

## REGRAS DE ASSOCIAÇÃO TEMPORAL EM CANCER DE MAMA

Rosyellen R. Szvarça (Psicóloga e mestranda em tecnologia em saúde – PUC-PR)

Sergio Ossamu Ioshii (Médico e Professor – PUC- PR)

Deborah Ribeiro Carvalho (Professor – PUC-PR) E-mail: ribeiro.carvalho@pucpr.br

Willian Felipe Sokoloski (Graduando em Engenharia da Computação – PUC-PR)

**Resumo:** *O objetivo deste artigo é apresentar o potencial do processo de Knowledge Discovery in Databases para melhor compreender a evolução do diagnóstico até a intervenção cirúrgica em beneficiárias de operadoras de planos de saúde com diagnóstico de cancer de mama. Para tal foram descobertas regras de associação, as quais foram pós-processadas para identificar possíveis situações de causa e efeito. Foi adotada uma base de dados compreendendo 50 mulheres submetidas a pelo menos um tratamento cirúrgico conservador ou mutilador no período de 1994 a 2013. As regras pós-processadas, considerando janelas temporais, mostraram que apesar das condutas cirúrgicas serem adotadas em períodos curtos de tempo, a decisão pelo tipo de procedimento depende do tempo para o diagnóstico, relacionado a rotinas mamográficas para detecção. O resultado suscitou a necessidade de ajuste no período de tempo entre procedimentos cirúrgicos e exames de detecção. O processo de Knowledge Discovery in Databases demonstrou ser interessante para apoiar sobre procedimentos que permitam diagnóstico e tratamento precoce, objetivando terapêuticas conservadoras e a diminuição das mutiladoras, as quais trazem grandes impactos psicológicos às mulheres.*

**Palavras-chave:** Cancer de mama, regras de associação, saúde, prevenção, mamografia.

## TEMPORAL ASSOCIATION RULES IN BREAST CANCER

**Abstract:** *This study aimed to present the potential of the process of Knowledge Discovery in Databases to better understand the evolution of diagnosis up to surgical intervention in women diagnosed with breast cancer who were beneficiaries of health insurance providers. Thus, association rules were discovered and were post-processed to identify possible cause-effect situations. A database of 50 women who underwent at least one conservative or mutilating surgery from 1994 to 2013 was adopted. The post-processed rules, which considered temporal windows, showed that although surgical treatments are performed in short periods of time, the decision on the type of procedure to be used depends on the time to diagnosis, related to mammographic detection routines. The result prompted the need for adjustment in the period of time between surgical procedures and screening tests. The process of Knowledge Discovery in Databases was found to be useful to support procedures that allow early diagnosis and treatment, aimed to favor the use of conservative treatments and reduce mutilating treatments that cause significant psychological impacts on women.*

**Keywords:** Breast Cancer, Association Rules, Healthcare, Prevention, Mammography

### 1. INTRODUÇÃO

O desenvolvimento do câncer de mama é dependente de múltiplos fatores e eventos determinados pelo ambiente ao longo da vida da mulher bem como pela suscetibilidade genética. A interação destes fatores tem sido objeto de estudo, entre outros, pelo National Institute of Environmental Health Sciences (NIEHS) e observadas que, dentre os fatores de risco identificados para câncer de mama, incluem a presença de um parente de primeiro grau

com histórico do problema, idade da primeira menstruação, idade do período de menopausa e os riscos ambientais como o uso de pesticidas e químicos sintéticos; fatores hormonais endógenos e ou exógenos; alimentação, uso de tabaco e álcool e exposição excessiva a radiação e ondas magnéticas (JOHNSON-THOMPSON e GUTHRIE, 2000).

As características e estágios desta neoplasia são definidas pelo tamanho do tumor e grau histológico, possuindo dentre os tratamentos, a cirurgia como procedimento invasivo, subdividido em cirurgia conservadora, em que há a retirada apenas do tumor e possíveis margens e Não conservadora ou Mutiladora, subdividida na retirada apenas da glândula mamária (mastectomia subcutânea), retirada da mama com pele e complexo aréola-papilar (mastectomia simples ou total), retirada com preservação de um ou dois músculos peitorais (mastectomia radical modificada) e retirada dos músculos peitorais (mastectomia radical) (BARROS, et al.; 2011).

A indicação de cirurgias conservadoras relacionam-se a diagnóstico precoce e ações de rastreamento por meio da mamografia (MMG). Em THULER e MENDONÇA (2005) é exemplificado que tanto o câncer de mama quanto o de colo do útero, são de bom prognóstico quando tratados precocemente. Em contrapartida, quando diagnosticado em fase avançada, possui altas taxas de mortalidades ou conforme DE SOUSA PINHO, et al. (2009) a predominância de intervenções mutiladoras como a mastectomia, um dos tratamentos mais temidos entre as mulheres com cancer de mama, interferindo diretamente nos aspectos emocionais e sociais, devido a mutilação de uma região do corpo importante para a sexualidade, função materna e estética feminina.

Dentre as formas de detecção do câncer de mama duas estratégias são relevantes como: o diagnóstico precoce, através da conscientização de sinais e sintomas precoces, realização de consultas médicas e o Rastreamento Mamográfico, considerado padrão pelo seu potencial de detectar calcificações características da neoplasia, a partir da submissão dos indivíduos assintomáticos a exames de triagem, potencializando a identificação em estágios iniciais da doença e mudança no prognóstico. (SILVA e HORTALE, 2012).

A mamografia como padrão ideal para detecção de tumores é creditada por diversas organizações, dentre elas The European Society of Breast Cancer (DEL TURCO et al., 2010) órgão responsável por avaliar o desempenho clínico em unidades européias dedicadas ao tratamento do câncer de mama, que padronizou um guia de Indicadores em Qualidade no cancer de Mama destacando a MMG como um indicador da integralidade da elaboração do diagnóstico por imagem com nível de evidência grau III conduzidos por estudos descritivos bem projetados e precisos.

No Brasil, diretrizes clínicas na saúde suplementar em conjunto com a associação Médica Brasileira (AMB) e da Agência Nacional em Saúde Suplementar da prevenção secundária do câncer de mama, preconiza o exame de mamografia, como um forte indicador para diagnóstico, orientando o rastreamento mamográfico em intervalos regulares de 1 a 2 anos para mulheres acima dos 40 anos de idade, prevendo a redução na mortalidade por esta neoplasia ainda maior, acima de 50 anos de idade (BARROS, et al.; 2011).

Liderado em causas de morte mundialmente e considerado o mais frequente entre mulheres, com 23% de incidência do câncer de mama, entre todos os outros cânceres (JBILOU, J. et al.; 2013), justifica-se a padronização dos programas de Rastreamento Mamográfico locais e a elaboração de protocolos em diferentes países, conforme tabelado em CANADIAN TASK FORCE ON PREVENTIVE HEALTH CARE et al. (2011) mostrando a frequência e o intervalo recomendado para realização de MMG de acordo com a faixa etária que, elaborados pelas equipes de força tarefa locais para o controle do câncer de mama, possuem entre outros objetivos, a padronização dos indicadores para diagnóstico, prevenção, qualidade assistencial, tornando-se ferramentas essenciais de apoio à decisão do tratamento

desta neoplasia, especialmente a redução de intervenções cirúrgicas mutiladoras e a o aumento das conservadoras.

O aumento de novos casos da doença e a prevalência de intervenções mutiladoras, é evidenciado em CHAVES (2014), que verificou no período de 2009 a 2013 o número de 40.651 casos de doenças malignas de mama com 10% deste total submetidos a mastectomias radical e radical modificada, no Brasil. O estudo, pôde exemplificar o aumento de informações desta patologia e dos perfis dos indivíduos submetidos aos tratamentos, os quais estão contidos em bancos de dados das instituições de saúde, tornando-se demandas para novos estudos e conseqüentemente, informações relevantes aos gestores na tomada de decisão futuras em iniciativas de prevenção, tratamentos, cura, estadiamento, controle financeiro e principalmente o aumento da qualidade na assistência prestada.

No estudo bibliográfico sobre a padronização do Prontuário Eletrônico do Paciente (PEP) é discutido a importância da estruturação de recursos da informática como instrumentos auxiliares da organização de atendimentos clínicos, coleta e armazenamento dos dados, que associados a integração de sistemas de apoio à decisão tornam-se colaborativos ao diagnóstico e terapêutica (WECHSLER, R. et al.; 2003). Igualmente, no estudo de CHIMIESKI e FAGUNDES (2013) sobre diagnóstico médico de doenças como infarto do miocárdio, câncer de mama e colo do útero e do controle de infecção hospitalar, é relatado a condução de técnicas de aprendizado de máquina, como a utilização de Redes Neurais, no Apoio a Decisão às pessoas e grupos envolvidos no manejo clínico.

Sendo assim, observar e analisar o histórico de mulheres submetidas a tratamento cirúrgico de câncer de mama, considerando a associação dos eventos ocorridos bem como a relação de causa e efeito entre eles, pode melhor apoiar os estudos sobre a padronização de condutas ao diagnóstico precoce, a rotina de rastreamento e determinantes para determinados tipos de intervenção cirúrgica. Uma alternativa tecnológica que possibilita esta análise é o processo Knowledge Discovery in Database (KDD), o qual compreende três grandes etapas: pré-processamento, data mining e pós-processamento REZENDE, S.O.; 2005).

O objetivo deste artigo é apresentar um estudo de caso que descubra associações, considerando a janela temporal, sobre a rotina de rastreamento mamográfico, sua influência no diagnóstico que determinaram tratamentos conservadores ou mutiladores.

## **2. MATERIAIS E MÉTODOS**

Para a pesquisa foi utilizada uma base de dados secundária de uma operadora de plano de saúde, contendo dados referentes a 50 mulheres entre 36 e 81 anos. Como critério de seleção foi adotado o período de 1994 a 2013, a partir do qual foram selecionadas as 50 primeiras mulheres submetidas a pelo menos um tratamento cirúrgico referente ao câncer de mama, com vinculação à operadora por no mínimo 5 anos previamente à cirurgia propriamente dita.

Foram considerados dados referentes a exames de detecção, especialmente a MMG e tratamento cirúrgico e suas respectivas datas de realização. Os exames mamográficos foram tratados distintamente em MMG digital e convencional. Os tratamentos cirúrgicos foram agrupados em conservador ou mutilador. Como conservador foram consideradas quadrantectomia subdividida em linfadenectomia axilar (QLA) e com ressecção segmentar (QRS), além da mastectomia subcutânea com inclusão de prótese (MSIP). Como tratamentos mutiladores foram consideradas a mastectomia radical ou radical modificada (MRM) e mastectomia simples (MS).

Para a etapa de data mining foi adotado o algoritmo APRIORI (BORGELT, C) e para o pós-processamento foi adotado o ASSOCTEMP (SOKOLOSKI, CARVALHO E DALLAGASSA, 2014). Foi selecionada a tarefa de Descoberta de Regras de Associação tendo em vista a natureza do objetivo deste trabalho, ou seja a busca por associações entre os

eventos relacionados ao câncer de mama. A partir das regras de associação descobertas, foi importante inferir a possibilidade de causa e efeito entre os eventos associados. Situação possibilitada a partir da identificação da janela temporal entre as ocorrências associadas pelo algoritmo (SOKOLOSKI, CARVALHO e DALLAGASSA, 2014).

Para os experimentos foram adotadas as seguintes janelas temporais: 365 dias equivalente a (1 ano), 540 dias (1 ano e 6 meses), 730 dias (2 anos), 910 dias (2 anos e 6 meses), 1090 dias (3 anos) e 1825 dias (equivalente ao máximo de 5 anos).

Para a análise foram selecionadas regras para cada tipo de tratamento cirúrgico e a janela temporal de 730 dias (equivalente a 2 anos), intervalo máximo recomendado para realização de mamografias em mulheres a partir de 40 anos (BARROS et al., 2011). A seleção das regras para análise, segue os seguintes critérios. Entre as regras envolvendo MRM, MCB, QRS e MS foi selecionada aquela contemplando a maior frequência absoluta de ocorrências da regra. Considerando MSIP e QLA foi selecionada as regras que associaram juntamente MCB e MDB como antecedentes a MSIP ou QLA.

### 3. RESULTADOS E DISCUSSÃO

A leitura de R1, também como modelo para as seguintes regras, segue o padrão onde: em *A* é representado as medias de suporte e confiança geradas pelo algoritmo Apriori (BORGELT, C.) *B* representando o suporte e a confiança calculados em função das pacientes com ocorrência do MCB antes (antecedente) do MRM (consequente), dentro do intervalo de tempo estabelecido; *C* representando o suporte e confiança em função de todas as ocorrências nas quais o MCB ocorreu antes do MRM, dentro do intervalo de tempo estabelecido; e *D* que representa a frequência absoluta de ocorrências desta regra. Vale destacar que o item *D* foi adotado para o critério de seleção de regras envolvendo alguns tipos de tratamentos cirúrgicos.

*R1: MCB → MRM*

A - (288.9%, 19.2%)

B - (55.6%, 100.0%)

C - (183.3%, 54,5%)

D - (10.0)

Em R1, as medidas em *B* sinalizam que 55.6% das pacientes realizaram pelo menos um MCB antes do MRM, destas todas (100%) realizaram ambos, pelo menos uma vez dentro da janela de tempo especificada. Em relação as medidas em *C* cada uma das pacientes realizou quase duas (183.3% ) MCB antes da MRM, porém 54.5% destes eventos associados ocorreram dentro do período estabelecido.

*R2: MCB → MS*

A - (288.9%, 13.5%)

B - (38.9%, 100.0%)

C - (161.1%, 44.8%)

D - (7.0)

Para a descrição das proximas regras não será mais evidenciada a janela de tempo de 730 dias, por ser uma constante para todas. Em R2, as medidas em *B* sinalizam que 38.9% das pacientes realizaram pelo menos uma MCB antes da MS, destas todas (100%) realizaram ambas pelo menos uma vez dentro da janela especificada. Em relação as medidas em *C* cada uma das pacientes realizou em média uma vez e meia (161.1%) a MCB antes da MS, e apenas 44.8% destas dentro do período estabelecido.

*R3: MCB MDB → MSIP*

A - (22.6%, 14.3%)

B - (3.2%, 100.0%)

C - (19.4%, 50.0%)

D - (1.0)

Em R3, as medidas em *B* sinalizam que 3.2% das pacientes realizaram os exames MCB e MDB associados, antes da realização da MSIP. Destas todas (100%) apresentam a MCB e MDB antes da MSIP, pelo menos uma vez dentro da janela especificada. Em relação as medidas em *C*, em média duas em cada dez pacientes (19.4%) realizaram os exames MCB e MDB antes da MS, sendo que a metade destas (50.0%) ocorreram dentro do período estabelecido.

*R4: MCB MDB → QLA*

A - (22.6%, 14.3%)

B - (3.2%, 100.0%)

C - (12.9%, 100.0%)

D - (1.0)

Em R4 as medidas em *B* sinalizam que 3.2% das pacientes realizaram MCB e MDB associadas antes da QLA. Destas todas (100%) apresentaram MCB e MDB associadas a QLA, pelo menos uma vez, dentro da janela especificada. Em relação as medidas em *C*, em média uma em dez pacientes (12.9%) realizou MCB e MDB antes da QLA, sendo que somente 42.3%, destes eventos ocorreram dentro do período estabelecido.

*R5: MCB → QRS*

A - (387.1%, 15.8%)

B - (61.3%, 94.7%)

C - (251.6%, 42.3%)

D - (19.0)

Em R5 as medidas em *B* sinalizam que 61.3% das pacientes realizaram MCB antes da QRS. Destas 94.7% apresentaram a MCB antes da QRS, pelo menos uma vez dentro da janela especificada. Em relação as medidas em *C* cada uma das pacientes realizou em media 2,5 (251.6%) a MCB antes da QRS, sendo que 42.3% destas associações ocorreram dentro do período estabelecido.

Esta pesquisa constituiu-se de um estudo de caso, através da descoberta de regras de associação considerando a janela temporal no histórico de rotinas de exames mamográficos associados a tratamentos cirúrgicos para o câncer de mama em mulheres beneficiárias de uma operadora de plano de saúde em determinado período de tempo, onde se pôde comparar os procedimentos cirúrgicos mais realizados dentro do período.

Quando avaliadas as regras apresentadas, especialmente a medida *C*, que representa o suporte e confiança em função de todas as ocorrências nas quais MCB ou MDB (técnicas distintas de mamografias) antecede o procedimento cirúrgico, percebe-se que dentro do período de 730 dias (equivalente a dois anos) houve 54,5% de ocorrência para evolução de MRM, 44,8% para MS ambos tratamentos cirúrgicos Mutiladores; e 50.0% para MSIP; 100% para QLA e 42.3% para QRS (classificados como tratamentos cirúrgicos Conservadores. O resultado distingue um percentual relativamente alto, acima de 50.0% de mulheres submetidas aos tratamentos conservadores de MSIP e 100% para QLA, e pouco abaixo de 50%, com 42.3% para QRS, no período avaliado de 730 dias, o que pressupõe que a decisão pela técnica cirúrgica esteja relacionado ao diagnóstico precoce. Porém, considera-se relativamente alto a incidência para o mesmo período por tratamentos cirúrgicos mutiladores de MS e principalmente para MRM, o tipo cirurgico mais agressivo, com 54,5% de ocorrências, deixando considerações preocupantes, quanto as rotinas mamográficas destas mulheres, bem como o tempo para diagnóstico e tratamento. Tal resultado incita o aprofundamento das técnicas de Mineração de Dados nestes bancos de dados para novas descobertas.

No que ainda diz respeito ao câncer de mama, no Brasil, tem prevalecido taxas de mortalidades associadas ao retardamento do diagnóstico (SILVA e HORTALE, 2012). O resultado apresentado pelas regras de associação não corrobora em partes com esta literatura, visto os percentuais de intervenções via tratamentos conservadores ou mutiladores em curto espaço de tempo, onde é verificado percentuais relevantes de rotinas mamográficas seguidas de procedimentos cirúrgicos como terapêutica da doença, num período de 730 dias (equivalente a 2 anos).

Essa análise, portanto, compreende um simples recorte sobre a possibilidade da utilização de regras de associação em bases de dados de pacientes com diagnósticos de câncer de mama. E, como primeira aplicação do algoritmo neste contexto, verifica necessidade de reavaliação dos atributos e a distribuição mais detalhada das janelas temporais, frente uma doença que independe por exemplo da idade tanto para a submissão a tratamentos conservadores e ou mutiladores, mas depende significativamente do tempo para descoberta, diagnóstico e tratamento.

A relação ao período de tempo, ou seja, o menor possível, favorece maiores intervenções via tratamentos cirúrgicos conservadores e limita o impacto psicológico e pessoal dos tratamentos cirúrgicos mutiladores, estes portanto, relacionadas a períodos maiores de tempo de diagnóstico.

Sendo assim a investigação das práticas clínicas e terapêuticas regressas em base de dados, com auxílio do processo KDD (BORGELT, C), pôde trazer informações relevantes quanto as práticas já adotadas e estrategicamente o favorecimentos de novas intervenções focadas no comportamento preventivo para o diagnóstico precoce em câncer de mama, melhorando então a qualidade assistencial prestada pelas operadoras de planos de saúde e consequentemente o equilíbrio psicológico e a qualidade de vida da mulher com diagnóstico de câncer de mama.

#### **4. CONCLUSÕES**

Este experimento demonstrou o potencial das regras de associação considerando a janela temporal sobre os eventos ocorridos, como contribuição para otimizar o uso das bases de dados de operadoras de planos de saúde. Bases estas que contém milhares de informações relacionadas às condutas clínicas do histórico de pacientes com cancer de mama, o comportamento destas beneficiárias frente ao serviço de saúde ofertado, além do registro das demandas motivadas por outras doenças. Fato que por si só já representa uma fonte importante para apoiar processos decisórios relacionados não apenas a gestão em saúde, qualidade assistencial, mas também à qualidade de vida dos beneficiários.

A descoberta de regras de associação sobre beneficiárias portadoras de câncer de mama, constitui mais uma alternativa para comparação do percurso das condutas regressas não apenas frente às diretrizes clínicas de prevenção secundária da doença, mas também a outros protocolos estabelecidos e relevantes no controle da doença.

De forma geral, a utilização do processo KDD traz subsídios que contribuem para uma medicina focada na prevenção e diagnóstico precoce, que no caso de cancer de mama permite considerar práticas de diagnóstico precoce, e consequentemente acesso a tratamentos conservadores.

#### **AGRADECIMENTOS**

Os autores agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudo e ao Marcelo Rosano Dalagassa pela sistematização dos dados.

## REFERÊNCIAS

- BARROS, A. C. S. D. et al.** *Diagnóstico e tratamento do câncer de mama. São Paulo: Associação Médica Brasileira/Brasília: Conselho Federal de Medicina, 2011.*
- DE SOUSA PINHO, Luana et al.** *Câncer de mama: da descoberta à recorrência da doença. Revista eletrônica de enfermagem, v. 9, n. 1, 2009.*
- BORGELT, C.** *Apriori. Disponível em: <<http://www.borgelt.net/apriori.html>> Acesso em: 15 de Dezembro de 2014.*
- CANADIAN TASK FORCE ON PREVENTIVE HEALTH CARE et al.** *Recommendations on screening for breast cancer in average-risk women aged 40–74 years. Canadian Medical Association Journal, v. 183, n. 17, p. 1991-2001, 2011.*
- CHAVES, F. P.R. et al.** *Neoplasias malignas da mama no brasil e sua variação espaço-temporal no período de 2009 a 2013. Blucher Medical Proceedings, v. 1, n. 3, p. 15-15, 2014.*
- CHIMIESKI, B.F. e FAGUNDES, R.D.R.** *Association and classification data mining algorithms comparison over medical datasets. Journal of Health Informatics, v. 5, n. 2, 2013.*
- DEL TURCO, M.R. et al.** *Quality indicators in breast cancer care. European journal of cancer, v. 46, n. 13, p. 2344-2356, 2010.*
- JBILLOU, J. et al.** *Medical genetic counseling for breast cancer in primary care: a synthesis of major determinants of physicians' practices in primary care settings. Public health genomics, v. 17, n. 4, p. 190-208, 2013.*
- JOHNSON-THOMPSON, M.C. e GUTHRIE, J.** *Ongoing research to identify environmental risk factors in breast carcinoma. Cancer, v. 88, n. S5, p. 1224-1229, 2000.*
- REZENDE, S.O.** *Sistemas inteligentes: fundamentos e aplicações. Editora Manole Ltda, 2005.*
- SILVA, R.C.F. e HORTALE, V.A.** *Rastreamento do câncer de mama no Brasil: quem, como e por quê?. Rev. bras. cancerol, v. 58, n. 1, p. 67-71, 2012.*
- SOKOLOSKI, W.F.; CARVALHO, D.R. e DALLAGASSA, M.R.** *Regra de Associação Temporal. XIV Congresso Brasileiro de Informática em Saúde 2014: Santos, Brasil.*
- THULER, L. C. S. e MENDONÇA, G.A.** *Estadiamento inicial dos casos de câncer de mama e colo do útero em mulheres brasileiras. Rev bras ginecol obstet, v. 27, n. 11, p. 656-60, 2005.*
- WECHSLER, R. et al.** *A informática no consultório médico. J Pediatr (Rio J), v. 79, n. Supl 1, p. 3-12, 2003.*