Subgrouping of Iranian Women Based on Breast Cancer Risk Factors in A Population-Based Screening Program.

Mojtaba Sepandi  
Baqiyatallah University of Medical Sciences

Abbas Rezaianzadeh  
Shiraz University of Medical Sciences

Sima Afrashteh  
Shiraz University of Medical Sciences

Yousef Alimohamadi  
Iran University of Medical Sciences

Abbas Abbasi-Ghahramnaloo (✉ abbas.abbas49@yahoo.com)  
Ardabil University of Medical Sciences

Research Article

Keywords: latent class analysis, risk factors, breast cancer, cluster.

DOI: https://doi.org/10.21203/rs.3.rs-129617/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Breast cancer is the most common type of cancer in women worldwide. The aim of this study was to find the subgroups of women on the basis of clustering of breast cancer risk factors.

**Methods:** This cross-sectional study was performed within the framework of Shiraz breast cancer screening program between 2004 and 2013. Clinical breast examination, mammography or sonography, fine needle aspiration or biopsy, and surgery in case of indications were performed for all participants. Also, face-to-face interviews were performed by trained staff using a structured questionnaire to collect demographic information. Latent class analysis was used to achieve study's objectives.

**Results:** Four latent classes were indemnified; 1) general population risk (83.4% in healthy and 65.7% in diseased women), 2) moderate risk (7.8% in healthy and 20.5% in diseased women), 3) semi moderate risk [for healthy women (7.6%)] / low risk [for diseased women (8.9%)], and 4) low risk [for healthy women (0.9%)] / high risk [for diseased women (4.9%)]. This study showed that among healthy and diseased women, in the latent class 1 and 2 there was a similar pattern of breast cancer risk factors. However, in latent class 3 and 4 there were some different in the clustering of risk factors among two groups.

**Conclusions:** Results from the present study indicated that among healthy women, 15.4% of them fell under moderate risk or semi moderate risk classes, which stresses the necessity of implementing educational and preventive intervention for this stratum of women.

**Background**

Breast cancer is the most common type of cancer and the leading cause of cancer death in women worldwide. With 1.7 million new cases per year, it accounts for 25% of all cancers in women. According to estimates of global burden of cancer in 2018 (GLOBOCAN 2018), the incidence rate of breast cancer is quite high in Australia, Western Europe and Northern Europe. However, the rate of this cancer is increasing much faster in Asia than in Western countries. This rapid increase of incidence in developing countries may reflect cultural and socio-economic changes, such as delayed marriage and childbearing, having fewer children, increased obesity, and lack of knowledge and awareness towards breast cancer.

Although, breast cancer occurs all over the world, its prevalence, mortality and survival rates vary in different parts of the world, which can be due to different lifestyle of individuals, the population structure of countries, as well as genetic and environmental factors. For example, breast cancer accounts for 30% and 27% of all new female cancer cases in the United States and Europe, respectively. However, 44% of deaths and 39% of new cases of breast cancer occurs in Asia. Among Asian countries, China, the top ranked populous country of Asia, accounts for 25% of overall deaths due to breast cancer, especially in the younger population. Also, 25% of new cases of breast cancer occur in India. Iran is a multi-geographical, climatic, ethnic, regional and cultural country that make Iranian people exposed
to various risk factors for breast cancer(11). The breast cancer has been reported to affect Iranian women one decade younger than their counterparts in Western countries, which is responsible for 24.4% of malignancies among Iranian women(12). Metastatic breast cancer, also known as advanced or stage IV breast cancer, can spread to other organs in the body, such as bone, liver, lung and brain, thus early detection of the disease can lead to best improved prognosis and long-term survival of breast cancer patients(13). A number of factors, including premature menstruation, late-onset menopause, late age at first birth, breastfeeding period, socioeconomic status, smoking, low physical activity, obesity, and high-fat diet have been associated with increased risk of breast cancer(2, 14).

Cancer cluster is widely used to identify risk factors associated with breast cancer. The purpose of cancer cluster analysis is to identify potential unique subgroups of patients with specific behaviors and sensitivities as well as to determine the co-occurrence of these risk factors in order to detect the disease at early stages and promote health in a target population(15, 16). Latent class analysis (LCA) as a statistical modeling method, is used to detect heterogeneity in response patterns or clinical features of the classes in a population. This person-centered approach is also used in social and behavioral sciences(17). Since clustering of risk factors for breast cancer has not been previously investigated in Iran and also due to the uncertainty of epidemiological aspects of breast cancer among Iranian female patients(12), this study aimed to investigate the clustering of risk factors for breast cancer among female breast cancer patients versus healthy women in Iran using LCA analysis.

**Methods**

**Sample and setting**

We undertook a large study based on the results of breast cancer screening program in 11860 women referring to Shahid Motahhari breast clinic affiliated to Shiraz University of Medical Sciences between 2004 and 2013 in Iran. Clinical breast examination was performed for women participating in the screening program. Then, they underwent mammography or sonography, fine needle aspiration or biopsy, and surgery in case of indications, depending on the physician's decision. For all of the women, a face-to-face interviews were performed by trained staff using a structured questionnaire to collect information regarding family history of breast cancer, height, weight, marital status, education level, age at menarche, occupation, parity, age, past use of oral contraceptives (OCP), age at the first pregnancy, and lifetime breastfeeding.

The study was approved by the Ethics Committee of Shiraz University of Medical Sciences (IR.SUMS.REC.1391.S6422). Permission to conduct the study was obtained from this committee and all women had signed an informed consent form. In addition, all methods were performed in accordance with the relevant guidelines and regulations of this Committee.

**Statistical analysis**
A LCA model was used to detect the clustering of breast cancer risk factors among Iranian women using the disease status of participants as the grouping variable. According to this model, a number of observed variables were aggregated to represent a categorical latent variable. In order to select the best fitting model, we firstly calculated the G2 index. In addition, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were also calculated in order to identify the best model. For all of these indices, lower values indicate a better fit and parsimony for the model. Item response probabilities of > 0.5 were used to label each latent class and to describe characteristics of each. Eight dichotomous observable (i.e., indicators) were selected for creating the subgroups. These indicators were: obesity, age of menarche, first pregnancy age, parity, OCP use, family history of breast cancer, breast feeding, and marital status. SPSS 16 was used for reporting the frequency of the observables, while the LCA was performed using PROC LCA in SAS 9.2 software. In all analysis, a P-value < 0.05 was considered to be statistically significant.

**Results**

The mean age of participants was 41.12 ± 10.58 (41.18 among non-diseased and 40.25 among diseased women). Table 1 presents the participants status for each of the risk factors measured in this study and theirs relationship to disease status. This shows that the prevalence of OCP use was the highest among the all risk factors (56.4% in healthy women and 46.7% in diseased women). In total, 2580 (21.8%) participants were obese, while only 0.9% had a history of having higher age in the first pregnancy. This table shows that except obesity, premature menstruation and late age at first pregnancy, all of the risk factors studied had a significant relationship with disease status.

<table>
<thead>
<tr>
<th>Items</th>
<th>Breast cancer</th>
<th>P value</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No (11186)</td>
<td>No (674)</td>
<td></td>
</tr>
<tr>
<td>N(%)</td>
<td>N(%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being obese (BMI &gt; 30)</td>
<td>2432(21.7)</td>
<td>148(22.0)</td>
<td>0.572</td>
</tr>
<tr>
<td>Premature menstruation (under 12 y)</td>
<td>857(7.7)</td>
<td>54(8.0)</td>
<td>0.742</td>
</tr>
<tr>
<td>Late age at first pregnancy (up to 35 y)</td>
<td>103(0.9)</td>
<td>6(0.9)</td>
<td>0.935</td>
</tr>
<tr>
<td>Parity (don't have any child)</td>
<td>1626(14.5)</td>
<td>161(23.9)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>OCP use</td>
<td>6313(56.4)</td>
<td>315(46.7)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td>4146(37.1)</td>
<td>194(28.8)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Skipping breast feeding</td>
<td>1907(17.0)</td>
<td>173(25.7)</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Marital status (being single)</td>
<td>823(7.4)</td>
<td>96(14.2)</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
Eight binary variables (indicators) were used to perform LCA. We attempted to fit the LCA models with classes ranging from 1 to 7, using disease status as the grouping variable. The different measures of model selection are shown in Table 2. For each model, we computed $G^2$, AIC and BIC. According to these model selection indices, and the interpretability of the results, we concluded that a four-class model was most suitable for these women.

Table 2
Comparison of LCA Models With Different Latent Classes Based on Model Selection Statistics.

<table>
<thead>
<tr>
<th>Number of latent class</th>
<th>Number of parameters estimated</th>
<th>$G^2$</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>Maximum log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>14541.45</td>
<td>495</td>
<td>14573.45</td>
<td>14691.55</td>
<td>-39427.61</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>651.62</td>
<td>477</td>
<td>719.62</td>
<td>970.57</td>
<td>-32482.69</td>
</tr>
<tr>
<td>3</td>
<td>52</td>
<td>492.59</td>
<td>459</td>
<td>596.59</td>
<td>980.40</td>
<td>-32403.18</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>174.91</td>
<td>441</td>
<td>314.91</td>
<td>831.57</td>
<td>-32244.33</td>
</tr>
<tr>
<td>5</td>
<td>88</td>
<td>149.46</td>
<td>423</td>
<td>325.46</td>
<td>974.99</td>
<td>-32231.61</td>
</tr>
<tr>
<td>6</td>
<td>106</td>
<td>119.02</td>
<td>405</td>
<td>331.02</td>
<td>1113.40</td>
<td>-32216.39</td>
</tr>
<tr>
<td>7</td>
<td>124</td>
<td>93.08</td>
<td>387</td>
<td>341.08</td>
<td>1256.31</td>
<td>-32203.42</td>
</tr>
</tbody>
</table>

*Note. LCA = latent class analysis; AIC = Akaike information criterion; BIC = Bayesian information criterion.*

Table 3 presents the results of the LCA model for the breast cancer risk factors, based on disease status. This table include the latent class prevalence and item response probabilities. As can be seen in Table 3, among healthy women, the first (general population risk), second (moderate risk), third (semi moderate risk), and fourth (low risk) classes described 83.6%, 7.8%, 7.6%, and 0.9% of participants. Among diseased women, the prevalence of general population risk, moderate risk, low risk, and high risk classes was 65.7%, 20.5%, 8.9% and 4.9% respectively. Table 3 shows that in the latent classes 1 and 2 there was a similar pattern of breast cancer risk factors. In general population risk class, the probability of OCP use was high in healthy and diseased women. In the moderate risk class, parity status, skipping breast feeding, and being single had a high probability of being present, for both group. However, in latent class 3 there were several differences between healthy and diseased women. In healthy women, this class named as semi moderate risk class, because of high probability of parity and skipping breast feeding in this class. In contrast, among diseased women, latent class 3 named as low risk class. In this class, only obesity had a high probability of being present. In latent class 4 among healthy women, late age at first pregnancy had a high probability. But, in diseased women, parity, OCP use, and skipping breast feeding had a high probability of being present in class 4. With considering these probabilities we named these classes as low risk and high risk for healthy and diseased women respectively.
Table 3
The four Latent Classes Model of risk factors of breast cancer among healthy and diseased women

<table>
<thead>
<tr>
<th>Latent class</th>
<th>Healthy women</th>
<th>General population risk</th>
<th>Moderate risk</th>
<th>Semi moderate risk</th>
<th>Low risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent class prevalence</td>
<td>0.836</td>
<td>0.078</td>
<td>0.076</td>
<td>0.009</td>
<td></td>
</tr>
</tbody>
</table>

**Item-response probabilities**

<table>
<thead>
<tr>
<th>Item-response probabilities</th>
<th>Healthy women</th>
<th>General population risk</th>
<th>Moderate risk</th>
<th>Semi moderate risk</th>
<th>Low risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being obese (BMI &gt; 30)</td>
<td>0.247</td>
<td>0.160</td>
<td>0.193</td>
<td>0.222</td>
<td></td>
</tr>
<tr>
<td>Premature menstruation (under 12 y)</td>
<td>0.076</td>
<td>0.083</td>
<td>0.082</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>Late age at first pregnancy (up to 35 y)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
<td><strong>0.858</strong></td>
<td></td>
</tr>
<tr>
<td>Parity (don't have any child)</td>
<td>0.000</td>
<td><strong>0.999</strong></td>
<td><strong>0.879</strong></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>OCP use</td>
<td><strong>0.647</strong></td>
<td>0.002</td>
<td>0.267</td>
<td>0.298</td>
<td></td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td>0.375</td>
<td>0.359</td>
<td>0.323</td>
<td>0.493</td>
<td></td>
</tr>
<tr>
<td>Skipping breast feeding</td>
<td>0.017</td>
<td><strong>1.00</strong></td>
<td><strong>0.993</strong></td>
<td>0.205</td>
<td></td>
</tr>
<tr>
<td>Marital status (being single)</td>
<td>0.000</td>
<td><strong>0.931</strong></td>
<td>0.011</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

**Diseased women**

<table>
<thead>
<tr>
<th>Latent class prevalence</th>
<th>Diseased women</th>
<th>General population risk</th>
<th>Moderate risk</th>
<th>Low risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.657</td>
<td><strong>0.205</strong></td>
<td>0.089</td>
<td>0.049</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Item-response probabilities**

<table>
<thead>
<tr>
<th>Item-response probabilities</th>
<th>Diseased women</th>
<th>General population risk</th>
<th>Moderate risk</th>
<th>Low risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being obese (BMI &gt; 30)</td>
<td>0.220</td>
<td>0.140</td>
<td><strong>0.535</strong></td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td>Premature menstruation (under 12 y)</td>
<td>0.091</td>
<td>0.094</td>
<td>0.011</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Late age at first pregnancy (up to 35 y)</td>
<td>0.000</td>
<td>0.007</td>
<td>0.083</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Parity (don't have any child)</td>
<td>0.000</td>
<td><strong>0.998</strong></td>
<td>0.000</td>
<td><strong>0.681</strong></td>
<td></td>
</tr>
<tr>
<td>OCP use</td>
<td><strong>0.628</strong></td>
<td>0.001</td>
<td>0.199</td>
<td><strong>0.757</strong></td>
<td></td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td>0.337</td>
<td>0.207</td>
<td>0.189</td>
<td>0.150</td>
<td></td>
</tr>
<tr>
<td>Skipping breast feeding</td>
<td>0.004</td>
<td><strong>0.998</strong></td>
<td>0.002</td>
<td><strong>0.991</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Note. The probability of a “No” response can be calculated by subtracting the item-response probabilities shown above from 1.*

* Item-response probabilities > .5 in bold to facilitate interpretation.
<table>
<thead>
<tr>
<th>Latent class</th>
<th>0.000</th>
<th>0.692</th>
<th>0.000</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marital status (being single)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. The probability of a “No” response can be calculated by subtracting the item-response probabilities shown above from 1.*

* Item-response probabilities > .5 in bold to facilitate interpretation.

**Discussion**

The present study evaluated risk factors associated with incidence of breast cancer in two groups of female patients with breast cancer and healthy women. According to the results of this study, four latent classes were identified for patients and healthy women. In this study, the majority of subjects in both groups fell under general population risk class. OCP use in general population risk class, and skipping breastfeeding and being single in moderate risk class, were more likely in both study groups.

However, the prevalence of general population risk class was much higher among healthy women compared to breast cancer patients. In general, the pattern of risk factors between the two groups was not different in class 1 and 2, whereas, the pattern of risk factors was different between the two groups in class 3 and 4. In class 3, parity and skipping breastfeeding were more likely among healthy women. However, obesity was more likely to occur in breast cancer patients. In class 4, parity, OCP use, and skipping breastfeeding were more likely to occur in breast cancer patients, while late age at first birth was more likely to occur in healthy women. In this study, we also found that, obesity was exclusively associated with class 3 membership among breast cancer patients, while in healthy women, this variable showed no impact on risk factors clustering.

To the best of our knowledge, this has been the first study focusing on the clustering of breast cancer risk factors based on LCA approach in Iran. However, here we discuss a number of studies that used almost the same method used in our study. Lifestyle patterns in women with breast cancer were assessed using LCA. In this study, three latent classes were identified, including healthy behavior and diet pattern (Class 1), healthy behavior and unhealthy diet pattern (Class 2), and unhealthy behavior and diet pattern (Class 3)(18). Zhang et al. examined the relationship between body composition and risk of breast cancer in Chinese women, and identified four classes: Low Muscle Mass (Class 1), High Body Composition (Class 2), High Fat (Class 3), and Normal Body Composition (Class 4)(19).

The results of our study suggest that OCP use had important role in the clustering of breast cancer patients. This factor had high probability of being present in latent class 1 and 4. Bethea et al. have also shown that the OCP use, especially long-term use, was associated with an increased risk of breast cancer(20). Use of OCPs is associated with increased risk of breast cancer due to the direct elevation of estrogen levels and the role of progesterone in gaining weight indirectly (21). Our study indicated that healthy women in class 1 were also more likely to consume OCP. Therefore, it seems that this risk factor
can also cause problems over time and co-occur with other risk factors in healthy women, leading to development of breast cancer in these individuals.

We found a relationship between breast cancer and family history of this cancer. The results of previous studies showed that family history of breast cancer was a well-established and significant risk factor associated with breast cancer(22, 23). Family members share genes and common genetic factors, which may explain the link between family history of breast cancer and increased risk of this cancer. These genetic factors constituted of mutations in genes, such as BRCA1 and BRCA2. Family members also have environments and lifestyles in common, which can increase the risk of breast cancer in women with a family history(24). However, family history was not linked with the clustering of patients and healthy individuals, and probably the effects of other variables were more significant on the subgrouping of women in this study.

Increased parity may defer the onset of breast cancer and inhibit the metastasis of axillary lymph node (ALN)(25). Also, hormonal factors, including estrogen and progesterone may affect the risk of breast cancer, and ultimately changes in these hormones have a major impact on protection against this cancer(26). In an animal study, Yuri et al. found that exposure to estrogen and progesterone in animals during pregnancy protected breast tissue against cancer(27). The results of the present study showed that parity was significantly associated with risk of development of breast cancer. In the pattern of risk factors, this variable also played an essential role in the classification of patients, as the probability of the occurrence of parity was high in two classes. However, it should be noted that among healthy women, parity was more likely to be a warning signal for developing risk factors for this cancer in healthy women in the moderate risk and semi moderate risk classes.

Our findings indicate that skipping of breastfeeding was associated with development of breast cancer. This factor also was likely to occur among breast cancer patients in the moderate risk and high risk classes. The results of a review study showed that breastfeeding significantly reduced the risk of breast cancer in young women(28). Breastfeeding period has been also shown to be associated with hormonal changes and changes in breast tissue, which may reduce the risk of breast cancer, reduce the menstrual cycle and ovulation timing through reproductive life, resulting in reduced exposure to rapid hormonal changes(29, 30). However, skipping breastfeeding was more likely in healthy women in the moderate risk and semi moderate risk classes, thus the co-occurrence of this variable with other risk factors can increase the risk of breast cancer. Probably, the interaction and simultaneous influence of several factors is required for the disease occurrence.

We found that being single was associated with breast cancer. Many studies have proposed various reasons indicating married patients as a better prognosis of breast cancer. As such, these people generally have richer financial resources that support them to perform examinations earlier and more frequent. In addition, in married women, due to having more psychosocial support compared to singles, a better prognosis of the disease was observed, as reduced psychological support and stress is associated with tumor progression and immune deficiency(31, 32). In addition, single women may have poor health
habits and behaviors. For example, the results of a study showed that smoking and poor health examinations were more common in single women than in married women(33).

**Conclusion**

In this study, we reported the prevalence and pattern of risk factors of breast cancer by subgrouping a sample of Iranian women into four classes. Results of this study showed that the majority of healthy and diseased women fell under the latent class of general population risk. It should be noted that we found a high risk class among diseased women. However, among healthy women, there was not a high risk class. Among healthy women, 15.4% of them fell under moderate risk and semi moderate risk classes, which stresses the necessity of implementing educational and preventive intervention for this stratum of women.

**Abbreviations**

LCA
latent class analysis; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion.

**Declarations**

**Authors’ contributions:**

All authors contributed to the study conception and design. AAG analyzed the data and wrote the manuscript. MS, AR, SA, YA, and AAG collected the data and wrote manuscript. MS, YA and SA interpreted the data. All authors have read and approved the final manuscript.

**Founding:**

This research was supported by grant from Shiraz University of Medical Sciences. This university had no role in the study design, analysis, interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

**Availability of data and materials:**

The datasets generated during the current study are not publicly available since they will contain patient data and the informed consent agreement does not include sharing data publicly. However, the data are available from the corresponding author on reasonable request.

**Ethics approval and consent to participate:**

The study was approved by the Ethics Committee of Shiraz University of Medical Sciences (IR.SUMS.REC.1391.S6422). Permission to conduct the study was obtained from this committee.
Consent for publication:

Not applicable.

Competing interests:

None.

Author details:

1 Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran.

2 Department of Epidemiology and Biostatistics, Faculty of Health, Baqiyatallah University of Medical Sciences, Tehran, Iran.

3 Breast Diseases Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.

4 Student Research Committee, Shiraz University of Medical Sciences, Shiraz, Iran

5 Department of Epidemiology, school of health, Shiraz University of Medical Sciences, Shiraz, Iran

6 Pars Advanced and Minimally Invasive Medical Manners Research Center, Pars Hospital, Iran University of Medical Sciences, Tehran, Iran

7 Department of Public Health, School of Health, Ardabil University of Medical Sciences, Ardabil, Iran

References


