Improve Cancer Detection Using Computer-Aided Detection with Diagnostic and Screening Mammography: Prospective Study of 104 Cancers

Judy C. Dean¹
Christina C. Ilvento

OBJECTIVE. This study prospectively evaluated a computer-aided detection (CAD) device used with diagnostic and screening mammography by assessing cancers detected; tumor sizes, histology, and stage; positive predictive value (PPV) of biopsy recommendation; and recall rates before and after CAD introduction.

SUBJECTS AND METHODS. Interpretations of 9,520 consecutive mammograms were recorded without and then with CAD for a 28-month period. Cancer detections based on initial radiologist review and additional detections based on CAD findings were noted. Recall rates, tumor size and histology, and PPV of biopsy recommendation before and after the introduction of CAD were compared.

RESULTS. Cancers detected only with CAD assistance were 9.6% of all cancers (10 of 104); screening-detected cancers increased 13.3% with CAD assistance (four in addition to 30 screening-detected cancers). The 95% one-sided confidence boundary using binomial distribution is consistent with at least 5.3% for all cancers and 5.1% for nonpalpable cancers. The greatest impact was on ductal carcinoma in situ, for which CAD increased cancer detection by 14.2% (three added to 21). Similar percentages of cancers were detected only with CAD assistance in both screening (11.4%; 4 of 35) and diagnostic (9.5%; six of 63) studies. Additional cancers were detected using CAD in patients with implants and previous lumpectomy. The additional cancers detected with CAD were smaller ($p=0.01$ for all cancers, $p=0.03$ for nonpalpable invasive cancer). The screening recall rate increased from 6.2% to 7.8% after CAD, with a decrease in the biopsy rate and a nonsignificant increase in the biopsy PPV from 21.9% to 26.3%.

CONCLUSION. CAD resulted in detection of more cancers in screening and diagnostic patients, with an increased recall rate but no deterioration in PPV of biopsy. Additional cancers detected were significantly smaller.

Mammography is the most widely used examination for breast cancer screening and evaluation of breast symptoms such as masses, but not all breast cancers are detected with mammography. An additional review by a second radiologist (double reviewing) increases cancer detection [1–3]; however, this is not feasible in many facilities and may not become standard practice because of cost and lack of available radiologists. Recent efforts to reduce the number of false-negative examinations have focused on computer-aided detection (CAD). A study comparing sensitivity for cancer detection with a test set of 150 mammograms including 17 cancers showed a similar increase in sensitivity for reviewing with CAD compared with simulated double reviewing by experienced radiologists [4].

CAD reviews mammograms using software designed to alert the radiologist to potential findings associated with breast cancer, such as microcalcifications and focal densities. For film-screen mammograms, the films are labeled and scanned to create a digital file for analysis by the software, and results are displayed for radiologist review when the films are interpreted.

Several studies have validated CAD techniques using retrospective review of known cancer cases [5–11]. Retrospective studies have shown the ability of CAD to mark cancers with a high degree of accuracy, especially when microcalcifications are present. One study estimated that as many as 25% of cancers could have been detected an average of 14 months sooner using CAD [6]. In another retrospective review, the sensitivity of CAD was 75% for masses and 99% for microcalcifications [7].
More prospective studies are needed before it can be assumed CAD will perform as well in day-to-day practice. As noted by Sacks [12], it is not possible to calculate the actual increase in sensitivity for cancer detection from decreases in false-negatives shown by retrospective studies. As of this writing, almost 4,000 CAD systems are in clinical use in the United States (personal communications with R. A. Castellino of R2 Technology and M. Y. Sallam of iCAD), but there are few published studies on the prospective result of incorporating CAD in clinical practice [13–17] and none on the prospective use of CAD in symptomatic patients.

The purpose of this study was to evaluate the impact of CAD in a nonacademic setting, most notably its effect on cancer detection in both screening and diagnostic patients. The positive predictive value (PPV) of biopsy recommendation, biopsy rate, and recall rate before and after the introduction of CAD were compared. Size, stage, and histology of cancers detected with and without CAD findings were evaluated.

**Subjects and Methods**

During the study period, 9,520 film-screen mammograms were interpreted; 59.1% \((n = 5,631)\) of them were screening studies, 39.9% \((n = 3,803)\) were diagnostic mammograms, and less than 1% \((n = 86)\) were for consultation or second opinion of films from another facility. All mammograms were interpreted by a single radiologist, without knowledge of the first reviewer’s interpretation. The interpreting radiologist has devoted more than 90% of practice time to mammography for more than 15 years.

This was solely an observational study with no experimental or additional procedures, and institutional review board approval was not required. Data already collected to satisfy mammography facility certification required by the Mammography Quality Standards Act (MQSA) were correlated with the findings of a U.S. Food and Drug Administration (FDA)-approved commercially available CAD system. Additional observations on recall rates before and after introduction of CAD were collected from schedule records.

In April 2002, the SecondLook CAD system by iCAD was installed and has been used at the study facility for all mammograms performed since that date and for all film/screen studies presented for second opinion. The initial software version 3.2 was updated in June 2003 to version 3.9. According to the manufacturer, both versions place an average of 3.2 false marks per four-view mammogram. No film-screen mammograms were excluded from the study. The study period was from April 2002 to July 2004.

Initial instruction in the use of SecondLook CAD was provided by the manufacturer in the form of a manual and onsite training. In addition, the mammograms of 50 prior patients with known cancers and 50 prior mammograms with normal findings from the study facility were evaluated retrospectively with MammoReader CAD before beginning clinical use of the device. These cases were reviewed with and without the CAD findings to allow the radiologist to become familiar with the performance of the system on actual cases.

Screening and diagnostic mammograms generally consisted of two views of each breast. Three views per breast were performed for patients with a previous breast cancer and four views per breast for patients with breast implants. Supplemental views, such as exaggerated craniocaudal lateral views, were added by the technologist if deemed necessary for the inclusion of all the breast tissue. Magnification or other workup views were only done when directed by the radiologist for evaluation of specific findings after interpretation of the initial films.

In accordance with procedures recommended by the manufacturer, the CAD device was used in the following manner: initial interpretation of the mammogram films without CAD results, review of CAD findings, and re-review of the CAD-marked areas on the mammogram. For purposes of this study the radiologist recorded a “workup” or “no workup” decision before viewing the CAD findings and marked on the films any areas deemed to warrant additional workup before viewing CAD results. Additional findings for which CAD results led to workups were noted.

All findings initially noted by the radiologist for additional examination were evaluated whether or not the finding was also marked by CAD. The radiologist did have knowledge of relevant clinical information, such as whether the patient was referred for evaluation of a mass. However, no physical examinations or sonography studies were performed until after interpretation of the initial mammograms. Cancers detected because of findings deemed actionable only after viewing CAD results were counted as CAD-added detections. The cancer cases detected by radiologist interpretation with and without corroborating CAD marks were tabulated and compared.

All available prior mammograms were compared at the time of interpretation. Approximately half of all patients during the study period had at least one earlier mammogram at the study facility; one third of them had prior films from another facility compared; prior mammograms could not be located for 6% of patients; and 8% were baseline mammograms. Although we routinely recommend that patients older than 40 years return annually for mammography, only around half of patients in the study had mammograms at the recommended yearly interval.

Biopsy and surgical data were obtained from the routine outcome audit performed as part of the MQSA certification program. The size, stage, and histologic type of each cancer were recorded and compared for cancers detected with and without CAD assistance. Data for size, stage, and histologic type of the cancers were available from surgical pathology reports in all patients except three cases of ductal carcinoma in situ (DCIS). Cancer size recorded was the largest single dimension measured, or the sum of the largest single dimensions of individual sites for multifocal tumors. Size for tumors less than 5 mm was estimated from a combination of imaging and surgical pathology reports because presurgical core biopsy may have resulted in a significant reduction in the size of the tumor remaining at surgery in these cases. The size of the invasive tumor is reported for cancers with both invasive and in situ components. In the three cases of DCIS in which the patient refused surgery or the surgical pathology report was not available, tumor type and size were determined from only imaging and percutaneous biopsy reports.

The performance of the CAD system was evaluated for all cancers: DCIS, invasive cancer, invasive ductal cancer, invasive lobular cancer, and nonpalpable and palpable cancers. Comparison was also made between screening, diagnostic, and consultation patients. One-sided binomial distribution was used to assess the 95% confidence boundary of the detection rate data for statistical significance. Wilcoxon’s rank sum test was used to determine the \( p \) values for the median size of all cancers detected and in the various subsets.

The recall rates for screening patients were compared for a 6-month period immediately before installation of the CAD device, the initial 2 months of use, and the subsequent 19-month portion of the study period. Recall rate was also assessed for a 4-month period after the study ended. Screening mammography patients who underwent sonography or additional mammography were considered recalled patients, even when the studies were performed the same day. Recall data were analyzed for the separate periods using chi-square testing for two-by-two-tables with Yates continuity correction.

Positive predictive value of biopsy was compared for the 28-month study period and the 28-month period immediately before installation of the CAD system using data from routine biopsy audits.

**Results**

**Overall Cancer Detection With and Without Computer-Aided Detection**

During the 28 months of the study, 104 cancers were detected in 9,520 mammograms performed.
grams evaluated with CAD. Patients whose cancers were detected during the study period ranged from 34 to 92 years old, with a mean age of 58.5 years. Table 1 depicts the numbers of cases detected by the radiologist and marked by CAD (column A), the radiologist without corroborating CAD marks (column B), the CAD-added cases (column C), and cases not marked by CAD or detected by the radiologist on mammography (column D). Overall positive-mark rate of CAD for all cancers was 82.7% (the sum of columns A and C).

The 10 cancers (Table 1, column C) not apparent to the radiologist on initial review but marked by CAD comprised 9.6% of the total cancers detected. There were three DCIS and seven invasive cancers in this group. Four of the 10 cancers added because of CAD findings were nonpalpable cancers marked by calcification notations, including two DCIS and two invasive ductal cancers. Five of the 10 cancers added by CAD were marked by a density notation, and one had both calcification and density marks. Three of the 10 cancers detected by CAD but not apparent on initial radiologist review were palpable masses, including one noncalcified DCIS. Figure 1 shows the mammograms of a patient with breast implants and a 4.5-mm nonpalpable invasive ductal carcinoma marked by CAD but not initially seen by the radiologist. Five of the 10 CAD-added cancers were nonpalpable invasive ductal carcinomas.

The proportion of cancers detected by CAD alone is statistically significant. A 95% confidence boundary (Table 2) using a one-sided binomial analysis indicates this effect is at least 5.3% for the entire group of cancers. Because of the smaller number of cases in the subsets, statistical significance is lower, even though the percentage of additional cancers detected is actually higher for screening cases, nonpalpable cancers, and DCIS than for the group as a whole.

Similar percentages of nonpalpable cancers were detected by the radiologist (58 of 66/87.9%) and noted by CAD (56 of 66/84.8%). However the radiologist detected more palpable cancers (35 of 38/92.1%) compared with CAD (30 of 38/78.9%). The proportion of cases added by CAD is statistically significant, with the 95% confidence boundary indicating at least 5.08% for nonpalpable and 2.19% for palpable tumors.

DCIS was the only type of cancer for which CAD noted more cases than the radiologist, marking 22 of 24 cancers (91.7%) compared with 21 of 24 (87.5%) detected by the radiologist before viewing CAD findings.

There were 80 invasive cancers (76.9% of total cancers), which ranged in size from 0.5 mm to 70 mm. The overall positive mark rate of the CAD system in noting invasive cancers was 80% (64 of 80 cases). CAD was less sensitive for invasive lobular carcinoma, marking seven of 11 (63.6%) compared with 57 of 69 invasive ductal cancers (82.6%). Seven of the 69 invasive ductal cancers (10.1%) were not detected by the radiologist’s interpretation without CAD, and this was statistically significant. No additional invasive lobular cancers were added by CAD.

The one invasive cancer not detected on the mammogram by the radiologist or marked by CAD was a nonpalpable 8-mm mixed invasive lobular and ductal cancer detected by screening sonography. This case is listed with the lobular invasive cancers in the tables.

Thirty-five cancers were detected in 5,631 screening patients, 33 of which were nonpalpable. In the 3,803 diagnostic mammograms, 63 cancers were detected, of which 36 were palpable and 27 were nonpalpable. Six nonpalpable cancers were confirmed in 86 patients seen for consultation after mammograms performed at another facility. Rates of CAD-added cases not initially seen by the radiologist were similar in screening (four of 35/11.4%) and diagnostic patients (six of 63/9.5%).

### TABLE 1: Number of Cancers Detected With and Without Computer-Aided Detection Assistance

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Marked by CAD</td>
<td>Not Marked by CAD</td>
<td>Marked by CAD</td>
<td>Not Marked by CAD</td>
</tr>
<tr>
<td>All cancers (n = 104)</td>
<td>76 (73.1)</td>
<td>17 (16.3)</td>
<td>10 (9.6)</td>
<td>1 (&lt;1)</td>
</tr>
<tr>
<td>Nonpalpable (n = 66)</td>
<td>49 (74.2)</td>
<td>9 (13.6)</td>
<td>7 (10.6)</td>
<td>1 (1.5)</td>
</tr>
<tr>
<td>Palpable (n = 38)</td>
<td>27 (71.1)</td>
<td>8 (21.1)</td>
<td>3 (7.9)</td>
<td>0</td>
</tr>
<tr>
<td>DCIS (n = 24)</td>
<td>19 (79.2)</td>
<td>2 (8.3)</td>
<td>3 (12.5)</td>
<td>0</td>
</tr>
<tr>
<td>All invasive (n = 80)</td>
<td>57 (71.3)</td>
<td>15 (18.8)</td>
<td>7 (8.8)</td>
<td>1 (1.3)</td>
</tr>
<tr>
<td>Invasive lobular (n = 11)</td>
<td>7 (63.6)</td>
<td>3 (27.3)</td>
<td>0</td>
<td>1 (9.1)</td>
</tr>
<tr>
<td>Invasive ductal (n = 69)</td>
<td>50 (72.5)</td>
<td>12 (17.4)</td>
<td>7 (10.1)</td>
<td>0</td>
</tr>
<tr>
<td>Screening cases (n = 35)</td>
<td>27 (77.1)</td>
<td>3 (8.6)</td>
<td>4 (11.4)</td>
<td>1 (2.9)</td>
</tr>
<tr>
<td>Diagnostic cases (n = 63)</td>
<td>45 (71.4)</td>
<td>12 (19.0)</td>
<td>6 (9.5)</td>
<td>0</td>
</tr>
<tr>
<td>Consultation cases (n = 6)</td>
<td>4 (66.7)</td>
<td>2 (33.3)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note—Numbers in parentheses are percentages. CAD = computer-aided detection, DCIS = ductal carcinoma in situ.

Recall Rates

The percentage of screening patients recalled for additional evaluation (Table 3) increased from 6.2% before use of the CAD device to 7.8% after the initial 2 months of use. This represents a 26% increase in the recall rate, although that is only a 1.6-percentage-point increase in recalls of total screening patients. The difference is statistically significant with \(p = 0.0421\) (95% confidence interval [CI], 7.0–8.7%). The recall rate since the end of the study period, still using CAD, has fallen to 6.75%, based on 59 recalls in 874 screening mammograms during 4 months. During the initial 2 months of use, the recall rate more than doubled to 13.4%. This change in recall rate is also statistically significant, with \(p < 0.0001\) (95% CI 10.1–17.2%).
Interval Cancers

Interval cancer is defined as a cancer that becomes clinically apparent, usually because of a palpable mass, within a year of a negative mammogram. There were 10 interval cancers during the study period, half of which were in patients who had prior mammograms at the study facility. No additional interval cancers have been observed since the end of the study period. Comparison with the number of interval cancers immediately before using CAD showed no significant difference, with six interval cancers in 70 cases ($p = 1$). Retrospective CAD analysis of prior mammograms showed four of the 10 interval cancers had CAD marks at the cancer site on prior mammograms an average of 10 months earlier. Four patients had CAD during the first year of the study, and two of these had no CAD marks at the cancer site. Two of the 10 interval cancers were invasive lobular cancer, one was DCIS, and seven were invasive ductal carcinomas that included one grade 3 invasive ductal carcinoma.

Size, Stage, and Grade of Cancers Detected

Table 4 shows proportionately more stages 0 and I cancers added by CAD (column C), suggesting that the cancers detected because of using CAD may be smaller, earlier cancers. However, this interpretation is not statistically significant.

The median size of nonpalpable invasive cancers for CAD-added diagnoses was 5 mm ($n = 5$; range, 4 to 10 mm), compared with 10.5 mm ($n = 32$; range, 4 to 32 mm) for cases detected by the radiologist regardless of CAD assistance (Table 5). This difference is statistically significant, with $p = 0.030$. The median size of cancers for CAD-added cases is also smaller for all cancers ($p = 0.010$) and all invasive cancers ($p = 0.026$), again suggesting that the additional cancers detected by using CAD are earlier, smaller tumors. Median sizes of CAD-added DCIS and palpable tumors were also smaller but not statistically significant, possibly because of the smaller numbers of cases in these categories.

CAD marked 20 of 23 cancers less than 10 mm in size (87%) and 35 of 42 cancers less than 20 mm in size (83.3%). Of invasive cancers less than 10 mm, CAD detected 15 of 18 (83.3%) and 28 of 34 (82.4%) invasive tumors less than 20 mm.

The histologic grades of the additional cancers detected by using CAD were all grade 1. However, the number of cases of each grade was not sufficient to show statistical significance.

Positive Predictive Value of Biopsy and Biopsy Rate

For the 28-month period immediately before installation of the CAD system, 374 biopsies were performed and 82 cancers were diagnosed from 6,422 mammograms. During the 28-month study using CAD, 396 biopsies were performed and 104 cancers were diagnosed from 9,520 mammograms. The PPV of biopsy recommendation before CAD was,
TABLE 2: Cancer Detection Added by Computer-Aided Detection

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Total Cases</th>
<th>Cases Found by CAD, but Not by Radiologist</th>
<th>Proportion</th>
<th>95% One-Sided Lower Confidence Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cancers</td>
<td>104</td>
<td>10</td>
<td>0.0962</td>
<td>0.0531</td>
</tr>
<tr>
<td>Nonpalpable</td>
<td>66</td>
<td>7</td>
<td>0.1061</td>
<td>0.0508</td>
</tr>
<tr>
<td>Palpable</td>
<td>38</td>
<td>3</td>
<td>0.0789</td>
<td>0.0219</td>
</tr>
<tr>
<td>DCIS</td>
<td>24</td>
<td>3</td>
<td>0.125</td>
<td>0.0350</td>
</tr>
<tr>
<td>All invasive</td>
<td>80</td>
<td>7</td>
<td>0.0875</td>
<td>0.0418</td>
</tr>
<tr>
<td>Invasive lobular</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Invasive ductal</td>
<td>69</td>
<td>7</td>
<td>0.1014</td>
<td>0.0486</td>
</tr>
<tr>
<td>Screening cases</td>
<td>35</td>
<td>4</td>
<td>0.1143</td>
<td>0.0400</td>
</tr>
<tr>
<td>Diagnostic cases</td>
<td>63</td>
<td>6</td>
<td>0.0952</td>
<td>0.0423</td>
</tr>
<tr>
<td>Consultation cases</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note—CAD = computer-aided detection, DCIS = ductal carcinoma in situ.

Fig. 2—Nonpalpable 12-mm cancer in upper inner quadrant in 59-year-old woman, which was probably not marked by computer-aided detection (CAD) device because of far posterior location.
A, and B, Screening mammogram with lesion marked by half circle.
C, Spot magnification view with skin marker (white dot) over corresponding lesion (circled) seen with sonography.

therefore, 21.9%, and the biopsy rate was 5.8%. During the 28-month study period using CAD, the PPV of biopsy was 26.3%, and the biopsy rate 4.2%. The 4.4-percentage-point change in PPV of biopsy observed after the introduction of CAD represents a 20% increase, but was not statistically significant using chi-square testing \( p = 0.16 \). The decrease in biopsy rate from 5.8% to 4.2% represents a 38% change and was statistically significant \( p < 0.0001 \).

The biopsy data include numerous palpable fibroadenomata biopsied because of palpable masses without suspicious imaging findings. A subset of biopsies performed because of suspicious mammographic or sonographic findings (classes 4 and 5 interpreta-
Cancer Detection Using Computer-Aided Detection with Mammography

The recall rate of mammography patients was also analyzed for the same periods. Before the use of CAD, 157 biopsies in classes 4 and 5 patients yielded 73 (46.5%) cancers. With the use of CAD, 242 biopsies in classes 4 and 5 patients yielded 103 (42.6%) cancers. The 3.8-percentage-point decrease in PPV of recommended biopsies in classes 4 and 5 patients represented an 8% change and was not statistically significant ($p = 0.44$).

**TABLE 3: Recall Rates for Screening Mammography Patients**

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Recall Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months before CAD install</td>
<td>6.2% (65 recalls in 1,047 studies)</td>
</tr>
<tr>
<td>First 2 months of using CAD</td>
<td>13.4% (50 recalls in 374 studies)</td>
</tr>
<tr>
<td>Months 3–21 of study</td>
<td>7.8% (326 recalls in 4,157 studies)</td>
</tr>
<tr>
<td>Four-month follow-up period</td>
<td>6.75% (59 recalls in 874 studies)</td>
</tr>
</tbody>
</table>

Note: CAD = computer-aided detection.

Recall data includes months 1–21 of the study period and a 4-month follow-up.

**TABLE 4: Stage of Cancers**

<table>
<thead>
<tr>
<th>Cancer Stage (Total Cancers = 104)</th>
<th>Column A</th>
<th>Column B</th>
<th>Column C</th>
<th>Column D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Detected by Radiologist Not Marked by CAD Marked by CAD Not Marked by CAD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 0 (n = 24) (22.1)</td>
<td>19 (79.2)</td>
<td>2 (8.3)</td>
<td>3 (12.5)</td>
<td>0</td>
</tr>
<tr>
<td>Stage I (n = 48) (46.2)</td>
<td>29 (80.4)</td>
<td>12 (25.0)</td>
<td>6 (12.5)</td>
<td>1 (2.1)</td>
</tr>
<tr>
<td>Stage II (n = 29) (27.9)</td>
<td>25 (86.2)</td>
<td>3 (10.3)</td>
<td>1 (3.4)</td>
<td>0</td>
</tr>
<tr>
<td>Stage III (n = 3) (2.9)</td>
<td>3 (100)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses are percentages of total number of cancers. Numbers in brackets are percentages of the specific stage. CAD = computer-aided detection.

**TABLE 5: Median Size of Cancers Detected by Radiologist Compared with Cancers Added Using Computer-Aided Detection**

<table>
<thead>
<tr>
<th>Type of Cancer</th>
<th>Detected by Radiologist</th>
<th>Marked by CAD but Not Detected by Radiologist</th>
<th>$p^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cancers (n = 103)$d$</td>
<td>14 (n = 93)</td>
<td>7 (n = 10)</td>
<td>0.010</td>
</tr>
<tr>
<td>Invasive cancers (n = 79)$d$</td>
<td>13.5 (n = 72)</td>
<td>9 (n = 7)</td>
<td>0.026</td>
</tr>
<tr>
<td>Invasive nonpalpable (n = 44)</td>
<td>10.5 (n = 39)</td>
<td>5 (n = 5)</td>
<td>0.030</td>
</tr>
<tr>
<td>Invasive palpable (n = 35)</td>
<td>23.5 (n = 33)</td>
<td>14.5 (n = 2)</td>
<td>0.194</td>
</tr>
<tr>
<td>DCIS (n = 24)</td>
<td>15 (n = 21)</td>
<td>5 (n = 3)</td>
<td>0.100</td>
</tr>
</tbody>
</table>

Note: All sizes in millimeters. CAD = computer-aided detection, DCIS = ductal carcinoma in situ.

$d$: Regardless of CAD findings; sum of Columns A and B in Table 1.

c: CAD-added diagnoses, Column C in Table 1.

$p$: $p$-values less than 0.05 are regarded as statistically significant.

d: One cancer detected by sonography and not found on mammogram by radiologist or CAD is not included.

Discussion

A prospective study by Morton et al. [15] of more than 12,000 patients reported a 7.6% increase in cancer detection using CAD in a screening population. The additional cancers were earlier stage, and there was an increase in the recall rate of 11%. Birdwell et al. [16], reporting on 8,682 screening mammograms, showed 7.4% more cancers detected using CAD (2 of 29 cancers diagnosed) with an 8.2% increase in the recall rate. Cupples et al. [17] showed a 16.1% increase in cancers detected at screening using CAD. Nicholas et al. (presented at the 2004 meeting of the American Roentgen Ray Society) reported a 3% increase in cancer detection using CAD on 3,470 screening mammograms, with change in the recall rate from 12% to 14%. Freer and Ulissey [13], reporting on more than 12,000 mammograms, showed a prospective increase of 19% in cancer detection using CAD in a screening population from a community practice. Only 49 cancers were detected in their study of screening mammography, and the eight additional cancers detected because of using CAD were all stage 0 or I.

The current study also demonstrates increased cancer detection using CAD. Comparing the 10 cancers detected only by CAD with the 93 cancers detected on the mammograms regardless of CAD findings, the percentage increase using CAD is 10.8%. In the screening group, this outcome was 13.3% in our study (four CAD-detected cancers added to 30). Similar to the study by Freer and Ulissey [13], all the additional cancers detected using CAD were stage 0 or I except for one palpable stage II tumor marked in a diagnostic patient.

The higher rate of cancer in the current study (104 cancers in 9,520 mammograms, or almost 1.1%) reflects the inclusion of both screening and diagnostic patients and a relatively higher-risk population. It is frequently an arbitrary decision whether a patient schedules a screening or diagnostic study, which may depend on factors of no clinical relevance such as distance from her residence to the facility. It has been the policy in the study facility to schedule all patients with breast implants or previous lumpectomy as diagnostic mammography studies, although the actual purpose of the examination is usually screening. Although all patients are specifically questioned about palpable lumps when scheduling appointments, two cancers detected on screening mammograms during the study were palpable masses the patients were already aware of. Thus, considerable overlap...
occurs between screening and diagnostic patients in the clinical purpose for mammography. For these reasons, and because every diagnostic mammogram also includes screening for cancer, CAD was used for all patients during the study. Our results indicate this is a valid approach because the use of CAD did result in similar rates of increased cancer detection in both screening and diagnostic groups. A retrospective study by Butler et al. [8] showed that 7.6% of cancers diagnosed in women with palpable masses or other focal symptoms were located away from the symptomatic site, and CAD correctly marked 87% of those cancers.

A prospective study by Gur et al. [14] of 115,571 screening mammograms over a 3-year period showed no statistically significant change in cancer detection rate or recall rate. Reanalysis of the data from Gur et al. [14] performed early in the study period, and most of the CAD-assisted interpretations were later. The small changes in cancer detection for the group as a whole using CAD in the Gur et al. study may therefore be a reflection of lower prevalence of cancer and increased availability of comparison films in the later rounds of screening performed.

It is notable that the recall rate was essentially unchanged using CAD in the Gur et al. study [14] for the group as a whole, and lower for the subset of most experienced radiologists. Recall rate is expected to increase using CAD and has been shown to increase by the other authors in the prospective studies cited previously [13, 15–17]. An unchanged or falling recall rate with CAD suggests that the radiologists may be ignoring CAD markings altogether or dismissing findings unless they are corroborated by CAD. When correctly used, CAD findings should be expected to result in the recall of some additional patients and evaluation of some additional findings. If all other factors remain constant, recall rates are expected to be higher using CAD. Future versions of CAD may allow assessment of the likelihood of malignancy for a given finding or diagnosis, but for current FDA-approved CAD systems, the intent is only to detect potential signs of malignancy that have been overlooked by the radiologist.

In the study by Ciatto et al. [4], a group of 10 radiologists reviewed test films without and later with CAD results, and individual performances were compared. Without CAD, sensitivity ranged from 64.7% to 100%, with an 85.8% average. With CAD, sensitivity increased overall to an average of 90.0%. However, one inexperienced radiologist actually missed more cancer with CAD than interpreting without it. All but one had increased recall rates with CAD.

Cancer detections should increase if recalls are increased, whether as a result of CAD or any other reason. The radiologist may in effect be moving to another point on the same receiver operating characteristic curve without any real change in sensitivity. In actual practice, however, it can be difficult to consciously increase recall rate. Presumably all radiologists would like to detect more breast cancer, especially earlier and smaller tumors. CAD may provide a guide for deciding which additional patients should be recalled, and may help ensure that small cancers, especially with subtle microcalcifications, are not overlooked. The size of cancers detected can be anticipated to be smaller when more subtle findings are subjected to workup. As shown by this study, the additional cancers detected with CAD are significantly smaller, and thus presumably harder to detect. A lower yield of cancer for recalls generated by CAD might therefore be expected.

In a study of subtle findings seen on prior mammograms at the site of subsequently diagnosed breast cancer, Ikeda et al. [19] noted that 42% had positive CAD findings on earlier mammograms. Logan-Young et al. had similar findings in patients with repeat CAD studies and interval cancers (presented at the 2003 meeting of the Radiological Society of North America, Chicago, IL). Thus, it appears it is not unusual to find CAD markings on earlier studies at the site where cancer may subsequently be diagnosed. It is both reasonable and expected that the radiologist will sometimes dismiss such CAD findings because they do not appear suspicious enough to warrant workup or biopsy. If the four interval cancers with prior CAD-interpreted mammograms during the study are included, the total cancers in the study would be 108 instead of 104. Including these false-negatives yields a CAD sensitivity of 79.6% (86 of 108). The other interval cancers detected during the study period all had prior mammograms performed before the start of the study, thus without CAD, or at other facilities.

Although the use of CAD in this study was associated with increased recalls, the PPV of biopsy did not change significantly and the biopsy rate actually fell. We do not attribute the lower biopsy rate to the use of CAD, but rather to the lower rate of cancer during the study, probably related to fewer baseline and more repeat rounds of screening in the study mammograms.

The use of CAD in clinical practice ultimately depends on sorting out which of the thousands of marks reviewed warrant further evaluation. This is complicated by the relatively large number of extraneous CAD marks and is especially difficult at the outset. The doubling of recall rate in the first 2 months of using CAD in our study underscores this difficulty and perhaps suggests a strategy for introducing CAD in clinical practice. If initial use is focused on diagnostic patients instead of screening patients, it could allow workup of additional findings without necessitating the recall of additional patients or undue expense. It soon becomes apparent that most marks can be dismissed without further workup. As CAD algorithms are becoming more sophisticated, the number of nonsignificant marks is falling, but it is important for this type of system that sensitivity remain high so that subtle findings are marked. Sorting meaningful marks from unhelpful marks will therefore remain the crux of effectively using CAD.

In this study, CAD was effective in detecting additional cancers in both screening and diagnostic populations. The screening function of diagnostic mammography cannot be overemphasized because 43% (27 of 63) of the cancers detected in the diagnostic patients were nonpalpable. In this study, two patients with breast implants had cancers marked by the CAD, and one of these cancers would not otherwise have been detected (Fig. 1). Two recurrences of cancer after lumpectomy were also marked by CAD. These findings argue against excluding diagnostic, post-surgical, or implant patients from the use of CAD.

The detection of significant numbers of cancers not marked by the CAD (20% in our study) underscores the importance of using CAD markings only as a guide to additional findings and not to dismiss findings already noted by the radiologist. It is well shown that not all cancers are marked by currently available CAD systems, with sensitivity ranging from 76% to 94% [6, 9, 10, 15–17]. Therefore, it would be incorrect to rely on CAD to identify all relevant findings.
It is recommended by the manufacturer that the radiologist first interpret the films without review of CAD findings, and then re-check the areas marked by CAD for any potentially missed abnormality. Viewing CAD results first may bias the observer to give undue attention to only marked areas—in the current study this method could potentially have led to a significant number of missed cancers. Cancers not marked by CAD in our study included three cases with microcalcifications, three far posterior masses, two ill-defined masses, and several patients with masses in areas of extremely dense background parenchyma.

One concern about the study design used in this study, and by Freer and Ulissey [13], is that the radiologists may have been less diligent in initial review of the films because it was known that a final diagnosis would not be made until re-review after CAD findings were assessed. In fact, the opposite psychology may have led to even greater diligence, because no one wants to be “beaten by a machine.” Recall rates in our study may also have been increased in an effort to find all possible CAD-added cancers. This impression is supported by the observation that the recall rate fell in the 4 months after completion of the study period, although CAD was still in use. This may also reflect the decline in recalls as experience with using CAD increases. It should also be noted that the relative increase in recall rate was only 1.6 percentage points, increasing from 6.2 to 7.8%, and similar to the recall rates reported by Freer and Ulissey, which changed from 6.5 to 7.7% using CAD. Relative to the rate of cancers detected (6.2 per 1,000 in screening patients), these recall rates are reasonable. Despite more recalls, we also observed no significant change in PPV of biopsy, although the frequency of biopsy declined during the study period.

An obvious disadvantage of our study is that it included only one interpreting radiologist. However, the results do not appear to be unique. The rate of additional cancers added using CAD is within the range reported by the numerous authors cited previously. Recall rate and PPV of biopsy at the study facility are also well within recommended guidelines, and the number of interval cancers and the size and stages of cancers detected at the study facility indicate an effective mammography program. The higher-than-usual rates of cancer detection (16.6 per 1,000 in the diagnostic population and 6.2 per 1,000 in screening patients) probably reflect self-selection of a relatively higher-risk population because this is the only facility in the area where the radiologist is a dedicated breast imager.

This study was conducted by an experienced mammography interpreter, and it is possible that CAD might contribute a higher percentage of increased cancer detection in less specialized practice settings. Sickles et al. [20] showed higher cancer detection rates among experienced mammography interpreters compared with nonspecialist radiologists. The study by Ciatto et al. [4] suggests that different radiologists may respond to CAD markings quite differently. It is possible that the number of cancers detected by the radiologist regardless of CAD findings might also be different in less specialized practice settings.

Estimating the sensitivity of CAD is difficult, even with prospective data, because not all false-negative interpretations may be known, even with diligent follow-up. Because this was an observational study, many patients had prior mammograms at other facilities, and some will have subsequent mammograms performed elsewhere. The comparison of interval cancer rate and PPV of biopsy before and after introduction of CAD is also problematic because this is an observational study and not a controlled trial. Other factors, such as the quality of sonography, frequency of use of MRI, and improved biopsy technique, all may have contributed to the diagnosis of smaller tumors during the study.

The iCAD device has an upper size limit of mass the system will mark. Therefore, it is expected that CAD-marked cancers will be smaller than CAD-negative cancers, especially among palpable tumors. However, our study showed significantly smaller size for the additional nonpalpable invasive cancers detected using CAD. An increase in the proportion of small, early stage invasive cancers was also observed by Cupples et al. [16], and we expect it will be confirmed by others as well.

In this study five of the 10 cancers added by CAD were nonpalpable invasive tumors, and half of the invasive cancers not marked by CAD were palpable and therefore presumably would not have escaped detection. If the use of CAD only prompted the detection of larger or palpable tumors or only low-grade DCIS, using CAD might not impact patient survival or treatment options. Although CAD technology for mammography is still evolving, the results of this study show that its use can lead to the detection of more cancers in screening and diagnostic patients and significantly smaller, potentially earlier tumors.

Acknowledgment

We thank Professor Yuedong Wang, director of the statistical consulting laboratory at the University of California, Santa Barbara, for his help with the statistical analysis of this study.

References

13. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study
of 12,860 patients in a community breast center. *Radiology* 2001; 220:781–786


