Can missed breast cancer be recognized by regular peer auditing on screening mammography?

Huay-Ben Pan a,b,c, Tsung-Lung Yang a,b, Giu-Cheng Hsu d, Chia-Ling Chiang a, Jer-Shyung Huang a,b, Chen-Pin Chou a,b, Yen-Chi Wang a,b, Huei-Lung Liang a,b,* San-Kan Lee e, Yi-Hong Chou f, Kam-Fai Wong g

* Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan, ROC
b National Yang-Ming University School of Medicine, Taipei, Taiwan, ROC
c Department of Medical Imaging and Radiological Sciences, I-Shou University, Kaohsiung, Taiwan, ROC
d Department of Radiology, Tri-Service General Hospital, Taipei, Taiwan, ROC
e Department of Radiology, Taichung Veterans General Hospital, Taichung, Taiwan, ROC
f Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan, ROC
g Institute of Statistics, National University of Kaohsiung, Kaohsiung, Taiwan, ROC

Received December 9, 2011; accepted March 9, 2012

Abstract

Background: This study was conducted to investigate whether detectable missed breast cancers could be distinguished from truly false negative images in a mammographic screening by a regular peer auditing.

Methods: Between 2004 and 2007, a total of 311,193 free nationwide biennial mammographic screenings were performed for 50- to 69-year-old women in Taiwan. Retrospectively comparing the records in Taiwan’s Cancer registry, 1283 cancers were detected (4.1 per 1000). Of the total, 176 (0.6 per 1000) initial mammographic negative assessments were reported to have cancers (128 traditional films and 48 laser-printed digital images). We selected 186 true negative films (138 traditional films and 48 laser-printed ones) as control group. These were seeded into 4815 films of 2008 images to be audited in 2009. Thirty-four auditors interpreted all the films in a single-blind, randomized, pair-control study. The performance of 34 auditors was analyzed by chi-square test. A p value of <0.05 was considered significant.

Results: Eight (6 traditional and 2 digital films) of the 176 false negative films were not reported by the auditors (missing rate of 4.5%). Of this total, 87 false negatives were reassessed as positive, while 29 of the 186 true negatives were reassessed as positive, making the overall performance of the 34 auditors in interpreting the false negatives and true negatives a specificity of 84.4% and sensitivity of 51.8%. The specificity and sensitivity in traditional films and laser-printed films were 98.6% versus 43.8% and 41.8% versus 78.3%, respectively. Almost 42% of the traditional false negative films had positive reassessment by the auditors, showing a significant difference from the initial screeners (p < 0.001). The specificity of their reinterpretation of laser-printed films was obviously low.

Conclusion: Almost 42% of the false negative traditional films were judged as missed cancers in this study. A peer auditing should reduce the probability of missed cancers.

Keywords: breast cancer; false negative; peer auditing; screening mammography

1. Introduction

Early detection with screening mammography has the potential to reduce breast cancer mortality rates.1,2 To quantify the success of mammography screening in detecting early breast cancer, the mammographic medical audit is recognized
as a means of measuring the interpretive ability of screeners.\textsuperscript{3,4} Large-scale audits are normally conducted by governments,\textsuperscript{5} which have the financial and human resources to organize a structured body for overseeing the audit, segregating the diagnosis from screening data,\textsuperscript{6} and identifying false negative cases,\textsuperscript{7} so that we can set standards for evaluation, including the choice of hospitals, machines, screeners, and the content, frequency, and type of continuing education required for radiologists.\textsuperscript{8} The study of false negative reports contributes greatly to the evaluation of screeners’ ability to recognize signs of early breast cancer and the choice of images for continuing education purposes.\textsuperscript{3,4} However, this approach requires considering the threshold and subthreshold features of malignancy.\textsuperscript{4} If only false negative cases are used to emphasize the occult or subtle nonspecific mammographic findings, this might increase the number of false positive cases, an undesirable tradeoff.

Since 2002, Taiwan has provided free biennial mammography screening for its entire population of women between 50 and 69 years old. It was not until 2004 that the number of participants had increased to the extent that a large sample of films could be studied. Thus, Taiwan’s Bureau of Health Promotion has sponsored a retrospective annual audit of the previous year’s mammography screening. As immediate audits to confirm the readings are impossible, and positive results (the diagnoses of breast cancer) may be delayed for 2 years when they have been recorded on Taiwan’s Cancer Registry, and not only the screeners but also the auditing program need evaluation, thereafter, we conducted a single-blind, randomized, and pair-control study to evaluate the auditing program and identify the detectable but misinterpreted false negatives.

2. Methods

Data on participants in a population-based mammography-screening program were obtained from the Bureau of Health Promotion, which coordinates cancer screening in Taiwan. Each patient receives a standard screening examination consisting of a mediolateral oblique view and a craniocaudal view of each breast. To be qualified to interpret the mammograms, all screeners (board-certified radiologists with specific training in screening mammography) are required to have interpreted at least 1000 mammograms previously. Each screening mammogram is categorized using the Breast Imaging Reporting and Data System (BI-RADS) derived from the American College of Radiology criteria. When there is more than one lesion in a breast, only the highest BI-RADS assessment category is recorded. In addition to lesion category, screeners’ reports also include breast density and specific recommendations for each case. Mammograms assessed as BI-RADS 1, 2, or 3 are defined as negative, while those assessed as BI-RADS 0 (needs additional examination), 4, or 5 are defined as positive.

Our annual auditing design, instead of conducting a full audit, mainly focuses on the assessments of initial screeners who perform below benchmark. According to the stipulation of the Taiwan’s Bureau of Health Promotion, if the initial screener’s recall rate is higher than 20%, the selection will focus on positive assessment to find out avoidable false positive assessments; if the initial screener’s recall rate is less than 7%, the selection will focus on negative assessment to avoid false negative. The audited films are randomly submitted to the 34 auditors. To be auditors, they must have a minimum experience of 5 years, each having interpreted more than 3000 screening mammograms and each rated within the top 25% most accurate readers.

Between 2004 and 2007, there were 311,193 consecutive screenings performed by 230 qualified radiologists, with 1283 cancers detected (4.1 per 1000), while 176 patients (0.57 per 1000) with initial mammographic negative assessments were reported to have cancers in Taiwan’s Cancer Registry within the following 12 months. The latter were categorized as false negatives. These consisted of 128 traditional films (74 BI-RADS 1, 38 BI-RADS 2, and 16 BI-RADS 3) and 48 laser-printed digital images (19 BI-RADS 1; 19 BI-RADS 2; 10 BI-RADS 3). We created two BI-RADS-matched control groups, one using 138 true negative traditional films and the other 48 true negative laser-printed digital images; all of them were not reported to have cancers on the Cancer Registry for at least the following 3 years. The false negative study groups (n = 176) and control groups (n = 186) were then seeded within the 4815 mammograms of 2008 slated to be audited in 2009. The 48 false negative laser-printed films were further reread by three specialists with mammographic experience of more than 10 years with consensus to analyze the imaging features of the undetectable or missed cancers.

All demographic information belonging to the screened women was anonymized and replaced with dummy values to protect healthcare information. The protocol for this study was approved by the institutional review board at Kaohsiung Veterans General Hospital, and the study was performed according to the principles set forth in the Declaration of Helsinki.

The performance of the 34 auditors was analyzed by chi-square test. A p value <0.05 was considered significant. All statistical operations were performed using SPSS version 12.0.

3. Results

Of the 2008 audited films, 75 mammographic films were not responded by the auditors (miss rate of 1.5%). With regard to the 176 seeded false negative study films, six traditional and two laser-printed digital images were not responded (missing rate of 4.5%), leaving us with 168 study films to include in our analysis. Of this total, 87 false negative study films (51.8%) were assessed as positive by the auditors (Table 1), representing a detection rate of 4.4 per 1000 [(1283 + 87)/311,193] if the missed cancers were added. Meanwhile, 29 (15.6%) of the 186 true negative (control) images were assessed as positive, suggesting a significant difference in assessments by the auditors and those by the screeners (p < 0.001).

If the traditional and laser-printed digital study films were analyzed separately, we could find different performance of the auditors between the two films. For the traditional images,
the specificity of the auditors’ assessment of traditional films was 41.8%, with 52.8% false negatives still unrecognized (Table 2). The specificity and sensitivity of the auditors’ reassessments of the laser-printed digital images were 43.8% and 78.3%, respectively (Table 3). The auditors’ ability to interpret the laser-printed digital images showed less significant difference ($p = 0.023$). They reinterpreted 56.3% of the true negative films as positive, a much higher differential rate than their reinterpretations of the non-study and non-control group films, which was only 3.4%. As to the imaging features of the 48 laser-printed false negative mammograms, 22 (46%) mammograms were considered unpredictable because of dense breast tissue or with nonspecific mammographic sign to indicate malignancy, while 26 (54%) were judged as missed cancers. Of the latter, 20 patients (77%) had the breast density of extremely dense or heterogeneously dense. The lesions were located in the right breast in 16 patients and in 10 in the left breast. Two lesions could be seen in one view only. The mammographic findings in these 26 missed cancers include calcifications in 7 patients (27%), mass or focal asymmetry in 19 (73%) with architectural distortion found in four.

4. Discussion

In clinical practice, there are two kinds of false negatives on mammograms. The first is truly false negative, which is either rapid-growth cancers that have become clinically evident during a short-term follow-up or image-occult cancers that are not visible mammographically (e.g., dense breast tissue or machine limitation). The other false negative is the detectable cancers that are present at the time of screening but not recognized or misinterpreted by the screening radiologists. In this single-blind study, we defined missed cancers as ones which could be recognized by auditors during the peer auditing review to distinguish truly false negative films (undetectable cancers) from detectable missed cancers. Our study showed that 51.8% of the false negatives, being reassessed as positive by the auditors, were missed cancers. If the missed cancers can be included, the cancer detection rate will be 4.4 per 1000, making a 7.3% increase as compared with the initial detection rate of 4.1 per 1000. The increase rate is comparable to those reported (5–15%) in the literature.

If only traditional films were taken into account, the truly false negatives were even higher (58.2%). To avoid a “witch hunt,” effect on the screeners which may increase large number of false positives in later interpretation, the later educational programs should focus on the characteristics of avoidable missed cancers only. In this study, we did not consider the auditors’ competence to be a major issue since they had correctly reinterpreted most of the true negative traditional films as negative (specificity: 98.6%), and there found a significant difference between their assessments and those of the initial screeners.

As to the laser-printed films, the auditors reassessed 78.3% of 46 false negative films to be positive, but also 56.3% of the 48 true negatives to be positive, much higher than the rates found in their reinterpretation of traditional ones. As the diagnostic accuracy of digitized images was considered to be equivalent to that of traditional screen-film mammograms, these findings suggested that something about the laser-printed control films led the reviewers to sense they were problematic, probably making the auditors more suspicious and leading to what has been called intelligent failure. In fact, the seeded laser-printed films did look different from the other 4815 audited and 267 study/control traditional films, which probably gave rise to some suspicion. This would reduce the statistical power of differences in our analysis, meaning that the interpretations of the auditors were not completely objective, rendering the evaluation of the laser-printed film unreliable. Thus, the detectable cancers in this group cannot be utilized as teaching materials since that might lead to a compromise of the false positive tradeoff.

In this review, the false negative rate of screening mammography in Taiwan was 0.057%, which is comparable to that of Siegal et al’s series (0.098%). As to the laser-printed images reread by three specialists in our review, 54% of the false negatives were interpreted as missed cancers, which is slightly higher than that in Siegal et al’s series (39%). It may be due to more dense breasts tissue in our patients (77% vs. 60%). The majority (73%) of findings of missed cancers were asymmetries or masses, which is comparable to previous
reports of 70–72%. Califications constitute the rest (27–30%) of the mammographically missed cancers. From the peer review process, we can identify any radiologists who may require additional training, and the mammographic features of the missed cancers can be used for later education. This study has several limitations. First, there was the obvious blinding problem in the reassessment of the laser-printed films, as indicated by the statistics. Future such studies require devising a more effective way of blinding for laser-printed images, since their use is increasing nowadays. Second, a few reassessment sheets from the study group and the control group were not filled out, suggesting that the auditors may not have been able to make a decision. Third, we did not perform analysis of the relationship between pathology and mammogram findings. Further study to gain better insight into what features might lead to missing a detectable cancer is warranted.

In conclusion, we found a significant difference in the initial assessments of the screeners and the reassessments of the auditors, especially in the traditional films. The false negative films reassessed as positive by the auditors can be used for future continuing education courses for radiologists, especially if the factors leading to their miss can be identified. The false negative cases with cancers missed by both screeners and auditors cannot be used for training, as they are truly undetectable. The results of this study confirmed that the peer auditing design was able to reduce the false negative rate and improve our national screening program. We hope that this study might lead to implementation of such national screening programs for breast cancer in other countries, and if they already have such programs in place, then improvement of those programs.

References