Assessing an Emerging Nationwide Population-based Mammography Screening Program in Taiwan

Huay-Ben Pan¹,²,³ Giu-Chen Hsu⁴ Huei-Lung Liang¹,² Chen-Pin Chou¹,² Yen-Chi Wang¹,² San-Kan Lee⁵ Yi-Hong Chou²,⁶ Kam-Fai Wong⁷

Department of Radiology¹, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan
National Yang-Ming University², Taipei, Taiwan
Department of Medical Imaging and Radiological Sciences³, I-Shou University, Kaohsiung, Taiwan
Department of Radiology⁴, Tri-Service General Hospital, Taipei, Taiwan
Department of Radiology⁵, Taichung Veterans General Hospital, Taichung, Taiwan
Department of Radiology⁶, Taipei Veterans General Hospital, Taipei, Taiwan
Institute of Statistics⁷, National University of Kaohsiung, Kaohsiung, Taiwan

ABSTRACT

Breast cancer is the most common cancer and ranks fourth among cancer deaths in women in Taiwan. In 2004, Taiwan’s Bureau of Health Promotion, Department of Health and Radiology Society of Republic of China collaborated to implement a program to monitor the quality and trained the people involved in a nationwide biennially mammography screening program for women between 50 and 69 years of age, which was started from a small population trial in 2002. In this study, we assessed the effects of this program by comparing evaluations of the women’s mammographic images obtained from Taiwan’s Bureau of Health Promotion in Taiwan with whether or not they were listed on Taiwan’s National Cancer Registry from 2004 to 2007, to which all pathology confirmed tumors are reported. A total of 311,193 consecutive mammograms were performed over the four-year period. Although prevalence of confirmed breast cancer was found to 0.63% in the mammograms evaluated the first year (2004), our later findings suggested that actual prevalence to be about 0.48%. By 2007, screening was improved as evidenced by an increased positive predictive value of an abnormal mammography (4.4%), a decreased recall rate for further work-ups (9.3%) and improved sensitivity (84.9%). Sensitivity was unusually high the first year of screening (86.1% in 2004) because of lack of mammography specialists training and the confounding of screening mammography results with diagnostic mammography results. Recall rate was high in the second year of screening (11.9% in 2005). We concluded that the mammography screening program was improved by our quality assurance and education program and would improvement would continue as these efforts continue and mammography-audit recommendations are adjusted based on the performance indicators we studied. Our results may help focus our future training programs and serve as valuable reference for other regions that are implementing population-based mammography screening programs.

The incidence and mortality rate of breast cancer, already the most common cancer among women worldwide, continues to increase. It ranks fourth among cancer deaths among women in Taiwan where the number of cases jumped from 2,838 in 1995 to 6,895 in 2006, an 82% increase in incidence. There the number of breast cancer deaths also
increased by 14.4% from 918 in 1995 to 1,552 in 2007 [1]. It has been found in regions that do not have such programs in place, the mortality rate due to breast cancer is higher [2]. Therefore, these increases suggest efficient and accurate early screening is becoming more important. Mammography is the only tool that has been found to detect early asymptomatic breast cancer and has been found to increase survival rate by as much as 30% [3-6]. However, the accuracy of mammograms are affected by how they are used to make a diagnosis by the experience of the radiologists in recognizing and handling low positive rate of mammographic findings suggesting possible cancer. For example, in one study of a country’s screening program, it was found that the sensitivity and specificity of that program could be improved by providing special training to radiologists to improve their assessment of mammographic images. The proportion of cancers with definite preoperative diagnosis increased from 33.5% to 91.0% with structured radiological training from 2000 to 2005 [7]. Therefore, effective of screening not only depends on whether or not one receives a mammogram with good resolution but also on how it is accessed, i.e., quality of screening and assessment.

Since 2002, all women in Taiwan between the ages of 50 and 69 years old can receive a free mammographic examination every two years. In 2004, to improve the effectiveness of this program, Taiwan implemented a program to ensure the quality of equipment and experience and training of the radiologists and mammographers participating in the program. With such a large nationwide program, it is important to assess the effects of such a program and judge its effectiveness and identify areas that may need improvement. To do this, using the database of the Bureau of Health Promotion in Taiwan, we compared the screenee’s mammographic reports with whether or not they were listed on Taiwan’s National Cancer Registry from 2004 to 2007, to which all pathology confirmed tumors are reported. The findings of such a study may provide valuable information for future policy decisions, and quality assurance and training programs.

MATERIALS AND METHODS

Data collection

The study was approved by the institutional review board at Veterans General Hospital Kaohsiung and was performed according to the Declaration of Helsinki principles. We collected data from the results of a free nationwide population-based mammography screening program funded and coordinated by Taiwan’s Bureau of Health Promotion. Our study period was from July 2004 to December 2007. It was in July 2004 that the Bureau of Health Promotion and the Radiology Society of Republic of China (RSROC) implemented a quality control program to a mammography screening program that was begun 2002. In that program, women 50 to 69 years old were self-referred or referred from nationwide community health centers to receive mammography screening every two years. The information obtained from the quality promotion program would be used as feedback for radiologic technologists and radiologists to improve their methods performing and interpreting of mammograms

Strategy of screening mammography

In this nationwide screening program, asymptomatic women 50 to 69 years old received mammographic examinations at 138 participating hospitals. These screening women had to be checked at least 24 months after any other mammogram examinations to fit the criteria of screening mammography. Each mammographic examination included mammograms taken from both a mediolateral oblique (MLO) view and a craniocaudal (CC) view.

The assessments of the mammograms were performed by 293 board-certified radiologists with specific training in screening mammography. Mammography assessments were based on the Breast Imaging Reporting and Data Systems (BI-RADS) categorization system established by the American College of Radiology (ACR). In that system, category 0 represents need for further study, and 1 to 5, represents increasing risk of malignancy. Women with more than one lesion received only one BI-RADS assessment, the highest of the BI-RADS category was recorded. The examination reports included breast density, assessment category, and specific recommendations for follow-up. Women with BI-RADS category 0 were recommended or recalled to receive further diagnostic procedures, e.g., spot compression magnification or ultrasound examinations.

Definitions

Negative mammography: a mammogram classified as BI-RADS category 1, 2, or 3.
Positive mammography: a mammogram classified as BI-RADS category 0, 4 and 5.
Recall rate: the percentage of those receiving BI-RADS category 0, 4 and 5, who were recalled to the hospital for further study.
True-positive (TP) and false-positive (FN): defined by whether or not a participant with BI-RADS category 0, 4, of 5 was listed on Taiwan’s National Cancer Registry within 12 months of the mammography.
True-negative (TP) and false-negative (FN): defined by whether or not a participant with BI-RADS category 1, 2, or 3 were not listed on the National Cancer Registry within the same period.
Positive predictive value of recall (PPV1): the ratio of number of women who received BI-RADS category 0, 4, or 5 who were actually listed on the National Cancer Registry.
Breast cancer prevalence: rate of women who received a first screening who were actually listed on the National Cancer Registry, regardless of their BI-RADS scores.
Breast cancer incidence: rate of women who had a previous negative screening mammography who were actually listed on the National Cancer Registry, also regardless of their BI-RADS scores.

Statistical Analysis
The PPV1 and rates for false negative, recall, true positives, true positives plus false negatives, sensitivity, and specificity were calculated for the 2004 to 2006. Because the National Cancer Registry has not been published for 2007 as of the end of 2010, the 2007 values for false negative, breast cancer incidence, sensitivity of screening mammography, and specificity were estimated based on Baye’s rule and conditional probability. Cancer incidence for that year was estimated based on the previous year’s figures. Due to size of the population we study, we had a power of 99% at the 1% type I error rate for detect small differences, and thus there was little need to tests of statistical significance.

RESULTS

Patient population
A total of 311,193 mammography examinations for nationwide population-based breast cancer screening of women 50 to 69 years old were performed over the study period (2004-2007) (Table 1). The percentage of the total population of women in this age groups increased by at least one percent each year during the study period (range 0.85 to 5.48).

Indicators of mammography screening
Table 2 is a summary of the important indicators of mammography screening. During the four-year period, a total of 1359 of the mammograms were pathology proved to be breast cancer (true positives: 1283; false negatives: 176). Prevalence the first year, listed under incidence in Table 2, was 0.63%. The incidence rates for the following three years ranged between 0.48 to 0.49%. Cancer detection rate the first year was 0.54% and ranged between 0.40 to 0.41% the following three years. The positive predictive value of an abnormal mammography was 5.2% the first year and 3.4%, 4.1%, and 4.4% the next three years. Recall rate was 10% the in 2004, spiked to almost 12% in 2005 and then dropped to around 10% and 9% in 2006 and 2007, respectively. Sensitivity was highest in 2004 (86.1%) and lowest in .8191 in 2005, and rose moderately in 2006 and slightly in 2007 (83.9% and 84.9%, respectively. The most common breast densities were either type 2 or 3.

National Cancer Registry Report of Pathology Results
An annual average of 36% cases with pathologically proved breast cancer had final surgical staging of early breast cancer (carcinoma in situ or small invasive ductal carcinoma less than 1cm). From 2005 to 2007, they were 33.33%; 43.37%; 38.12% and 31.98% respectively. Our early cancer detection rate is significantly higher than 30%, which is recommended by ACR standard for performing screening mammography. An annual average of 58% cases with pathologically proved breast cancer had negative rate of axillary lymph nodes. From 2005 to 2007, they were 59.70%; 60.84%; 55.22% and 58.99% respectively. Our negative rate is significantly lower than 75%, which is recommended by ACR standard.

DISCUSSION

This study found that there was a great variance in PPV1 the first two years (2004, 5.2%; 2005, 3.4%), which leveled off to 4% the last two years. It found that recall rates, which ranged between nine and ten percent, were much higher the second year (almost 12%). It also found that sensitivity, which was unusually high the first year (86.1%), dropped to 81.9% the second year, and leveled out to about 84% the last two years. In addition, we found a higher PPV1 and incidence rate of pathologically proved cancer the first year and a decrease in false negatives starting in 2006. The specificities were all higher than 88.5%.

The nationwide screening rate of Taiwan’s total population of women between 50 and 69 years old steadily increased from 17,272 of 2,022,452 women (0.85%) to 153,870 of 2,410,281 women (6.38%). This study analyzed data collected from 2004 to 2007 nationwide screening.

Table 1. The records of mammography screening from year 2004 to year 2007

<table>
<thead>
<tr>
<th>Year</th>
<th>Total number in population</th>
<th>Number of screening (ratio)</th>
<th>Number of recall</th>
<th>Number of true positive</th>
<th>Number of false negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>2,022,451</td>
<td>17,272 (0.85%)</td>
<td>1,786</td>
<td>93</td>
<td>15</td>
</tr>
<tr>
<td>2005</td>
<td>2,117,198</td>
<td>61,897 (2.92%)</td>
<td>7,365</td>
<td>249</td>
<td>55</td>
</tr>
<tr>
<td>2006</td>
<td>2,220,218</td>
<td>105,411 (4.75%)</td>
<td>10,286</td>
<td>425</td>
<td>81</td>
</tr>
<tr>
<td>2007</td>
<td>2,310,669</td>
<td>126,613 (5.48%)</td>
<td>11,751</td>
<td>516</td>
<td>25*</td>
</tr>
</tbody>
</table>

*Not yet completely observed
We found proportionate increases in the detection of occult breast cancer, false negatives, and true positives. (Table 1, 2). We found a higher PPV1 in the first year of screening than following years. However, during the first year, due to possible less orientation and training programs for primary care givers, the results were confounded by the inadequate inclusion of mammograms of clinically suspicious breast cancer cases, e.g., those self-referring themselves for screening mammography because they found breast lumps. The contaminations of symptomatic women may also reflect an increased breast cancer detection rate, an increased incidence of breast cancer from screening women, and the high sensitivity of screening mammography in the first year (Table 1). In the first year of screening, RSROC gave continuous education programs for screening providers and set up on-line check-up for mammogram records. In the following three years, more women with clinically suspicious breast cancer were excluded from our screening program. The cancer incidence rate in our first and second years of screening mammography actually indicated the prevalence of breast cancer in the Taiwan’s women and the incidence rate of first and second years should be higher than the incidence rate of subsequent screening[8]. In a population undergoing annual screening, incidence screenings are much more representative because after the first prevalence screening and incidence screenings are the ones occurring over the remainder of the lifetime of the patient [9]. The reason of lower PPV1 in the second year may be explained by that we excluded the patients with clinically suspicious breast cancer. Therefore, the second year prevalence (0.49%) may be more realistic figure.

Because the program enrollment policy (no prior mammogram within two years and no clinical breast symptoms) may not have been fully followed by primary care providers who included diagnosis mammograms and other insufficiency of complementary measurement in 2004, the 2005 prevalence rate (0.49%) may be more realistic, not only because diagnostic mammograms were excluded but also because the participating radiologists were involved in continuing education programs and had one more year of experience. For that year, although there was a higher recall rate, there were no increases in true positives or false positives. Sensitivity slightly declined. At the beginning of the 2004 screening program, the participating radiologists were still in the process of learning to interpret screening mammography, had much lower positive detection rates than actual incidence and prevalence of pathologically proved cancer. In response, a series of continuing education efforts were begun to improve the rate since 2004. While the gap between positive detection rates and incidence narrowed, the recall rate increased sharply. The increase in recall probably resulted from the possibility that the interpreters trying to improve their detection rates over scrutinized such subtle findings such as asymmetry tissue and asked that the participant be called back for further study. The result was increase unnecessary callbacks. There was also an increase in the false positive rate and a large decrease in PPV1 (Table 2).

Once the participating radiologists had attended continuing education programs since 2004, they could have better consensus about BI-RADS categorizing system. Dense breast is common in Asia’s women, and some radiologists miscategorized dense breast as BI-RADS category 0 and recommended further work-up with breast ultrasound. After radiologists understood categorization of BI-RADS category 0, overcompensation effect in recall that occurred, PPV1, and sensitivity in 2005 was corrected in 2006 (2005 vs. 2006: 0.1190 vs. 0.0976; 0.0338 vs. 0.0413; and 0.8191 vs. 0.8399, respectively) (Table 2). The fact that recall rate was reduced by more than 2% and that sensitivity was

<table>
<thead>
<tr>
<th>Year</th>
<th>PPV1</th>
<th>False negative rate from screening</th>
<th>Recall rate</th>
<th>Cancer detection rate from screening</th>
<th>Cancer incidence rate from screening</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>0.0521</td>
<td>0.0010</td>
<td>0.1034</td>
<td>0.0054</td>
<td>0.0063</td>
<td>0.8611</td>
<td>0.9014</td>
</tr>
<tr>
<td>2005</td>
<td>0.0338</td>
<td>0.0010</td>
<td>0.1190</td>
<td>0.0040</td>
<td>0.0049</td>
<td>0.8191</td>
<td>0.8845</td>
</tr>
<tr>
<td>2006</td>
<td>0.0413</td>
<td>0.0009</td>
<td>0.0976</td>
<td>0.0040</td>
<td>0.0048</td>
<td>0.8399</td>
<td>0.9060</td>
</tr>
<tr>
<td>2007</td>
<td>0.0439</td>
<td>0.0008*</td>
<td>0.0928</td>
<td>0.0041</td>
<td>0.0048**</td>
<td>0.8490*</td>
<td>0.9108*</td>
</tr>
</tbody>
</table>

*Values assume to be same as the previous year.
**Value calculated according to Baye’s rule.

Cancer incidence rate = 1000 X number of cancer / person-time at risk = Delay-adjusted cancer incidence rates

Cancer detection rate = 1000 X Number of true positive / Total screening cases

Cancer incidence rate from screening = Number of cancer (true positive + false negative) / Total screening cases that regardless of their BI-RADS categories.

Recall rate = Cancer detection rate from screening / Number of cancer

**Table 2. Important indicators of mammography screening from year 2004 to year 2007**
increased indicated that the training program offered by RSROC improved the screening program.

Although outcomes did not change appreciatively in 2007, compared to 2006, the outcomes in 2007 did not change appreciatively, though there was a slight increase in PPV1 and decrease in recall rate (0.0413 vs. 0.0439 and 0.0976 vs. 0.0928, respectively), suggesting continued improvements in the quality of screening mammography program and continued improvements in the interpretative skills of the radiologists. The false negative cases have not completely obtained by delay response of results [10]. Complete Cancer Registry Data (pathology proved cases) have not been released on December 2011, so we estimated the rate of false negatives for 2007. Due the relatively short period, it was assumed that the cancer detection and incidence rates would remain stable. PPV1 and recall rate for 2007 could be observed through medical audits for screening mammography results. These two rates could be used to predict the sensitivity and specificity by assuming that cancer incidence rate of 2007 was the same as 2006 and performing the following calculation: sensitivity = PPV1 × Recall rate/Breast cancer incidence rate from screening. Based on this estimation, sensitivity increased from 84% in 2006 to 84.9% in 2007, suggesting continued, albeit mild improvement.

According to WHO criteria for planning screening programs, a desirable screening test should have a high sensitivity, a high specificity, disease should preferably be highly prevalent in the population, and also considers the expenditure and be continued over extended periods of time [11]. The Quality Determinants of Mammography Guidelines published in 1994 by the Agency for Healthcare Policy and Research have been widely used for implementing breast cancer screening programs. The main monitoring factors of screening quality showed that the mammography screening in Taiwan is a well-organized screening program and measurable goals were almost achieved. First, we found the incidence of breast cancer in Taiwan's screening women is appropriate with the other screened countries[1]. Second, the cost of screening supported by Taiwan Bureau of Health Promotion and the cost of screening would be economically balanced in related to more cases with early breast cancer detection and less expenditure on medical care for advanced breast cancer. Third, the additional work-ups and cost for recall cases were also supported by the government named Bureau of National Health Insurance (NIH). Taiwan is the country implement National Health Insurance System to provide health care of citizens. Since 1995, when Taiwan implemented universal national health insurance legislation, coverage has increased from 57% to 98% of the population[12]. Screened positive cases will be transferred to this National Health Insurance System for diagnostic and therapeutic processes. And the last, the quality should be assured with good accuracy by powerful or workable organism or association [13, 14]. There is, take reliable data from public and medical community awareness [15]. The aim of quality assurance program launched by the collaboration of Bureau of Health Promotion and the Radiology Society of Taiwan (RSROC) is expectantly achieving higher rates of early cancer detection, increasing the population coverage and controlling the quality of the screening mammography via continuing education and monitored the performance of equipment, mammographers and radiologists.

The outcomes for mammography have been extensively reported in many countries and several performance benchmarks [16, 17] and updated on an official website [18]. In 2004, Taiwan’s Bureau of Health Promotion collaborated the Radiology Society of Taiwan (RSROC) in an effort to increase the number being screened, improve the quality of the quality of the screening mammography and improve the rates of early cancer detection in an ongoing screening program that had been previously established in 2002. They attempted to do this by monitoring performance of the equipment used and the skill of the technicians and radiologists and offering continuing education courses. Using benchmark indicators, we evaluated the effectiveness of Taiwan's screening mammography program and its quality assurance efforts.

Desirable goals and benchmarks of minimal cancer detection rate and node positively rate for screening mammogram are >30% and <25%, respectively. On average thirty-six percent of the breast cancer within screening programs were found to be DCIS and invasive cancers <1cm (range 31.98% in 2007 to 43.37% in 2005), all above 30 percent and indicating that the mammography screening in Taiwan reached the desirable measurable goals [19]. Although detection of early breast was significant, the rate of negative axillary lymph node finding was far below the ACR threshold. A “M-shape” distribution of breast cancer staging due to good performance of small cancer detection and also advanced cases with ignorant axillary lymph node metastases contaminated in this screening was found. The reason for this discrepancy was mostly related to case enrollment technique or unsatisfied diagnostic accuracy. Alternatively, this discrepancy may indicate that there were still some cases with positive clinical history receiving screening mammography. Further evaluation of the screening program enrollment policy and mammogram interpretation is needed to explain this discrepancy.

This study has several limitations. The false negative rate of screening mammography was defined based on pathological findings obtained from the National Cancer Registry one year after screening only. In fact, more than a year would be needed to determine whether a negative finding would be subsequently pathologically proved to
be positive. There may be a small proportion of our study population made up of symptomatic women, which might explain increased incidence in higher axillary lymph node metastasis. The cancer detection rate was 4.1 per thousand in 2007, higher than it was in 2005 and 2006. This discrepancy may suggest we still have not reached a sample with enough power, due to enrolling less than ten percent of the eligible women in this program, to indicate actual prevalence and that still many of them may have had no prior mammogram. Therefore, studies involving a longer period of observation, a larger proportion of Taiwan’s population of women in this age group, and a more detailed analysis may be needed to adequately address this question.

We will select more eligible women, especially high-risk cases, for screening mammography results in this program. False negative rate averaged around 9.1 per ten thousand the first three years of the study and the estimated false negative rate is 8 per ten thousand for 2007. Therefore, continuing medical education is needed to improve this rate and monitor mammography indicators.

CONCLUSION

In conclusion, our findings indicate a substantial and continued improvement in our emerging national screening mammography program. The results of first two years were confounded by lack of mammogram interpretation experience, an over-emphasis on tumor detection, and some inclusion of diagnostic mammogram results with screening mammogram results, and therefore any conclusions based on these two years should be made with caution. Nevertheless, it remains clear that our efforts to improve the program by modifying enrollment and future training improved outcome measure and our continuous attention to these details will further improve the program. We believe that findings of this study and our continuous study can better focus the direction of our training programs and serve as a reference for other countries seeking establish population-based mammography screening programs as well as other cancer radiology-based cancer screening programs.

ACKNOWLEDGEMENT

This study was partial granted by the Bureau of Health promotion (contrast number: 9910027C), Department of Health, Executive Yuan, Taipei, Taiwan.

REFERENCES