Diagn Interv Radiol 2007; 13:183–187

ORIGINAL ARTICLE

Clinical outcome assessment in mammography: an audit of 7,506 screening and diagnostic mammography examinations

Işıl Tunçbilek, Ayşegül Özdemir, Serap Gültekin, Törel Oğur, Rabia Erman, Cemal Yüce

PURPOSE
To perform an audit of our routine mammographic practice and to compare our results to performance benchmarks.

MATERIALS AND METHODS
We analyzed the outcomes of 7,506 consecutive examinations performed in 1 year. Screening and diagnostic cases were evaluated separately and mammographic assessments were based on the Breast Imaging Reporting and Data System (BI-RADS) classification.

RESULTS
In 6,858 (91%) screening and 648 (9%) diagnostic cases, outcomes varied substantially. The recall rate was 10.9%. Estimated sensitivity and specificity were similar (100% vs. 98% and 88% vs. 94%) in the screening and diagnostic groups. Positive predictive values (PPV1, PPV2, and PPV3) were higher in the diagnostic group compared to the screening group (64%, 65%, and 68% vs. 4.9%, 33%, and 39%, respectively). Cancer outcomes in the screening and diagnostic groups were, respectively, as follows: cancer detection rate, 6.1‰ vs. 86.4‰; mean invasive cancer size, 15.7 mm vs. 24.5 mm; minimal cancers, 38% vs. 19%; stage 0–1 cancers, 50% vs. 21%; and lymph node negativity, 76% vs. 29%.

CONCLUSION
The measures of our screening outcomes were concordant with the literature and the performance benchmarks for screening mammography; however, in our diagnostic group, the reasons for the higher PPV, higher cancer detection rate, and the diagnosis of cancer in a more advanced stage compared to the performance benchmarks should be investigated with more detailed periodic audits.

Key words: • *outcome assessment* • *mammography* • *breast*

From the Department of Radiology (I.T. ✉ *isiltuncbilek@yahoo.com*), Gazi University School of Medicine, Ankara, Turkey.

Received 13 April 2007; revision requested 3 July 2007; revision received 9 July 2007; accepted 6 August 2007.

A medical audit is a compilation of patient outcomes over a certain period of time, generally 1 year. Mammographic auditing provides an objective criterion of the appropriateness and accuracy in image interpretation, and is the best measure of a mammographer's performance (1–3). Regular auditing is a teaching activity to promote the mammographer and the mammography unit's progress, so that correct interpretations are more likely and the compliance of patients and clinicians to screening mammography is increased (1, 4).

The audit assesses 3 important outcomes: i) detection of a high percentage of cancers in a population, ii) finding these cancers while they are still curable (small and node negative), iii) finding these cancers through an acceptably low number of recalls and biopsies (5, 6).

Outcomes for mammography have been extensively reported in many countries (2, 4, 5, 7–12) and several performance benchmarks have been published (13, 14). The largest data come from the Breast Cancer Surveillance Consortium-US (BCSC), which is a research group consisting of 7 population-based research centers and a statistical coordinating center that collects and analyzes mammographic and pathological data in defined populations (15). The data of 3,020,471 screening and 448,225 diagnostic mammographic examinations are published (13, 14) and updated on the official website (http://breastscreening.cancer.gov/data/benchmarks.2006), serving as a reference for performance benchmarks. Interest in mammographic auditing is more recent in Turkey, and there's only one published article from Turkey concerning the subject (16).

This study presents detailed audit data involving 7506 consecutive examinations performed in 1 year. Screening and diagnostic cases were separated and a thorough benchmarking was performed.

Materials and methods

Data collection and study group definitions

The data in our breast imaging unit were collected from January 3, 2005 to December 31, 2005. The study was approved by our institutional review board, and it was performed according to the Declaration of Helsinki principles.

The patients were referred from several clinics (mainly surgery, gynecology, and check-up) by physicians who were asked to report a detailed clinical breast examination (CBE). Every patient was asked to complete a questionnaire that collected medical history and demographic data, including personal or family history of breast cancer, previous biopsies, and the presence of recent breast symptoms. Personal history of breast cancer, previous biopsy with atypical and lobular intraepithelial neoplasia, and family history of breast cancer were recorded as specific risk factors.

Screening mammographies were defined as those performed in asymptomatic women with a negative CBE. A standard screening examination consisted of a mediolateral oblique (MLO) view and a craniocaudal (CC) view of each breast.

Diagnostic mammographies were defined as those performed in symptomatic women with one or more signs or symptoms related to breast disease, such as a lump, focal breast pain, nipple discharge, or palpable breast thickening. It is a problem-solving examination using additional mammographic projections and ultrasonography (US) when necessary.

Assessment of the mammograms was performed by the same 2 mammography fellows conjointly, using the Breast Imaging Reporting and Data System (BI-RADS) (17). When there was more than one lesion in a breast, only the highest BI-RADS assessment category was recorded. The breast density, assessment category, and specific recommendations for each lesion were noted. All necessary cases were consulted to one breast imaging specialist for a final decision.

Mammographic outcome evaluation was based on comparisons of the abovementioned data to the final pathology reports in cases that were biopsied. In cases that were not biopsied, the records were maintained for the following year's screening mammography evaluation, or for short interval follow-up.

Management recommendations according to the final BI-RADS categories are: annual mammography for negative (category 1) and benign (category 2) findings, 6-month follow-up for probably benign (category 3) findings, and biopsy for suspicious (category 4) and highly suggestive of malignancy (category 5) findings (17). Incomplete assessment (category 0) was used only for screening examinations, and additional imaging work-up by mammography and/or US was completed the same day, although these cases were still referred to as recalls. Biopsy-proven cases (category 6) in the diagnostic group were included in the patient list, but were not included in statistical analysis.

Mammograms assessed as BI-RADS 1, 2, and 3 were considered negative. Mammograms assessed as BI-RADS 4 and 5, and BI-RADS 0 in the screening group, were considered positive. Data for both screening and diagnostic examinations were recorded daily on a Microsoft Excel worksheet, and weekly on a computer database.

Histopathological and/or cytological diagnoses of the lesions categorized as BI-RADS 4 or 5 were obtained from the pathology database or from patient files. Patients who did not follow recommendations for biopsy or follow-up were determined with telephone calls and reminders were left.

Histopathological tumor type, sonographic tumor size (mammographic size was recorded if the lesion was not visible sonographically), lymph node status, and cancer stage were recorded. Staging was based on tumor size, lymph node status, and metastases, according to the Manual of the American Joint Committee on Cancer (18). Invasive cancers ≤10 mm and ductal carcinomas in situ were considered as minimal cancers.

The cancer detection rate was calculated as the overall number of cancers detected per 1000 patients examined with mammography.

Outcome measurements

True-positive and false-positive results were defined as positive mammographic interpretations with (true-positive result) or without (false-positive result) a cancer diagnosis reported within 12 months. A false-negative result was defined as a negative mammographic interpretation with cancer diagnosed within the next 12 months (interval cancer). A true-negative result was a negative mammographic interpretation with no diagnosis of cancer within the next 12 months. We used these data to calculate the following statistics:

i. Sensitivity = true positive/true positive + false negative. It is impossible to know all the interval cancers constituting the false negatives and the real sensitivity can be measured only after a performance assessment of several years. Therefore, our sensitivity in this medical audit based on a 1-year evaluation is actually referred to as estimated sensitivity.

ii. Specificity = true negative/true negative + false positive.

iii. Positive Predictive Value (PPV) = true positive/true positive + false positive. The PPV is the percentage of requested biopsies that result in a diagnosis of breast cancer. However, 3 separate PPV calculations actually exist: PPV_1 is the ratio of cancers in all abnormal examinations, PPV_2 is the ratio of cancers in all recommended biopsies, and PPV_3 is the ratio of cancers in all performed biopsies. PPV_1 is, in particular, more relevant in screening mammography, and it is considered a measure of one's perceptive skills at screening, whereas PPV_2 and PPV3 are measures of analytical skills used in diagnostic mammography (6). All 3 types of PPV can be used in facilities performing screening and diagnostic mammography simultaneously, and were calculated in this audit.

Results

Patient population

Of the 7,506 examinations, 6,858 (91%) were performed for screening and 648 (9%) were performed for diagnostic purposes, with a ratio of 91:9.

The mean age of diagnostic mammography patients (49.7 years) was 1.1 years younger than the mean age of screening mammography patients (50.8 years), although the difference was not statistically significant ($P>0.005$).

The difference in risk factor positivity was not statistically significant between the 2 groups (22% in the screening vs. 23% in the diagnostic group) ($P>0.005$).

American College of Radiology (ACR) breast patterns were not statistically different and were mostly type 2 or 3 (76% in the screening group and 79% in the diagnostic group) ($P>0.005$).

Mammography outcomes

Mammography assessments in the screening and diagnostic groups are shown in Table 1. Most examinations ended with a negative or benign assessment (86.1% in the screening group vs. 72.7% in the diagnostic group) (Table 1).

The recall rate (BI-RADS 0 assessment) in the screening group was 10.9% (n = 754); the final assessment was based on US in 80% of the cases, on additional mammographic views in 6%, and on both additional mammography and US examinations in 14% of the cases. US, with or without

mammographic additional views, was the method of final decision-making in 94% of the patients.

Clinical outcomes and performance measurements

Clinical outcomes in the screening and diagnostic groups are shown in Table 2. The number of performed biopsies in the diagnostic group was about 8 times more than in the screening group. The cancer rate in biopsies in the diagnostic group was nearly twice that in the screening group (Table 2).

The estimated sensitivity, specificity, and PPV are outlined in Table 3. We had one false-negative case in the diagnostic group. This lesion was classified as BI-RADS 3 and was biopsied due to a high level of patient anxiety (resultant histopathological diagnosis: ductal carcinoma in situ [DCIS] of 7 mm).

PPV_1, PPV_2, and PPV_3 were higher in the diagnostic group when compared to that of the screening group; however, the difference was significant for PPV_1 (64% compared to 4.9%) (Table 3).

Cancer outcomes

Cancer outcomes in the screening and diagnostic groups are summarized in Table 4. The cancer diagnosis rate for diagnostic mammography was 14 times higher than for screening mammography (6.1‰ vs. 86.4‰) (Table 4). The mean size of invasive tumors was significantly higher in diagnostic mammography (24.5 mm and 15.7 mm, respectively).

The rate of minimal cancers (38% vs. 19%), stage 0 and 1 cancers (50% vs. 21%), and lymph node negativity (76% vs. 29%), which define early-stage breast cancer, were about twice as common in the screening group. Data were missing for 15 screening and 17 diagnostic mammographically-detected cancers.

Discussion

The most important function of the audit is to evaluate the mammographer's success in detecting very small cancers, which is the main goal of mammography practice (1–4). If the outcome results are within the expected limits, the mammographer's confidence is increased, improving diagnostic accuracy. This also increases the compliance of patients and referring physicians with screening mammogra-

Table 1. Comparative assessments in the screening and diagnostic mammography groups

BI-RADS category	Screening mammography n (%)	Diagnostic mammography n (%)
0. Incomplete	754 (10.9)	-
1. Negative	5,070 (74.0)	324 (50.0)
2. Benign	821 (12.1)	147 (22.7)
3. Probably benign	91 (1.3)	72 (11.0)
4. Suspicious	85 (1.2)	35 (5.4)
5. Suggestive of malignancy	37 (0.5)	51 (7.9)
6. Biopsy-proven malignancy[a]	-	19 (3.0)

BI-RADS: Breast Imaging Reporting and Data System
[a] Cases not included in statistical analysis

Table 2. Clinical outcomes in the screening and diagnostic mammography groups

	Screening mammography n (%)	Diagnostic mammography n (%)
Positive mammography	876 (12.6)	86 (13.3)
Performed biopsies	103 (11.7)	77 (89.5)
Detected cancers in biopsies	42 (40.7)	56 (72.7)

Table 3. Performance outcomes in the screening and diagnostic mammography groups

	Screening mammography	Diagnostic mammography
Sensitivity (%)	100	98
Specificity (%)	88	94
PPV (%)		
PPV_1	4.9	64
PPV_2	33	65
PPV_3	39	68

PPV: positive predictive value

Table 4. Cancer outcomes in the screening and diagnostic mammography groups

	Screening mammography	Diagnostic mammography
Cancer diagnosis rate (per 1000)	6.1	86.4
Invasive cancer mean size (mm)	15.7	24.5
Number of cancer types [n (%)]		
DCIS	5 (12)	6 (10)
IDC	29 (69)	41 (73)
ILC	2 (5)	1 (2)
Mixed invasive cancer	3 (7)	4 (7)
Other types	2 (5)	2 (4)
Unknown	1 (2)	2 (4)
Minimal cancers [n (%)]	16 (38)	10 (19)
Stage 0 and 1 cancers [n (%)]	21 (50)	12 (21)
Lymph node negativity [n (%)]	32 (76)	16 (29)

DCIS: ductal carcinoma in situ
IDC: invasive ductal carcinoma
ILC: invasive lobular carcinoma

phy guidelines, which is an important issue, since screening should be performed at regular intervals to be effective (1, 4, 19). If the outcome results are not within the expected range, it is possible to identify the causes and to take corrective measures, assessing the results in subsequent audits (3, 4, 20).

Data collection is an important issue and minimizing data collection is necessary to facilitate the statistical calculations and outcome analysis. Although data collection can be performed manually, this is frequently time-consuming and tedious work (1, 3). For this reason, special computer programs have been designed (1, 21). In our outcome analysis, we did not use a specially designed computer software for data collection, but we combined our manual collecting method with our computerized reporting system. The perfection of an audit can only be possible with the help of a software program especially designed for time-saving and easy data collection (1, 22).

In published reports, clinical outcomes for diagnostic mammography are different from those for screening mammography (4, 10–14, 23–25); therefore, we collected and analyzed our screening and diagnostic data separately.

The demographic characteristics of our patients were similar in the screening and diagnostic groups (mean age: 50.8 vs. 49.7 years; personal/family history of breast cancer: 22% vs. 23%). In previously reported series, diagnostic mammography patients were 3.0–3.3 years younger than the screening patients (10–14, 23). The reason of our similar age populations in the 2 groups may have been the lack of a national screening program for older age groups in Turkey.

The cancer detection rate, which is an important parameter reflecting the quality of mammography practice, is affected by several variables, including age, indication of mammography examination, and size of the tumor. For example, mammography performed for palpable lumps in an older population will reveal higher cancer detection rates. Our cancer detection rates of 6.1‰ and 86.4‰ in the screening and diagnostic groups, respectively, were higher than the average values in the literature (4.8‰ and 12.3‰ in the screening and diagnostic groups, respectively). The relatively high cancer detection rate in our screening group is probably due to the prevalence of cancers related to the inclusion of data from the first-time screened population into the data of the subsequently-screened population, and is actually expected. On the other hand, our remarkably higher cancer detection rate in the diagnostic group may be primarily due to the trend of late referral of patients in Turkey. Another factor may be that our breast imaging unit functions as a center for patients referred by clinicians with suspected breast cancer and other breast pathologies.

Tumor size, stage, and lymph node status are important prognostic factors in breast cancer. Small tumor size, a high percentage of minimal cancers, and a high rate of lymph node negativity in a screening population audit indicate the ability to detect disease at an early stage. In our screening group, the mean size of invasive cancers (15.7 mm) and lymph node negativity (76%) were similar to published studies and performance benchmarks (16.7 mm, and 79.8%), whereas the percentage of minimal cancers (38%) and stage 0–1 cancers (50%) were moderately lower than the desired goals (51.9%, and 74.1%) (2, 5, 6, 8–10, 12, 14, 25). In our diagnostic group, mean size of invasive cancers (24.5 mm) was higher, although lymph node negativity (29%), minimal cancers (19%), and cancers in stages 0 or 1 (21%) were somewhat lower than the published reports and benchmarks of 20.9 mm, 71.9%, 39%, and 61%, respectively (10–14). These results may be due to the lack of an official breast carcinoma screening program, low participation rate of women in routine screening examinations, and the tendency to seek medical care only when the disease is advanced.

Sensitivity is the most difficult information to obtain, which requires an exact number of false-negative cases (interval cancers) for accurate calculation, necessitating the presence of a national tumor registry (2, 7, 8). Sensitivity depends on patient age, screening interval, and the scheduling of follow-up for negative mammograms (6, 8). Our sensitivity of 100% and 98% (screening vs. diagnostic) is higher compared to the reported performance benchmarks of 77.4% and 79.8% (screening vs. diagnostic) (13, 14); however, it may be misleading to conclude that our results indicate superior performance. Lack of a national tumor registry and the requirement of a rather short follow-up for this first audit probably had some impact on our figures.

PPV_1 (abnormal exam), an indicator of perceptual skills, was 4.9% in the screening group, which is in accordance with the recommended range of 5%–10% (1, 2, 8, 25). PPV_2 (biopsy recommended: 33%) and PPV_3 (biopsy performed: 39%) in the screening group, indicating analytical skills, were also concordant with the literature (range of 25%–40% for PPV_2 and PPV_3) (2, 5, 6, 8–10, 12, 14, 25). In the diagnostic group, PPV_2 and PPV_3 were 65% and 68%, respectively. These values are substantially higher than the mean values of 30.3% and 37.3% for the performance benchmarks of PPV_2 and PPV_3. This may be explained by the same factors responsible for our higher cancer detection rate and higher number of advanced cases in the diagnostic group.

Detecting cancers through an acceptably low rate of recall is one of the main goals of screening mammography (5, 6). The recall rate is defined as the percentage of screening patients requiring additional mammographic views or US. Our recall rate of 10.9% is in accordance with the 10.3% screening performance benchmark (14).

To the best of our knowledge, this is the first detailed audit in Turkey analyzing screening and diagnostic mammographic examinations separately. In a previous mammography audit of 3,048 screening cases reported by Türk and Arıbal, the cancer detection rate (6.9 per 1000) and lymph node negativity (66.7%) were similar to ours; however, a higher recall rate and PPV_1, and lower rates of PPV2, PPV3, minimal cancers, and stage 0–1 cancers were reported (16). It is not clear if they had implemented US into their routine examinations, as we did, which might have been the cause underlying some of the lower rates of detection they reported. It is also impossible to know to what extent US might have affected our diagnostic mammography assessments and management recommendations; however, it has already become a standard approach as the last edition of BI-RADS guidelines also recommends their integrated use (17).

There are several limitations of our study. First, due to the lack of an institutional and national tumor regis-

try, false-negative cases were difficult to identify. Second, our study period was short. Studies with longer time intervals may contribute to better and more complete, as well as more realistic, outcomes in subsequent audits. We assume that our continuing audit analyses will provide more reliable data compared to the past study period.

One difficulty we experienced in this study was related to the method of data collection. We would suggest that a practical data collection method with a dedicated computer program designed for outcome analysis is mandatory for avoiding certain methodological difficulties in such studies.

Our data suggest that the measures of our screening outcomes, primarily recall rate, cancer detection rate, and PPV, are concordant with the literature and the performance benchmarks for screening mammography. In our diagnostic group, higher PPV, higher cancer detection rate, but larger tumor size and the higher number of advanced cancers compared to the literature demonstrate the need for more detailed outcome analysis in diagnostic mammography. Efficient auditing systems, as well as a national tumor registry, are essential for confident and standardized data collection. We think that mammography facilities will benefit from auditing their data regularly, which would help them to improve their outcomes and achieve greater success in diagnosing early-stage breast cancer.

References

1. Sickles EA. Quality assurance: how to audit your own mammography practice. Radiol Clin North Am 1992; 30:265–275.
2. Robertson CL. A private breast imaging practice: medical audit of 25,788 screening and 1,077 diagnostic examinations. Radiology 1993; 187:75–79.
3. Spring DB, Kimbrell-Wilmot K. Evaluating the success of mammography at the local level: how to conduct an audit of your practice. Radiol Clin North Am 1987; 25:983–992.
4. Sickles EA, Ominsky SH, Sollitto RA, Galvin HB, Monticciolo DL. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. Radiology 1990; 175:323–327.
5. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Gröntoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin North Am 1992; 30:187–210.
6. Linver MN. The medical audit: statistical basis of clinical outcomes analysis. In: Diagnosis of diseases of the breast. 2nd ed. Pennsylvania: Elsevier Saunders, 2005; 135–148.
7. Linver MN, Paster S, Rosenberg R, et al. Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38.633 cases. Radiology 1992; 184:39–43.
8. Burhenne LJW, Hislop TG, Burhenne HJ. The British Columbia mammography screening program: evaluation of the first 15 months. AJR Am J Roentgenol 1992; 158:45–49.
9. Perry M, Broeders M, de Wolf C, Törnberg S, eds. European guidelines for quality assurance in mammography screening. 3rd ed. Luxembourg, European Commission, 2001.
10. Dee KE, Sickles EA. Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently. AJR Am J Roentgenol 2001; 176:729–733.
11. Sohlich RE, Sickles EA, Burnside ES, Dee KE. Interpretating data from audits when screening and diagnostic mammography outcomes are combined. AJR Am J Roentgenol 2002; 178:681–686.
12. Poplack SP, Tosteson AN, Grove MR, Wells WA, Carney PA. Mammography in 53,803 women from the New Hampshire Mammography network. Radiology 2000; 217:832–840.
13. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. Radiology 2005; 235:775–790.
14. Rosenberg RD, Yankaskas BC, Abrahams LA, et al. Performance benchmarks for screening mammography. Radiology 2006; 24:55–66.
15. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol 1997; 169:1001–1008.
16. Türk A, Arıbal E. Medical audit of mammography screening examinations. Marmara Medical Journal 2002; 15:27–32.
17. D'Orsi CJ, Bassett LW, Berg WA, et al. Breast Imaging Reporting and Data System: ACR BI-RADS. 4th ed. Reston, Va: American College of Radiology, 2003; 229–251.
18. American Joint Committee on Cancer. Manual for staging of cancer. 5th ed. Philadelphia: Lippincott, 1997.
19. Linver M, Osuch J, Brenner J. The mammography audit: a primer for the mammography quality standards act (MQSA). AJR Am J Roentgenol 1995; 165:19–25.
20. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992; 184:613–617.
21. Haug PJ, Tocino IM, Clayton PD, Bair TL. Automated management of screening and diagnostic mammography. Radiology 1987; 164:747–752.
22. Monticciolo DM, Sickles EA. Computerized follow-up of abnormalities detected at mammography screening. AJR Am J Roentgenol 1990; 155:751–753.
23. Rosenberg RD, Lando JF, Hunt WC, et al. The New Mexico Mammography Project: Screening mammography performance in Albuquerque, New Mexico, 1991–1993. Cancer 1996; 78:1731–1739.
24. Bassett LW, Liu TH, Giuliano AE, Gold RH. The prevalence of carcinoma in palpable vs. impalpable, mammographically detected lesions. AJR Am J Roentgenol 1991; 157:21–24.
25. Rosen EL, Sickles EA, Keating D. Ability of mammography to reveal nonpalpable breast cancer in women with palpable breast masses. AJR Am J Roentgenol 1999; 172:309–312.