Radiology

Jay A. Baker, MD
Joseph Y. Lo, PhD
David M. Delong, PhD
Carey E. Floyd, PhD

Index terms:
Breast neoplasms, localization, 08.112, 08.115
Breast radiography, quality assurance, 08.112, 08.115
Computers, diagnostic aid

Published online before print
10.1148/radiol.2332031200
Radiology 2004; 233:411–417

Abbreviation:
CAD = computer-aided detection

[1] From the Departments of Radiology and Biomedical Engineering, Duke University Medical Center, Erwin Rd, Durham, NC 27710. From the 2003 RSNA scientific assembly. Received July 28, 2003; revision requested October 6; revision received January 24, 2004; accepted March 2. **Address correspondence to** J.A.B. (e-mail: *jay.baker@duke.edu*).

Authors stated no financial relationship to disclose.

Author contributions:
Guarantor of integrity of entire study, J.A.B.; study concepts, J.A.B., J.Y.L., C.E.F.; study design, all authors; literature research, J.A.B., J.Y.L.; clinical and experimental studies, J.A.B.; data acquisition, J.A.B., J.Y.L.; data analysis/interpretation, all authors; statistical analysis, all authors; manuscript preparation and definition of intellectual content, all authors; manuscript editing, J.A.B., J.Y.L., C.E.F.; manuscript revision/review and final version approval, all authors

Computer-aided Detection in Screening Mammography: Variability in Cues[1]

PURPOSE: To evaluate the variability of true-positive and false-positive cues by using a commercially available computer-aided detection (CAD) system for analysis of 50 malignancies in a screening population.

MATERIALS AND METHODS: Fifty breast cancers detected at screening were analyzed by using a commercially available CAD system. Mean patient age was 62.2 years. Each set of mammograms (craniocaudal and mediolateral oblique views) was digitized and analyzed by the CAD system 10 times. One radiologist compared CAD output with the location of the malignancy at mammography and determined whether each lesion was marked accurately in one mammographic view, both views, or neither. Sensitivity and reproducibility of the CAD system were determined for both case- and image-based analysis.

RESULTS: Overall sensitivity of the CAD system when at least one of the two mammographic views was marked correctly (case-base sensitivity) was 82.4%. Sensitivity when each mammographic view was considered separately (image-based sensitivity) was 61.1%. For case-based analysis, variability in true-positive CAD cues was demonstrated for 14 of 50 (28%) cases. For image-based analysis, inconsistency in CAD output was observed in 33 of 100 (33%) mammographic views that contained malignancies detected at screening. However, the CAD system consistently detected 40–43 of the 50 breast cancers in each of the 10 CAD runs. Variability for false-positive marks was significantly greater than that for true-positive marks.

CONCLUSION: Inconsistency was demonstrated for CAD analysis of breast cancers detected at screening. However, the CAD system was reasonably consistent in the overall number of cancers identified from run to run. Greater variability of the CAD system was also demonstrated for false-positive marks, as compared with true-positive marks.

Interpretation of mammograms is a difficult task that results in a wide variation in ability, even between expert breast imagers (1–3). Because of the subtle appearance of some breast cancers, combined with the speed at which a large number of screening images must be interpreted, the false-negative rate for screening mammography has been reported to be approximately 20% (4,5). Computer-aided detection (CAD) systems have proved useful in reducing the frequency of "missed" breast cancers and improving the sensitivity of screening mammography (4,6–8). A retrospective study by Warren Burhenne et al (4) indicates that commercially available CAD systems can be used to successfully identify 77% of overlooked breast malignancies, while a prospective study by Freer and Ulissey (6) demonstrates that routine use of one CAD system may increase the number of cancers detected at screening mammography by up to 20%.

Because CAD systems use computer algorithms, they are presumed to offer extremely high or virtually perfect reproducibility in their analysis of mammograms (9). Indeed, this reputed reproducibility has been publicized as one of the many benefits of double-reading with a CAD system (9,10). However, a prior study performed in 2000 (11) demonstrated surprising variability in the output of one CAD system. The authors concluded at that time that such systems were insufficiently reproducible for routine clinical use. A recent study

(12) also demonstrated inconsistency in a more contemporary version of the same CAD system.

In contrast to the two published studies on CAD system reproducibility (11,12), two manufacturers' unpublished studies reported on a marketing Web site *(www.r2tech.com/prf/prf001.html#3)* and a Food and Drug Administration device labeling submission (iCAD device labeling, 2003) report excellent or near-perfect reproducibility. An important limiting characteristic in all of these prior reports is the patient population studied. In several of the studies in which the patient population can be determined, the study population likely includes symptomatic lesions rather than lesions in a typical screening population.

Further, in three of the four studies, each set of mammographic images was digitized and analyzed only three times, which may result in overestimation of the consistency of the CAD system. Thus, the purpose of our study was to evaluate the variability of true-positive and false-positive cues by using a commercially available CAD system for analysis of 50 malignancies in a screening population.

MATERIALS AND METHODS

Study Population

Institutional review board approval was obtained for this study. Informed consent was not required for this review.

Between November 1, 2001, and January 31, 2003, 67 biopsy-proved breast malignancies were detected in 12,789 screening mammographic examinations performed at our institution. As in prior similar studies, to simplify statistical analysis, multifocal and multicentric cases were excluded (11). The first 25 consecutive malignant masses and the first 25 consecutive malignant calcification clusters were used to constitute a study population of 50 breast cancers detected at screening. All mammograms were obtained by using standard screen-film technique with one of eight mammography systems qualified according to the Mammography Quality Standards Act (four Mammomat 3000 Nova systems, Siemens, Erlangen, Germany; two Mammomat 3000 systems, Siemens; one Senographe DMR+ system, GE Medical Systems, Milwaukee, Wis; and one MIII system, Lorad, Danbury, Conn).

Patients ranged in age from 40 to 90 years (mean, 62.2 years). Of the 50 malignancies, 17 (34%) represented ductal carcinoma in situ, 13 (26%) represented invasive ductal carcinoma, and 18 (36%) represented invasive ductal carcinoma and ductal carcinoma in situ. Two of the 50 cases (4%) represented invasive lobular carcinoma. Each malignancy was visible in both the craniocaudal and mediolateral views.

Study Design

Each craniocaudal and mediolateral oblique image in the 50 screening cases was digitized and analyzed 10 times by using the hardware and software provided with a commercially available CAD system (ImageChecker M1000, version 3.2; R2 Technology, Sunnyvale, Calif). The CAD system is designed to assist radiologists in the detection of breast cancer by identifying groups of bright specks that are suggestive of calcification clusters and by identifying densities with or without radiating lines that are suggestive of breast masses or foci of architectural distortion (8).

The output of the CAD algorithm was displayed on an 18-inch flat-panel display as a low-resolution mammographic image with small triangles placed to mark locations of possible calcification clusters and small asterisks placed to mark locations of possible breast masses. The CAD system used for this study had been in service for 4 months, and all quality-control tests recommended by the manufacturer were performed daily and weekly, as appropriate.

The output of the CAD system was reviewed by one dedicated breast radiologist (J.A.B.), who was qualified according to the Mammography Quality Standards Act and had 6 years of mammography experience. This radiologist traced the outline of the low-resolution image of each breast on an overlay, and the location of each CAD mark was recorded on the overlay. CAD marks for possible masses—displayed by the CAD system as asterisks—were differentiated on the overlay from triangle marks, which indicated potential clusters of calcifications.

The radiologist used all available mammographic views—including additional or special views—to determine the outline of the actual breast cancer (ie, margins of a mass or extent of calcifications) on each screening image. The radiologist then determined whether each CAD mark indicated the location of the malignancy. While there is no absolute rule to determine whether a CAD mark is sufficiently close to a lesion to represent a true-positive mark, previously published guidelines were used to determine true-positive and false-positive marks (13).

A CAD mark was labeled as true-positive if the mark was within the boundary that outlined the mass margin or within the extent of the calcifications. All CAD marks that did not mark the known malignancy were defined as false-positive marks for the purposes of this study.

The radiologist recorded whether each malignancy was marked correctly in the craniocaudal view, medolateral oblique view, both views, or neither view for each of the 10 CAD runs for all 50 cases. Therefore, a total of 500 CAD runs constituting 1000 mammographic images with a visible breast cancer in 50 screening cases were evaluated for this study. The reviewing radiologist recorded how many times and in which views each true-positive and each false-positive location was marked for the 10 CAD runs in each case. By using this system, the number and location of each true-positive and each false-positive mark were determined for the craniocaudal and medolateral oblique views for each of the 10 runs for the 50 cases.

Statistical Analysis

All statistical analysis was performed by using statistical software (SAS, version 8.2; SAS, Cary, NC). For the purpose of this study, each digitization and CAD analysis of the craniocaudal and medolateral oblique views of a single case were termed as one CAD run. Therefore, for this study, 10 CAD runs were performed in each of 50 cases, resulting in a total of 500 CAD analyses performed.

The sensitivity and reproducibility of the CAD system were determined for all 50 cases and were also calculated separately for malignant masses and malignant calcification clusters. In addition, sensitivity and reproducibility were calculated for these three populations by using case-based and image-based analysis. For case-based analysis, a successful mark of the cancer in either the craniocaudal or mediolateral oblique view was considered a true-positive identification of the cancer for that case. For image-based analysis, the craniocaudal and mediolateral oblique views were considered separately, resulting in two CAD analyses considered separately for each malignant case. CAD analysis could be true-positive in one view, both views, or neither view for image-based analysis.

The mean number of false-positive marks per image was calculated for all 50 cases. False-positive marks were defined

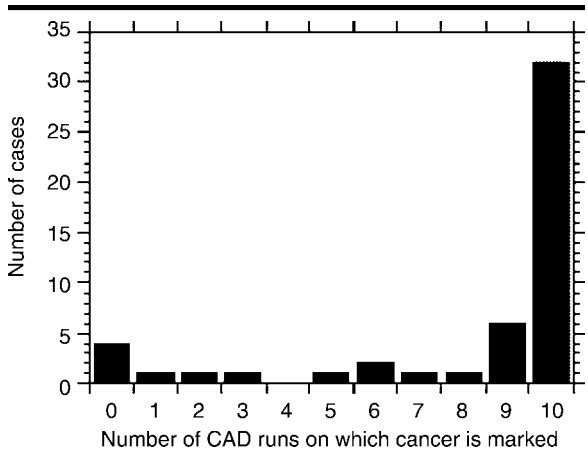


Figure 1. Histogram depicts the number of true-positive analyses out of 10 CAD runs for 50 breast cancers detected at screening by using case-based evaluation (ie, true-positive CAD cue in either mammographic view qualifies as true-positive for the case). Fourteen cases were marked between one and nine times but not all 10 times. Therefore, histogram illustrates that the CAD system provided inconsistent results for 14 of 50 (28%) malignancies in the present study.

TABLE 1
Breast Cancers Detected in Each of 10 CAD Runs for 50 Malignancies Detected at Screening

CAD Run	No. of True-Positive Cases
1	42 (84)
2	41 (82)
3	41 (82)
4	41 (82)
5	42 (84)
6	43 (86)
7	41 (82)
8	41 (82)
9	40 (80)
10	40 (80)

Note.—Numbers in parentheses are percentages. All lesions were visible in both craniocaudal and mediolateral oblique mammographic views.

as marks that the radiologist determined did not require further imaging or biopsy evaluation on the basis of the imaging characteristics. Reproducibility of false-positive marks was determined by using image-based analysis. The craniocaudal and mediolateral oblique views were considered separately (ie, image-based analysis) because false-positive marks cannot always be correlated between the two views. Reproducibility of false-positive marks was compared with the reproducibility of the true-positive marks by using image-based analysis.

Comparison of variance measures of the false-positive and true-positive marks in the craniocaudal and mediolateral oblique views was performed by using the Wilcoxon signed rank test. Comparison of sensitivities of the mediolateral oblique and craniocaudal views was also performed by using the Wilcoxon signed rank test. Comparison of sensitivity of the CAD system for detection of malignant microcalcification clusters and malignant masses was performed by using the Wilcoxon rank sum test. Only locations actually marked as false-positive in at least one view were considered for this comparison. Significance level was defined as $P < .05$.

RESULTS

Sensitivity of CAD Output

The overall case-based sensitivity (ie, true-positive finding defined as a lesion detected in either craniocaudal or mediolateral oblique view) for the 500 total case runs (50 cases, 10 CAD runs each) was 82.4% (412 correctly marked in 500 total case runs). The image-based sensitivity in which the craniocaudal and medolateral oblique views were considered separately was 61.1% (611 correctly marked images in 1000 total CAD runs).

Sensitivity for the craniocaudal view was compared with sensitivity for the mediolateral oblique view by using the Wilcoxon signed rank test. The sensitivity of all lesions in the 50 craniocaudal views analyzed was 62.0% ± 6.2 (standard error), which was not significantly different than the 60.2% ± 6.1 sensitivity of the same 50 lesions analyzed with the mediolateral oblique view ($P > .10$).

Sensitivity was also evaluated separately for masses and clusters of calcifications. No cases of masses with associated calcifications were included in this study. The case-based sensitivity for malignant masses detected at screening was 78.4% ± 7.0 (392 of 500 runs), and the case-based sensitivity for screening-detected malignant calcifications was 86.4% ± 5.9 (432 of 500 runs) over all 10 CAD runs.

With the Wilcoxon rank sum test, the sensitivities for detection of masses and calcifications were not significantly different ($P > .10$). The image-based sensitivity was 52.2% ± 6.0 (261 of 500 images) for screening-detected malignant masses and 70.0% ± 6.6 (350 of 500 images) for malignant calcification clusters. With the Wilcoxon rank sum test, the image-based sensitivity for malignant calcifications was significantly better than that for masses ($P = .04$).

Variability of CAD Output

While the case-based sensitivity was 82.4% ± 4.6 (412 of 500) for the combined 10 CAD runs in all 50 cases, the range in sensitivity for the 10 separate CAD analyses of the 50 cases was 80%–86%. That is, in 50 malignancies detected at screening, the CAD system detected 40–43 cancers in at least one routine mammographic view for each of the 10 runs (Table 1). Because there were differences in which cancers were identified in each of the 10 CAD runs, a total of 46 of the 50 cancers (92%) were identified on at least one of the 10 runs.

The clinical effect of variability in CAD analysis is demonstrated in a histogram that details the number of times each cancer was detected in the 10 CAD runs by using case-based determination of true-positive cases (Fig 1). The far right column of this histogram graphically demonstrates that 32 of the 50 cancers (64%) were detected in at least one mammographic view in all 10 CAD runs. The far left column indicates that four of the 50 malignancies (8%)—two masses and two groups of calcifications—were not detected in either view in any of the CAD runs.

In contrast, the remaining columns combined demonstrate that the cancer in 14 of the 50 cases (28%) was detected in at least one mammographic view between one and nine times out of the 10 CAD runs (Figs 2, 3). Ten of the 14 cases that demonstrated variability in CAD detection were masses, and four were malignant calcifications.

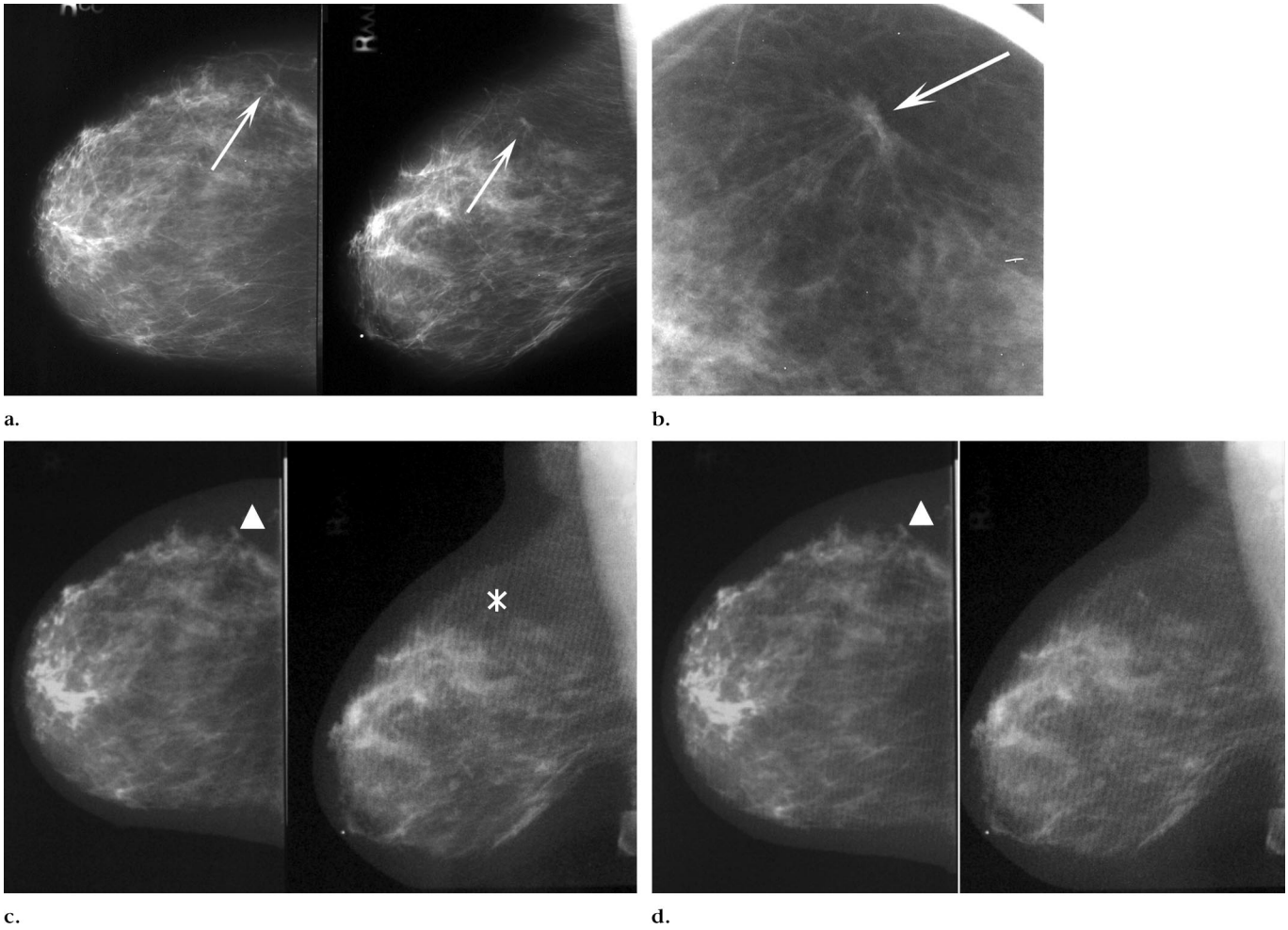


Figure 2. **(a)** Craniocaudal (left) and mediolateral oblique (right) mammograms in a 64-year-old woman with invasive ductal carcinoma detected as a 9-mm spiculated mass (arrows) at screening mammography. **(b)** Spot compression magnification view demonstrates the spiculated mass (arrow) in **a** to better advantage. **(c)** Photograph of computer monitor display of CAD system output. Asterisk overlying spiculated mass in the mediolateral oblique image (right) confirms accurate detection by the CAD system. Spiculated mass is noncalcified, and triangle overlying lateral right breast in craniocaudal view (left) indicates location of an artifact. **(d)** Photograph of computer monitor display of repeat CAD analysis. An asterisk does not overlie the mass in either projection, indicating false-negative analysis by the CAD system. This mass was detected in eight of 10 CAD runs.

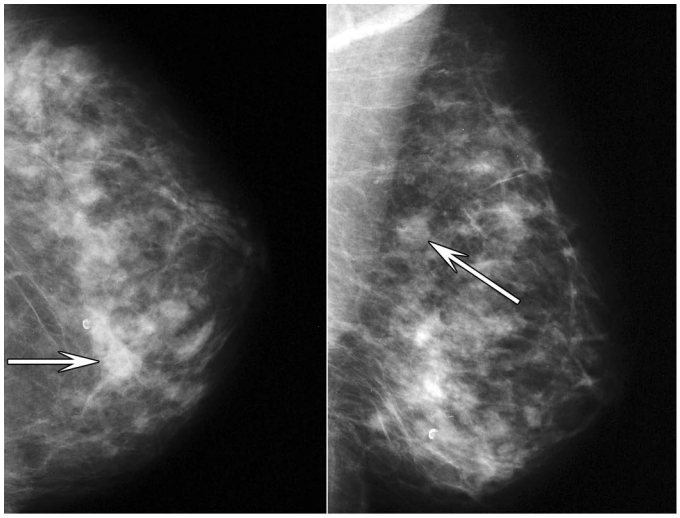
The image-based sensitivity was somewhat more variable. Image-based sensitivity for all 10 CAD runs combined was 61.1% (611 of 1000 images), but the sensitivity for the 10 separate CAD analyses for 100 mammographic views (ie, 50 cases with two routine views considered separately) ranged from 57% to 67%. That is, 57 to 67 cancers were detected in each of the 10 CAD runs for the 100 mammographic views (Table 2). Again, because of variability in which cancers were identified in each of the CAD runs, a total of 74 cancers were identified in at least one CAD run in the 100 mammographic views (74%).

Variability in the use of image-based analysis is best demonstrated in a histogram that illustrates the number of cancers identified in each of the 100 mammographic views (ie, two views for each of 50 cases) (Fig 4). For image-based true-positive analysis, the malignant lesion was detected in all 10 CAD runs for 41 of the 100 mammographic projections (far right column). The lesion was never detected by the CAD system in any of the 10 runs in 26 mammographic views (far left column). Columns one through nine combined demonstrate CAD inconsistency for the remaining 33 views (33%) in which the malignant lesion was identified correctly at least once but less than all 10 times.

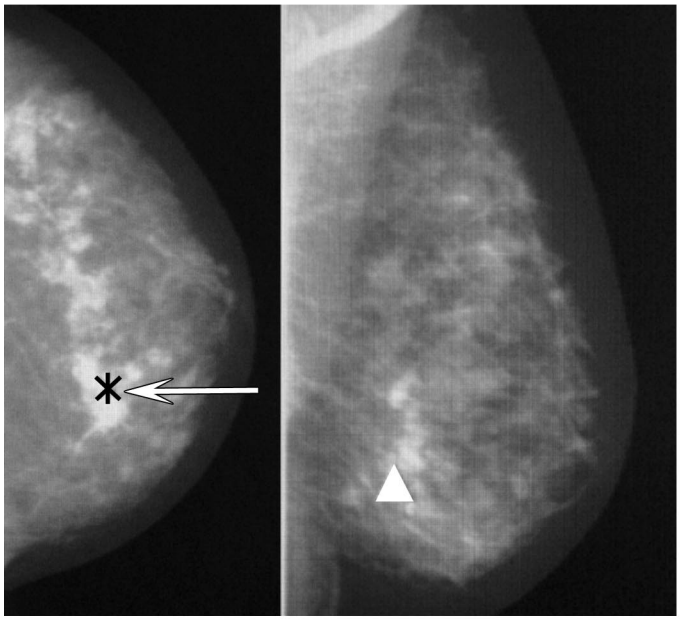
Reproducibility of False-Positive CAD Marks

Because many sites of false-positive findings—such as film artifacts or overlapping tissues in one view only—have no correlative false-positive findings in the accompanying mammographic view, false-positive marks were evaluated by using only image-based (ie, single-view) analysis in our study. Forty of the 100 mammographic views (17 craniocaudal, 23 mediolateral oblique) demonstrated no false-positive marks in any of the 10 CAD runs. The maximum number of false-positive marks identified over the 10 CAD runs combined for a single mammographic view was nine. However, a maximum of only five false-positive marks was identified in any one of the 10 CAD runs for this case; four additional false-positive marks were identified in the remaining nine CAD runs for this case.

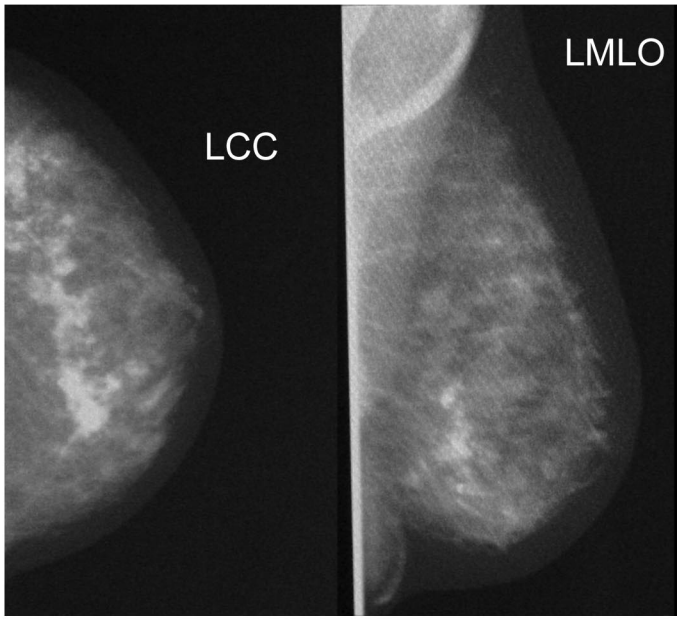
A total of 133 false-positive locations



a.



b.



c.

Figure 3. **(a)** Craniocaudal (left) and mediolateral oblique (right) mammograms in a 69-year-old woman with invasive ductal carcinoma detected as a 10-mm mass with an obscured margin (arrows) at screening mammography. **(b)** Photograph of computer monitor display of CAD system output. Asterisk (arrow) overlying breast mass in the craniocaudal projection (left) confirms accurate detection by the CAD system. Triangle overlying the inferior breast on the mediolateral oblique view (right) indicates location of an artifact. **(c)** Photograph of computer monitor display of repeat CAD analysis. An asterisk does not overlie the mass in either projection, indicating false-negative analysis by the CAD system. This mass was detected in seven of 10 CAD runs. *LCC* = left craniocaudal, *LMLO* = left mediolateral oblique.

were identified from the 10 CAD runs of the 100 mammographic images analyzed. Some of the false-positive locations were identified in only a single CAD run, while others were identified several times. A total of 660 false-positive marks were recorded by the CAD system for the 10 CAD runs of the 100 mammographic images for an average 0.66 (660 of 1000) false-positive marks per image. Only 24 false-positive locations in the 100 mammographic images were identified consistently in all 10 CAD runs.

Comparison of variability of the true-positive marks and false-positive marks by using the Wilcoxon signed rank test indicates that false-positive marks were significantly more variable for both craniocaudal views ($P < .001$) and mediolateral oblique views ($P = .022$).

DISCUSSION

While a cursory examination of CAD systems suggests that CAD analysis might be perfectly reproducible, prior studies have demonstrated some inconsistency in CAD cues (11,12). In the first published study on CAD reproducibility, to our knowledge, Malich et al (11) analyzed mammograms from 100 cases of breast cancer. Images from each examination were analyzed three times by using an older version of the same CAD system evaluated in the present study. Malich et al (11) determined that the CAD system provided identical true-positive and false-positive cues in the three analyses for only 18 of the 100 studies. When evaluating reproducibility of true-positive cues, the authors of that study compared only whether a lesion was marked consistently in both views of the three CAD runs. Since a true-positive cue in either mammographic projection may be sufficient to help a radiologist detect a cancer, however, this method of analysis may lead to overestimation of the potential clinical effect of CAD variability. Despite assertion that the high sensitivity of a CAD system could reduce the number of false-negative mammographic interpretations, the authors concluded, "be-

TABLE 2
Correctly Marked Breast Cancers on 100 Images (for Each of 10 Separate CAD Analyses)

CAD Run	No. of True-Positive Marks
1	60 (60)
2	57 (57)
3	59 (59)
4	63 (63)
5	64 (64)
6	67 (67)
7	61 (61)
8	59 (59)
9	63 (63)
10	58 (58)

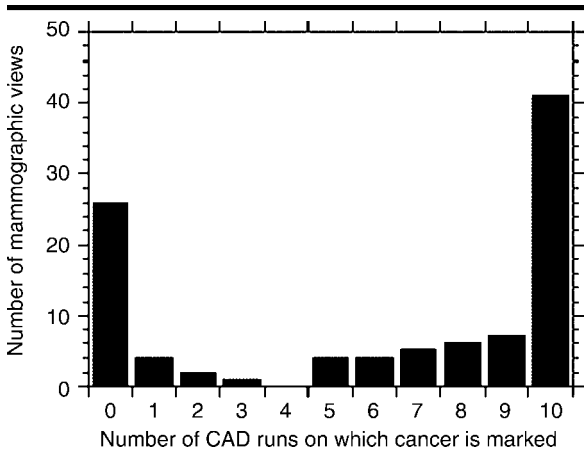


Figure 4. Histogram depicts the number of true-positive analyses in 10 CAD runs for 100 mammographic views of 50 breast cancers detected at screening by using image-based evaluation (ie, each mammographic view considered separately). Because of inconsistency in CAD output, a malignancy was marked correctly between one and nine times but not all 10 times for 33 of the 100 mammographic views (33%).

cause of technical limitations, the system used in the study does not reach sufficient values of reproducibility for the clinical routine" (11).

In the study of Zheng et al (12), in which a more recent version of the same CAD system was used, the authors also reported inconsistency in CAD sensitivity when 100 cases were evaluated three times. In that study, the authors reported that variability had decreased substantially with 78.8% of malignant masses and 93.5% of malignant calcification clusters identified consistently by using case-based analysis in three CAD runs. In the present study, only 52% of malignant masses and 76% (19 of 25) of malignant calcification clusters were identified consistently in all 10 CAD runs. Variability appears substantially higher in our analysis in part because CAD evaluation was performed 10 times in the present study, compared with only three times in the two reports published previously. These additional CAD runs provide substantially greater opportunity to detect CAD variability and, therefore, provide a more exhaustive evaluation of the potential effect of this inconsistency in actual clinical practice.

In addition to the two studies published previously with regard to CAD reproducibility, the manufacturers of the two most widely available commercial CAD systems have reported reproducibility studies that have not been published in the medical literature, including one study listed on a manufacturer's Web site *(www.r2tech.com/prf/prf001.html#3)* and the other reported in the manufacturer's submission for Food and Drug Administration device labeling for their CAD system (iCAD device labeling, 2003). Each of these unpublished studies report excellent or virtually perfect reproducibility. However, the design of these unpublished studies, as well as that of the reports published previously, may not precisely reflect the clinical effect of inconsistency in CAD output. Three of four prior studies limit analyses to three CAD runs per examination (11,12). Investigators in at least one prior published study (11) and one unpublished study *(www.r2tech.com/prf/prf001.html#3)* used older versions of commercial CAD systems. However, we believe the most important limiting characteristic of prior studies is the patient population used.

Because CAD is most commonly used as a "double read" of screening mammograms, a population of screening-detected malignancies should be emphasized in studies on the reproducibility of CAD systems. Although the precise patient population cannot be determined for any of the four prior CAD reproducibility studies, prior studies appear to have included symptomatic patients in the study population. In one manufacturer's study *(www.r2tech.com/prf/prf001.html#3)*, only "well-characterized" cancers were evaluated. Given the previously published studies and the data presented here, the near-perfect reproducibility reported in this manufacturer's study suggests that the cancers used may have been readily apparent.

The original study on CAD reproducibility (11), as well as the most recent report (12), do not state clearly whether palpable—and potentially more apparent—cancers were excluded. Just as sensitivity of a CAD system may be overestimated when conspicuous symptomatic cancers are analyzed, consistency of CAD may also be overestimated when the study population includes conspicuous symptomatic cancers.

In the present study, a current, well-maintained CAD system was used to analyze consecutive breast cancer cases detected at screening mammography. Although it is not known whether a CAD system can mark lesions in different ways depending on maintenance, it is possible that debris in the system could affect the digitization process adversely and therefore affect cancer detection. Each mammogram was analyzed by the CAD system 10 times. These factors suggest that the CAD variability presented here most closely reflects the inconsistency that can be encountered in daily clinical practice.

The potential clinical effect of CAD variability is considerable; in one-third of screening mammographic images with a visible breast cancer, the CAD system was inconsistent in its ability to detect and mark the malignant lesion. Indeed, in more than one case in four (28%) in which the cancer was only sporadically marked in at least one view, inconsistency in the output of the CAD system had the potential to directly affect whether a cancer was detected. Further, the cases that were marked with the least consistency—subtle masses—are the same cases for which radiologists would be most likely to benefit from consistent CAD assistance.

While inconsistency of the CAD system may directly affect whether an individual breast cancer is detected, less vari-

ability was noted for CAD analysis of the entire group of malignant cases. For the entire group, the sensitivity of the CAD system was more reproducible. That is, while individual cases were often marked inconsistently in the 10 CAD runs, the CAD system did consistently detect 40–43 of the 50 cancers present. This level of consistency was achieved because different cancers were detected or missed in different CAD runs. While an individual cancer may have been missed in a particular CAD run, a different cancer—missed previously—was detected in that run, maintaining the sensitivity over the study population in a moderately narrow range.

The results of this study indicate that while there is considerable inconsistency in the ability of the CAD system to detect any single cancer, this variability averages out, and the CAD system is reasonably consistent in its sensitivity when an entire population of cases is considered. The source of variability in CAD systems has been explored previously and is likely due to inconsistency in the initial digitization of the mammographic film and concomitant electronic noise caused by the digitization process (11). The present generation of CAD systems requires digitization of the mammographic film as the first step in analysis (14). This initial digitization step is subject to inconsistency, since the film advances into the digitizer in a slightly different lateral position and angle (11). For lesions at the threshold of detectability, this slight difference appears to be sufficient to alter the ability of the CAD system to detect the lesion. Since false-positive regions are likely to be closer to the threshold of detectability, slight differences in film positioning during digitization likely account for the greater variability measured for false-positive cues, as compared with true-positive cues (12).

In the present study, it is interesting to note that the overall sensitivity of the CAD system could have been improved by 10%, from an average of 82.4% to a maximum of 92.0%, by combining the CAD outputs for all 10 CAD runs. This approach to improving CAD sensitivity is clearly not feasible at the present speed of digitization. Further, combining the outputs of all 10 CAD runs would have doubled the false-positive marks from an average of 0.66 marks per film to a maximum of 1.33 marks per film.

There are three noteworthy limitations to the study presented here. First, in this study, we evaluated only one CAD model from only one of several commercial manufacturers. It is not known whether CAD systems developed by other manufacturers with different digitizers and detection algorithms would perform similarly. Second, this study was performed by using only a single example of the CAD model. Although the system was new and well maintained, it is possible that other units of the same model could perform dissimilarly. Finally, this study demonstrates how the CAD system performs independent of the interpreting radiologist. While the CAD system alone may have substantial interrun variability, the skill of the interpreting radiologist may mitigate some of this variability and result in overall greater reproducibility than the CAD system alone.

In summary, the results of this study indicate that state-of-the-art commercially available CAD systems for mammography can suffer from greater inconsistency than that reported previously when used to evaluate a screening population. The performance of the system is more reproducible when a population of cases is considered instead of a single mammographic examination. However, the importance of CAD reproducibility should be considered in conjunction with *(a)* prior studies that clearly demonstrate the ability of CAD systems to detect more than three-quarters of breast cancers overlooked by radiologists alone (4) and *(b)* studies that have demonstrated the ability of these systems to substantially improve the sensitivity of interpreting radiologists in actual clinical practice (6). These studies, combined with the data presented here, indicate that CAD systems can indeed help radiologists improve patient care, although CAD systems—like their human counterparts—may offer different opinions each time a mammogram is reviewed.

References

1. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. AJR Am J Roentgenol 2000; 174:1769–1777.
2. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer Inst 1998; 90:1801–1809.
3. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994; 331:1493–1499.
4. Warren Burhenne LJ, Wood S, D'Orsi C, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology 2000; 215:554–562.
5. Burrell H, Sibbering D, Wilson A, et al. Screening interval breast cancers: mammograhic features of prognostic factors. Radiology 1996; 199:811–817.
6. Freer T, Ulissey M. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology 2001; 220:781–786.
7. Birdwell R, Ikeda D, O'Shaughnessy K, Sickles E. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 2001; 219:192–202.
8. Evans W, Burhenne LW, Laurie L, O'Shaughnessy K, Castellino R. Invasive lobular carcinoma of the breast: mammographic characteristics and computer-aided detection. Radiology 2002; 225: 182–189.
9. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. AJR Am J Roentgenol 1994; 162: 699–708.
10. Working Party of the Radiologists Quality Assurance Coordinating Group. Computer aided detection in mammography. Sheffield, England: NHSBSP Publication, 2001.
11. Malich A, Azhari T, Bohm T, Fleck M, Kaiser W. Reproducibility: an important factor determining the quality of computer aided detection (CAD) systems. Eur J Radiol 2000; 36:170–174.
12. Zheng B, Hardesty L, Poller W, Sumkin J, Golla S. Mammography with computer-aided detection: reproducibility assessment—initial experience. Radiology 2003; 228:58–62.
13. Baker JA, Rosen EL, Lo JY, Gimenez EI, Walsh R, Soo MS. Computer-aided detection (CAD) in screening mammography: sensitivity of commercial CAD systems for detecting architectural distortion. AJR Am J Roentgenol 2003; 181:1083–1088.
14. Roehrig J, Castellino RA. The promise of computer aided detection in digital mammography. Eur J Radiol 1999; 31:35–39.