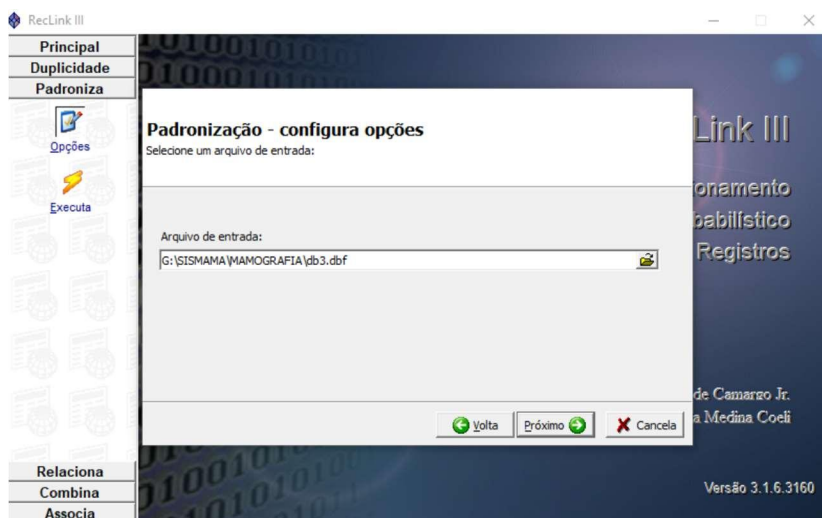
table(db08$F_ETARIA)
exames_CIR1<- as.data.frame(table(db08$CIR_COD, db08$CIR, db08$F_ETARIA=='50 a 59
anos'))
names(exames_CIR1)<-c("CIR_COD", "CIR", "CONDIÇÃO", "N_EXAMES")
exames_CIR1<-exames_CIR1[exames_CIR1$N_EXAMES>0,]
exames_CIR1<- subset(exames_CIR1,(CONDIÇÃO=="TRUE"))
write.csv(exames_CIR1,"exames_CIR1.CSV")

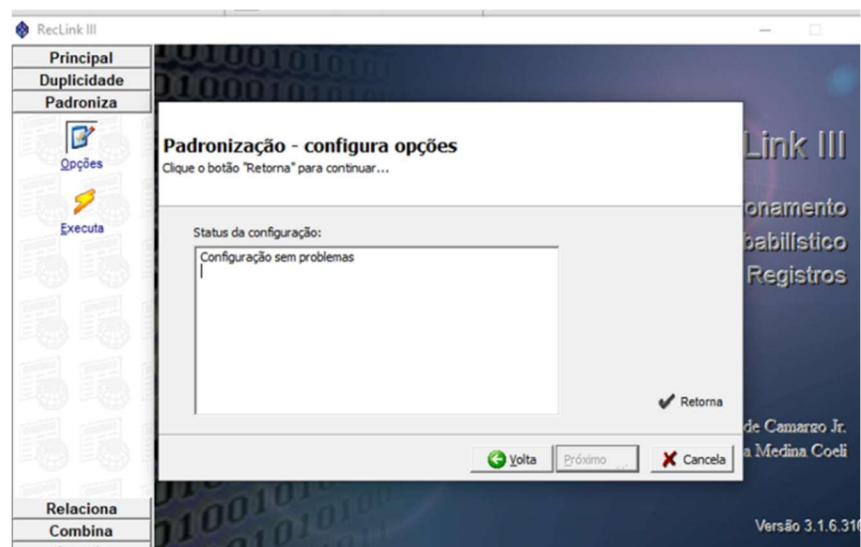
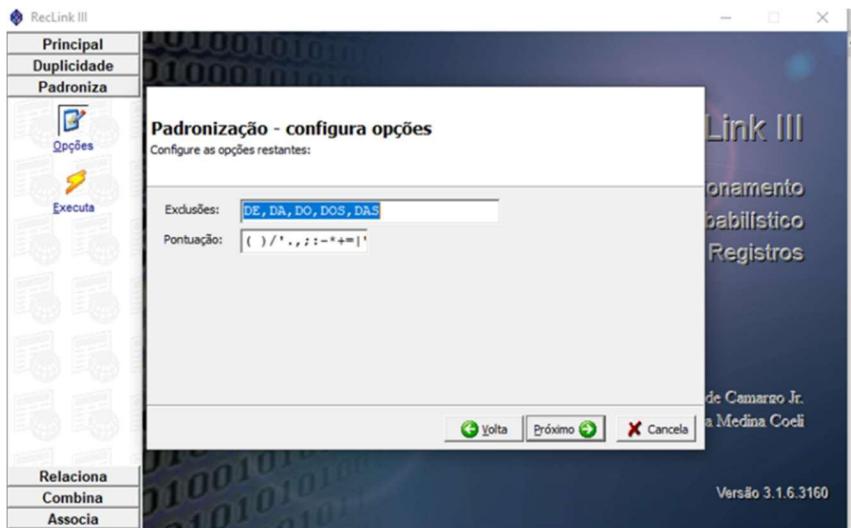
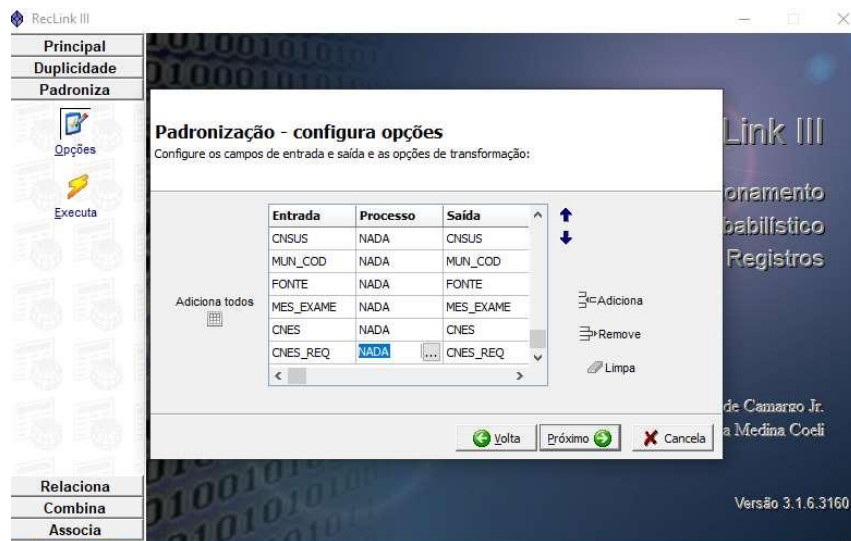
exames_CIR2<- as.data.frame(table(db08$CIR_COD, db08$CIR, db08$F_ETARIA=='60 a 69
anos'))
names(exames_CIR2)<-c("CIR_COD", "CIR", "CONDIÇÃO", "N_EXAMES")
exames_CIR2<-exames_CIR2[exames_CIR2$N_EXAMES>0,]
exames_CIR2<- subset(exames_CIR2,(CONDIÇÃO=="TRUE"))
write.csv(exames_CIR2,"exames_CIR2.CSV")

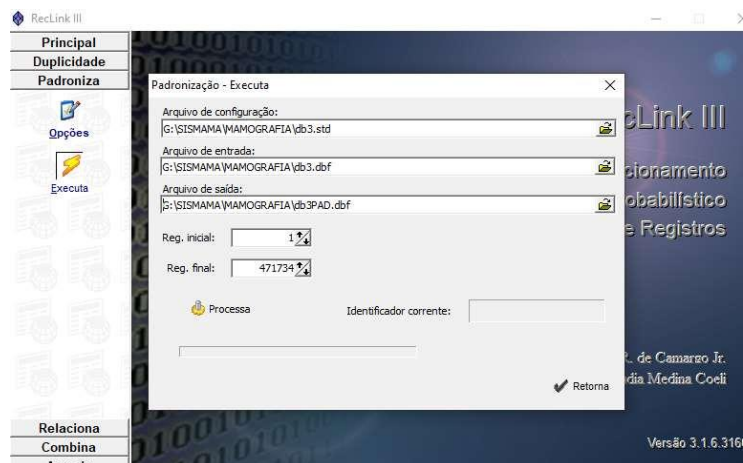
```

APÊNDICE III - SCRIPTS EXECUTADOS NO RECLINK

1. Padronização da base referência



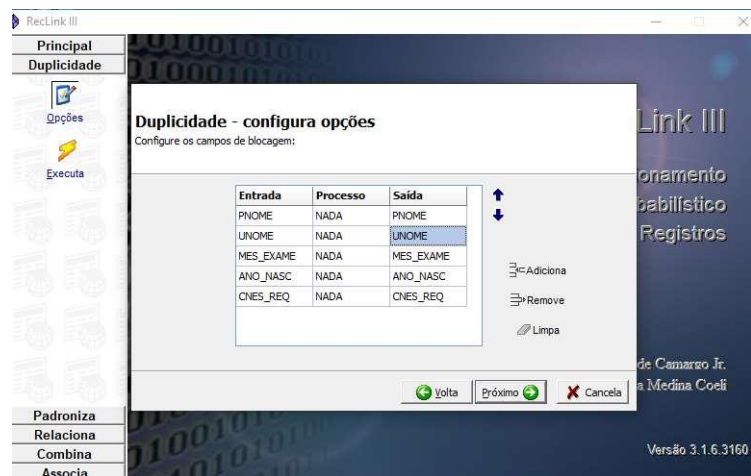
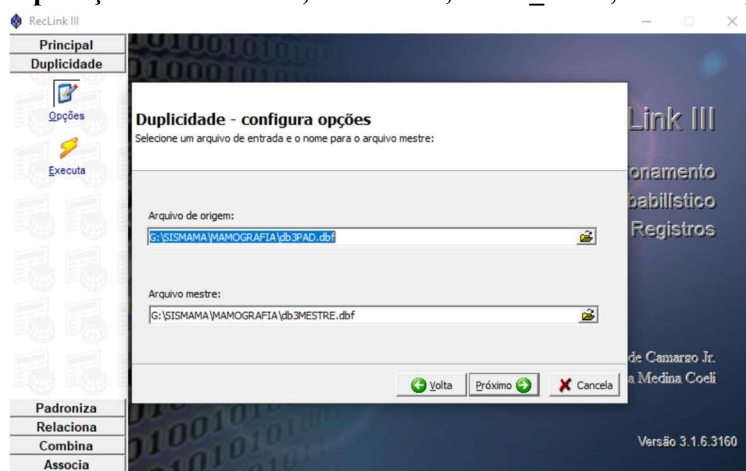


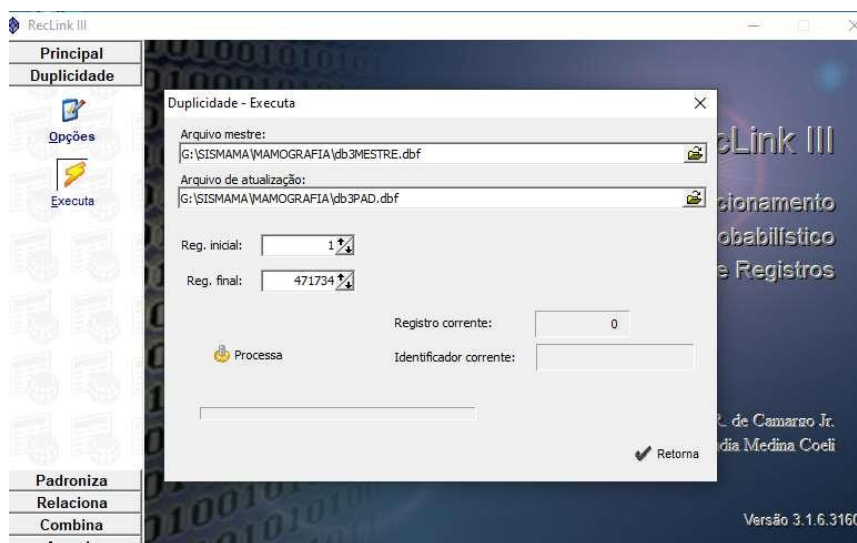
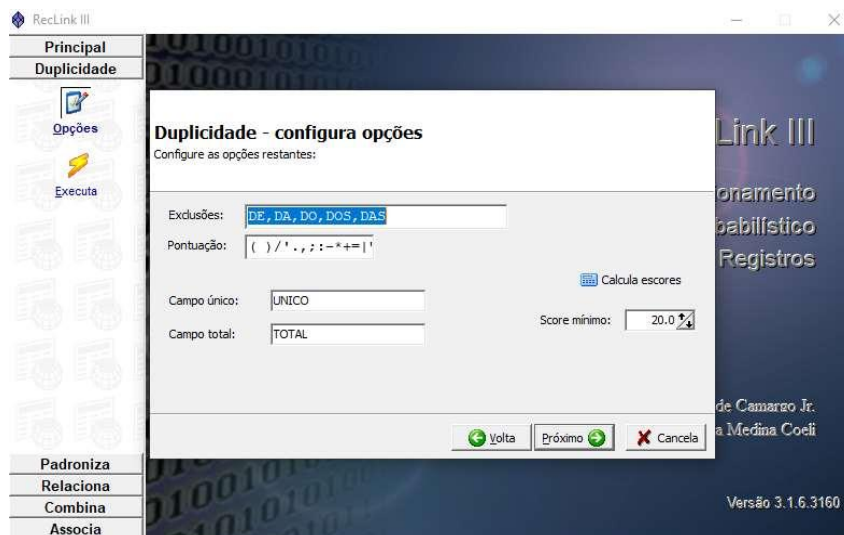
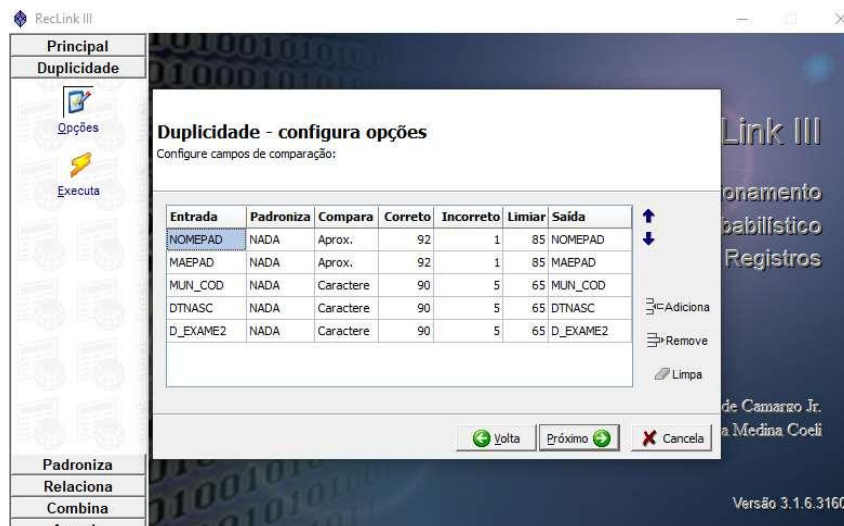


Observação: Os bancos Exames Alterados, Seguimento e Cobertura foram Padronizados da mesma forma.

2. Retirada de Duplicidades da base referência– Duplicidade 2

- Usar na blocagem: **PNOME, UNOME, MÊS_EXAME, ANO_NASC, CNES_REQ**
- Usar na comparação: **NOMEPAD, MAEPAD, MUN_COD, DTNASC, D_EXAME2**

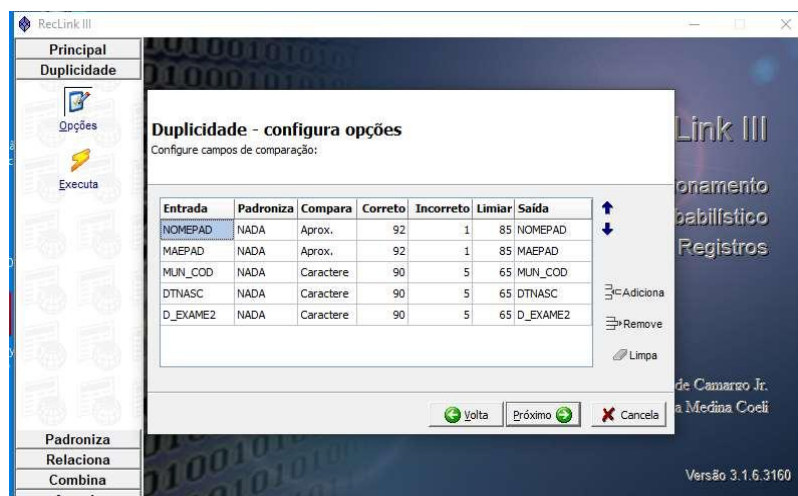
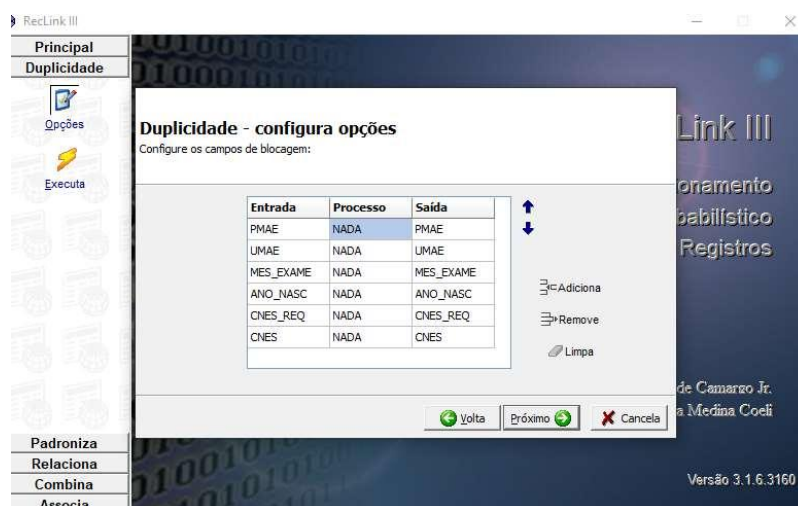
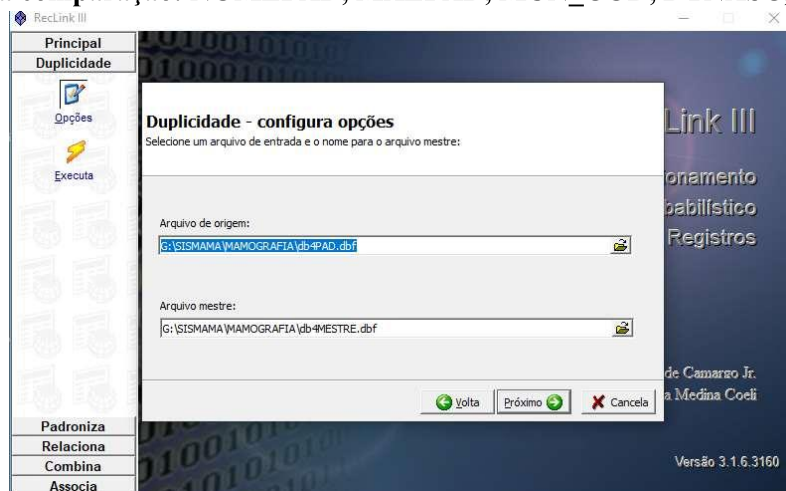


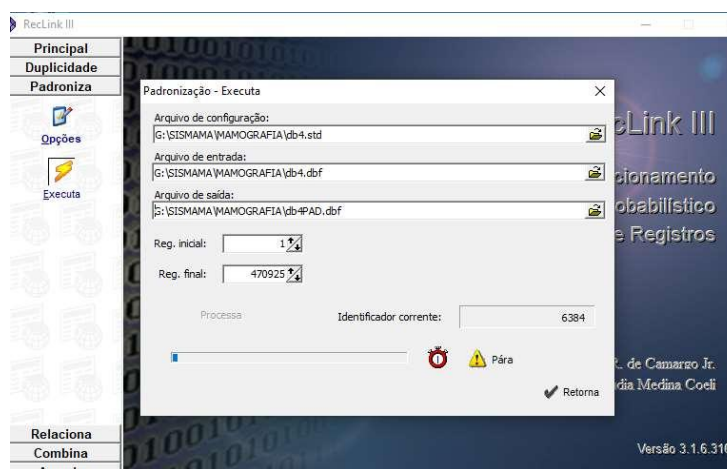
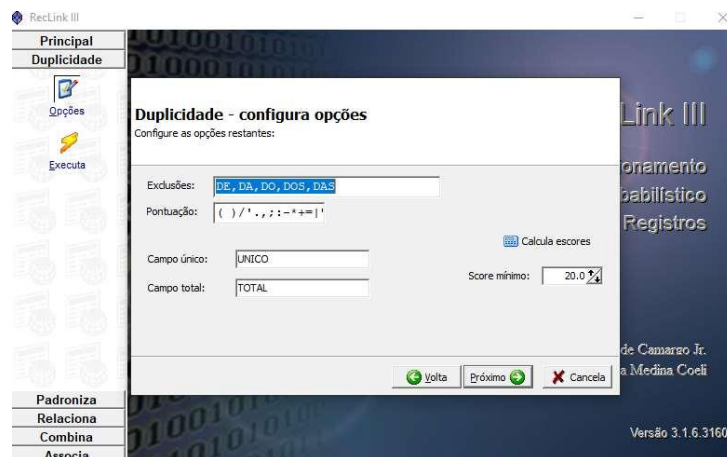


Observação: Os bancos Exames Alterados, Seguimento e Cobertura passaram pelo processo de retirada de duplicidades 2.

3. Retirada de Duplicidades da base referência – Duplicidade 3

- Usar na blocagem: PMAE, UMAE, MÊS_EXAME, ANO_NASC, CNES_REQ, CNES
- Usar na comparação: NOMEPAD, MAEPAD, MUN_COD, DTNASC, D_EXAME2



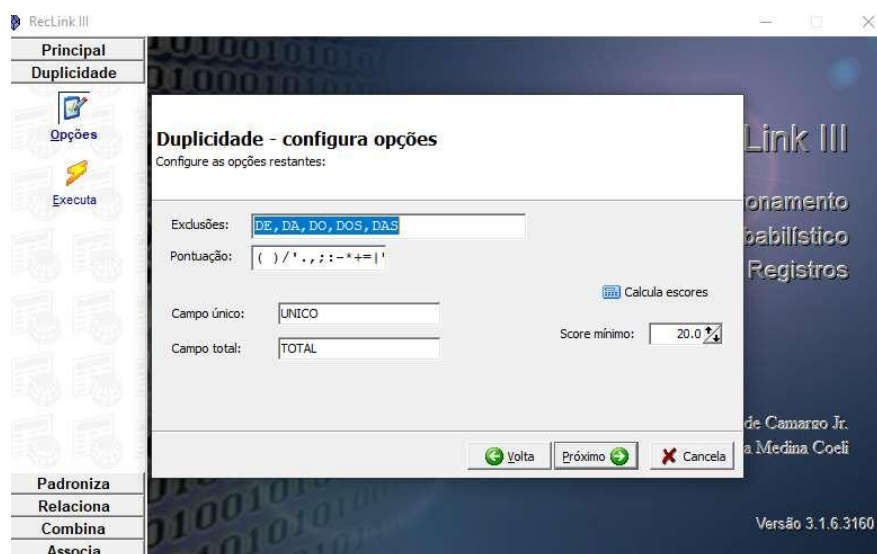
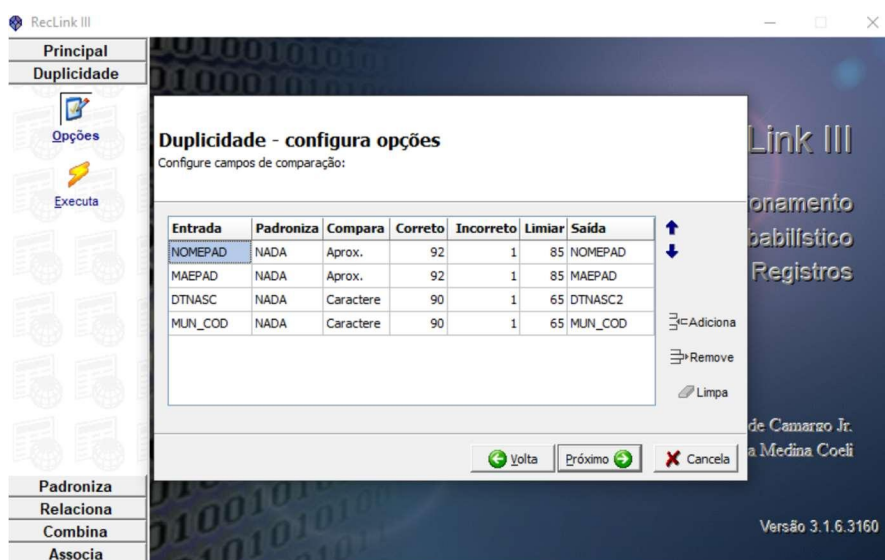
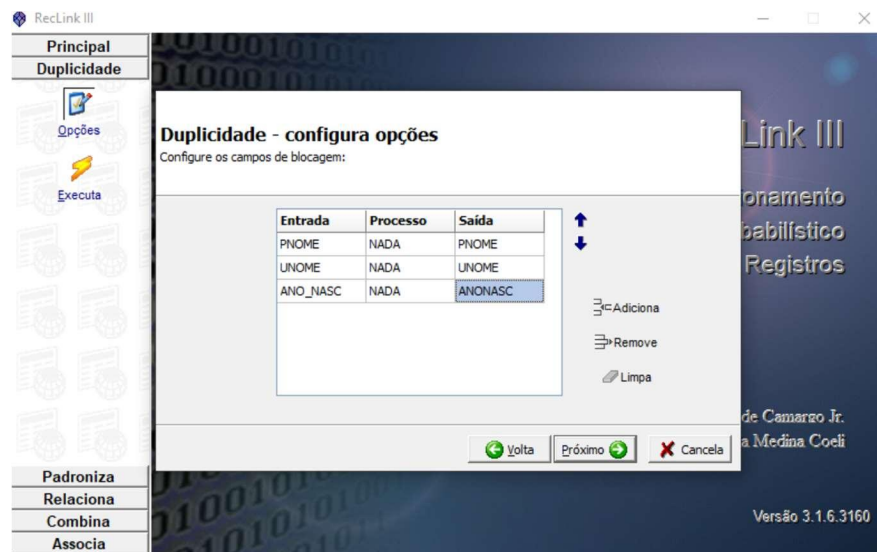


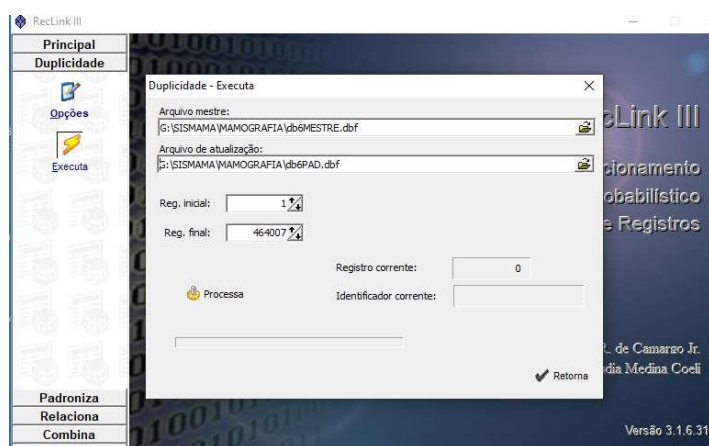
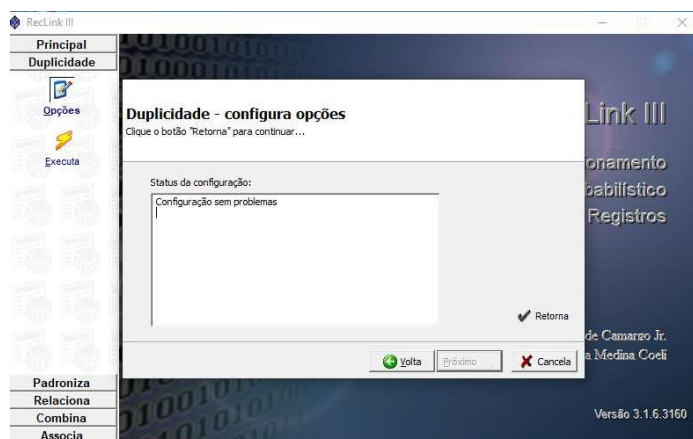
Observação: Os bancos Exames Alterados, Seguimento e Cobertura passaram pelo processo de retirada de duplicidades 3.

4. Retirada de Repetições da base referência – Repetição 2

- **BLOCAÇÃO:** PNOOME, UNOME, ANO_NASC
- **PAREAMENTO:** NOMEPAD, MAEPAD, DTNASC, MUN_COD





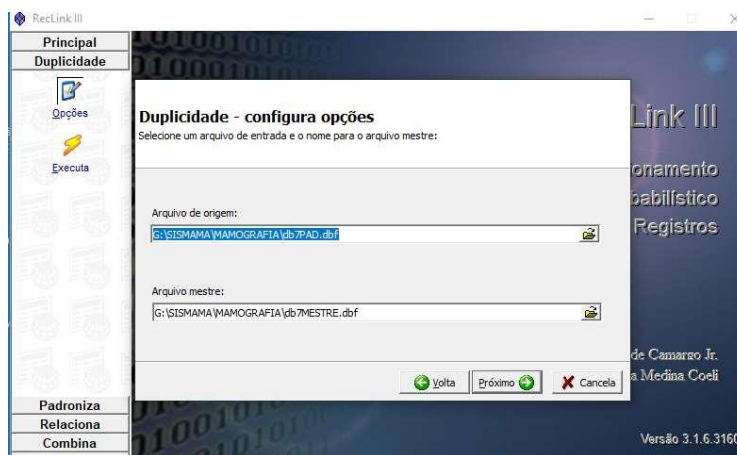


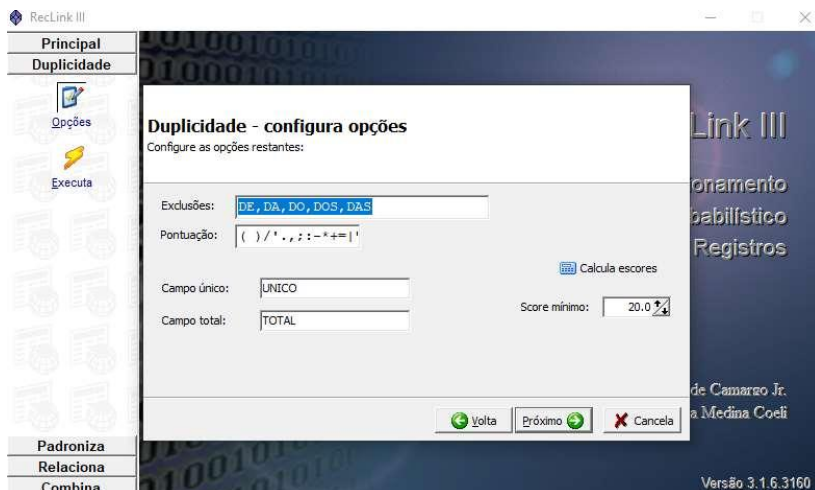
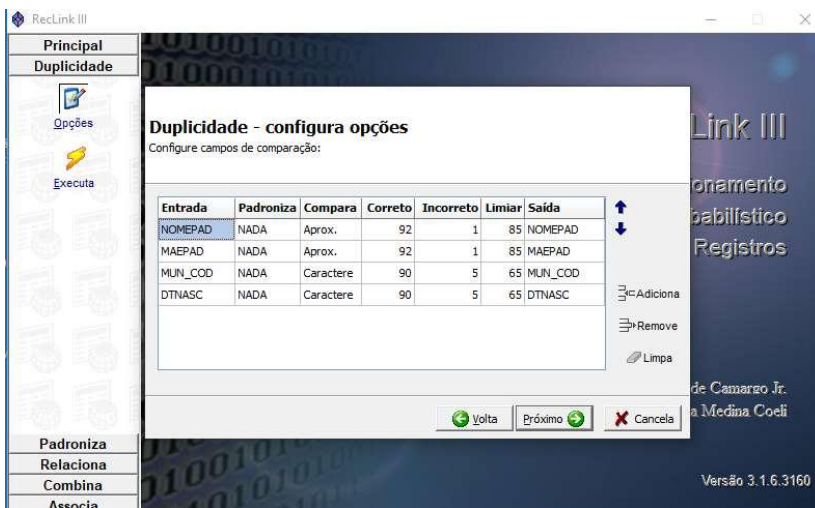
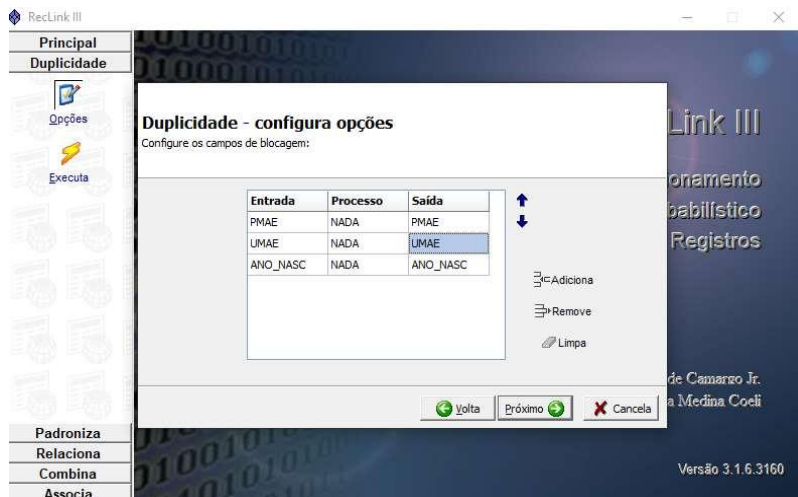
Observação: Os bancos Exames Alterados e Cobertura passaram pelo processo de retirada de repetição 2.

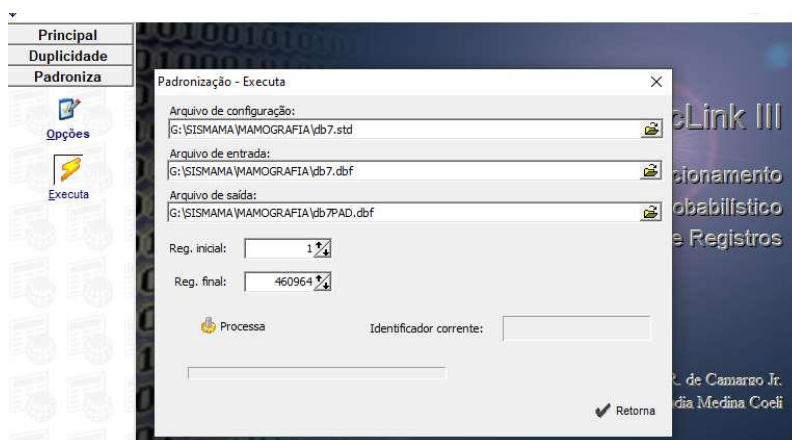
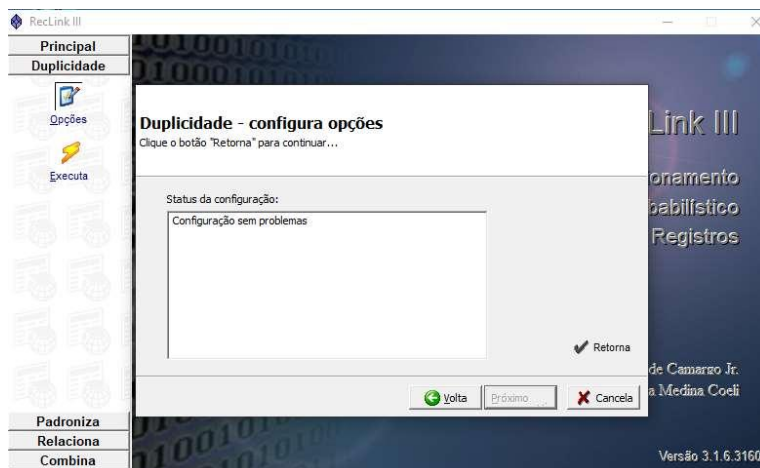
5. Retirada de Repetições da base referência – Repetição 3

BLOCAÇÃO: PMAE, UMAE, ANO_NASC

PAREAMENTO: NOMEPAD, MAEPAD, DTNASC, COD_MUN





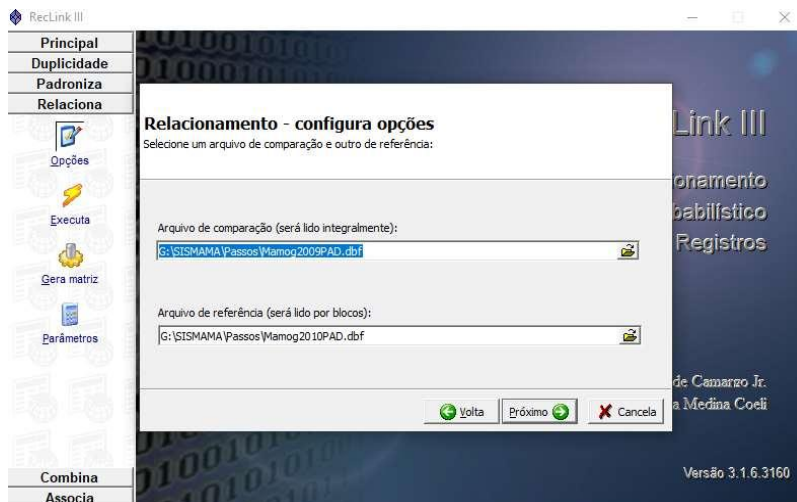


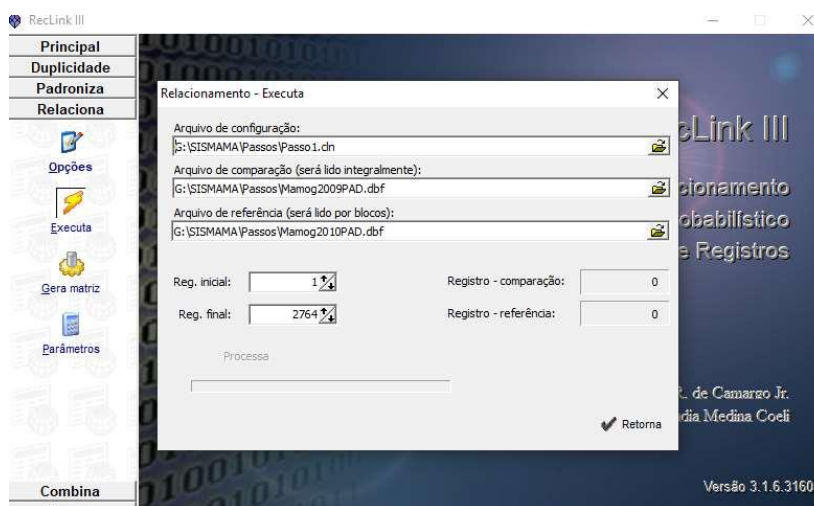
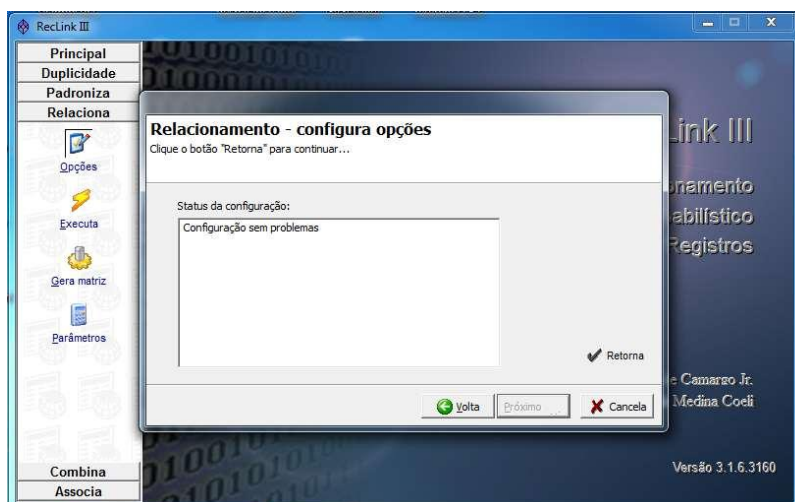
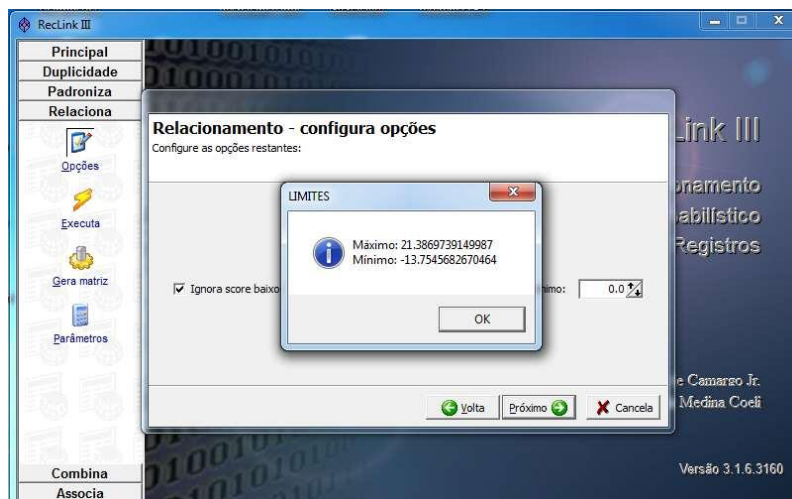
Observação: Os bancos Exames Alterados e Cobertura passaram pelo processo de retirada de repetição 3.

6. Relacionamento da base referência com o banco de exames alterados

a) Passo 1







Observação: O relacionamento entre as bases referência e exames alterados foi feito em 16 passos. Entre as bases referência e seguimento com apenas 3 passos (16, 9 e 12).

Passos, Variáveis de blocagem e pareamento, escores, pares formados e pares verdadeiros do relacionamento dos bancos do SISMAMA 2009 e 2010.

Passos	Variáveis de blocagem	Variáveis de pareamento	Escore
1	PN + UN + AN	N+DN+M+MUN	-13 a 21
2	PN + UN + AN	N+DN	-6 a 10
3	CNS	N + DN + M	-10 a 17
4	PN + UN + AN + PM + UM	N + DN + M	-10 a 17
5	PN + UN + AN	N + DN + M	-10 a 17
6	PN + AN + PM + UM	N + DN + M	-10 a 17
7	PN + AN + UM	N + DN + M	-10 a 17
8	PN + AN + PM	N + DN + M	-10 a 17
9	PN + AN	N + DN + M	-10 a 17
10	PN + UN + PM + UM	N + DN + M	-10 a 17
11	PN + UN + PM	N + DN + M	-10 a 17
12	PN + UN	N + DN + M	-10 a 17
13	PN + PM + UM	N + DN + M	-10 a 17
14	PN + UM	N + DN + M	-10 a 17
15	UN + AN	N + DN + M	-10 a 17
16	AN	N + DN + M	-10 a 17

AN: Ano de Nascimento; CNS: Cartão Nacional do SUS; DN: Data de Nascimento; M: Mãe; MUN: Município; N: Nome; PM: Primeiro Nome da Mãe; PN: Primeiro nome; UM: Último Nome da Mãe; UN: Último nome.