Iran Red Crescent Med J. 2016 March; 18(3): e23131. doi: 10.5812/ircmj.23131

Published online 2016 March 01. **Research Article**

Application of Canonical Correlation Analysis for Detecting Risk Factors Leading to Recurrence of Breast Cancer

Farahnaz Sadoughi,[1] Hadi Lotfnezhad Afshar,[1,2,*] Asiie Olfatbakhsh,[3] and Neda Mehrdad[3]

[1]Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, IR Iran
[2]Department of Health Information Technology, School of Paramedical, Urmia University of Medical Sciences, Urmia, IR Iran
[3]Breast Cancer Research Center (BCRC), The Academic Center for Education, Culture and Research (ACECR), Tehran, IR Iran

*Corresponding Author: Hadi Lotfnezhad Afshar, Department of Health Information Technology, School of Paramedical, Urmia University of Medical Sciences, Urmia, IR Iran. Tel: +98-44332752300, Fax: +98-4432770047, E-mail: hadi.afshar@gmail.com

Received 2014 August 29; **Revised** 2014 October 14; **Accepted** 2014 October 22.

Abstract

Background: Advances in treatment options of breast cancer and development of cancer research centers have necessitated the collection of many variables about breast cancer patients. Detection of important variables as predictors and outcomes among them, without applying an appropriate statistical method is a very challenging task. Because of recurrent nature of breast cancer occurring in different time intervals, there are usually more than one variable in the outcome set. For the prevention of this problem that causes multicollinearity, a statistical method named canonical correlation analysis (CCA) is a good solution.

Objectives: The purpose of this study was to analyze the data related to breast cancer recurrence of Iranian females using the CCA method to determine important risk factors.

Patients and Methods: In this cross-sectional study, data of 584 female patients (mean age of 45.9 years) referred to Breast Cancer Research Center (Tehran, Iran) were analyzed anonymously. SPSS and NORM softwares (2.03) were used for data transformation, running and interpretation of CCA and replacing missing values, respectively. Data were obtained from Breast Cancer Research Center, Tehran, Iran.

Results: Analysis showed seven important predictors resulting in breast cancer recurrence in different time periods. Family history and loco-regional recurrence more than 5 years after diagnosis were the most important variables among predictors and outcomes sets, respectively.

Conclusions: Canonical correlation analysis can be used as a useful tool for management and preparing of medical data for discovering of knowledge hidden in them.

Keywords: Breast Neoplasms, Neoplasm Recurrence, Data Mining, Statistics as Topic

1. Background

Breast cancer is the most common type of diagnosed and fatal cancers in females of the most areas of the world, especially in Iran (1, 2). Iran is located in the western part of Asia where breast cancer in women is the first leading cause of death (3). In comparison with developed countries, breast cancer is diagnosed nearly one decade sooner in Iran and ages of incidence are often in the range of 40 to 49 years (2, 4, 5).

Identification of important predictors that prognosis (chance of recovery) of breast cancer depends on them, is a challenging task. There are many prognosis variables in paper and computerized medical records of breast cancer patients. Extraction of the most important predictors among them, regarding outcome variable(s), without using a proper statistical technique may be difficult (5, 6).

A common statistical technique used to find relationships between predictors and outcome is multiple regression analysis (MRA). This technique is suitable when the outcome side has one variable, but applying of it in the scenarios of more than one variable leads to wrong results (7). Another important issue that must be concerned in multiple regression models is multicollinearity. Existence of the strong correlations between predictors is a major cause of it. As multicollinearity increases, it will be difficult to assess the importance of individual predictors (8).

In MRA, identification of important predictors is based on their beta weights. Since beta weights are affected by multicollinearity, using an alternative approach ignoring multicollinearity is necessary. Canonical correlation analysis (CCA) developed by Hotelling (1936) is an approach applying structure coefficients as indices for selecting important predictors. Contrary to beta weights, structure coefficients reflect the direct contribution of one predictor to the outcome variable, regardless of the multicollinearity (9).

Because breast cancer recurs at any time -mainly during the first five years- after the primary treatment,

time is considered as an essential factor in the analysis of breast cancer recurrence (6). The Cox regression method is a traditional statistical one that is suitable for handling events such as cancer recurrence happening during different times (10). However, since cancer recurrences have different outcomes, the CCA method is preferable.

2. Objectives

The purpose of this study was to identify the most important risk factors leading to breast cancer recurrence during different time intervals by using the CCA method as a new technique in the clinical domain. The CCA method was applied to data set of breast cancer patients in the capital of Iran (Tehran).

3. Patients and Methods

3.1. Data Source

In this cross-sectional study, data of 584 female patients (mean age, 45.9 years) were analyzed anonymously. Data were obtained from Breast Cancer Research Center (BCRC) in Tehran, Iran. This center, located in capital of Iran, has multiple clinics related to breast cancer therapy and research. After consulting with oncologists of the research center and studying the literature in the domain (7, 11, 12), two sets of predictors and outcomes (see Box 1.) were selected.

3.2. Data Preprocessing

We obtained the data of this study in the format of Excel file from the BCRC in June 2012 anonymously; so, any special ethical certification was not needed. The original file contained demographic and clinical information of 843 breast cancer patients diagnosed in the BCRC. In the current study, only female patients that had been followed-up 5 years after diagnosis were selected as study samples. As mentioned criteria, 4 male patients and 255 ones that had a follow-up period less than 5 years (60 months) were excluded from the study. As a rule of thumb for minimum sample size calculation in the multivariate techniques that CCA is a subcategory of them, at least 30 observations for each variable are often used (13) that our study has met this criterion ($584 > 570$).

The raw data were transformed and converted as illustrated in Table 1. Two variables such as LN positive and LN removed were dichotomized based on their presence or absence in the patients. Other variables, except age, were already categorized. The variable of age, continuous before, was transformed to the dichotomous one.

Although, canonical correlation analysis can accommodate any variable without the strict assumption of normality (14), the normality tests were done and non-normal variables (surgery and pathology of tumor) fixed by the log transformation.

Variables with more than 50% of missing values were removed from dataset. Based on this criterion, three variables, age of menopause, tumor margin and Her2 were deleted. Missing values of other variables with less than 50% were substituted using the expectation maximization (EM) algorithm (15). The EM algorithm is an efficient repetitive procedure to compute the maximum likelihood (ML) estimate in the presence of missing or hidden data. In ML, estimation of the model parameter(s) for which the observed data are the most likely is performed. The E-step and the M-step are two processes that combine iterations of the EM algorithm. The E-step estimates the missing data through the observed data and current estimate of the model parameters and the M-step maximizes the likelihood function under the assumption that the missing data are known (16).

3.3. Canonical Correlation Analysis

Since two sets of variables existed and outcome set included more than one variable, the CCA method was performed. Figure 1 illustrates the fundamental principle behind CCA. Simply, the variable relationships in a hypothetical CCA with three predictor and two outcome variables have been shown. The combination of the observed variables in both sets into one unobserved variable is essential for evaluating the simultaneous relationship between several predictor and several outcome variables. In CCA, a linear equation is applied separately to the observed predictor and dependent variables to create one unobserved variable for each set. The reason these two equations are generated is that they yield the largest possible correlation between the two unobserved variables. The canonical correlation between the two unobserved variables is the most basic statistic in a CCA and it almost is a Pearson r (see Figure 1). Maximization of this simple correlation is the main purpose of the CCA (17).

Furthermore, in a CCA, the number of canonical functions is equal to the number of variables in the smaller set (e.g., two functions for the example in Figure 1). The first canonical correlation is the highest possible correlation between any synthetic predictor variable and synthetic outcome variable and is the most proper candidate for interpretation. The criterion for choosing the important variables in each canonical function is the structure coefficients, the bivariate correlation between an observed variable and a synthetic variable. As a rule of thumb for meaningful structure coefficients, an absolute value equal to or greater than 0.45 is often used (17).

SPSS version 16 for Windows (Chicago, SPSS Inc.; USA)

and NORM software (2.03) (18) were used for data transformation, running and interpretation of CCA and replacing missing values, respectively.

4. Results

A canonical correlation analysis was conducted using the thirteen attachment variables as predictors of the 6 outcome variables to evaluate the multivariate shared relationship between the two variable sets. The analysis yielded six functions with squared canonical correlations (R^2_c) of 0.48, 0.18, 0.1, 0.09, 0.04, 0.02 for each successive function. Collectively, the full model across all functions was statistically significant using the Wilks's $\lambda = 0.32$ criterion, $F(192, 3235.78) = 3.57$, $P < 0.001$. Because Wilks's λ represents the variance unexplained by the model, $1 - \lambda$ yields the full model effect size in a r^2 metric. Thus, for the set of six canonical functions, the r^2 type effect size was 0.68 indicating that the full model explained a substantial portion, about 68%, of the variance shared between the variable sets.

The dimension reduction analysis allows the researcher to test the hierarchal arrangement of functions for statistical significance. As noted, the full model (Functions 1 to 6) was statistically significant. Functions 2 to 6 and 3 to 6 were also statistically significant, $F(155, 2711.9) = 1.75$, $P < 0.001$. $F(120, 2181.04) = 1.26$, $P < 0.05$. Functions 4, 5 and 6, with $F(87, 1643.55) = 1.06$, $P = 0.329$, $F(56, 1100) = 0.67$, $P = 0.968$ and $F(27, 551) = 0.44$, $P = 0.994$ respectively, did not explain a statistically significant amount of shared variance between the variable sets.

Given the R^2_c effects for each function, the first function was considered noteworthy in the context of this study (48% of shared variance). The last five functions only explained 18%, 9.9%, 9.1%, 4.5% and 2.1%, respectively from the remaining variance in the variable sets after the extraction of the prior functions.

Table 2 presents the standardized canonical function coefficients and structure coefficients for Function 1. The squared structure coefficients are also given for each variable. Looking at the coefficients of the outcome set, one sees that important variables were LRR, more than 5 years and LRR, 3 - 5 years, respectively. This conclusion was supported by the squared structure coefficients. These variables also tended to have the larger canonical function coefficients. All of these variables' structure coefficients had the negative sign, indicating that they were all negatively related.

Regarding the predictor variable set in Function 1, family history, estrogen receptor, pathology of tumor (LCIS), type of surgery (bilateral BCS), tumor size (> 2), pathology of tumor (IDC), and hormone therapy (combined) had the highest coefficients, respectively. Structure coefficients of all of them were negative, except for variables of estrogen receptor and pathology of tumor (IDC) that had positive sign.

Box 1. List of Variables in Both Sets

Variables
Predictor Set
Age
Family history
Tumor size
Number of involved LN
LN positive
Number of removed LN
Pathology of tumor
Type of surgery
Tumor grade
Estrogen receptor
Progesterone receptor
Radiotherapy
Hormone therapy
Outcome Set [a]
DM, first three years
DM, 3 - 5 years
DM, more than 5 years
LRR, first three years
LRR, 3 - 5 years
LRR, more than 5 years

Abbreviations: DM, distant metastasis; LN, lymph node; LRR, locoregional recurrence.
[a]All periods are time after diagnosis.

Table 1. Transformation Rules and the Study Population Characteristics

Variables	Coding	Values [a]
Age, y		
> 50	0	219 (37.5)
≤ 50	1	365 (62.5)
Family history		
No	0	485 (83)
First degree	1	99 (17)
Tumor size, cm		
(not) < 2	(0) 1	84 (14.4)
(not) 2 - 5	(0) 1	268 (45.9)
(not) > 5	(0) 1	232 (39.7)

Number of involved LN		
(not) Nothing	(0) 1	231 (39.6)
(not) 1 - 3	(0) 1	187 (32)
(not) 3 - 9	(0) 1	112 (19.2)
(not) > 9	(0) 1	54 (9.2)
LN positive		
No	0	174 (29.8)
Yes	1	410 (70.2)
Number of removed LN		
Zero	0	29 (5)
One or more	1	381 (95)
Pathology of tumor		
LCIS	(0) 1	47 (8)
DCIS	(0) 1	72 (12.3)
IDC	(0) 1	272 (46.6)
ILC	(0) 1	95 (16.3)
Medullary	(0) 1	53 (9.1)
Micro invasive	(0) 1	24 (4.1)
Paget disease	(0) 1	4 (0.7)
Inflammatory	(0) 1	1 (0.2)
Other	(0) 1	16 (2.7)
Type of surgery		
MRM	(0) 1	399 (68.3)
BCS	(0) 1	135 (23.1)
Bilateral MRM	(0) 1	25 (4.3)
Bilateral BCS	(0) 1	23 (3.9)
MRM + BCS	(0) 1	1 (0.2)
Combined	(0) 1	1 (0.2)
Tumor grade		
First grade	1	114 (19.5)
Second grade	2	319 (54.6)
Third grade	3	151 (25.9)
Estrogen receptor		
Negative	0	241 (41.3)
Positive	1	343 (58.7)
Progesterone receptor		
Negative	0	242 (41.4)
Positive	1	342 (58.6)
Radiotherapy		
No	0	220 (37.7)
Yes	1	364 (62.3)
Hormone therapy		
No	(0) 1	41 (7)
Tamoxifen	(0) 1	173 (29.6)
Raloxifene	(0) 1	21 (3.6)
Letrozole	(0) 1	46 (7.9)
Aromasin	(0) 1	18 (3.1)
Megace	(0) 1	16
Combined	(0) 1	269

Abbreviations: BCS, breast conserving surgery; DCIS, ductal carcinoma in situ; IDC, invasive ductal carcinoma; ILC, invasive lobular carcinoma; LCIS, lobular carcinoma in situ; MRM, modified radical mastectomy; P, preservation.
[a]Values are presented as No. (%).

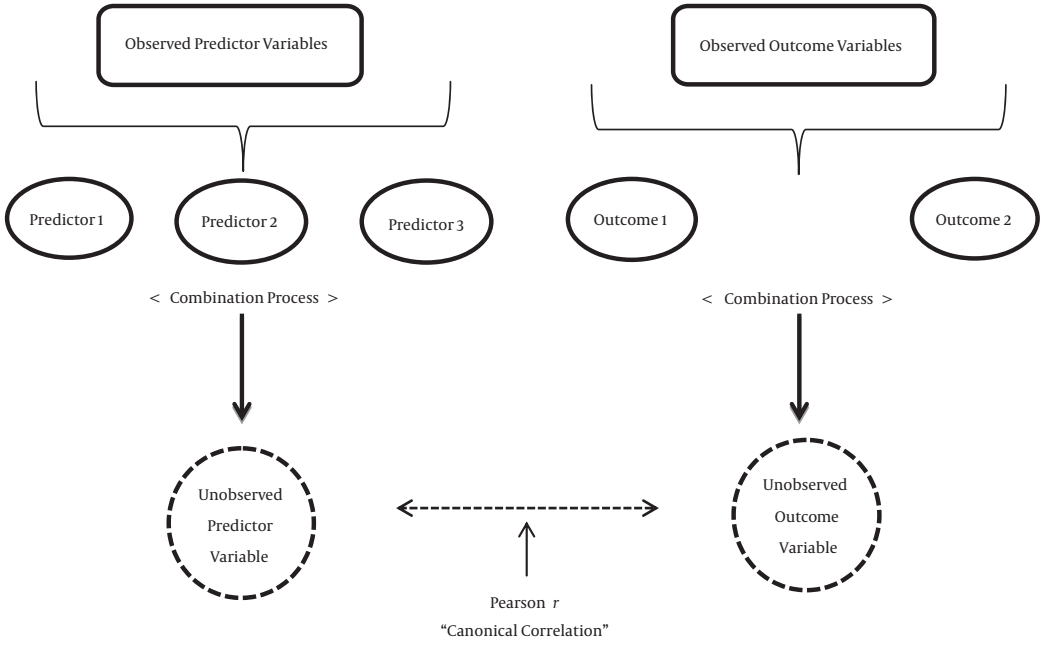

Figure 1. Illustration of the First Function in a Canonical Correlation Analysis With Three Predictors and Two Criterion Variables

Table 2. Canonical Solution for Function 1

Variables	Coef	R_3	R^2_s, %
Age	-0.015	-0.163	2.65
Family history	-0.436	-0.795[a]	63.2
Tumor size (< 2)	0.007	0.126	1.58
Tumor size (2 - 5)	0.085	0.414	17.13
Tumor size (> 2)	0	-0.512[a]	26.21
Number of involved LN (Nothing)	0.139	0.293	8.58
Number of involved LN (1 - 3)	-0.021	-0.081	0.65
Number of involved LN (3 - 9)	0.066	0.04	0.16
Number of involved LN (> 9)	0	-0.419	17.55
LN positive	0.061	-0.307	9.42
Number of removed LN	-.021	-0.155	2.4
Pathology of tumor (LCIS)	-.408	-0.562[a]	31.58
Pathology of tumor (DCIS)	-0.259	0.122	1.48
Pathology of tumor (IDC)	-0.373	0.492[a]	24.2
Pathology of tumor (ILC)	-0.545	-0.302	9.12
Pathology of tumor (Medullary)	-0.332	-0.041	0.16
Pathology of tumor (Micro invasive)	-0.191	-0.045	0.2
Pathology of tumor (Paget disease)	-0.153	-0.091	0.82
Pathology of tumor (Inflammatory)	-0.164	0.04	0.16
Pathology of tumor (Other)	0	0.019	0.03
Type of surgery (MRM)	1.569	0.285	8.12
Type of surgery (BCS)	1.317	0.054	0.29
Type of surgery (Bilateral MRM)	0.881	-0.212	4.49
Type of surgery (Bilateral BCS)	0.491	-0.558[a]	31.13
Type of surgery (MRM + BCS)	0.144	0.019	0.03
Type of surgery (Combined)	0	-0.119	1.41
Tumor grade	-0.01	0.033	0.1
Estrogen receptor	0.414	0.599[a]	35.88
Progesterone receptor	-0.111	0.198	3.92
Radiotherapy	-0.019	-0.069	0.47
Hormone therapy (Nothing)	0.161	0.112	1.25

Hormone therapy (Tamoxifen)	0.147	0.278	7.72
Hormone therapy (Raloxifene)	0.041	-0.025	0.06
Hormone therapy (Letrozole)	0.113	0.218	4.75
Hormone therapy (Aromasin)	0.056	0.06	.36
Hormone therapy (Megace)	-0.012	0.025	0.06
Hormone therapy (Combined)	0	-0.451[a]	20.34
R^2_c			9.12
DM, first three years	0.173	0.209	4.36
DM, 3 - 5 years	0.175	0.2	4
DM, more than 5 years	0.074	-0.058	0.33
LRR, first three years	-0.437	-0.356	12.67
LRR, 3 - 5 years	-0.564	-0.479[a]	22.94
LRR, more than 5 years	-0.739	-0.686[a]	47.05

Abbreviations: Coef, standardized canonical function coefficient; r_s, structure coefficient; r_s^2, squared structure coefficient, R^2_c, squared canonical correlations.
[a]Structure coefficients (r_s) greater than 0.45 are underlined.

5. Discussion

In this study, applying of the CCA method leads to identification of variables: family history, estrogen receptor, pathology of tumor, type of surgery, tumor size and hormone therapy, as important factors in predicting LRR, more than 5 years and LRR, 3 - 5 years.

Within the general linear models (i.e. CCA), r^2 type effect sizes are the first point for considering (19). Reporting results only with P values (without effect sizes) has little or no information about the importance of results (20). Our study's statistical significance and effect sizes demonstrate that there is a remarkable relationship between our variable sets.

As Sherry (17), structure coefficients are answer to the question "what variables are contributing to the relationship between the variables set across the functions?" So, they are critical for deciding what variables are useful for the model (9).

The variable of family history, which got the highest structure coefficient, is a common and important predictor for the prognosis of breast cancer (21, 22). The second variable among predictors set (based on its structure coefficient) is estrogen receptor. This variable is a significant one that has an inverse relation with outcome variables. In patients who have ER positive breast cancers local recurrence occurs less common. The estrogen receptor positive cancers, need to estrogen in order to grow and multiply, are less aggressive than negative ones and have better prognosis (23-25). Pathology plays an important part in determining the treatment strategy for women with breast cancer, with the evaluation of breast specimens determining the surgical and the oncological therapeutic options used (26, 27). In this study, variable of pathology of tumor values got third and sixth places among predictors. Type of surgery is a pertinent risk factor of breast cancer affecting mortality rate of this disease and has different mental and physical consequences on various age categories of patients (28, 29). Tumor size (24, 30, 31) and hormone therapy (32), were also found to be important predictors in the present study.

Some researchers have detected progesterone receptor as a prognostic factor in predicting breast cancer recurrence, especially in accompanying with estrogen receptor (7, 33), whereas this variable did not get enough structure coefficients for reporting as important predictor in our study. Lyman (34), Paik (35) and Arvold (36) have showed that variable of age is an important factor in detecting the breast cancer recurrence, although age has not been determined as an important one in the present study.

Lymph node removed, number of involved LN, LN positive, radiotherapy and tumor grade were not detected as predictors of breast cancer recurrence in this research. Razavi (7) in a nearly similar study to the current study determined DM during the first four years and LRR during the first two years after diagnosis as outcome variables related to the predictor variables. In our study, LRR between three and five years and LRR more than five years were detected as outcome variables.

Dissimilarities between detected variables in our study and any other similar studies could be due to: geographical difference, nature of population studied and finally the way of data preprocessing.

There are two important limitations in the current study. First, some important variables that probably had great potential to exist among other significant risk factors were removed in our study because they had not enough values for analyzing; so, some valuable information were missed. Second, this study was performed in a single institution and consequently the generalization of its results is weaker than population-based studies.

In this study, the CCA was applied on the variables selected after consulting with experienced clinicians to achieve the medical meaningfulness and identification of most important risk factors leading to breast cancer recurrence during different time intervals.

In the medical complicated cases, for example the breast cancer disease that contains considerable numbers of vari-

ables (risk factors) in the predictor set and usually more than one variable in the outcome set, applying CCA as a new solution for detecting important ones in both sets is an appropriate selection. Detected breast cancer risk factors in this study were consistent with clinical guidelines.

Acknowledgments

The author would like to thank Dr Khalkhali (Ph.D. of statistics) for giving valuable comments to us.

Footnotes

Authors' Contribution: Study concept and design: Farahnaz Sadoughi, and Hadi Lotfnezhad Afshar; acquisition of data: Hadi Lotfnezhad Afshar, Asiie Olfatbakhsh, and Neda Mehrdad; analysis and interpretation of data: Hadi Lotfnezhad Afshar; drafting of the manuscript: Farahnaz Sadoughi, Hadi Lotfnezhad Afshar, and Asieh Olfatbakhsh; critical revision of the manuscript for important intellectual content: Farahnaz Sadoughi, and Asiie Olfatbakhsh; statistical analysis: Hadi Lotfnezhad Afshar; administrative, technical, and material support: Farahnaz Sadoughi; study supervision: Farahnaz Sadoughi.

Funding/Support: This study was funded and supported by Iran university of medical sciences; grant no: 90-04-136-15619.

References

1. American Cancer Society. *Global Cancer Facts & Figures.* 2nd ed. Atalanta: American Cancer Society; 2011.
2. Ahmadinejad N, Movahedinia S, Movahedinia S, Holakouie Naieni K, Nedjat S. Distribution of breast density in Iranian women and its association with breast cancer risk factors. *Iran Red Crescent Med J.* 2013;**15**(12):e16615. doi: 10.5812/ircmj.16615. [PubMed: 24693398]
3. Pourhoseingholi MA, Vahedi M, Pourhoseingholi A, Ashtari S. Bayesian Analysis of Breast Cancer Mortality to Reduce the Effects of Misclassification. *Razavi Int J Med.* 2013;**1**(1):22–5.
4. Mousavi SM, Montazeri A, Mohagheghi MA, Jarrahi AM, Harirchi I, Najafi M. Breast cancer in Iran: an epidemiological review. *Breast J.* **13**(4):383–91. [PubMed: 17593043]
5. Sirous M, Ebrahimi A. The Epidemiology of Breast Masses among Women in Esfahan. *Iran J Surg.* 2008;**16**(3):51–7.
6. Razavi AR. *Applications of Knowledge Discovery in Quality Registries - Predicting Recurrence of Breast Cancer and Analyzing Non-compliance with a Clinical Guideline.* Sweden: Linköpings universitet; 2007.
7. Razavi AR, Gill H, Stal O, Sundquist M, Thorstenson S, Ahlfeldt H, et al. Exploring cancer register data to find risk factors for recurrence of breast cancer–application of Canonical Correlation Analysis. *BMC Med Inform Decis Mak.* 2005;**5**:29. doi: 10.1186/1472-6947-5-29. [PubMed: 16111503]
8. Field A. *Discovering Statistics Using SPSS (Introducing Statistical Methods).* Philadelphia: 2009.
9. Courville T, Thompson B. *Use of Structure Coefficients in Published Multiple Regression Articles: B Is Not Enough.* Texas, United States: Sam Houston State University; 2001.
10. Walters SJ. *What is a Cox Model?* . United States: Hayward Medical Communications; 1999. Available from: http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/cox_model.pdf.
11. Bonadonna G, Gabriel NH, Pinuccia V. *Textbook of Breast Cancer.* Informa Health Care; 2006.
12. Kim KS, Kim W, N KY, Park JM, Kim J-S, Lee K. New recurrence prediction model for breast cancer by data mining.; 7th European Breast Cancer Conference.; Barcelona, Spain. Elsevier Science Ltd; 2010. p. 98.
13. Van Voorhis CW, Morgan BL. Statistical rules of thumb: What we don't want to forget about sample sizes. *Psi Chi J Undergraduate Res.* 2001;**6**(4):139–4.
14. Hair JF, Anderson RE, Tatham RL, Black W. *Multivariate data analysis.* New Jersey: Upper Saddle River; 1998.
15. Scheffer JA. *An Analysis of the Missing Data Methodology for Different Types of Data: A Thesis Presented in Partial Fulfilment of the Requirements for the Degree of Master of Applied Statistics at Massey University, Albany New Zealand.* Albany: Massey University; 2000.
16. Borman S. *The expectation maximization algorithm: A short tutorial. 2004.* Available from: http://www.seanborman.com/publications.
17. Sherry A, Henson RK. Conducting and interpreting canonical correlation analysis in personality research: a user-friendly primer. *J Pers Assess.* 2005;**84**(1):37–48. doi: 10.1207/s15327752jpa8401_09. [PubMed: 15639766]
18. Schafer J. *NORM.* The Pennsylvania State University; 1997. Available from: http://sites.stat.psu.edu/~jls/misoftwa.html.
19. Thompson B. Rejoinder: Editorial Policies Regarding Statistical Significance Tests: Further Comments. *Educ Res.* 1997;**26**(5):29–32. doi: 10.3102/0013189x026005029.
20. Leach LAF. *Bias and Precision of the Squared Canonical Correlation Coefficient Under Nonnormal Data Conditions.* Denton, Texas: 2006.
21. Mesli Taleb-Bendiab F, El Kebir FZ. Family history of breast cancer in western Algeria: an investigation [France]. *Afr J Cancer.* 2012;**5**(1):27–31.
22. Mizota Y, Yamamoto S. Prevalence of breast cancer risk factors in Japan. *Jpn J Clin Oncol.* 2012;**42**(11):1008–12. doi: 10.1093/jjco/hys144. [PubMed: 22988038]
23. Bellenir K. *Breast Cancer Sourcebook.* Detroit: Omnigraphics; 2009.
24. Abdalla FB, Markus R, Buhmeida A, Boder J, Syrjanen K, Collan Y. Estrogen receptor, progesterone receptor, and nuclear size features in female breast cancer in Libya: correlation with clinical features and survival. *Anticancer Res.* 2012;**32**(8):3485–93. [PubMed: 22843935]
25. Omranipour R, Alipour S, Hadji M, Fereidooni F, Jahanzad I, Bagheri K. Accuracy of estrogen and progesterone receptor assessment in core needle biopsy specimens of breast cancer. *Iran Red Crescent Med J.* 2013;**15**(6):515–8. doi: 10.5812/ircmj.10232. [PubMed: 24349751]
26. Fisher ER, Costantino J, Fisher B, Palekar AS, Paik SM, Suarez CM, et al. Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) Protocol B-17. Five-year observations concerning lobular carcinoma in situ. *Cancer.* 1996;**78**(7):1403–16. doi: 10.1002/(SICI)1097-0142(19961001)78:7<1403::AID-CNCR6>3.0.CO;2-L. [PubMed: 8839545]
27. Hanby AM. The pathology of breast cancer and the role of the histopathology laboratory. *Clin Oncol (R Coll Radiol).* 2005;**17**(4):234–9. [PubMed: 15997917]
28. Roohan PJ, Bickell NA, Baptiste MS, Therriault GD, Ferrara EP, Siu AL. Hospital volume differences and five-year survival from breast cancer. *Am J Public Health.* 1998;**88**(3):454–7. [PubMed: 9518982]
29. Rowland JH, Desmond KA, Meyerowitz BE, Belin TR, Wyatt GE, Ganz PA. Role of breast reconstructive surgery in physical and emotional outcomes among breast cancer survivors. *J Natl Cancer Inst.* 2000;**92**(17):1422–9. [PubMed: 10974078]
30. Fisher B, Slack NH, Bross ID. Cancer of the breast: size of neoplasm and prognosis. *Cancer.* 1969;**24**(5):1071–80. [PubMed: 5353940]
31. Gasparini G, Weidner N, Bevilacqua P, Maluta S, Dalla Palma P, Caffo O, et al. Tumor microvessel density, p53 expression, tumor size, and peritumoral lymphatic vessel invasion are relevant prognostic markers in node-negative breast carcinoma. *J Clin Oncol.* 1994;**12**(3):454–66. [PubMed: 7509851]
32. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet.* 2005;**365**(9472):1687–717. doi: 10.1016/s0140-6736(05)66544-0. [PubMed: 15894097]
33. Bardou VJ, Arpino G, Elledge RM, Osborne CK, Clark GM. Pro-

gesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J Clin Oncol.* 2003;**21**(10):1973–9. doi: 10.1200/JCO.2003.09.099. [PubMed: 12743151]

34. Lyman GH, Lyman S, Balducci L, Kuderer N, Reintgen D, Cox C, et al. Age and the Risk of Breast Cancer Recurrence. *Cancer Control.* 1996;**3**(5):421–7. [PubMed: 10764500]
35. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med.* 2004;**351**(27):2817–26. doi: 10.1056/NEJMoa041588. [PubMed: 15591335]
36. Arvold ND, Taghian AG, Niemierko A, Abi Raad RF, Sreedhara M,, Nguyen PL, et al. Age, Breast Cancer Subtype Approximation, and Local Recurrence After Breast-Conserving Therapy. *J Clin Oncol.* 2012;**29**(29):3885–91. doi: 10.1200/JCO.2011.36.1105. [PubMed: 21900114]