Computer-aided Detection Output on 172 Subtle Findings on Normal Mammograms Previously Obtained in Women with Breast Cancer Detected at Follow-Up Screening Mammography

PURPOSE: To evaluate, by using a computer-aided detection (CAD) program, the nonspecific findings on normal screening mammograms obtained in women in whom breast cancer was later detected at follow-up screening mammography.

MATERIALS AND METHODS: Four hundred ninety-three mammogram pairs—an initial negative screening mammogram and a subsequently obtained screening mammogram showing cancer—were collected. The mean interval between examinations was 14.6 months. In 169 cases, in which 172 cancers were later depicted, findings on the initial mammogram were subtle enough that either none or only one or two of five blinded radiologists recommended screening recall. On the initial negative mammograms, of the 172 areas where cancer later developed, 137 (80%) had subtle nonspecific findings and were retrospectively judged as having a benign or normal appearance. The mammograms with these subtle findings were evaluated with a commercially available CAD program, and the numbers of CAD marks on these nonspecific findings were analyzed.

RESULTS: Of the 172 cancers, 129 (75%) were invasive and 43 (25%) were ductal carcinoma in situ. The CAD program marked 72 (42%) of the 172 findings that subsequently developed into cancer: 24 (29%) of 82 findings recalled by none, 25 (49%) of 51 findings recalled by one, and 23 (59%) of 39 findings recalled by two of the five radiologists. Among the 137 areas with nonspecific normal or benign findings, 41 (30%) areas where cancer subsequently developed were marked by the CAD program.

CONCLUSION: A subset of cancers have perceptible but nonspecific mammographic findings that may be marked by a CAD program, even when the findings do not warrant recall as judged at blinded and unblinded radiologist review. The authors believe failure to act on such nonspecific but CAD-marked findings prospectively does not constitute interpretation below a reasonable standard of care.

© RSNA, 2004

Computer-aided detection (CAD) of breast lesions involves the use of computer schemes to mark suspicious findings on mammograms, and the use of CAD to help improve breast cancer detection at mammographic screening has been proposed (1–3). Retrospective CAD studies of prior negative mammograms have been very sensitive in the marking of suspicious findings that were present but not prospectively recalled at screening. These studies have focused on the true-positive marks on suspicious findings that may have been missed. We recently reported on the nonspecific findings seen on the initial normal screening mammograms.
mammograms obtained in women in whom breast cancer was later detected at follow-up screening (4). In our report, we described perceptible but normal benign findings that would not be recalled for further evaluation, even in retrospect. In all cases, cancer was evident on follow-up screening mammograms obtained 9–14 months later. Thus, the purpose of our study was to evaluate, by using CAD, the nonspecific findings on normal screening mammograms obtained in women in whom breast cancer was later detected on follow-up screening mammograms.

**MATERIALS AND METHODS**

**Case Set**

The methods of mammogram collection in the case set have been previously described (1,2,4). Thirteen Mammography Quality Standards Act–certified facilities (eg, community-based hospitals, health maintenance organizations, and academic mammography centers) in the United States provided 1,083 consecutive cases of biopsy-proved cancer that was detected on screening mammograms obtained in asymptomatic women between 1994 and 1996 (1,2,4). Each institution gave institutional review board approval for use of these cases in this retrospective case-collection study, which was conducted in 1997. Informed consent was waived because any patient-identifying information was removed from all study materials. The average patient age at breast cancer detection was 62.3 years (range, 40–86 years). The 1,083 screening mammograms showing cancer were evaluated by one of the 13 facility radiologists, who, with knowledge of the biopsy-proved cancer location, marked this site on the screening mammograms by using transparent film overlays, one for each view.

The previously obtained negative screening mammograms (also referred to as prior or initial mammograms) from 493 cases were available for review. The mean time between the initial and follow-up screening examinations was 14.6 months (range, 9–24 months). Sixty-two of the 493 cases were excluded because of prior breast surgery that resulted in scars or findings marked by metallic skin markers. Four other cases were excluded because the original film hard copies were needed at the facility site before the end of the study. Thus, a total of 427 cases comprised the study cohort.

One of three board-certified radiologists, reviewed the 427 cases to determine if the cancers were visible in retrospect on the prior mammograms. One radiologist reviewed 242, one radiologist reviewed 103, and one radiologist reviewed 82 mammograms. Each radiologist used the previously created film overlays to locate the cancer on the prior mammograms. If a perceptible finding was deemed visible on the prior negative mammogram, the radiologist marked the location of the retrospectively visible finding by using a second set of transparent film overlays, thus establishing a reference-standard location of the subsequently detected cancer on the prior mammograms.

In 286 (67%) of the 427 cases, there were findings that were judged to be visible on the prior negative mammograms in locations where cancer later developed. The 286 prior negative mammograms were divided into four sets of approximately 75 mammograms each. Forty-five additional mammograms were added to each case set: five mammograms on which no abnormalities could be seen, 20 mammograms with small subtle cancers, and 20 negative mammograms, as confirmed on the basis of at least one subsequent mammographic examination with negative results during a 2-year follow-up period.

Blinded and Unblinded Radiologist Case Review

To determine if the findings on the prior negative mammograms should have been further evaluated, four panels of radiologists, each consisting of five members, performed a blinded review of the respective four case sets. The radiologists reviewed the original film hard copies and were unaware of the study purpose or the case mix. These panel radiologists (half of whom had a primary work focus in mammographic interpretation) were Mammography Quality Standards Act certified, had practiced radiology for a mean time of 17 years (range, 3–35 years), and had read a mean of 300 screening mammograms per month (range, 40–1,000 mammograms per month).

Each panel member independently assessed approximately 120 cases and categorized them according to American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) assessment codes (5). Lesions assigned BI-RADS codes 1 and 2 were considered normal or benign, and those assigned BI-RADS codes 0, 4, or 5 were considered abnormal. The use of BI-RADS code 3 was discouraged; however, the data showed that there were 16 cases with BI-RADS code 3 classifications, which for the purposes of this study were grouped with the BI-RADS codes 1 and 2 cases.

For clarity, we will refer to the cases that were assessed by the majority of five radiologists as BI-RADS codes 0, 4, or 5 as abnormal, meaning that the finding required immediate action. We will refer to mammograms showing findings that led the majority of radiologists to judge the case as BI-RADS code 1, 2, or 3 as normal, meaning that the finding was negative, benign, or not requiring immediate action. The panel radiologists were given the patients’ ages, they were shown only the prior negative mammograms (ie, mammograms obtained 9–24 months before the cancer was diagnosed at screening mammography), and no earlier obtained mammograms were presented. The panel radiologists had a mean sensitivity of 84% (mean of 16.8 of 20 cases) for cancer detection in the 20 cases of subtle cancer findings added to each case set and a mean specificity of 81% (mean of 16.2 of 20 cases) for diagnosing the 20 normal cases added to the case set.

Mammograms that were judged by three or more of the five blinded panel radiologists to show abnormal findings at the reference location were considered missed cancers, the rationale being that if the majority of radiologists at blinded review interpreted the mammogram as requiring immediate work-up, then the finding had been prospectively missed (1,2). One hundred twelve cases were judged to be abnormal by using these criteria (ie, by the majority of the panel) and thus were excluded from this study.

Three or more of the five blinded panel radiologists judged the remaining 174 cases to be negative, benign, or requiring no immediate work-up. We classified these cases as having nonspecific findings by using the rationale that if the majority of radiologists at blinded review interpreted a mammogram as normal, then the finding was very subtle, normal, or benign in appearance. Five of the 174 nonspecific finding cases were excluded: four cases in which the cancer location indicated on the mammogram was judged to be inconsistent with the pathologic diagnosis or the cancer location and one symptomatic case that should have been excluded in the original study. The remaining 169 cases, in which 172 cancers were depicted at subsequent follow-up screening, comprised our final study group or case set.

All mammograms were digitized at 50-μm spatial resolution capability by us-
ing a Lumisys LS85 digitizer (Lumisys, Sunnyvale, Calif) and downloaded to an Imation HQ969 laser printer (Imation Enterprises, St Paul, Minn) at 12 bits per pixel and 100-μm spatial resolution at the time of the initial study. These digitized images were then available on screen and printed to film hard copies for subsequent case review for this study. The digitized images were also archived for later CAD analysis.

The purpose of the unblinded review was to have the findings independently assessed by breast imaging specialists who knew the reference location of the subsequently detected cancer, to retrospectively reconfirm the assigned BI-RADS categories, to categorize each finding appearing, and to determine the reasons why the findings were nonspecific to the extent that they were categorized prospectively as BI-RADS category 1, 2, or 3 by the majority of five blinded radiologists. Two radiologists who specialize in breast imaging and have 14 (D.M.I.) and 19 (R.L.B.) years of experience in interpreting mammograms jointly reviewed the 169 cases in an unblinded review to categorize the findings and assess the possible reasons for the nondetection of and/or the nonaction on these findings. The digital-copy mammograms printed on film were used for this part of the study.

To ensure that the digital-copy mammograms were of sufficient quality for analysis, 20 original mammograms that included both masses (n = 12) and calcifications (n = 8) were recalled from the sites and compared side by side with the corresponding digital copies on dedicated mammography alternators. The two radiologists rated the image quality of the original and digital-copy mammograms by using a scale from 1 to 5—1 meaning unable to read, 3 meaning acceptable quality, and 5 meaning good quality—and a narrative description of mass or calcification visibility. The average quality ratings for the original (4.5) and digital-copy (4.4) mammograms were similar. The narrative descriptions revealed no cases in which the quality of the digital copy compromised the detection or characterization of a mass or calcification, further supporting the acceptability of using digital-copy mammograms for our study.

To assess the mammographic characteristics of the visible findings, the 169 prior negative digital-copy mammograms, the subsequently obtained follow-up mammograms on which cancer was detected, and the corresponding reference-location clear overlays were reviewed on a two-tiered dedicated motorized mammography alternator (RADX MS-604A; S&S X-Ray Products, Houston, Tex), with bright lights and magnifying lenses available for use. The four-view prior negative mammogram and its corresponding reference-location overlay were displayed on the top row, and the mammogram obtained 9–24 months later showing the cancer and its corresponding overlay were displayed on the bottom row. At the time of case review, although the locations of the subsequent cancers were evident from the follow-up mammograms, no patient information, examination date, or pathologic data were available to the unblinded reviewers.

The perceptible finding identified by using the reference-location overlay on the prior negative mammogram (reviewed on digital copies of the films) was analyzed according to finding type, size, location, and depth in the breast. The visibility of the lesion on each appropri-
ments of the abnormal findings on the mammograms. Of the invasive cancers, 112 were T1 tumors and 17 were T2 or higher-stage tumors. Of the 104 women with invasive cancer and known axillary node status, 22 (21%) had lymph nodes that were positive for metastatic disease.

All five of the blinded panel radiologists rated nearly half ($n = 80$ [47%]) of the 169 mammograms as normal at review. Fifty-one (30%) of the 169 mammograms were rated as normal by four of the five blinded panel radiologists, and 38 (22%) were rated as normal by three of these radiologists.

At unblinded review, with knowledge of the subsequent cancer location, the two breast imaging specialists would have recalled 35 (20%) of the 172 findings, rating them as BI-RADS 0 or 4 tumors. At unblinded review, the remaining 137 (80%) of the 172 findings were considered to be nonspecific, even in retrospect, by the two unblinded radiolo-

Figure 1. Mammograms obtained in 61-year-old woman show a BI-RADS 1 lesion seen at unblinded rereview with CAD output. Normal-appearing (a) craniocaudal and (b) mediolateral-oblique screening mammograms obtained 13 months prior to diagnosis of a 1.2-cm, grade II invasive ductal carcinoma are shown. When the location of the subsequently developing cancer is noted, in retrospect, there is a focal island of normal tissue in the lower part of the right breast that was interpreted as normal by the two breast imaging specialists at unblinded review. (c) Craniocaudal view subsequently obtained at the time of cancer diagnosis shows an oval obscured mass (arrow) in the outer part of the right breast that is denser than the tissue seen in a. (d) Mediolateral oblique view obtained at the time of cancer diagnosis shows the same mass (arrow) in the lower part of the right breast (Fig 1 continues).
gists, and all of these findings were judged to be BI-RADS 1 and 2 tumors with the radiologists having knowledge of the subsequent cancer location.

**CAD Evaluation**

The 169 original mammograms (ie, the high-spatial-resolution digital images copied from the original mammograms) showing the 172 findings were processed through a CAD system, and the number of CAD marks on each image was recorded. The CAD system (V2.3; R2 Technology, Los Altos, Calif) that was used consists of a laser digitizer, a computer that uses proprietary signal-processing algorithms, and a customized motorized viewer with video display monitors. The CAD algorithm involves the use of a neural network that identifies features of microcalcifications (ie, clusters of bright spots) and marks them with triangles. The program identifies masses or architectural distortions as regions of high density with or without radiating lines and marks them with asterisks. This software version has been found to make an average of two marks per case.

In general, cases with findings have more marks per case owing to the necessary addition of one or two marks per finding, depending on whether the finding is marked on one or on both standard mammographic views. A low-spatial-resolution (640 × 480 pixels) version of the marked digital mammogram is displayed on a small monitor on a motorized viewer directly below the digital mammograms when the unit is prompted. In normal clinical use, the marked images would be displayed only after the original film hardcopy images had been reviewed.

The 169 original mammograms were digitized and analyzed by the CAD program, and the resulting low-spatial-resolution images with marked areas were reviewed (by D.M.I., R.L.B., K.F.O.). The CAD-marked areas were directly compared with the findings on the digital-copy prior negative mammograms, and the corresponding reference-location overlays, described by the two unblinded breast imaging specialists. Each finding was judged to be either unmarked or marked by the CAD system. Marked findings included marks of the correct type on any part of the finding on either view, marks of either type on calcified masses, and calcification marks on any part of a calcification cluster. Unmarked findings were marks of the incorrect type on the finding or no marks on the findings. All other marks were recorded and counted as marks that were unrelated to the findings in this study.

Marks were recorded as “marks” or as “marks unrelated to the findings under study” and were compared with the pathology results, with each finding, and with the ratings assigned the findings at review by the five blinded panel radiologists and the two unblinded radiologists.

**RESULTS**

The numbers of cases and findings recalled by the five blinded radiologists as compared with the numbers of cases recalled by the two unblinded breast imaging specialists—data that indicate the number of times the CAD system correctly marked each finding—are summarized in Table 1. None of the five blinded panel radiologists recalled the cases associated with 82 (48%) of the 172 findings; CAD marked 24 (29%) of these 82 findings. The CAD system marked increasingly higher percentages of findings when one
or more of the five blinded radiologists recalled the case—specifically, it marked 25 (49%) of 51 findings recalled by one and 23 (59%) of 39 findings recalled by two of the five radiologists.

The data in Table 1 also show that of the 137 findings rated as normal or benign (ie, BI-RADS 1 or 2) by the two radiologists at unblinded retrospective review, 41 (30%) were marked by the CAD system. Specifically, CAD marked 41 findings that were judged to be not worthy of immediate action by both the majority of five blinded radiologists and the two unblinded radiologists (Figs 1, 2). CAD also marked the majority of findings (31 [89%] of 35) that were judged to be abnormal (BI-RADS 0 or 4) by the two radiologists at unblinded review.

The data in Table 2 are the two unblinded radiologists’ BI-RADS ratings of the findings, the pathologic diagnoses categorized by mammographic finding type, and the numbers of times the CAD system marked the findings. The CAD system frequently but not invariably marked nonspecific findings in locations where both invasive and noninvasive cancers later developed. Specifically, the system marked findings where 47 (36%) of the 129 invasive cancers and 25 (58%) of the 43 DCIS lesions later developed. The data in Table 2 also show the types of
TABLE 2
Findings Marked by CAD System Categorized by Finding Type, BI-RADS Rating, and Diagnosis at Follow-up Screening

<table>
<thead>
<tr>
<th>Finding Type</th>
<th>No. of Findings</th>
<th>Diagnosis at Follow-up*</th>
<th>BI-RADS Rating†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focal island of normal tissue</td>
<td>65</td>
<td>13/63</td>
<td>2/2</td>
</tr>
<tr>
<td>Benign-appearing calcifications</td>
<td>44</td>
<td>17/21</td>
<td>18/23</td>
</tr>
<tr>
<td>Few benign calcifications</td>
<td>24</td>
<td>4/9</td>
<td>3/15</td>
</tr>
<tr>
<td>Mass</td>
<td>11</td>
<td>6/11</td>
<td>0</td>
</tr>
<tr>
<td>Density (only on one view)</td>
<td>9</td>
<td>3/8</td>
<td>1/1</td>
</tr>
<tr>
<td>Mass with calcifications</td>
<td>3</td>
<td>2/2</td>
<td>1/1</td>
</tr>
<tr>
<td>Other</td>
<td>16</td>
<td>2/15</td>
<td>1/1</td>
</tr>
<tr>
<td>Total</td>
<td>172</td>
<td>47/129</td>
<td>25/43</td>
</tr>
</tbody>
</table>

* Data are numbers of findings marked by the CAD system and subsequently determined to be cancer/total number of findings.
† Diagnoses determined at follow-up screening mammography.
‡ BI-RADS rating assigned to findings seen on initial negative screening mammograms by two nonblinded breast imaging specialists.

TABLE 3
Number of Mammographic Views on Which a Finding Was Seen and CAD Results, Stratified according to BI-RADS Ratings

<table>
<thead>
<tr>
<th>No. of Views and CAD Results*</th>
<th>No. of Findings</th>
<th>BI-RADS 0 or 4†</th>
<th>BI-RADS 1 or 2†</th>
</tr>
</thead>
<tbody>
<tr>
<td>No view, none marked</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>One view</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marked</td>
<td>18</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Not marked</td>
<td>46</td>
<td>2</td>
<td>44</td>
</tr>
<tr>
<td>Two views</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both marked</td>
<td>23</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>One marked</td>
<td>31</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>None marked</td>
<td>48</td>
<td>2</td>
<td>46</td>
</tr>
<tr>
<td>All marked views‡</td>
<td>72</td>
<td>31</td>
<td>41</td>
</tr>
<tr>
<td>Total</td>
<td>172</td>
<td>35</td>
<td>137</td>
</tr>
</tbody>
</table>

* Numbers of views on which findings were seen and marked by the CAD system.
† Data are numbers of findings with the given BI-RADS classification, as assigned by two nonblinded breast imaging specialists.
‡ Total numbers of findings seen on marked views.

DISCUSSION

In most retrospective mammographic CAD studies, emphasis is placed on research of the true-positive marks on suspicious mammographic findings that require immediate action. Our study was focused on CAD marks on nonspecific findings that are perceptible on initial negative screening mammograms and subsequently develop into cancer. In the present study, a subset of 137 such nonspecific findings were not considered suspicious and did not warrant recall, even in retrospect. The CAD system marked 41 (30%) of these nonspecific findings, including 22 (47%) of 47 benign and benign-appearing calcifications and up to one-fifth (19 [21%] of 90) of other findings such as nonspecific islands of normal-appearing fibro glandular tissue. The CAD system marked nonspecific findings in locations where invasive cancer and DCIS subsequently developed.

Our finding that the CAD system marks normal or benign findings in locations where cancer later develops prompts several questions: When CAD is used on screening mammograms in a normal clinical setting, what is the clinical importance of CAD marks on nonspecific mammographic findings? Do all CAD marks on nonspecific findings warrant recall? Do all CAD marks on areas where cancer later develops warrant recall prospectively, even when the finding is nonspecific and cannot be distinguished from other nonspecific findings?

The CAD system marks a percentage of suspicious findings that might represent cancer—usually spiculated or irregular masses or pleomorphic calcifications. The benefit of CAD manifests when the radiologist’s attention is drawn to a find-
ing that is interpreted as abnormal, the patient is recalled at the screening, and the work-up results in a diagnosis of breast cancer that is treated adequately, and, thus, an improved prognosis for the woman is expected. With this scenario, one assumes that the radiologist interprets the marked finding correctly. However, mammographic interpretation is based on the mammographic features of the finding and the experience of the radiologist and not specifically on the CAD marks.

In clinical practice, nonspecific findings are often seen on mammograms. We have shown in prior work (4) that these nonspecific findings do not necessarily warrant recall, and it is our opinion that failure to act on these findings does not deviate from the appropriate standard of care. On the basis of standard medical practice, nonspecific findings do not require recall at screening (6,7) because these findings are ordinarily not reported. Thus, the CAD system may mark nonspecific findings, but these marks should be subordinated to standard interpretation. If nonspecific findings are correctly interpreted as benign and returned to screening in routine clinical practice, then it is reasonable that these findings will be returned to screening even when they are marked by a CAD system.

CAD algorithms are specifically designed to have a high stand-alone sensitivity for cancer detection, with a corresponding high number of false marks per case. Radiologists who purchase CAD systems are informed of the average number of marks per negative case and that the majority of marks do not indicate breast cancer. It is neither necessary nor reasonable for a radiologist to call back all findings marked by a CAD system. The CAD system used in this study made approximately two false-positive marks per mammographic examination. To calculate how many benign findings might be marked in this scenario, let us assume that there was a prevalence of six cancers per 1,000 screening mammograms and a (high) rate of false positives in a given screening mammogram of 0.2%. The predominant cause for false positive detections is the extent of digitization of a CAD system. Thus, the CAD systems in this study yield approximately 2 per 1,000 screening mammograms.

This decision is subject to the legal test of reasonableness—not of accuracy or certainty—so that a mistake is not tantamount to breaching the requisite standard of care (16).

Our study results indicate that CAD systems mark a finite percentage of nonspecific mammographic findings in areas where cancer subsequently develops. However, that percentage is neither sufficiently high nor designed for specificity to the extent that it can be an independent variable in deciding the reasonableness of recall. Similarly, if a finding that would be reasonable to recall is not recalled, then the presence (or absence) of a CAD mark is subordinate to the analysis of the specific mammographic features of that area. Although research of ways to reduce the variability of diagnoses by using CAD is underway (17), CAD technology has not been advocated with respect to recall rates, the determinants of which have been reported in other studies (18,19). Our study results indicate that when a nonspecific area is identified by using CAD in a detection setting, the clinical importance—and especially the legal importance—of the mark with respect to the decision to recall are moot.

There were several limitations in our study. The first limitation was our use of digitized-copy mammograms, which never have the spatial resolution or lesion conspicuity that completely matches those of the original films. Standard parameters were used to print the digital-copy films. However, we performed a matched evaluation of a subset of original and digital-copy mammograms, which revealed no substantial difference between the two images, validating our results.

A second possible limitation was our use of digitized mammograms from analog films. The fact that most commercially available CAD systems are used on digitized film, however, justifies the technique used in this study despite the noise and image degradation produced by the digitization process. In the future, other studies might be performed by using CAD on directly acquired digital mammograms.

Another limitation was the use of the five-radiologist panel that performed experimental blinded readings, as compared with truly prospective blinded readings, to define the set of nonspecific findings. It is well known that retrospective readings may yield more positive findings than prospective readings (13). The unblinded two-radiologist readings used to characterize and assess the nonspecific findings were another potential
but unavoidable limitation of our study, because complete case knowledge was necessary to truly know whether a finding was located where the cancer later developed. Also, since this was a retrospective study, whether the CAD marks would have prompted a change in prospective assessments was not evaluated, because CAD was not available at the time of the original interpretations. Although determining radiologists’ responses to CAD marks would have been interesting from a psychological aspect, it was beyond the scope of this study.

Our study results show that a subset of cancers have perceptible but nonspecific mammographic findings that do not warrant recall, as judged by a majority of blinded radiologists and by two unblinded radiologists, and that these findings may be marked by a CAD program. We believe failure to act on the nonspecific findings marked by a CAD program does not necessarily constitute mammographic interpretation below a reasonable standard of care.

References