RESEARCH ARTICLE

Clinical Risk Factor Analysis for Breast Cancer: 568,000 Subjects Undergoing Breast Cancer Screening in Beijing, 2009

Lei Pan1, Li-Li Han2, Li-Xin Tao1, Tao Zhou1, Xia Li1,3, Qi Gao1, Li-Juan Wu1, Yan-Xia Luo1, Hui Ding2*, Xiu-Hua Guo1*

Abstract

Objectives: Although there are many reports about the risk of breast cancer, few have reported clinical factors including history of breast-related or other diseases that affect the prevalence of breast cancer. This study explores these risk factors for breast cancer cases reported in Beijing in 2009. Materials and Methods: Data were derived from a Beijing breast cancer screening performed in 2009, of 568,000 women, from 16 districts of Beijing, all aged between 40 and 60 years. In this study, multilevel statistical modeling was used to identify clinical factors that affect the prevalence of breast cancer and to provide more reliable evidence for clinical diagnostics by using screening data. Results and Conclusion: Those women who had organ transplants, compared with those with none, were associated with breast cancer with an odds ratio (OR) = 65.352 [95% confidence interval (CI): 8.488-503.165] and those with solid breast mass compared with none had OR = 1.384 (95% CI: 1.022-1.873). Malignant tendency was strongly associated with increased risk of breast cancer, OR = 207.999 [95% CI: 151.950-284.721]. The risk of breast cancer increased with age, OR = 2.759 (95% CI: 1.837-4.144, 56-60 vs. 40-45), OR = 2.047 (95% CI: 1.394-3.077, 51-55 vs. 40-45), OR = 1.668 (95% CI: 1.145-2.431). Normal results of B ultrasonic examination show a lower risk among participants, OR = 0.136 (95% CI: 0.085-0.218). Those women with ductal papilloma compared with none were associated with breast cancer, OR=6.524 (95% CI: 1.871-22.746). Therefore, this study suggests that clinical doctors should pay attention to these high-risk factors.

Keywords: Multilevel statistical model - breast cancer screening - risk factors

Asian Pac J Cancer Prev, 14 (9), 5325-5329

Introduction

Breast cancer is one of the most common malignancies of women in the world today. Of every four deaths in Beijing households since 2007, one woman dies of cancer, and cancer has been ranked the leading cause of death in Beijing residents. Among common tumors, breast cancer ranked first in Beijing women from 2000 to 2005, the prevalence of breast cancer is increasing year by year, and the risk of breast cancer has an average growth rate of 16.14% (Liu et al., 2006). Several studies in the past have shown that active smoking may play an important role in breast cancer etiology (Reynolds et al., 2004), and a relationship exists between alcohol consumption and breast cancer (Hamajima et al., 2002). Most of the well-known risk factors for breast cancer are related to the reproductive life of women (Ross et al., 2000; Becher et al., 2003; Ravichandran et al., 2009; Schonfeld et al., 2011): early menarche, nulliparity or number of children, late age at first birth, shorter duration of breastfeeding, and late menopause. Moreover, there is excess mortality from breast and ovarian cancer among teachers, nurses, secretaries, librarians, retail sales clerks, and religious workers (MacArthur et al., 2007). It is also well known that genetic mutations in BRCA1 or BRCA2 have been identified as breast cancer risk factors (Bordeleau et al., 2011; Caruso et al., 2011). Although case control studies and cohort studies have already focused on the relationship between risk factors and breast cancer, these risk factors are mostly about lifestyle and reproductive or food factors (John et al., 1999; Verloop et al., 2000; Lillberg et al., 2003; Taylor et al., 2007); therefore, a large-scale breast cancer screening (Zakharova et al., 2011) was performed in 16 districts in Beijing, which aimed at early detection of patients with breast cancer for early treatment; furthermore, it aimed at identifying clinical factors including history of disease and breast-related diseases that affect the prevalence of breast cancer and at providing more reliable evidence for clinical diagnostics by using screening data.

The multilevel statistical model (McMahon et al., 2006) was first used in the field of pedagogy and then used in psychology, sociology, economics, organizational behavior, management science, and other fields, and was gradually applied in the fields of medicine and public health.

1Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, 2Beijing Obstetrics and Gynecology Hospital, Beijing, China, 3Department of Epidemiology and Public Health, University College Cork, Fourth Floor, Western Gate Building, Cork, Ireland 4Equal contributors *For correspondence: statguo@ccmu.edu.cn, dinghui1107@sina.com
Materials and Methods

Subjects

Fourteen districts and two counties of Beijing were included in the screening: Dongcheng District, Xicheng District, Chaoyang District, Haidian District, Fengtai District, Shijingshan District, Mentougou District, Fangshan District, Daxing District, Tongzhou District, Shunyi District, Pinggu District, Huairou District, Miyun County, Changping District, and Yanqing County. Data were collected from women between April 1, 2009, and November 30, 2009, who were between the ages of 40 and 60 years, by providing free breast examinations. In addition, informed consent was obtained from all potentially eligible participants. A total of 568,000 data points were collected when screening was completed on a voluntary basis.

Study content and methods

The large-scale screening surveyed basic population information including age, occupation, educational level, disease history, and breast-related diseases. We had great interest in whether disease history and breast-related diseases may contribute to a woman’s risk of developing breast cancer. We used ultrasound method B for breast screening. Screening of positive cases found in all suspected medical institutions would have been designated as referral diagnoses. All participants in the screening were examined by a unified medical staff trained in the screening examination.

Dependent variables and measurement of covariates

Breast cancer was the dependent variable in this study. Disease history and breast cancer related diseases were dichotomous variables. After approval by each subject’s physician, potential participants were interviewed by a trained interviewer, using a standardized, structured questionnaire to obtain information on well-established risk factors. Education level was categorized into three groups: junior high school and below, high school or college, and university level and above. Pregnancy frequency was collected as 0–3 times and >3 times. Education level was categorized into three groups: junior high school and below, high school or college, and university level and above. Pregnancy frequency was collected as 0–3 times and >3 times. Breast cancer risk factors were analyzed by age group at diagnosis (40-45, 46-50, 51-55, and 56-60 years).

Diagnosis of breast cancer

We examined the breast and axillary lymph nodes using ultrasound B-scans; furthermore, positive cases were examined by X-ray and biopsy. Histopathologic diagnosis results showed that precancerous breast lesions indicated breast tissue dysplasia or BIDP. Early breast cancer was indicated by LCIS (lobular carcinoma in situ), intraductal carcinoma in situ, early invasive carcinoma with point like basement membrane, and one breast cancer with a tumor diameter of less than or equal to 0.5 cm.

Multilevel statistical models

Many kinds of data, including observational data collected in the human and biological sciences, have a hierarchical or clustered structure. The existence of such data hierarchies is neither accidental nor ignorable. Nevertheless, classical statistical models assume that individuals and random error terms are independent, which is apparently not suitable for the above-mentioned data (Haneuse et al., 2011).

In this study, we refer to a hierarchy as consisting of units grouped at different levels. Thus participants are the Level 1 units in a two-level structure while the Level 2 units are the districts.

All multilevel statistical models were fit using PROC GLIMMIX. All analyses were completed using SAS 9.2(SAS Institute Inc., Cary, NC, USA). P < 0.05 indicates statistical significance.

Results

Years after May 2009, a total of 568,000 Beijing women attended a free breast screening, and 266 cases of breast cancer were reported, with a detection rate of 46.83/100,000; the distribution of participants’ characteristics was examined by χ² or Fisher’s exact test. There was a significant difference between the detection rate of various districts and counties (χ² = 94.355, P < 0.001), and the highest detection rate of Yanqing County, up 148.91/100,000, far exceeded the average detection rate in Beijing. The prevalence of breast cancer in each district is shown in Figure 1.

The detection rate for all age groups was significantly different (χ² = 14.082, P = 0.003), and the 56- to 60-year age group had the highest detection rate of 58.60/100,000: of the 105,805 participants, 62 women developed breast cancer during the study period. The lowest detection rate was in the <45 year age group, 45 people in 151,362 attended a free breast screening, and 266 cases were confirmed.

We observed that the occurrence of breast cancer was related to education level, and the detection rate increased with increase in education level (χ²= 6.423, P=0.040), and

it is possible that the stress that higher-educated women experience in their daily work may be greater than that of lower-educated women, making it difficult to adjust for this and resulting in a higher detection rate of breast cancer in those women. For those with a Bachelor’s degree or above (48,319), 34 were detected with breast cancer, with a detection rate of 70.37/100,000, whereas for those with a secondary education and below, 169 in 384,940 were reported to have breast cancer with a detection rate of 43.90/100,000; and the 63 patients with a high school diploma had a detection rate of 46.76/100,000 (Table 1). Agency personnel had the highest detection rate of breast cancer, up to 61.89/100,000. The lowest detection rate was 24.75/100,000, and an insignificant difference was found among various occupations ($\chi^2 = 13.120, P = 0.069$; Table 1).

The prevalence of relevant medical history (including malignant tumor, breast cancer, breast mass) in each age group was significantly different, with P values of less than 0.001, and the largest proportion (56- to 60-year olds) suffered from malignancies in the ratio 135.15/100,000. The history of breast cancer prevalence was also highest in the largest age group of 56- to 60-year olds (382.19/100,000). The women who had a higher distribution of breast masses than others are in the 46- to 50-year olds group (3799.07/100,000; Table 2).

In women aged 45 years and below, it was easier to report cystic mass (2,412.76/100,000), breast hyperplasia (30.99%), and fibroadenoma (796.10/100,000). Moreover, more women between the ages of 46 and 50 years reported solid masses (5,516.10/100,000) and other benign diseases (87.65/100,000; Table 2).

We performed a multilevel logistic regression analysis to assess the effect of history of diseases and breast-related diseases on the risk of breast cancer in the Beijing Women’s Free Breast Screening Program. Risks were estimated using two-level logistic regression analysis adjusted for known confounders. Those women with a previous history of organ transplantation (compared with those without) were associated with breast cancer, OR = 65.352 (95% CI: 8.488-503.165), as were those women associated with solid masses compared with none, OR = 1.384 (95% CI: 1.022-1.873). Malignant tendency was strongly associated with increased risk of breast cancer, OR = 207.999 (95% CI: 151.950-284.721). The risk of breast cancer also increased with age, OR1 = 2.759 (95% CI: 1.837-4.144, 56-60 vs. 40-45), OR2 = 2.047 (95% CI: 1.394-3.077, 51-55 vs. 40-45), OR3=1.668 (95% CI: 1.145-2.431). It was found that normal ultrasonic B-scan results indicated a lower risk among participants, OR = 0.136 (95% CI: 0.085-0.218). Those women with ductal papilloma compared with those without were associated with breast cancer, OR = 6.524 (95% CI: 1.871-22.746). (Table 3)

**Discussion**

We performed a two-level logistic regression analysis to assess the effect of history of diseases and breast-related diseases on the risk of breast cancer in the Beijing Women’s Free Breast Screening Program for the first time which eliminating area clustered and the large sample size (n = 568,000) guarantee the reliability of detection rate of breast cancer. The advantage of using multilevel modeling is that it takes the hierarchical structure of the data into account by specifying random effects at each level of analysis, and thus results in a more conservative inference for the aggregate effect (Wang et al., 2010).

In benign breast tumors, breast lumps are not

---

**Table 1. The Distribution of Detection Rates for Different Factors**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Screening Detected</th>
<th>Detected Detected Proportion $\chi^2$ $P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-45</td>
<td>151,362</td>
<td>45</td>
</tr>
<tr>
<td>46-50</td>
<td>156,306</td>
<td>83</td>
</tr>
<tr>
<td>51-55</td>
<td>154,527</td>
<td>76</td>
</tr>
<tr>
<td>56-60</td>
<td>105,805</td>
<td>62</td>
</tr>
<tr>
<td>Total</td>
<td>56,8000</td>
<td>266</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior high school and below</td>
<td>384,940</td>
<td>169</td>
</tr>
<tr>
<td>High school or college 134,741</td>
<td>63</td>
<td>46.76</td>
</tr>
<tr>
<td>University and above 48,319</td>
<td>34</td>
<td>70.73</td>
</tr>
<tr>
<td>Total</td>
<td>56,8000</td>
<td>266</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agency personnel</td>
<td>12,927</td>
<td>8</td>
</tr>
<tr>
<td>Technical staff</td>
<td>15,120</td>
<td>9</td>
</tr>
<tr>
<td>Staff</td>
<td>10,821</td>
<td>6</td>
</tr>
<tr>
<td>Service personnel</td>
<td>20,201</td>
<td>5</td>
</tr>
<tr>
<td>Agricultural workers 208,911</td>
<td>79</td>
<td>37.82</td>
</tr>
<tr>
<td>Equipment operators 63,536</td>
<td>26</td>
<td>40.92</td>
</tr>
<tr>
<td>Others</td>
<td>26,161</td>
<td>14</td>
</tr>
<tr>
<td>Unemployed</td>
<td>182,598</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>540,021</td>
<td>254</td>
</tr>
</tbody>
</table>

---

**Table 2. The Distribution of History of Disease and Breast-related Diseases in Each Age Group**

<table>
<thead>
<tr>
<th>Age groups (years)</th>
<th>$\leq$45</th>
<th>46-50</th>
<th>51-55</th>
<th>56-60</th>
<th>$\chi^2$ $P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of disease</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malignant tumor</td>
<td>91/151362 (60.12)</td>
<td>154/156306 (98.52)</td>
<td>156/154527 (100.95)</td>
<td>143/105805 (135.15)</td>
<td>37.814 &lt;0.001</td>
</tr>
<tr>
<td>Organ transplantation</td>
<td>10/151362 (6.61)</td>
<td>22/156306 (14.07)</td>
<td>17/154527 (11.00)</td>
<td>10/105805 (9.45)</td>
<td>4.275 0.233</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>193/122347 (157.75)</td>
<td>495/130374 (3799.07)</td>
<td>423/125287 (3377.05)</td>
<td>237/84917 (2800.38)</td>
<td>160.925 &lt;0.001</td>
</tr>
<tr>
<td>Breast mass</td>
<td>4401/124185 (3543.91)</td>
<td>4953/130374 (3799.07)</td>
<td>4231/125287 (3377.05)</td>
<td>2378/84917 (2800.38)</td>
<td>160.925 &lt;0.001</td>
</tr>
</tbody>
</table>

---

**Table 3.**
uncommon, with the most common being breast fibroadenoma. These tumors are common in young women over 40. Solid tumors often have a tough quality, like a complete capsule, are smooth and slippery to the touch, are dynamic, and generally do not adhere to the skin or cause nipple retraction. Intraductal papillomas are tumors which are often very small, easily palpable, and slightly larger than those in palpable nodules around the areola, and discharge from the nipples is the main clinical symptom. Intraductal papilloma has a risk of breast cancer in patients compared with no risk of catheter (OR = 6.524, 95% CI: 1.871-22.746). When clinicians diagnose patients with breast duct papilloma, the prognosis of patients should be closely observed if they are more likely to develop breast cancer in the future (Inumaru et al., 2011). Solid mass in the breast is also a kind of benign breast tumor. Solid breast mass in patients carries a higher risk of breast cancer compared with no solid mass (OR = 1.384, 95%CI: 1.022-1.873), suggesting that clinicians who discover solid breast masses in patients should be alert to the risk of breast cancer (Satake et al., 2011). Organs from those whose occupations were staff and government had the highest detection rate of breast cancer (Satake et al., 2011). Malignant tumor 0.9201 0.8181 2.12 0.0356 1.384 1.022 1.873. The present study presents several strengths, among which are the breast screening design, the large sample size (n = 568,000), and the detailed information regarding many risk factors. However, some limitations should be addressed. We actually put 263 patients instead of 266 into the model because some of the independent variables had missing information. Although we included a large set of risk factors, we did not account for genetic factors. It is well known that genetic mutations in BRCA1 or BRCA2 have been identified as breast cancer risk factors (Warner et al., 2011). However, we adjusted for breast cancer, compared with those who had an abnormal ultrasound (OR = 0.136, 95%CI: 0.085-0.218).

In our research, we did not put variables such as reproductive factors, including the age when menstruation began and the number of days in the menstrual cycle, into the two-level logistic regression. Others reported that a significant association was observed between early onset of menarche and risk of luminal disease (Millikan et al., 2008). Moreover, there were no significant differences associated with other reproductive factors such as parity, age at first live birth, breastfeeding history, age at menopause, or synthetic hormone use (Yanhua et al., 2012; Amaro et al., 2013).

The present study presents several strengths, among which are the breast screening design, the large sample size (n = 568,000), and the detailed information regarding many risk factors. However, some limitations should be addressed. We actually put 263 patients instead of 266 into the model because some of the independent variables had missing information. Although we included a large set of risk factors, we did not account for genetic factors. It is well known that genetic mutations in BRCA1 or BRCA2 have been identified as breast cancer risk factors (Warner et al., 2011). However, we adjusted for breast cancer history among relatives indirectly, accounting for this risk factor in a large scale of breast screening, but it is not significant in the model. Finally, this was a large-scale distribution of breast screening in 16 counties of Beijing, each screening unit had high and low levels of technology, therefore, there may be undetected cancer patients or some false-positive results (Mai et al., 2009).

**Acknowledgements**

This Study was supported by the Program of Beijing Municipal Science & Technology Commission: Common Cancer Screening of Women (Serial Number: D09050703570901).
This work was supported by the Beijing Obstetrics and Gynecology Hospital and by maternity and child care centers at all levels in Beijing. In addition, we would like to thank all participants for eagerly fulfilling the breast screenings. The author(s) declare that they have no competing interests.

References


