Purpose: To prospectively assess the effect of computer-aided detection (CAD) on screening mammogram interpretation in an academic medical center to determine if the outcome is different than that previously reported for community practices.

Materials and Methods: Institutional review board approval was granted, and informed consent was waived. During a 19-month period, 8682 women (median age, 54 years; range, 33–95 years) underwent screening mammography. Each mammogram was interpreted by one of seven radiologists, followed by immediate re-evaluation of the mammogram with CAD information. Each recalled case was classified as follows: radiologist perceived the finding and CAD marked it, radiologist perceived the finding and CAD did not mark it, or CAD prompted the radiologist to perceive the finding and recall the patient. Lesion type was also recorded. Recalled patients were tracked to determine the effect of CAD on recall and biopsy recommendation rates, positive predictive value (PPV) of biopsy, and cancer detection rate. A 95% confidence interval was calculated for cancer detection rate. Pathologic examination was performed for all cancers.

Results: Of 8682 patients, 863 (9.9%) with 960 findings were recalled for further work-up (Breast Imaging Reporting and Data System category 0). After further diagnostic imaging, it was recommended that biopsy or aspiration be performed for 181 of 960 findings (19%); 165 interventions were confirmed to have been performed. Twenty-nine cancers were found in this group, with a PPV for biopsy of 18% (29 of 165 findings) and a cancer detection rate of 3.3 per 1000 screening mammograms (29 of 8682 patients). CAD-prompted recalls contributed 8% (73 of 960 findings) of total recalled findings and 7% (two of 29 lesions) of cancers detected. Of 29 cancers (59%), 17 manifested as masses and 12 (41%) were microcalcifications. Ten (34%) cancers were ductal carcinoma in situ, and the remaining cancers had an invasive component. Both cancers found with CAD manifested as masses, and both were invasive ductal carcinoma.

Conclusion: Prospective clinical use of CAD in a university hospital setting resulted in a 7.4% increase (from 27 to 29) in cancers detected. Both cancers were nonpalpable masses.


Screening for breast cancer in asymptomatic women is challenging. In a clinical environment, where a large volume and rapid, efficient interpretation provide the most successful business model (1), the prevalence of disease is low, the organ structure is complex, the findings are subtle, and radiologist fatigue is a very real factor. The screening examination binary interpretable decision is to separate the mammograms into two groups of findings—negative or positive. There is a need, however, both economically and ethically, to limit the number of women recalled for diagnostic evaluation. Recalls from screening mammography occur in approximately 5%–10% of patients, with breast cancer ultimately confirmed in only three to 10 of 1000 women screened (2).
If we look only at studies of missed cancers that were later detected at screening (eliminating those studies including interval cancers), most (57%–75%) of these missed lesions have some finding visible on mammograms in retrospect. Of these visible findings, however, only 27%–36% are interpreted as “actionable” (either warranting a recall of the patient or worrisome enough to consider biopsy recommendation) when evaluated in a blinded fashion (3,4). Experience has shown that prospective double reading of screening mammograms increases the detection rate of cancer from 4% to 15% (5). Error rates have been shown to decrease with multiple readings, whether by humans (5–8) or by humans aided by means of machines (9,10).

Computer-aided detection (CAD), when used as a “spell check” type of system, has been shown to mark missed mammographic lesions later shown to be cancer in both retrospective and prospective studies. The goal of our study, therefore, was to prospectively assess the effect of CAD on the outcome of screening mammogram interpretation in an academic medical center to determine if this outcome is different than that previously reported for community practice experience (10).

**MATERIALS AND METHODS**

**Patients and Image Interpretation**

Institutional review board approval was granted for this study, which included a waiver of informed consent. During the 19-month period from January 2001 through July 2002, 8682 screening mammographic examinations were performed in the same number of women, with a median age of 54 years (range, 33–95 years). All mammograms were obtained with one of two mammography systems (600T or 800T; GE Medical Systems, Milwaukee, Wis) with screens (MIN-R; Eastman Kodak, Rochester, NY) and film (MIN-R; Eastman Kodak). Each examination consisted of a standard two-view mammogram obtained by experienced technologists certified to perform mammography. The images were evaluated with a CAD system (ImageChecker V2.2; R2 Technology, Sunnyvale, Calif), with which the images were digitized and analyzed in a freestanding unit. These images were linked to a modified mammogram viewer by means of a bar code. After the standard analog images were loaded on the screen-film viewing panels, the radiologist initially interpreted the images by making an assessment as to whether any abnormal findings were present. The CAD information was then presented to the radiologist on low-resolution monitors embedded in the control panel that displayed the CAD marks on low-resolution images of the mammogram. The CAD system displayed an asterisk for densities with or without crossing lines (masses or architectural distortions) or a triangle for collections of bright spots (calcifications).

The image interpretation protocol included preliminary evaluation of approximately half of the mammograms by a 2nd-, 3rd-, or 4th-year radiology resident and final interpretation by one of seven attending radiologists (two who specialized in breast imaging [R.L.B., D.M.I.] and five with special interest in breast imaging) who met the requirements of the Mammography Quality Standards Act of 1992. The years of experience for all breast image readers ranged from 10 to 30 years. The analog images were interpreted in a standard fashion, including the use of clinical history, a magnifying glass, and any available images from previous examinations. After the attending radiologist made the final assessment, the CAD marks were displayed. At this point, the radiologist reviewed the areas marked by the CAD system and assigned final Breast Imaging Reporting and Data System (BI-RADS) (11) categories for the cases.

For the purposes of this study, the detection methods for those mammograms assessed as BI-RADS category 0 (recall cases) were classified as follows: (a) the radiologist initially detected a finding and this area was marked by the CAD system on at least one of the two mammographic views (radiologist and CAD group); (b) the radiologist detected a finding and the area was not marked by the CAD system (radiologist-only group); and (c) the radiologist initially assessed the mammogram as negative but was prompted by the CAD marks to look again at an area (or areas) and then judged the finding as warranting a recall (CAD-only group). Each radiologist categorized the recall cases at interpretation into one of these three groups and recorded each finding separately into a logbook that allowed for more than one lesion per patient.

**Data and Statistical Analysis**

The results of the recall cases were collected and tracked on the basis of the systems put in place because of requirements of the Mammography Quality Standards Act of 1992. Of the 8682 women examined with screening mammography and CAD, the following data were collected (R.L.B., P.B.): the number of women recalled from screening; the type of lesion prompting the recalls and how these lesions were detected (radiologist and CAD, radiologist only, CAD only); and the results of follow-up diagnostic imaging, including the number of examinations in which additional imaging resulted in a negative reading, the number of biopsies recommended, and the number of women lost to follow-up. Also collected were the types of lesions recommended for biopsy and how these lesions were detected; biopsy methods used and information pertinent to the cancers diagnosed, including lesion type and method of detection; and cancer sizes, invasive versus in situ histologic characteristics, and grades. The callback rate was calculated and compared with that from a prior time period.

The predicted improvement in the detection rate was computed as the ratio of cancers detected with the CAD system alone over those detected without CAD. Because this ratio is necessarily greater than zero if even one cancer is detected with CAD alone, the effect is described in terms of a 95% confidence interval rather than a hypothesis test (12). All analyses were conducted by using SAS software (version 8.2; SAS Institute, Cary, NC). Our investigation was not a paired study in that we had no control group. Therefore, we performed no paired study analyses.

**RESULTS**

**Recalls**

Of the 8682 women who underwent screening examinations, 863 women with 960 findings were recalled for diagnostic imaging. Table 1 indicates both the lesion type and how the lesion was detected (radiologist and CAD, radiologist only, CAD only). The most common lesion type was a mass (583 of 960 lesions, 61%); the other lesion types were characterized as microlcalfications (345 of 960 lesions, 36%) and architectural distortions (32 of 960 lesions, 3%). Follow-up diagnostic imaging led to a recommendation for patients to return to screening for 718 of the 960 (75%) findings, and biopsy was recommended for 181 of the 960 findings (19%). For 61 of
Eight hundred sixty-three of the 8682 women (9.9%) were recalled during the study period. This recall rate is slightly higher than that from a similar period of time before the use of CAD (1021 of 12,147 women, 8.3%). Recall rates at the finding level (on the basis of the 960 abnormal findings) revealed a CAD-only contribution of 0.84% (73 of 8682 women), with most findings (10.2%, 887 of 8682 women) recalled by the radiologist (either radiologist-only or radiologist and CAD). Therefore, the use of CAD increased the recall rate at the finding level by 8.2% (73 of 887 women). Sixty-nine women were recalled on the basis of CAD findings alone, which resulted in a CAD-only contribution to the patient recall rate of 0.79% (69 of 8682 women).

**Biopsies**

Of the 181 lesions recommended for intervention (in 159 women), procedures were completed for 165 lesions (91%). Thirty-eight women with 16 findings (9%) did not comply with biopsy recommendations because they were lost to follow-up, they relocated, or they had insurance changes dictating care at a different facility.

## DISCUSSION

The stated purpose of the CAD systems that are presently approved by the U.S. Food and Drug Administration is to provide a “spell check” while looking at screening mammograms, with the intention of decreasing the oversight errors caused by those fallibilities inherent to the radiologist: fatigue, distraction, and inattention.

The positive predictive value (PPV) for all interventional methods was 18% (29 of 165 lesions). Twenty-eight of the 45 aspirations were performed in complicated cysts. If this group of benign lesions is removed from the analysis, the PPV is 24% (29 of 120 cancers). Furthermore, if the 54 lesions with benign core biopsies that were recommended for 6-month imaging follow-up (without any reported false-negative findings) are removed from analysis, the PPV for surgical biopsy is 44% (29 of 66 lesions).

### TABLE 1

<table>
<thead>
<tr>
<th>Lesion Type</th>
<th>Radiologist and CAD</th>
<th>Radiologist Only</th>
<th>CAD Only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>273 (52)</td>
<td>274 (57)</td>
<td>36 (55)</td>
<td>583 (61)</td>
</tr>
<tr>
<td>Calcification</td>
<td>244 (46)</td>
<td>72 (20)</td>
<td>29 (40)</td>
<td>345 (36)</td>
</tr>
<tr>
<td>Distortion</td>
<td>13 (2)</td>
<td>11 (3)</td>
<td>8 (11)</td>
<td>32 (3)</td>
</tr>
<tr>
<td>Total</td>
<td>530 (55)</td>
<td>357 (37)</td>
<td>73 (8)</td>
<td>960 (100)</td>
</tr>
</tbody>
</table>

Note.—Numbers in parentheses are percentages.

### TABLE 2

<table>
<thead>
<tr>
<th>Lesion Type</th>
<th>Radiologist and CAD</th>
<th>Radiologist Only</th>
<th>CAD Only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>55 (30)</td>
<td>45 (25)</td>
<td>3 (2)</td>
<td>103 (57)</td>
</tr>
<tr>
<td>Calcification</td>
<td>57 (31)</td>
<td>14 (8)</td>
<td>6 (3)</td>
<td>77 (43)</td>
</tr>
<tr>
<td>Distortion</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>112 (62)</td>
<td>60 (33)</td>
<td>9 (5)</td>
<td>181 (100)</td>
</tr>
</tbody>
</table>

Note.—Numbers in parentheses are percentages.

### TABLE 3

<table>
<thead>
<tr>
<th>Lesion Type</th>
<th>Radiologist and CAD</th>
<th>Radiologist Only</th>
<th>CAD Only</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass</td>
<td>10 (34)</td>
<td>5 (17)</td>
<td>2 (7)</td>
<td>17 (59)</td>
</tr>
<tr>
<td>Calcification</td>
<td>11 (38)</td>
<td>1 (3)</td>
<td>0 (0)</td>
<td>12 (41)</td>
</tr>
<tr>
<td>Total</td>
<td>21 (72)</td>
<td>6 (21)</td>
<td>2 (7)</td>
<td>29 (100)</td>
</tr>
</tbody>
</table>

Note.—Numbers in parentheses are percentages.
to fatigue. In addition to the complex breast structure and the need to detect subtle findings, an additional stress facing radiologists might be found in the effort to limit the recall rate. This limitation may contribute to the opportunity for false-negative interpretations. For most missed lesions (those cases of screening-detected lesions in which images from the most recent previous examination are available for comparison), some finding is visible in retrospect 57%–75% of the time; however, when these visible findings are evaluated in a blinded fashion, only 27%–36% are interpreted as “actionable” (ie, either warranting a recall of the patient or worrisome enough to consider a biopsy recommendation) (4).

CAD systems are designed to be used after independent (ie, unaided) case assessment by the interpreting radiologist who, after taking into account any CAD marks, assigns a final assessment to the case. One of the challenges inherent to the initial use of a CAD system by radiologists is to become comfortable with the number of predominantly false marks; false marks average 2.0 per four-view negative mammogram for the CAD system used in this study (13). It appeared to the interpreting radiologists that there was a learning curve when working with the CAD system marks, in that it took more time per screening mammogram interpretation early in our use of CAD than it does at the present time. Fortunately, with experience, the overwhelming majority of false CAD marks are readily dismissed.

At the time of preparing this article, there were only two published reports of prospective studies, to our knowledge, in which the effect of CAD on screening-detected breast cancer was reported. Both the study designs and results in these reports differ substantially. In 2001, Freer and Ulissey (10) reported on their prospective study, in which they used a study design similar to the one we followed, but their study was performed in a community practice setting and without the assistance of in-house staff. They interpreted 12,860 mammograms, recalled 986 women with 1026 findings, and diagnosed 49 cancers (10). Eight of the 49 cancers were detected with CAD only, which increased the cancer detection rate by 19.5% (eight of 49 cancers). In 2004, Gur et al (14), in a historical comparison study, reported no statistically significant difference in breast cancer detection rates for both groups (3.55% vs 3.49% per 1000 screening examinations).

In that study, 24 radiologists interpreted screening mammograms with \( (n = 59,139) \) and without \( (n = 56,432) \) CAD and found similar breast cancer detection rates for both groups (3.55% vs 3.49% per 1000 screening examinations).

Although Gur et al (14) reported no statistically significant increase in cancer detection between those radiologists using CAD and those reading screening mammograms without CAD, Feig et al (15) noted differences when separating seven high-volume radiologists—who interpreted 71% (82,128 of 115,571) of the cases—from 17 low-volume radiologists. Although the high-volume readers recorded no statistically significant difference in the recall or cancer detection rates with CAD compared with without CAD (11.05 vs 11.62 and 3.49 vs 3.61,
respectively), the cases in which images were interpreted by the low-volume readers did show differences in both recall and cancer detection rates with CAD and without CAD (12.00 vs 10.52 and 3.65 vs 3.05, respectively). The increase in the cancer detection rate for low-volume readers was approximately 19%; these increases in recall and cancer detection rates are similar to those reported by both Freer and Ulissey (10) and Cupples (16) with regard to community-based practices.

The only specific lesion type detection analysis in the study by Gur et al (14) was that of clustered microcalcifications alone, where they reported 1.44 per 1000 mammograms with CAD and 1.35 per 1000 mammograms without CAD. This finding differs from that of Freer and Ulissey (10), who saw an increase of 53% (80 of 150 cases) for those cases in which CAD results prompted the radiologist to detect the actionable calcifications. Differences in these two reported experiences raise the possibility that the Freer and Ulissey practice might have had a greater reliance on CAD and were more likely to recall those cases in which the CAD system marked calcifications. This may be due to the known performance of the CAD algorithm, which has consistently demonstrated higher sensitivity for calcifications (98%) than for masses (86%) (13).

Gur et al (14) suggest that our study and studies such as that by Freer and Ulissey (10) “may have been affected by the fact that our results of mammographic interpretations without and with CAD were reported on the same cases.” Their concern centers on the fact that the interpreting radiologists may have had a lower level of vigilance “because they knew that computer-aided detection would be available to them for the final recommendation and that the initial interpretation did not constitute a formal clinical recommendation.” However, we think it more likely that because the radiologists participating in the sequential study design used by us and by Freer and Ulissey knew that any “misses” on their part would be recorded, they were likely to be overly vigilant in their initial (pre-CAD) reading.

Our finding of a 7.4% increase in cancers detected with the use of CAD is similar to the 6.4% increase reported by Morton et al (17), but both are less than the 19.5% increase reported by Freer and Ulissey (10). Speculations as to the causes of these differences in cancer detection rates might be related to practice setting, the volume of cases interpreted per radiologist, the number of radiologists dedicated to interpreting breast images and teaching about breast imaging, and the effect of interpretations made in conjunction with the in-house radiology staff (double reading) before engaging a CAD system.
We were somewhat surprised to note that both the cancers detected in the CAD-only group were masses rather than microcalcifications. CAD systems are known to be very sensitive for marking microcalcifications (9,13). Although two of 29 (7%) additional cancers diagnosed because of CAD were masses, one of 12 (8%) cancers was perceived by a radiologist and not marked by the CAD system. The result that radiologists found 27 (93%) of the 29 cancers and the CAD system marked 23 (79%) emphasizes the fact that CAD is indeed an adjunct method—a reminder to “take another look.” CAD is not a first-line assessment of screening mammograms where the negative mammograms (those mammograms without CAD marks) can be ignored by the radiologist and all of the physician’s attention can be focused on those mammograms with CAD marks. At the current level of performance, a CAD system cannot and should not replace the radiologist as either a first or a final look.

Possible study limitations include the fact that each interpreting radiologist recorded his or her own recall findings at the prospective interpretation. The presence of a nonbiased individual in the reading room to record all findings might have resulted in a more strict assessment of how each of the lesions was categorized as having been detected (radiologist and CAD, radiologist only, CAD only). The study also did not record the exact contribution of the preliminary evaluations by residents to the interpretations. It is also possible that the results from the 13 women with biopsy recommendations who were lost to follow-up could have had an effect on the overall results reported herein. Finally, the study was not designed for follow-up of patients into the next screening interval. Therefore, the cancer detection rate and PPV at biopsy may have been affected by those recalled patients who were returned to screening after undergoing diagnostic imaging work-up alone.

Our study was designed to assess the effect of a CAD system on the detection of breast cancer in an academic medical center. We found a 7.4% increase in the detection rate of breast cancer with CAD. Our findings did not suggest a shift toward the assistance of CAD systems in the detection of microcalcifications. Our recall rate showed a modest increase when compared with that from a similar time period when we interpreted screening mammograms without the use of CAD systems. Comparing our experience with those from the other two published analyses, we suggest that the use of CAD systems may be more or less beneficial depending on practice type, practice volume, number of dedicated breast imagers interpreting the mammograms, the addition of human-human (attending radiologist and house staff) double interpretation before engaging the CAD system, and the experience of the radiologists with the CAD system.

Figure 3. Images in a 49-year-old nulliparous woman. (a) Mediolateral oblique mammogram shows a density (arrow) in the upper part of the left breast. (b) Craniocaudal mammogram shows that although this lesion was found to be medial in location, the density (arrowheads) was actually anterior to the mass and represented noncompressed breast tissue. The lesion was marked at CAD on a, and the radiologist “looked again” and recommended recall (CAD-only finding). (c) Sonogram obtained at the 2 o’clock position reveals an irregular hypoechoic mass (arrows) suspicious for carcinoma. Results of biopsy showed a 5-mm grade I infiltrating ductal carcinoma.
Acknowledgment: The authors thank Dave P. Miller, MS, for his assistance in performing the data statistical analyses.

References