World Cancer Research Journal WCRJ 2018; 5 (1): e1016

IRANIAN BREAST CANCER RISK ASSESSMENT STUDY (IRBCRAS): A CASE CONTROL STUDY PROTOCOL

H. SALEHINIYA[1], S. HAGHIGHAT[2], M. PARSAEIAN[1], R. MAJDZADEH[3], M.A. MANSOURNIA[1], S. NEDJAT[3]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[2]Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran
[3]Department of Epidemiology and Biostatistics, School of Public Health, Knowledge Utilization Research Center, Tehran University of Medical Sciences, Tehran, Iran

Abstract – *Background:* *Breast cancer is the most common cancer among Iranian women. Therefore, its early diagnosis can reduce death and health costs. There is no specific method or model for detecting individuals at a high-risk of breast cancer in Iran. Thus, a native model for detecting groups at higher risk of breast cancer seems necessary. This research was designed to develop a predictive model for detecting women with high-risk of breast cancer in Iran.*

Patients and Methods: *With a case control study, we will recruit 1000 women diagnosed with breast cancer from Tehran's cancer centers and another 1000 healthy controls from all regions of Tehran City. Data will be collected by a valid and reliable questionnaire consisting of 8 sections: 1) demographic and socio-economic information; 2) data related to pregnancy, medical records and life events; 3) family history of cancer and gynecological diseases; 4) history of drug use and occupational exposures; 5) weight and height measurement; 6) physical activity; 7) use of cigarettes, hookah, alcohol, narcotics; 8) food consumption. Data analysis will be done with logistic regression models; to check the appropriateness of the model, both ROC and net reclassification index will be applied. Moreover, h-fold cross-validation and bootstrap techniques will be applied to examine the validity of the model.*

Conclusions: *We intend to build a suitable prediction model for detecting high-risk groups of breast cancer in Iran. The results can be useful to estimate the risk of breast cancer for health service providers in healthcare centers.*

KEYWORDS: *Breast cancer, Risk assessment, Iran, Case control study.*

LIST OF ABBREVIATIONS: *ICC: Intra-class correlation coefficient, ROC: Receiver operating characteristic, NRI: Net reclassification improvement, AUC: Area under the curve.*

INTRODUCTION

Nowadays, cancer is one of the main causes of morbidity and mortality and is considered one of the most important health issues globally[1]. Among different types of cancers, breast cancer is the second most common cancer in the world, affecting over 1.67 million people annually[2]. Worldwide, it is one of the most common malignancies among women[3-5]. This increase is seen especially in Asian countries where its rate of mortality is raising[6]. Breast cancer has been one of the most common types of cancer in Iran[7], and has been a major cost of the healthcare system[8,9]. The incidence, morbidity and mortality of breast cancer in Iran are estimated at 28.3 and 4.33 in every 100000 women, respectively[10]. Furthermore, the epidemi-

Corresponding Author: Saharnaz Nedjat, MD, PhD; e-mail: nejatsan@tums.ac.ir

ologic pattern of breast cancer in Iran differs from the rest of the world; patients have been diagnosed at more advanced stages of disease[7] at a lower mean age than the global average rate[7,11]; thus, a growing incidence in young women has been detected[12].

The most important ways to reduce morbidity, mortality and treatment costs of breast cancer are the prevention and its early diagnosis, and the treatment in the preliminary stages of disease[13]. However, screening the entire population in order to identify high-risk individuals is neither cost-effective nor affordable[13,14]. Therefore, predictive models are needed to detect groups at a higher risk of breast cancer[15]. Such models have been introduced in various countries across the world[16]. However, given different epidemiologic profiles of patients with breast cancer in different countries, make models hardly to be replicable across countries[17].

Hence, various studies have been conducted to develop models for detecting groups at high risk of breast cancer around the world; each reserach has recommended a model appropriate to the population under study. The models constructed in European countries and United States for example, are not appropriate for Asian countries, and each country has presented a model bearing in mind its own circumstances[17-19].

In Iran, there is no method or model with which groups at a higher risk of breast cancer can be detected, based on which they can be screened. Moreover, the population, risk factors and epidemiologic distribution of breast cancer differ from the rest of the world[7,11-12]. Thus, the models constructed elsewhere are not suitable for the Iranian population and a native model is necessary.

Cohort studies are more suitable for detecting high-risk groups and classifying individuals into ill and healthy. However, they are expensive, time-consuming, and complex. Conversely case control studies are affordable and cost-effective in detecting high risk groups[20]. As a result, this study will be conducted by a case-control design to introduce a suitable model for detecting groups at higher risk of breast cancer in Iran, based on receiver operating characteristic curve (ROC curve) and Net Reclassification Index (NRI). Moreover we aim to calculate the sensitivity, specificity and validity of the constructed model.

PATIENTS AND METHODS

Patients

This study will be conducted in 2018 as a case control study in Tehran (Iran), a city where reside all Iranian ethnicities, with a population exceeding 8 million people, approximately 10% of the country's entire population.

Inclusion criteria: Women, Iranian, 25-75 years old, resident of Tehran, which gave their consent to participate in the study.

Exclusion criteria: Pregnant women, a woman who has any other type of cancer – other than breast cancer, a woman who is undergoing preventive treatment for breast cancer (including, mastectomy, tamoxifen, estrogen blockers).

Definition and selection of cases

Here, the cases are women who have a malignant growth in the breast tissue that has originated from breast tissue and has been confirmed by mammography, biopsy, and pathologic tests, and/or confirmed by a specialist as a breast cancer patient (new cases: diagnosed within the last year). The sampling of these individuals will be an ongoing process wherein patients will be selected from breast cancer diagnostic centers in Tehran. Patient selection from diagnostic and treatment centers will be appropriate to the share of each center from the total number of diagnosed breast cancer patients. To this end, the patients diagnosed in the past year in the main diagnostic centers of breast cancer in Tehran, which are located in different areas of the city, will be examined, and appropriate to each center's share of all patients, ongoing sampling will take place from each center. Sampling will continue up to completion in each center.

Selection of controls

The controls will be selected from the general population of Tehran who are not affected by breast cancer. The absence of breast cancer will be affirmed by the person's self-report (due to the low prevalence of breast cancer, high costs of diagnosis of the disease and the harmful nature of diagnostic techniques for the control group's self-statement will be chosen as the grounds for selection. However, if a woman is suspected to have breast cancer (self-stated) and yet no definite diagnosis has been given, she will be excluded from the control group).

Selection of controls will be population – based and probability proportional to size as follows: Tehran will be divided into 22 districts. Stratified sampling will be used to select the controls; each district will be considered as stratum, and based on 2011's population data, the population of each district will be estimated. Appropriate to the size of each district, individuals will be selected as samples. In each district, the number of blocks appropriate to the population of that district will be selected and 20 women will be selected from each block. Within the blocks, one house will be randomly selected based on the houses number, and then sampling will continue up to 20 houses. After selecting the house, if an eligible person will exist in that house,

she will be included in the study; if not, that house will not be included. If there are two or more eligible individuals in that house, then one will be randomly selected by Kish Method (in this case the number of eligible individuals will be noted down and considered during weighting). In case the eligible person is absent from the house, three more visits will be made to the house to complete the questionnaire.

Assessment and measurement

The questionnaire includes: 1) demographic information and socio-economic status; 2) data related to pregnancy, medical records and life events; 3) family history of cancer and gynecological diseases; 4) history of drug use and occupational exposures; 5) weight and height measurement; 6) physical activity; 7) use of cigarettes, hookah, alcohol and narcotics; 8) food consumption (Table 1).

To construct the questionnaire, first, studies on breast cancer risk factors by using comprehensive search in international and national database and variables from other existing models for the detection of groups at a higher risk of breast cancer were examined; the questions and framework for the questionnaire were outlined. After preparing questions and contents of the questionnaire, this was given to experts in the field of breast cancer (oncologist), an epidemiologist and a methodologist in the form of an assessment form. They were then asked to give their opinions regarding the content of the questionnaire, its comprehensiveness, and clarity of the questions. Furthermore, the questionnaire was given to a number of individuals possessing the inclusion criteria to obtain their opinions on the content and face validity of the questions. Their opinions were then applied to the questionnaire, and the content validity of it was approved through qualitative approach.

Test-retest was applied to examine the reliability of the questionnaire. To this end, once the validity was approved, the final questionnaire was given to 30 women who possessed the inclusion criteria to be completed. Two weeks later, they were given the questionnaire to be completed again. Using the Intra-class Correlation Coefficient (ICC) and kappa, the questionnaire was analyzed, and all the questions that had an ICC and/or kappa greater than 0.7 were considered acceptable and were included[21]; questions with low ICCs or kappa (less than 0.7) were revised.

Statistical Analysis

In this study, the logistic regression model will be used to construct the model, and different logistic regression models will be examined. Each model will be assessed in comparison to its previous model. For every new risk factor, one coefficient will be added to the model and the predicted probability will be estimated based on the logistic model[22]. To examine the appropriateness of each model, net reclassification improvement and ROC will be used. The net reclassification improvement (NRI) method is a method for assessing what percentage of individuals are correctly, or incorrectly, classified from one risk group to another after adding a risk factor.

The NRI and ROC methods have been based on predicted probabilities. The probabilities predicted in logistic models of case control studies are incomprehensible. In this study, the baseline risk or intercept will be adjusted to estimate the risk by inverse probability weighting.

To adjust the baseline risk or intercept, we will consider the latest incidence of breast cancer in Tehran and the annual percent rise in the final incidence rate for the year of executing this study.

TABLE 1. Details of the questionnaire.

Section	*Subsection Questions*
Demographic information and socio-economic status	Age, marital status, education level, job, ethnicity, insurance status, socioeconomic status,)
Data related to pregnancy, medical records and life events	History of related disease such as: carcinoma *in situ*, fibrocystic breast etc…
	Life events: such as debt, migration and etc…
	History of mammography, breast self-examination, biopsy etc…
Family history of cancer and gynecological diseases	History of breast cancer in the first, second and third relatives, history of ovarian cancer in first and second degree relatives, history of other cancers in the family
History of drug use and occupational exposures	Drugs to prevent abortion, medication for regulating menstrual cycle, medication for the treatment of infertility, menopause replacement therapy etc…
Weight and height measurement	Weight, height, body mass index
Physical activity	International Physical Activity Questionnaire (IPAQ)
Use of cigarettes, hookah, alcohol and narcotics	Cigarette, hookah, alcohol, drugs
Food consumption	Tea, fruit, greens, meat, beans, snacks etc..

The following formula will be used to estimate net reclassification improvement[23].

$$NRI=P(up \mid event) - p(down \mid event) + p(down \mid nonevent) - p(up \mid nonevent)$$

Here, upward movement refers to the predicted risk change to a higher group, based on the model, and downward movement refers to the opposite change in risk[23].

Finally, the model with the best prediction (based on ROC and the significance of the NRI) will be introduced as the final appropriate model for detecting groups at higher risk of breast cancer. Then, to ensure its appropriateness for other individuals and groups, its validity will also be examined.

Examining the validity of the final model

The appropriateness of the model will be determined based on AUC; thus, upon examining the validity, the AUC will be compared. Here, the h-fold cross-validation method in which h is equal to 5 will be used, such that the data will be divided into 5 random and equal sets. At each stage, one of the sets will be put aside and the model will be fitted on the remaining data. In each stage cross-validation will be calculated for each subset and this procedure will be repeated ten times until the AUC is calculated for the cross-validated set and will be compared with the ROC of the data[24]. Furthermore, we will use the bootstrap validation technique to examine validation in this study. In this method, the original data set is used to create new random sets by random sampling with replacement. Usually, a large number of new data sets are created. Here, we will have a maximum of 200 separate sets of healthy and ill individuals. The model will be fitted separately for each of these 200 sets, and then the AUC will be calculated 200 times and compared with the model's AUC[25].

All the codes required to fit the model, calculate the diagnostic values and perform cross-validation will be written by Stata and R software. p-value < 0.05 is considered as significant.

Ethical considerations

This study complies with the Declaration of Helsinki's principles of treating human subjects and has been approved by Tehran University of Medical Sciences' Ethics Board under grant number IR.TUMS.VCR.REC.1395.1630. We will obtain the subjects consent to enter the study and the results will be published as group data; the subjects' personal data will remain confidential throughout the study.

DISCUSSION

This study will develop a model for the classification of women based on their risk of becoming affected with breast cancer wherein individuals are divided into high-risk and low-risk groups. Breast cancer is on the rise in Iran, and its resultant morbidity and mortality are high[7] since patients used to be diagnosed at more advanced stages of the disease[11,12]. Thus, having a simple model for detecting high-risk groups and focusing screening and preventive services on those groups can lead to a reduction in health costs and disease complications.

This study will be conducted as a case control population – based study by adjusting base rate and will employ a comprehensive questionnaire and use a large sample size. The latter two are the study's strengths. However, being a case control, it has its own shortcomings. Moreover, people's genetic characteristics will not be examined, which can also be considered as limitation. Nevertheless, an appropriate model for primary screening should be cheap and based on variables that can be easily evaluated with the least amount of costs. Therefore, this limitation can be overlooked. Despite the shortcomings, the results of this study can provide a local model for detecting groups at a higher risk of breast cancer in Iran.

CONCLUSIONS

This study intends to build a suitable prediction model for detecting high-risk groups of breast cancer in Iran. The results can be useful to estimate the risk of breast cancer for health service providers in healthcare centers.

Conflict of Interest:

The authors declare that they have no conflict of interest.

Authors' contributions:

All authors contributed to the design of the research. All authors drafted the manuscript and approved final version.

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015; 136: 359-386.
3. Almasi Z, Mohammadian-Hafshejani A, Salehiniya H. Incidence, mortality, and epidemiological aspects of cancers in Iran; differences with the world data. J BUON 2016; 21: 994-1004.
4. Jara-Lazaro AR, Thilagaratnam S, Tan PH. Breast cancer in Singapore: some perspectives. Breast Cancer 2010; 17: 23-28.

5. Babu GR, Lakshmi SB, Thiyagarajan JA. Epidemiological correlates of breast cancer in South India. Asian Pac J Cancer Prev 2013; 14: 5077-5083.
6. Ghoncheh M, Momenimovahed Z, Salehiniya H. Epidemiology, incidence and mortality of breast cancer in Asia. Asian Pac J Cancer Prev 2016; 17: 47-52.
7. Rafiemanesh H, Salehiniya H, Lotfi Z. Breast cancer in Iranian woman: incidence by age group, morphology and trends. Asian Pac J Cancer Prev 2016; 17: 1393-1397.
8. Mousavi SM, Mohagheghi MA, Mousavi-Jerrahi A, Nahvijou A, Seddighi Z. Burden of breast cancer in Iran: a study of the Tehran population based cancer registry. Asian Pac J Cancer Prev 2006; 7: 571-574.
9. Daroudi R, Akbari Sari A, Nahvijou A, Kalaghchi B, Najafi M, Zendehdel K. The economic burden of breast cancer in Iran. Iran J Public Health 2015; 44: 1225-1233.
10. Sharifian A, Pourhoseingholi MA, Emadedin M, Rostami Nejad M, Ashtari S, Hajizadeh N, Firouzei SA, Hosseini SJ. Burden of breast cancer in Iranian women is increasing. Asian Pac J Cancer Prev 2015; 16: 5049-5052.
11. Vostakolaei FA, Broeders MJ, Rostami N, van Dijck JA, Feuth T, Kiemeney LA, Verbeek AL. Age at diagnosis and breast cancer survival in iran. Int J Breast Cancer 2012; 22: 517976.
12. Keramatinia A, Mousavi-Jarrahi SH, Hiteh M, Mosavi-Jarrahi A. Trends in incidence of breast cancer among women under 40 in Asia. Asian Pac J Cancer Prev 2014; 15: 1387-1390.
13. Barfar E, Rashidian A, Hosseini H, Nosratnejad S, Barooti E, Zendehdel K. Cost-effectiveness of mammography screening for breast cancer in a low socioeconomic group of Iranian women. Arch Iran Med 2014; 17: 241-245.
14. Yoo KB, Kwon JA, Cho E, Kang MH, Nam JM, Choi KS, Kim EK, Choi YJ, Park EC. Is mammography for breast cancer screening cost-effective in both Western and asian countries?: results of a systematic review. Asian Pac J Cancer Prev 2013; 14: 4141-4149.
15. Huang Y, Pepe MS. Assessing risk prediction models in case-control studies using semiparametric and nonparametric methods. Stat Med 2010; 29: 1391-1410.
16. Howell A, Anderson AS, Clarke RB, Duffy SW, Evans DG, Garcia-Closas M, Gescher AJ, Key TJ, Saxton JM, Harvie MN. Risk determination and prevention of breast cancer. Breast Cancer Res 2014; 16: 446.
17. Min JW, Chang MC, Lee HK, Hur MH, Noh DY, Yoon JH, Jung Y, Yang JH. Validation of risk assessment models for predicting the incidence of breast cancer in korean women. J Breast Cancer 2014; 17: 226-235.
18. Park B, Ma SH, Shin A, Chang MC, Choi JY, Kim S, Han W, Noh DY, Ahn SH, Kang D, Yoo KY, Park SK. Korean risk assessment model for breast cancer risk prediction. PLoS One 2013; 8: e76736.
19. Challa VR, Swamyvelu K, Shetty N. Assessment of the clinical utility of the Gail model in estimating the risk of breast cancer in women from the Indian population. Ecancermedicalscience 2013; 7: 363.
20. Huang Y, Pepe MS. Semiparametric methods for evaluating risk prediction markers in case-control studies. Biometrika 2009; 96: 991-997.
21. Buckens CF, de Jong PA, Mol C, Bakker E, Stallman HP, Mali WP, Van der Graaf Y, Verkooijen HM. Intra and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography. PLoS One 2013; 8: e71204.
22. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Stat Med 2011; 30: 11-21.
23. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. Net reclassification indices for evaluating risk prediction instruments: a critical review. Epidemiology 2014; 25: 114-121.
24. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics: linear, logistic, survival, and repeated measures models: Springer Science and Business Media, 2011.
25. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, Omar RZ. How to develop a more accurate risk prediction model when there are few events. BMJ 2015; 351: h3868.