TEMPORAL ASSOCIATION RULES IN BREAST CANCER

Rosyellen Rabelo Szvarça, Sergio Ossamu Ioshii, Deborah Ribeiro Carvalho, Willian Felipe Sokoloski
Pontifícia Universidade Católica do Paraná (PUC-PR)
E-mails: rosyellenpsico@hotmail.com; drdrcarvalho@gmail.com

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 beneficiaires 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 to the patients.

Keywords: Breast Cancer; Association Rules; Healthcare; Prevention; Mammography.

1. INTRODUCTION

The development of breast cancer depends on many factors and environmental events determined throughout the woman's life, as well as by genetic susceptibility. For Johnson-Thompson and Guthrie (2008), the interaction of these factors has been investigated, among others, by the National Institute of Environmental Health Sciences (NIEHS) and it has been found that the risk factors identified for breast cancer include the presence of a first-degree relative with history of the problem, age of the first menstrual period, age at menopause period and environmental risks such as the use of pesticides and synthetic chemicals; endogenous and/or exogenous hormone factors; eating habits, use of tobacco and alcohol and excessive exposure to radiation and magnetic waves.

The characteristics and stages of this neoplasia are defined by tumor size and histological degree, and the available treatments are surgery, an invasive procedure, subdivided into conservative surgery where there is only excision of the tumor and possible margins, and non-conservative or mutilating surgery, which may involve removal of the mammary gland only (subcutaneous mastectomy), removal of breast tissue along with the overlying skin and nipple-areolar complex (simple or total mastectomy), excision of breast tissue with preservation of one or two pectoral muscles (modified radical mastectomy) and removal of pectoral muscles (radical mastectomy) (GEBRIM, et al, 2011). The indication of conservative surgeries is related to early diagnosis and screening actions by means of mammography (MMG).

Of the most common types of screening for breast cancer, two strategies deserve mention, namely early diagnosis, through awareness of early signs and symptoms and medical consultations; Mammography characterized by screening tests of non-symptomatic

individuals to identify the early stages of the disease, improving the chances of successful treatment.
It is known that both the breast and cervical cancers have a better prognosis when treated early. However, they have high rates of mortality when diagnosed at an advanced stage according Thuler and Mendonça (2005).

Among the imaging techniques for screening used, MMG is considered standard due to its potential to detect calcifications common in breast cancer (Silva, Correa and Hortale, 2005). Likewise, the European Society of Breast Cancer (Rosselli Del Turco et al., 2010) is responsible for assessing the clinical performance in European breast cancer treatment, where dedicated units has standardized a guide of Quality Breast Cancer Performance Indicators, with emphasis to MMG as a very reliable and accurate digital imaging technique with a higher level evidence (III).

In Brazil, clinical guidelines on supplementary health services, as well as the Brazilian Medical Association (AMB) and the National Agency for Supplementary Health Services regarding the prevention of breast cancer, recommend mammographic screening at regular intervals of 1 to 2 years for women over 40 years of age, for reducing mortality from this disease, with a more significant reduction observed after 50 years of age (GEBRIM, et al, 2011).

Being a leading cause of death worldwide and considered the most frequent cancer in women (breast cancer has an incidence rate of 23% compared to all other types of cancer), the standardization of local programs of mammographic screening is justified.(JBILOU et al, 2013). The protocols of these mammographic screening programs in different countries established the frequency and intervals for conducting MMG, according to the age range. Prepared by local task force teams to control breast cancer, these protocols aimed at the standardization of these guides for control of prevention and care quality indicators, which provide the foundations for the decisions regarding the most appropriate treatments for this neoplasia, for example, the canadian protocol (CANADIAN TASK FORCE ON PREVENTIVE HEALTH CARE, 2011).

Health professionals must adhere to the requirements and protocols of medical guidelines in their decision making regarding the specific treatments for each individual, seeking to reduce the use of mutilating invasive interventions and increase the use of conservative treatments. Also, with the growing number of patients diagnosed with breast cancer who seek health services, there is more information available on the disease and on the profiles and medical charts of individuals treated for this neoplasia. Such information organized in databases assist health managers in their programs of prevention, treatments, cure, staging, financial control and, particularly, increasing the quality of the care delivered.

The bibliographic study from Wechsler, Anção, Campos and Sigulem (2003) focus on the standardization of the Electronic Patient Record (EPR) and discusses the use of information technology in the organization of medical appointments, collection and storage of data that associated to the integration of decision support systems collaborate with the diagnosis and treatment of the disease.

Another study from Chimieski and Fagundes (2013) reports on the medical diagnosis of diseases such as myocardial infarction, breast and cervical cancer and on the control of hospital infection suggesting the use of machine learning techniques such as Neural

Networks, in the Support to Decision Making of people and groups involved in clinical management.

The observation and analysis of the history of women who underwent surgical treatment for breast cancer, considering the association of events occurred, as well as the cause and effect relationship between them can support the standardization of conducts regarding early diagnosis, the screening routine and the decision of the most appropriate surgical intervention. A technological alternative that allows this analysis is the Knowledge Discovery in Database (KDD) process. The KDD comprises three main steps: preprocessing, data mining and post-processing (Rezende, 2005).

The purpose of this article is to present a case study that discovers associations, considering the temporal window, on the routine of mammographic screening, its influence on the diagnosis that determined conservative or mutilating treatments.

4. METHODOLOGY

A secondary database of a health insurance provider was used in the study, which contained data from 50 women aged 36-81 years. The selection criterion was the period of 1994 to 2013, and based on this criterion, the first 50 women who underwent at least one surgical treatment for breast cancer and who have been beneficiaries of the health insurance provider for at least 5 years before the referred surgery were selected.

Data related to screening tests, especially MMG and surgical treatment and the dates on which they occurred was considered. The mammographies were classified as digital and conventional MMG. The surgical treatments were grouped into conservative or mutilating. The conservative treatments were quadrantectomy subdivided into axillary lymphadenectomy (QLA) and dissection lymphadenectomy (QRS), as well as subcutaneous mastectomy and prosthesis insertion (MSIP). The mutilating treatments considered were radical or modified radical mastectomy (MRM) and simple mastectomy (MS).

For the data mining step the APRIORI algorithm was adopted (Borgelt, 2013) and for post-processing the ASSOCTEMP (Sokoloski, Carvalho and Dallagassa, 2014) was adopted. Considering the purpose od this study, that is, the search for associations between the events related to breast cancer, the task Discovery of Association Rules was selected. Based on the association rules discovered, it was important to infer the possibility of cause and effect between the associated events. This has been possible after the identification of the temporal window between the occurrences associated by the algorithm (Sokoloski, Carvalho and Dallagassa, 2014).

The following temporal windows were adopted for the experiments: 365 days equivalent to (1 year), 540 days (1 year and 6 months), 730 days (2 years), 910 days (2 years and 6 months), 1090 days (3 years) and 1825 days (equivalent to a maximum of 5 years).

For the analysis, rules were selected for each type of surgical treatment, and the temporal window of 730 days (equivalent to 2 years), maximum interval recommended for mammogrophies in women over 40 years, according Gebrim et al (2011). The selection of rules for the analysis met the following criteria.: the rules were selected by frequency of occurrence with the categories MRM, MCB, QRS and MS. Considering MSIP and QLA, the rules selected were those that jointly associated MCB and MDB as antecedents to MSIP or QLA.

5. RESULTS

The reading of R1, also as a model for the following rules, follows the pattern in which: *A* represents the measures of support and confidence generated by the Apriori algorithmi (Borgelt, 2013) *B* represents the support and confidence whose calculations are associated to those of the patients with occurrence of MCB before (antecedent) MRM (consequent), within the time interval established entro do intervalo de tempo estabelecido; *C* representing the support and confidence associated to the occurrences in which MCB occurred before MRM, within the time interval established, and *D* representing the absolute frequency of occurrences of this rule. It should be stressed that item D was adopted for the selection criteria of rules involving some types of surgical treatments.

R1: MCB → MRM
A - (288.9%, 19.2%)
B - (55.6%, 100.0%)
C - (183.3%, 54,5%)
D - (10.0)
In R1, the measures in *B* indicate that 55.6% of the patients underwent at least one MCB before MRM, and all (100%) of these patients underwent both procedures at least once within the specified temporal window. Regarding the measures in *C* each one of the patients underwent almost two (183.3%) MCB before MRM. However, 54.5% of these associated events occurred within the established time period.

R2: MCB→ MS
A - (288.9%, 13.5%)
B - (38.9%, 100.0%)
C - (161.1%, 44.8%)
D - (7.0)
The temporal rule of 730 days will not be evidenced for the description of the subsequent rules because it is a constant for all measures. In R2, the measures in *B* indicate that 38.9% of the patients underwent at least one MCB before MS .All (100%) of these patients underwent all measures at least once in the specified window. Regardignt the measures in *C* one of the patients underwent in average one and a half time (161.1%) MCB before MS, and only 44.8% of these tests occurred within the established period.

R3: MCB MDB→ MSIP
A - (22.6%, 14.3%)
B - (3.2%, 100.0%)
C - (19.4%, 50.0%)
D - (1.0)
In R3, the measures in *B* indicate that 3.2% of the patients underwent MCB and MDB associated, before MSIP. All (100%) of these patients underwent MCB and MDB before MSIP, at least onece within the specified window. Regarding the measures in *C*, in average two of each ten patients (19.4%) underwent MCB and MDB before SM, with half of these tests (50.0%) occurring within the established period.

R4: MCB MDB→ QLA
A - (22.6%, 14.3%)
B - (3.2%, 100.0%)

C - (12.9%, 100.0%)
D - (1.0)
In R4 the measures in *B* indicate that 3.2% of the patients underwent MCB and MDB associated before QLA. All (100%) of these patients underwent MCB and MDB associated to QLA at least once within the specified window. Regarding the measures in *C*, in aaverage one of each ten patients (12.9%) underwent MCB and MDB before QLA, with only 42.3% of these events (tests) occurring within the established period.

R5: MCB→ QRS
A - (387.1%, 15.8%)
B - (61.3%, 94.7%)
C - (251.6%, 42.3%)
D - (19.0)
In R5 the measures in *B* indicate that 61.3% of the patients underwent MCB before QRS. Of these, 94.7% underwent MCB before QRS, at least once within the specified window. Regarding the measures in *C* each one of the patients underwent in average 2.5 (251.6%) MCB before QRS, with 42.3% of these associations occurring within the established period.

6. DISCUSSION

This research consists in a case study through the discovery of association rules considering the temporal window in the history of routine mammographies associated to surgial treatments of breast cancer in women who were beneficiaries of a health insurance provider within a given time period, and the most frequent surgical procedures performed in this time period were compared.

When the rules discovered were assessed, especially measure C that represents the support and confidence regarding the occurrences in which MCB or MDB (different mammographic techniques) precede the surgical procedure, it can be seen that within the period of 730 days (equivalent to two years) there was na incidence rate of 54.5% for evolution of MRM, 44.8% for MS (both being mutilating surgical treatments) and 50.0% for MSIP; 100% for QLA and 42.3% for QRS (classified as conservative surgical treatments). The result expresses a relatively high percentage, above 50.0% of women who underwent conservative treatments of MSIP and 100% for QLA, and sligthly below 50%, with 42.3% for QRS, in the assessed period of 730 days, which implies that the choice of the surgical treatment is related to early diagnosis.

However, the incidence rate for the same period of mutilating surgical treatments of MS and particularly MRM, the most aggressive surgery, are considered relatively high with 54.5% of the occurrences, which gives rise to concern with the mammographic screening routines of these women, as well as with the time to diagnosis and treatment. So, Data Mining techniques of these databases should be further explored for new discoveries.

Still regarding breast cancer, in Brazil, according Thuler and Mendonça (2005), the mortality rates have been associated to late diagnosis. The result presented by the association rules does not corroborate in part with this literature, considering percentages of interventions with conservative or mutilating treatments in a short period of time, where there are significant rates of mammographies followed by surgical treatments within a period of 730 days (equivalent to two years).

This analyis is merely a view of the possibility of use of association rules in databases of patients diagnosed with breast cancer. And since this is the first application of the algorithm in this context, these attributes need to be reassessed in the future, and a more detailed distribution of the temporal windows is required, in the case of breast cancer, the choice of conservative or mutilating treatments does not depend on the age of the patients, but is significantly associated to the time of discovery, diagnosis and treatment.

The relationship with the time period, that is, the least possible time, favors conservative surgical treatments and limits the psychological and personal impact of mutilating surgical treatments, which are related to longer time to diagnosis.

Thus, the investigation of clinical and therapeutic practices in databases, using the KDD process (Borgelt, 2013) can provide significant information on the practices adopted and on the strategic importance of new interventions focused on prevention , with early diagnosis in breast cancer, improving the quality of care delivered by health insurance providers and, thus, the quality of life of women diagnosed with breast cancer.

5. CONCLUSIONS

This experiment demonstrated the potential of the association rules, considering the temporal window, on the events described, as a contribution to optimize the use of databases of health insurance providers. Such databases include thousands of pieces of information related to clinical practices from the medical records of patients with breast cancer, the the attitudes/views of these beneficiaries regarding the health care delivered by insurance providers, as well as the record of demands motivated by other diseases. This constitutes a valuable source to support decision making processes related not only to health management and care quality, but also to the quality of life of the beneficiaires.

The discovery of association rules on patients diagnosed with breast cancer who are beneficiaires of health insurance providers allows comparing former guidelines not only to the medical guidelines for secondary prevention of the disease, but also to other relevant protocols in the control of the disease.

In general, the use of the KDD process contributes to a medical practice focused on prevention and early diagnosis, and in the case of breast cancer, early diagnosis practices can be considered, with the encouragement of conservative treatments.

REFERENCES

BORGELT C. **Apriori.** Disponível in: <http://www.borgelt.net/apriori.html> Acess em: 15 de Dezembro de 2013.

CANADIAN TASK FORCE ON PREVENTIVE HEALTH CARE. **Recommendations on screening for breast cancer in average-risk women aged 40–74 years.** *Canadian Medical Association Journal* 2011: 183 (n. 17): 1991-2001

CHIMIESKI B. F.; FAGUNDES R. D. R. **Association and classification data mining algorithms comparison over medical datasets**. *Journal of Health Informatics* 2013: 5 (n. 2).

GEBRIM L. H. Et al. **Câncer de mama: Prevenção secundária.** Associação Médica Brasileira e Agência Nacional de Saúde Suplementar, 2011.

JBILOU 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* 2013:17 (n. 4): 190-208.

JOHNSON-THOMPSON M. C.; GUTHRIE J. **Ongoing research to identify environmental risk factors in breast carcinoma**. *Cancer* 2000: 88 (S5):1224-1229.

REZENDE O.S. **Sistemas Inteligentes: Fundamentos e Aplicações**. São Paulo: Manole; 2005.

ROSSELLI DEL TURCO M. Et al. **Quality indicators in breast cancer care**. *European journal of cancer* 2010: 46 (n. 13): 2344-2356.

SILVA, R. C. F. D. **Rastreamento do câncer de mama no Brasil: quem, como e por quê?** *Rev. bras. Cancerol* 2005: 58 (n. 1): 67-71.

SOKOLOSKI, F. W.; CARVALHO, R. D.; DALLAGASSA, R. M. **Regra de Associação Temporal.** XIV Congresso Brasileiro de Informática em Saúde 2014: Santos, Brasil.

THULER, L. C. S.; MENDONÇA, G. A. **Estadiamento inicial dos casos de câncer de mama e colo do útero em mulheres brasileiras**. *Rev bras ginecol obstet* 2005: 27 (n.11).

WECHSLER, R.; ANÇÃO, M.S.; CAMPOS, C.J.R.D.; SIGULEM, D. **A informática no consultório médico.** Rio de Janeiro: *J Pediatr* 2003: 79 (Supl 1): 3-12.