Impact of Computer-Aided Detection Systems on Radiologist Accuracy With Digital Mammography

Elodia B. Cole[1], **Zheng Zhang**[2], **Helga S. Marques**[3], **R. Edward Hendrick**[4], **Martin J. Yaffe**[5], and **Etta D. Pisano**[1]

[1]Department of Radiology and Radiological Science, Medical University of South Carolina, 96 Jonathan Lucas St, Ste 210, MSC 323, Charleston, SC 29425.

[2]Department of Biostatistics and Center for Statistical Science, Brown University School of Public Health, Providence, RI.

[3]Center for Statistical Science, Brown University, Providence, RI.

[4]Department of Radiology, School of Medicine, Anschutz Medical Campus, University of Colorado–Denver, Aurora, CO.

[5]Sunnybrook Health Sciences Centre, Toronto, ON, Canada.

Abstract

OBJECTIVE—The purpose of this study was to assess the impact of computer-aided detection (CAD) systems on the performance of radiologists with digital mammograms acquired during the Digital Mammographic Imaging Screening Trial (DMIST).

MATERIALS AND METHODS—Only those DMIST cases with proven cancer status by biopsy or 1-year follow-up that had available digital images were included in this multireader, multicase ROC study. Two commercially available CAD systems for digital mammography were used: iCAD SecondLook, version 1.4; and R2 ImageChecker Cenova, version 1.0. Fourteen radiologists interpreted, without and with CAD, a set of 300 cases (150 cancer, 150 benign or normal) on the iCAD SecondLook system, and 15 radiologists interpreted a different set of 300 cases (150 cancer, 150 benign or normal) on the R2 ImageChecker Cenova system.

RESULTS—The average AUC was 0.71 (95% CI, 0.66–0.76) without and 0.72 (95% CI, 0.67–0.77) with the iCAD system ($p = 0.07$). Similarly, the average AUC was 0.71 (95% CI, 0.66–0.76) without and 0.72 (95% CI 0.67–0.77) with the R2 system ($p = 0.08$). Sensitivity and specificity differences without and with CAD for both systems also were not significant.

CONCLUSION—Radiologists in our studies rarely changed their diagnostic decisions after the addition of CAD. The application of CAD had no statistically significant effect on radiologist AUC, sensitivity, or specificity performance with digital mammograms from DMIST.

Keywords

AUC; computer-aided detection; mammography; sensitivity; specificity

---

Address correspondence to E. B. Cole (coleeb@musc.edu)..

The American College of Radiology Imaging Network's Digital Mammographic Imaging Screening Trial (DMIST), which accrued women presenting for screening mammography from 2001 to 2003, included the acquisition of both digital and film-screen mammograms [1] and showed overall low sensitivity for both modalities [2]. Computer-aided detection (CAD) for film-screen mammography was introduced in the late 1990s. This tool was designed to aid radiologists in detecting breast cancers by functioning as a second reader, prompting an interpreting radiologist to take a second look at an algorithmically determined suspicious ROI.

Since the introduction of CAD into clinics, one of the chief complaints by radiologists using the technology has been the number of false-positive marks it produces. Each ROI marked by CAD has to be looked at a second time by the radiologist, so even 0.5 false-positive marks per image (i.e., two false-positive marks per four-view mammogram) leads to an increase in interpretation time. Concerns about high false-positive rates affecting study results, the limited clinical use of the technology, the lack of availability of the technology for digital mammography at that time, and the limited interest in CAD by radiologists on the DMIST trial design team are among the reasons why CAD was not used in DMIST. In the 4 years after the publication of DMIST study results, there was a significant increase in adoption of digital mammography, and CAD systems for digital mammography became available. This made it easier to integrate CAD into clinical work flows in mammography clinics. As a result, between 2004 and 2008, the use of CAD in screening mammography increased from 39% to 74% [3].

Although CAD was in widespread use in the United States by 2008, opinions expressed in articles published since remain mixed on the impact of this technology on radiologist performance in the detection of breast cancer [4–6]. Given the low sensitivity to breast cancers on both film-screen (0.41) and digital (0.41) mammography in DMIST [2] and the results of stand-alone assessment of the sensitivity of the two CAD systems, iCAD (0.74) and R2 (0.74), to cancers in DMIST (including interval cancers and those found at 1-year follow-up) [7], we anticipated that application of CAD to the challenging screening cases in DMIST should result in improved radiologist per formance. We therefore decided to examine radiologist performance without and with CAD applied to DMIST cases in a retrospective reader study.

Materials and Methods

Institutional review board approval was obtained before conducting these reader studies. Cases were obtained from DMIST, which consisted of 49,528 asymptomatic women presenting for screening mammography who enrolled at 33 participating institutions in the United States and Canada between 2001 and 2003. All DMIST cases were obtained with informed consent and HIPAA compliance. Only those cases with both proven cancer status on the basis of either biopsy or 1-year follow-up and available digital mammo-grams were included in this study.

Equipment and Case Selection

Two commercially available CAD systems were tested: R2 ImageChecker Cenova (version 1.0, Ho-logic); and iCAD SecondLook (version 1.4, iCAD). Cases were first grouped into BI-RADS categories [1–5], then randomly selected from within each category to mirror the distribution of BI-RADS assessments seen in the DMIST study. The cases were not consecutively chosen. Cancer cases were selected depending on the digital machine type for which the CAD algorithm could be applied, making an effort to retain similar proportions of film mammography BI-RADS scores within each cancer and noncancer group as the original DMIST data.

For the R2 study, 300 digital mammograms were selected from those available in Hologic Selenia (Hologic), GE 2000D (GE Healthcare), Fischer SenoScan (Fischer Imaging), and Fuji CR (Fuji Medical) systems. Another 300 digital mammograms were selected from those available in Hologic Selenia, GE 2000D, and Fuji CR systems for the iCAD study. The iCAD algorithm was not trained for use on Fischer SenoScan mammograms; therefore, this algorithm could not be applied to the Fischer cases acquired in DMIST. A random sample of cases, stratified first by cancer status and then by the BI-RADS category assigned to the film mammogram, was obtained from the initial DMIST screening. For each dataset, there were 150 cancer and 150 noncancer cases; noncancer cases included both normal and benign cases proven by biopsy or confirmed negative by 1-year follow-up mammography and without cancer 15 months after initial screening. In the iCAD study, the noncancer case set consisted of 90 normal and 60 benign cases. There were 95 normal and 55 benign cases included in the R2 study noncancer dataset. The R2 and iCAD reader studies had 206 cases in common.

Image Preparation

Both CAD systems required digital images to be presented in the for-processing DICOM format. In addition, the for-presentation DICOM format was required for display of the clinical images, along with the CAD marks overlaid on CAD manufacturer-specified mammography review workstations. The for-presentation format was previously applied to the DMIST-archived for-processing images in preparation for several previously published DMIST reader studies [8–10]. Each of the four digital mammography systems generated DICOM images. Some corrections had to be made to Fuji and Fischer DICOM headers, however, because they were missing DICOM values that were required by the CAD systems. These corrections were made using customized software written specifically for DMIST Fuji and Fischer cases. All corrected DICOM images included in this CAD study were added to the American College of Radiology Imaging Network archive for future use.

Two dedicated CAD reading rooms—one for the iCAD reader study and one for the R2 reader study—were set up in the breast imaging research laboratory at the University of North Carolina at Chapel Hill. CAD manufacturer-qualified soft-copy (i.e., digital image) review workstations were used for this study, with only one reader per CAD system reading at a time. Readers were not kept from seeing cases they may have seen in the original DMIST trial. The readers in our study were not informed of the cancer prevalence in the datasets.

Reader Experience

A total of 14 radiologists with clinical CAD experience using the iCAD system participated in the iCAD reader study. The iCAD readers had an average of 20.6 years of mammography experience (range, 6–40 years' experience) and read an average of 207 mammograms per week (range, 40–300 mammograms per week). All iCAD readers were currently using CAD in their mammography practice. The 14 iCAD study readers had an average of 5.7 years of iCAD experience (range, 2–10 years' experience) and an average of 4.6 years of digital mammography experience (range, 2–7 years' experience). Of these readers, 42.9% (6/14) had completed breast imaging fellowships, 78.6% (11/14) spent at least 20 hours per week (range, 3–50 hours per week) reading mammograms, and 21.4% (3/14) were original readers in the DMIST study.

A total of 15 radiologists with clinical CAD experience using the R2 system participated in the R2 reader study. The R2 readers had an average of 16.6 years of mammography experience (range, 3.5–32 years' experience) and read an average of 210 mammograms per week (range, 100–400 mammograms per week). Radiologists in the R2 study read an average of 207 mammo-grams per week (range, 40–500 mammograms per week). For the R2 study, all readers were currently using CAD in their mammography practice. The 15 readers had an average of 4 years of R2 experience (range, 4 months' to 8 years' experience) and an average of 4.8 years of digital mammography experience (range, 2 months' to 13 years' experience). Of these readers, 60% (9/15) had completed breast imaging fellowships, 73.3% (11/15) spent at least 20 hours per week (range, 10–48 hours per week) reading mammo-grams, and 40% (6/15) were original readers in the DMIST study.

Each reader was asked to identify the presence of actionable lesions, whether callback for the subject was recommended, and probability of malignancy (using DMIST 7-point scale [1]) per breast, first without CAD and then with CAD. Each reader reviewed each digital case without CAD marks and provided a forced BI-RADS assessment [1–5] for each breast. The reader then applied CAD to the digital case by pressing a toggle button on the mammography review workstation control keypad, allowing display of the CAD structured report with CAD marks overlaying the mammo-graphic images. The readers reviewed each case with the CAD marks and provided a second forced BI-RADS assessment and probability-of-malignancy score for each breast.

Statistical Methods

Our goal with this study was to determine whether the use of CAD could improve the sensitivity, specificity, and AUC of digital mam mography using screening cases obtained in the DMIST study. The DMIST 7-point probability-of-malignancy scale was used to score each breast. For this study, analysis was performed on the subject level. Readers had to correctly specify the laterality of the breast cancer and score that breast 4 or higher to get credit for finding a cancer. For bilateral cancer or noncancer cases, we used the larger of the two breast probability-of-malignancy scores. The empirical AUC of digital mammography with and without CAD was calculated. Sensitivity and specificity also were calculated. For each CAD system studied, the AUC was estimated for each reader, and the reader average AUCs with and without CAD were compared using a mixed model approach where the

modality is a fixed effect and the reader is a random effect [11]. Analyses of sensitivity and specificity were done in a similar manner using a generalized linear mixed model approach [12].

Results

Patient and Case Characteristics

The cases used in the two reader studies were selected from the original DMIST dataset to retain the distribution of film BIRADS assessments, and stratified sampling was performed on the basis of the true cancer status. Table 1 provides patient (age, breast density, and menopausal status) and case (manufacturer, workup, tumor histology, tu mor size, and lesion type) characteristics. Of all cancer cases included in this study, 72.7% and 70.7% were invasive in the iCAD and R2 reader studies, respectively. The stand-alone sensitivities of the two CAD systems tested were 0.75 for iCAD and 0.73 for R2 for cancer cases included in this study. False-positive detection rates for the two CAD systems were 0.79 for iCAD and 0.77 for R2 for the noncancer cases included in this study.

AUC Results

Nine of the 14 readers from the iCAD study showed improved AUC with CAD, and the differences were significant for four of these nine readers. However, the average AUC was 0.71 (95% CI, 0.66–0.76) without CAD and 0.72 (95% CI, 0.67–0.77) with CAD. The difference was not statistically significant ($p = 0.07$). Similarly, nine of the 15 readers from the R2 study showed improved AUC with CAD, and the differences were significant for four of these nine readers. However, the average AUC was 0.71 (95% CI, 0.66–0.76) without CAD and 0.72 (95% CI, 0.67–0.77) with CAD. Again, the difference was not statistically significant ($p = 0.08$) (Table 2).

Sensitivity Results

For the iCAD study, 13 of the 14 readers had better sensitivity with CAD, and four of those reached statistical significance. The average sensitivity was 0.49 (95% CI, 0.40– 0.57) without CAD and 0.51 (95% CI, 0.43– 0.60) with CAD ($p = 0.09$) (Table 3). For the R2 study, 12 of the 15 readers had better sensitivity with CAD, and four of those reached statistical significance. The average sensitivity was 0.51 (95% CI, 0.46–0.56) without CAD and 0.53 (95% CI, 0.48–0.58) with CAD ($p = 0.18$).

Specificity Results

For the iCAD study, the average specificity was 0.89 (95% CI, 0.83–0.93) without CAD and 0.87 (95% CI, 0.81–0.92) with CAD ($p = 0.15$). For the R2 study, the average specificity was 0.87 (95% CI, 0.83–0.91) without CAD and 0.86 (95% CI, 0.82–0.90) with CAD ($p = 0.24$) (Table 4).

Score Discordance Results

Of the 4191 case reviews in the iCAD study, the radiologists did not change their 7-point cancer probability rating in 4091 (97.6%) cases after CAD prompts were provided. An

example of a case in which cancer presenting as a mass was missed without and with both CAD systems by the majority of readers is shown in Figure 1. Of the 100 (2.4%) instances where the cancer probability rating was changed after CAD use, 61 were upgraded and 39 were downgraded. The average number of cases with discordant scores without or with CAD was 7.1 per reader and ranged from 1 to 23 cases. There were 35 unique cancer cases that were correctly upgraded by one or more of the 14 iCAD readers; of these cases, 46% (16/35) included calcification CAD marks overlaying the known cancer location, and 54% (19/35) included mass CAD marks overlaying the known cancer locations. An example of a cancer case presenting as calcifications that was correctly upgraded on the basis of CAD for four iCAD readers is shown in Figure 2. There were two cancer cases that were incorrectly downgraded. For the normal cases, 37 were correctly downgraded and six were incorrectly upgraded.

Of the 4494 case reviews in the R2 study, the radiologists did not change their 7-point probability rating in 4403 (98.0%) cases before and after CAD prompts were provided. The average number of cases with discordant scores without or with CAD was 6.1 per reader and ranged from 0 to 22 cases. Of the 33 unique cancer cases that were correctly upgraded by one or more of the 15 R2 readers, 45% (15/33) included calcification CAD marks overlaying the known cancer location and 55% (18/33) included mass CAD marks overlaying the known cancer location. There were six instances of incorrect downgrades of cancer cases with addition of CAD marks.

For the normal cases, 31 were correctly downgraded and four were incorrectly upgraded.

Discussion

In this study, we assessed the performance of CAD for digital mammography in a clinically realistic mix of cancer and noncancer (benign and normal) cases, albeit with an enriched ratio of cancer to noncancer cases, using 29 CAD-experienced readers. Our results are similar to those of others who have looked at the sensitivity of CAD with digital mammography [6, 13–16], finding no difference in radiologists' performance with and without CAD overall. Four of 14 readers (28.6%) achieved significant improvement in sensitivity with the iCAD system, and four of 15 readers (26.7%) achieved significant improvement in sensitivity with the R2 system. In our study, there was a wide range of radiologist performance without CAD, with sensitivity scores ranging from 0.23 to 0.79 for iCAD readers and 0.32 to 0.67 for R2 readers.

Variability in breast imaging radiologists' performance has been widely reported, first by Beam and Sullivan [17] and, more recently, by the Breast Cancer Surveillance Consortium [18]. Consortium data indicate a wide range of screening performance in a clinical setting, with sensitivities ranging from 60.6% to 97.8% [18]. Lower sensitivities have been reported in laboratory reader studies compared with radiologist clinical performance; this is probably because, unlike in clinical practice, there is no penalty for making a wrong decision in reader studies [17, 18]. The lack of prior mammo-grams has also been shown to result in lower sensitivity [19]. Although there was room for improvement in our radiologists' interpretations without CAD, few readers showed improved sensitivity after the application

of CAD to digital mammograms. The readers in the present study rarely changed their diagnostic decision after the addition of CAD.

High numbers of CAD marks per image are considered the primary reason for limited improvement in sensitivity with CAD in clinical practice. In the present study, there was an average of 0.78 marks per image for both iCAD and R2 datasets. Mahoney and Meganathan [20] suggested that most false-positive marks on normal mammograms are readily dismissed by radiologists and do not impact performance. In our study, there was a cancer prevalence rate of 50%, which is over one hundred times the prevalence rate that would be experienced in practice, which is approximately five cancers per 1000 women. Although we told the readers that the study was enriched with cancers, we did not specify by how much. The dismissal of so many true-positive CAD marks could be attributed in part to radiologists' behavioral differences in clinical environments, where they see far fewer abnormal mammograms, versus laboratory environments, where from necessity they see many more cancers and benign (but still abnormal) findings.

A second factor described in the literature as having an impact on radiologists' performance with CAD is training. Luo et al. [21] showed improved radiologist performance after 4 weeks of training. Although we did not recruit radiologists who underwent specific types of CAD training, our study did not include readers without clinical CAD experience, instead allowing only radiologists with iCAD experience to read in the iCAD study and only radiologists with R2 experience to read in the R2 study. A study conducted by Gilbert et al. [22] that compared single reading with CAD to double reading of screening mammograms showed improved detection performance with single reading with CAD over double reading for readers with 2 months of CAD training. Our study readers had at least 4 months of experience in using CAD in their clinical practices. We found no trends in our data suggesting that, after several months of experience, additional clinical experience with CAD had any impact on AUC or sensitivity (Figs. 3A and 3B).

A third factor cited in the literature as contributing to radiologists' improvement with CAD is lack of mammography interpretation experience. In the literature, the largest improvement in performance with CAD has been for radiologists with the least experience in mammography, such as junior breast imaging radiologists [23] or general radiologists with limited mammography interpretation in their practice [24]. Although we did not collect data on type of practice (academic vs nonacademic radiologists) or type of radiologist (general radiologist vs mammography specialist), our study found no correlation between mammography interpretation experience and the addition of CAD leading to an improvement in sensitivity or AUC (Figs. 3C and 3D). We suspect that as radiologists' confidence in their ability to correctly interpret findings increases over time, it is less likely that their opinion will change regarding the significance of a finding solely on the basis of the presence of a CAD mark.

There were a few limitations to this study. We did not establish a threshold for clinical relevance before starting the reader study. We did not disclose the prevalence of cancers in our two datasets to the readers, and this may have had an impact on their interpretation behavior. The readers did not have prior mammograms available during the study, which

could have had an impact on sensitivity and specificity. We did not require readers to localize the cancer beyond breast laterality. It is possible that requiring more detailed lesion localization could have increased the difference in cancer detection between interpretations without and with CAD. In our study, readers could have identified a suspicious area in the mammogram of a woman who had breast cancer, but the suspicious area might not have been the actual cancer. This error may have persisted in the CAD reading or may have been corrected if CAD correctly identified the cancer.

Conclusion

Radiologists in our study rarely changed their diagnostic decision after the addition of CAD, regardless of which CAD system was used. The application of CAD had no statistically significant effect on radiologist performance in interpreting digital mammograms from DMIST.

Acknowledgments

References

1. Pisano ED, Gatsonis CA, Yaffe MJ, et al. American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial: objectives and methodology. Radiology. 2005; 236:404–412. [PubMed: 15961755]

2. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med. 2005; 353:1773–1783. [PubMed: 16169887]

3. Rao VM, Levin DC, Parker L, Cavanaugh B, Fran-gos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? J Am Coll Radiol. 2010; 7:802–805. [PubMed: 20889111]

4. Fenton JJ, Abraham L, Taplin SH, et al. Effectiveness of computer-aided detection in community mammography practice. J Natl Cancer Inst. 2011; 103:1152–1161. [PubMed: 21795668]

5. Nishikawa RM, Schmidt RA, Linver MN, Edwards AV, Papaioannou J, Stull MA. Clinically missed cancer: how effectively can radiologists use computer-aided detection? AJR. 2012; 198:708–716. [PubMed: 22358014]

6. van den Biggelaar FJ, Kessels AG, van Engelshoven JM, Boetes C, Flobbe K. Computer-aided detection in full-field digital mammography in a clinical population: performance of radiolo-gist and technologists. Breast Cancer Res Treat. 2010; 120:499–506. [PubMed: 19418215]

7. Cole EB, Zhang Z, Marques HS, et al. Assessing the stand-alone sensitivity of computer-aided detection with cancer cases from the digital mammographic imaging screening trial. AJR. 2012; 199[web]:W392–W401.

8. Hendrick RE, Cole EB, Pisano ED, et al. Accuracy of soft-copy digital mammography versus that of screen-film mammography according to digital manufacturer: ACRIN DMIST retrospective multireader study. Radiology. 2008; 247:38–48. [PubMed: 18372463]

9. Nishikawa RM, Acharyya S, Gatsonis C, et al. Comparison of soft-copy and hard-copy reading for full-field digital mammography. Radiology. 2009; 251:41–49. [PubMed: 19332845]

10. Pisano ED, Acharyya S, Cole EB, et al. Cancer cases from ACRIN digital mammographic imaging screening trial: radiologist analysis with use of a logistic regression model. Radiology. 2009; 252:348–357. [PubMed: 19703878]

11. Obuchowski NA. Multireader, multimodality receiver operating characteristic curve studies: hypothesis testing and sample size estimation using an analysis of variance approach with dependent observations. Acad Radiol. 1995; 2(suppl 1):S22–S29. discussion, S57–S64, S70–S71. [PubMed: 9419702]

12. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc. 1993; 88:9–25.

13. Bolivar AV, Gomez SS, Merino P, et al. Computer-aided detection system applied to full-field digital mammograms. Acta Radiol. 2010; 51:1086–1092. [PubMed: 20883182]

14. James JJ, Gilbert FJ, Wallis MG, et al. Mammographic features of breast cancers at single reading with computer-aided detection and at double reading in a large multicenter prospective trial of computer-aided detection: CADET II. Radiology. 2010; 256:379–386. [PubMed: 20656831]

15. Scaranelo AM, Crystal P, Bukhanov K, Helbich TH. Sensitivity of a direct computer-aided detection system in full-field digital mammography for detection of microcalcifications not associated with mass or architectural distortion. Can Assoc Radiol J. 2010; 61:162–169. [PubMed: 20137883]

16. Sadaf A, Crystal P, Scaranelo A, Helbich T. Performance of computer-aided detection applied to full-field digital mammography in detection of breast cancers. Eur J Radiol. 2011; 77:457–461. [PubMed: 19875260]

17. Beam C, Sullivan D. Variability in mammogram interpretation. Adm Radiol J. 1996; 15:47, 49–50, 52. [PubMed: 10162612]

18. Breast Cancer Surveillance Consortium. Smoothed plots of sensitivity (among radiologists finding 30 or more cancers), 2004–2008—based on BCSC data through 2009. National Cancer Institute; website. breastscreening.cancer.gov/data/benchmarks/screening/2009/figure10.html. Last modified July 9, 2014. [July 10, 2014]

19. Ichikawa LE, Barlow WE, Anderson ML, et al. Time trends in radiologists' interpretive performance at screening mammography from the community-based breast cancer surveillance consortium, 1996–2004. Radiology. 2010; 256:74–82. [PubMed: 20505059]

20. Mahoney MC, Meganathan K. False positive marks on unsuspicious screening mammography with computer-aided detection. J Digit Imaging. 2011; 24:772–777. [PubMed: 21547517]

21. Luo P, Qian W, Romilly P. CAD-aided mammo-gram training. Acad Radiol. 2005; 12:1039–1048. [PubMed: 16087097]

22. Gilbert FJ, Astley SM, McGee MA, et al. Single reading with computer-aided detection and double reading of screening mammograms in the United Kingdom national breast screening program. Radiology. 2006; 241:47–53. [PubMed: 16990670]

23. Balleyguier C, Kinkel K, Fermanian J, et al. Computer-aided detection (CAD) in mammography: does it help the junior or the senior radiologist? Eur J Radiol. 2005; 54:90–96. [PubMed: 15797297]

24. Hukkinen K, Pamilo M. Does computer-aided detection assist in the early detection of breast cancer? Acta Radiol. 2005; 46:135–139. [PubMed: 15902887]
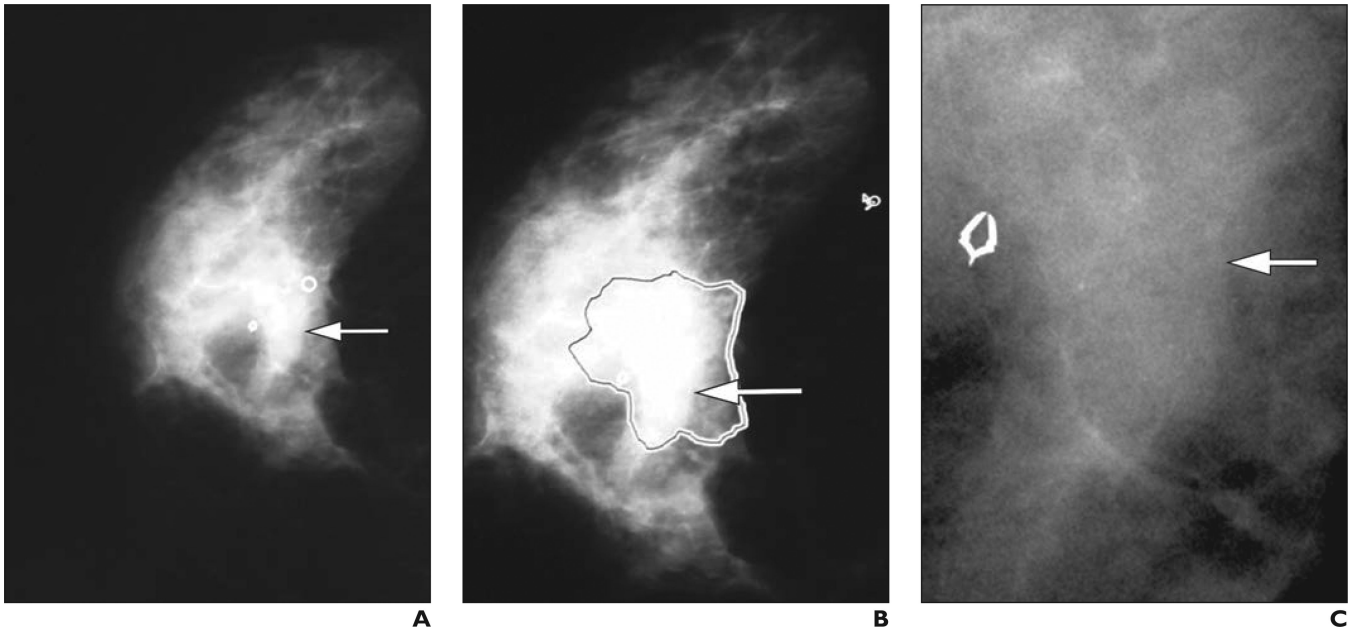
Fig. 1.
59-year-old woman with cancerous mass (*arrows*) missed by majority of readers without and with CAD.

A, Right craniocaudal digital mammogram shows cancerous subtle kidney-shaped mass, with iCAD mass mark (*circle*) adjacent to location of cancer, that was missed by nine of 14 iCAD readers without CAD. Presence of CAD mark changed impression of one reader; remaining eight did not change their initial impression.

B, R2 CAD mark (mass area outlined) is overlaid on same mass that was missed by 12 of 15 R2 readers without CAD. Presence of CAD mark did not change any of 12 readers' impressions.

C, High-resolution close-up shows mass.
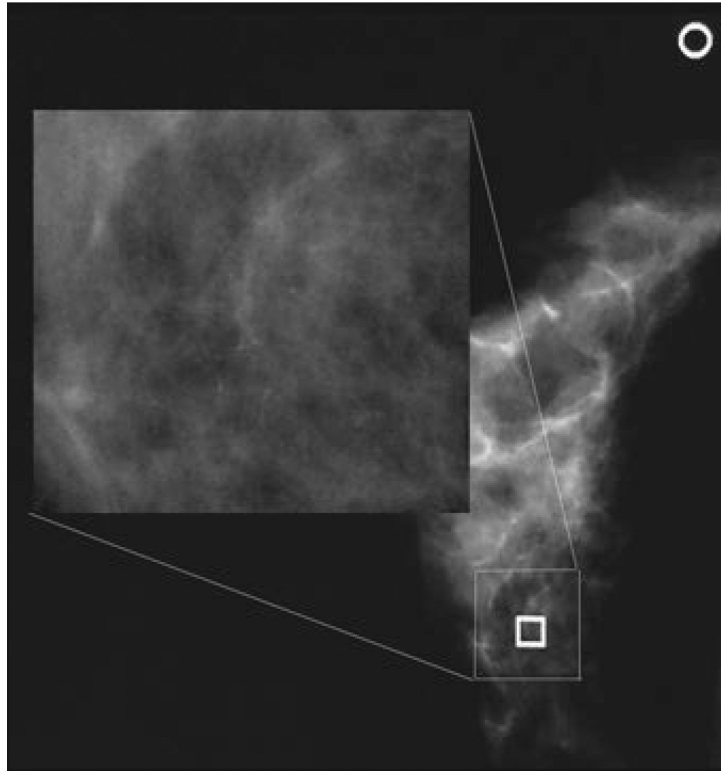
Fig. 2.
49-year-old woman with calcifications. Application of iCAD system resulted in change of reader impression regarding calcifications. Square is computer-aided detection (CAD) true-positive mark denoting cluster of calcifications, and circle is CAD false-positive mark denoting mass. In this cancer case, six of 14 readers missed calcification cluster on initial review without CAD. Once CAD was applied, five of six readers who initially specified no significant findings in right breast changed their impression to significant findings in right breast with probability of malignancy (rating of 4 on 7-point scale).
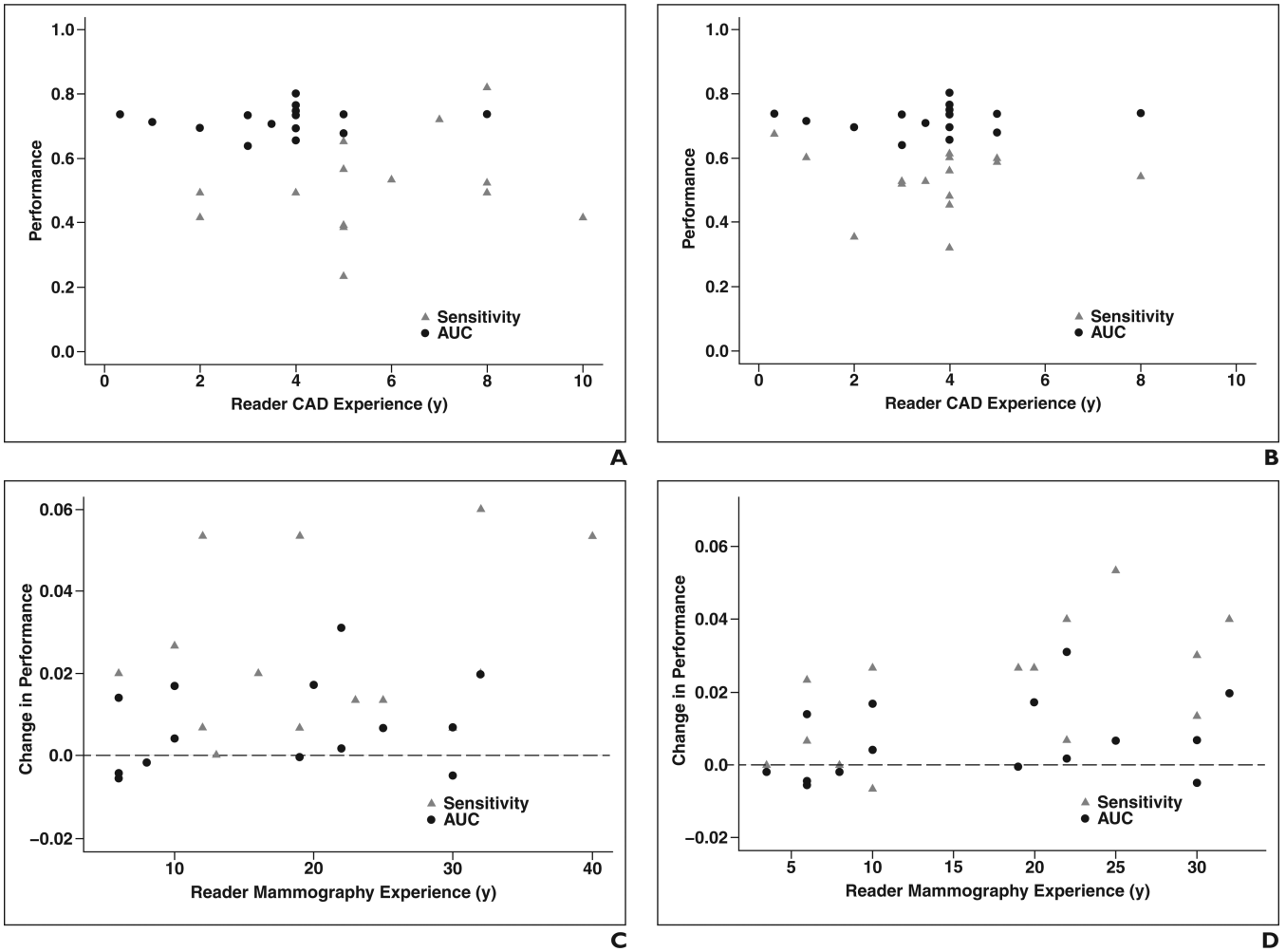
Fig. 3.

Scatterplots of reader experience versus AUC and sensitivity performance.

A and **B,** Scatterplots graph reader computer-aided detection (CAD) experience against AUC and sensitivity performance with iCAD (**A**) and R2 (**B**) systems.

C and **D,** Scatterplots graph reader mammography experience against improvement in AUC and sensitivity performance with iCAD (**C**) and R2 (**D**) systems. Dashed horizontal lines at 0.0 indicate no change in performance.

TABLE 1

Patient and Case Characteristic Distributions

	CAD System					
	iCAD			**R2**		
Parameter	**Noncancer (n = 150)**	**Cancer (n = 150)**	**All (n = 300)**	**Noncancer (n = 150)**	**Cancer (n = 150)**	**All (n = 300)**
Age (y)						
< 50	44 (29.3)	29 (19.3)	73	38 (25.3)	26 (17.3)	64
50–64	74 (49.3)	76 (50.7)	150	74 (49.3)	80 (53.3)	154
65	32 (21.3)	45 (30.0)	77	38 (25.3)	44 (29.3)	82
Breast density						
Dense	75 (50.0)	69 (46.0)	144	77 (51.3)	75 (50.0)	152
Fatty	75 (50.0)	81 (54.0)	156	73 (48.7)	75 (50.0)	148
Menopausal status						
Pre- or perimenopausal	49 (32.7)	40 (26.7)	89	38 (25.3)	42 (28.0)	80
Postmenopausal	96 (64.0)	106 (70.7)	202	105 (70.0)	105 (70.0)	210
Unknown	5 (3.3)	4 (2.7)	9	7 (4.7)	3 (2.0)	10
Manufacturer						
Fischer	0 (0)	0 (0)	0	46 (30.7)	47 (31.3)	93
Fuji	37 (24.7)	18 (12.0)	55	27 (18.0)	13 (8.7)	40
GE	113 (75.3)	127 (84.7)	240	76 (50.7)	86 (57.3)	162
Lorad-Selenia (Hologic)	0 (0)	5 (3.3)	5	1 (0.7)	4 (2.7)	5
Digital workup?						
No	138 (92.0)	65 (43.3)	203	139 (92.7)	63 (42.0)	202
Yes	12 (8.0)	85 (56.7)	97	11 (7.3)	87 (58.0)	98
Histology[a]						
DCIS		41 (27.3)			43 (28.7)	
Invasive cancer		109 (72.7)			106 (70.7)	
Other		0 (0)			1 (0.7)	
Tumor size[a] (mm)						
5		17 (11.3)			17 (11.3)	
> 5–10		29 (19.3)			26 (17.3)	
> 10		70 (46.7)			66 (44.0)	
Unknown		34 (22.7)			41 (27.3)	
Lesion type[a]						
Mass		65 (43.3)			65 (43.3)	
Asymmetric density		10 (6.7)			7 (4.7)	
Calcification		52 (34.7)			48 (32.0)	
Architectural distortion		8 (5.3)			11 (7.3)	
Unknown		15 (10.0)			19 (12.7)	

Note—Data shown are number of cases with percentages in parentheses. CAD = computer-aided detection, iCAD = SecondLook CAD system (version 1.4, iCAD), R2 = ImageChecker Cenova CAD system (version 1.0, Hologic), DCIS = ductal carcinoma in situ.

^aPresented for positive cases when available from pathology reports.

TABLE 2

Empirical AUCs for Computer-Aided Detection (CAD) Study

Vendor and Reader	Without CAD	With CAD	Difference	*p*
iCAD				
1	0.72 (0.67–0.78)	0.74 (0.69–0.79)	0.01 (0.00–0.03)	0.075
2	0.75 (0.70–0.81)	0.77 (0.72–0.82)	0.02 (0.00–0.03)	0.038
3	0.75 (0.70–0.81)	0.75 (0.70–0.81)	−0.001 (−0.02 to 0.02)	0.933
4	0.70 (0.64–0.76)	0.70 (0.65–0.76)	0.002 (−0.02 to 0.03)	0.872
5	0.77 (0.71–0.82)	0.79 (0.74–0.84)	0.02 (0.00–0.05)	0.027
6	0.67 (0.61–0.73)	0.67 (0.61–0.73)	0.004 (0.00–0.01)	0.318
7	0.59 (0.53–0.65)	0.62 (0.56–0.68)	0.03 (0.01–0.05)	0.007
8	0.78 (0.73–0.83)	0.78 (0.73–0.83)	−0.0002 (−0.01 to 0.01)	0.974
9	0.68 (0.62–0.74)	0.71 (0.65–0.76)	0.03 (0.00–0.06)	0.056
10	0.65 (0.60–0.71)	0.64 (0.58–0.70)	−0.02 (−0.03 to 0.00)	0.098
11	0.70 (0.65–0.76)	0.72 (0.66–0.77)	0.02 (0.00–0.03)	0.034
12	0.70 (0.64–0.76)	0.71 (0.65–0.77)	0.01 (−0.01 to 0.04)	0.210
13	0.76 (0.70–0.81)	0.75 (0.70–0.81)	−0.002 (−0.01 to 0.01)	0.625
14	0.76 (0.71–0.81)	0.76 (0.70–0.81)	−0.003 (−0.03 to 0.02)	0.829
Average	0.71 (0.66–0.76)	0.72 (0.67–0.77)	0.01 (−0.001 to 0.02)	0.070
R2				
1	0.70 (0.65–0.76)	0.73 (0.68–0.79)	0.04 (0.01–0.05)	0.008
2	0.70 (0.64–0.76)	0.69 (0.64–0.75)	−0.005 (−0.02 to 0.01)	0.584
3	0.68 (0.62–0.73)	0.68 (0.62–0.74)	0.002 (−0.02 to 0.02)	0.877
4	0.64 (0.59–0.70)	0.64 (0.58–0.70)	−0.01 (−0.02 to 0.01)	0.605
5	0.72 (0.66–0.78)	0.73 (0.68–0.79)	0.01 (0.00–0.03)	0.132
6	0.80 (0.75–0.85)	0.80 (0.75–0.85)	−0.001 (−0.02 to 0.02)	0.940
7	0.74 (0.69–0.79)	0.74 (0.68–0.79)	−0.01 (−0.02 to 0.01)	0.537
8	0.73 (0.68–0.78)	0.74 (0.68–0.79)	0.01 (−0.01 to 0.03)	0.493
9	0.75 (0.69–0.80)	0.76 (0.71–0.82)	0.02 (0.00–0.03)	0.041
10	0.71 (0.65–0.76)	0.71 (0.65–0.77)	0.01 (−0.02 to 0.04)	0.653
11	0.72 (0.66–0.77)	0.74 (0.68–0.79)	0.02 (0.00–0.04)	0.025
12	0.73 (0.68–0.78)	0.75 (0.70–0.80)	0.02 (0.00–0.03)	0.043
13	0.69 (0.63–0.75)	0.69 (0.63–0.75)	0.004 (−0.01 to 0.02)	0.576
14	0.66 (0.60–0.72)	0.66 (0.60–0.71)	−0.002 (−0.01 to 0.00)	0.319
15	0.71 (0.65–0.77)	0.71 (0.65–0.76)	−0.002 (−0.01 to 0.00)	0.450
Average	0.71 (0.66–0.76)	0.72 (0.67–0.77)	0.01 (−0.001 to 0.01)	0.079

Note—Except where otherwise indicated, data shown are AUCs with 95% CIs in parentheses. iCAD = SecondLook CAD system (version 1.4, iCAD), R2 = ImageChecker Cenova CAD system (version 1.0, Hologic).

TABLE 3

Sensitivities for Computer-Aided Detection (CAD) Study

| Vendor and Reader | Without CAD | | With CAD | | p |
	Raw Fraction	Estimate	Raw Fraction	Estimate	
iCAD					
1	60/149	0.40 (0.32–0.49)	62/149	0.42 (0.34–0.50)	0.500
2	82/150	0.55 (0.46–0.63)	85/150	0.57 (0.48–0.65)	0.250
3	76/149	0.51 (0.43–0.59)	78/149	0.52 (0.44–0.61)	0.500
4	65/150	0.43 (0.35–0.52)	74/150	0.49 (0.41–0.58)	0.004
5	100/150	0.67 (0.59–0.74)	108/150	0.72 (0.64–0.79)	0.008
6	61/149	0.41 (0.33–0.49)	62/149	0.42 (0.34–0.50)	1.000
7	34/149	0.23 (0.16–0.30)	35/149	0.23 (0.17–0.31)	1.000
8	95/150	0.63 (0.55–0.71)	98/150	0.65 (0.57–0.73)	0.250
9	66/150	0.44 (0.36–0.52)	74/150	0.49 (0.41–0.58)	0.008
10	58/150	0.39 (0.31–0.47)	58/150	0.39 (0.31–0.47)	1.000
11	71/150	0.47 (0.39–0.56)	74/150	0.49 (0.41–0.58)	0.250
12	72/150	0.48 (0.40–0.56)	80/150	0.53 (0.45–0.62)	0.008
13	58/150	0.39 (0.31–0.47)	59/150	0.39 (0.31–0.48)	1.000
14	119/150	0.79 (0.72–0.86)	123/150	0.82 (0.75–0.88)	0.219
Average		0.49 (0.40-0.57)		0.51 (0.43-0.60)	0.089
R2					
1	62/150	0.41 (0.33–0.50)	68/150	0.45 (0.37–0.54)	0.031
2	52/150	0.35 (0.27–0.43)	53/150	0.35 (0.28–0.44)	1.000
3	87/150	0.58 (0.50–0.66)	88/150	0.59 (0.50–0.67)	1.000
4	75/149	0.50 (0.42–0.59)	77/149	0.52 (0.43–0.60)	0.500
5	73/150	0.49 (0.40–0.57)	79/150	0.53 (0.44–0.61)	0.031
6	86/150	0.57 (0.49–0.65)	90/150	0.60 (0.52–0.68)	0.125
7	100/150	0.67 (0.59–0.74)	101/150	0.67 (0.59–0.75)	1.000
8	88/150	0.59 (0.50–0.67)	90/150	0.60 (0.52–0.68)	0.625
9	88/150	0.59 (0.50–0.67)	92/150	0.61 (0.53–0.69)	0.125
10	82/150	0.55 (0.46–0.63)	90/150	0.60 (0.52–0.68)	0.021
11	75/150	0.50 (0.42–0.58)	81/150	0.54 (0.46–0.62)	0.031
12	80/150	0.53 (0.45–0.62)	84/150	0.56 (0.48–0.64)	0.125
13	73/150	0.49 (0.40–0.57)	72/150	0.48 (0.40–0.56)	1.000
14	48/150	0.32 (0.25–0.40)	48/150	0.32 (0.25–0.40)	1.000
15	79/150	0.53 (0.44–0.61)	79/150	0.53 (0.44–0.61)	1.000
Average		0.51 (0.46–0.56)		0.53 (0.48–0.58)	0.183

Note—Except where otherwise indicated, data shown are sensitivities with 95% CIs in parentheses. iCAD = SecondLook CAD system (version 1.4, iCAD), R2 = ImageChecker Cenova CAD system (version 1.0, Hologic).

TABLE 4

Specificities for Computer-Aided Detection (CAD) Study

Vendor and Reader	Without CAD		With CAD		p
	Raw Fraction	Estimate	Raw Fraction	Estimate	
iCAD					
1	141/148	0.95 (0.90–0.98)	140/148	0.95 (0.90–0.98)	1.000
2	137/149	0.92 (0.86–0.96)	137/149	0.92 (0.86–0.96)	1.000
3	135/150	0.90 (0.84–0.94)	132/150	0.88 (0.82–0.93)	0.250
4	139/150	0.93 (0.87–0.96)	134/150	0.89 (0.83–0.94)	0.063
5	114/150	0.76 (0.68–0.83)	112/150	0.75 (0.67–0.81)	0.625
6	126/150	0.84 (0.77–0.89)	126/150	0.84 (0.77–0.89)	1.000
7	137/150	0.91 (0.86–0.95)	137/149	0.92 (0.86–0.96)	1.000
8	118/150	0.79 (0.71–0.85)	115/150	0.77 (0.69–0.83)	0.250
9	142/150	0.95 (0.90–0.98)	139/150	0.93 (0.87–0.96)	0.250
10	131/150	0.87 (0.81–0.92)	128/150	0.85 (0.79–0.91)	0.250
11	129/150	0.86 (0.79–0.91)	129/150	0.86 (0.79–0.91)	1.000
12	130/150	0.87 (0.80–0.92)	127/150	0.85 (0.78–0.90)	0.250
13	147/150	0.98 (0.94–1.00)	147/150	0.98 (0.94–1.00)	1.000
14	88/150	0.59 (0.50–0.67)	78/149	0.52 (0.44–0.61)	0.049
Average		0.89 (0.83–0.93)		0.87 (0.81–0.92)	0.146
R2					
1	140/150	0.93 (0.88–0.97)	140/150	0.93 (0.88–0.97)	1.000
2	140/150	0.93 (0.88–0.97)	139/150	0.93 (0.87–0.96)	1.000
3	109/149	0.73 (0.65–0.80)	107/148	0.72 (0.64–0.79)	1.000
4	120/150	0.80 (0.73–0.86)	115/150	0.77 (0.69–0.83)	0.063
5	131/150	0.87 (0.81–0.92)	130/150	0.87 (0.80–0.92)	1.000
6	138/150	0.92 (0.86–0.96)	135/150	0.90 (0.84–0.94)	0.250
7	111/149	0.74 (0.67–0.81)	108/149	0.72 (0.65–0.79)	0.250
8	125/150	0.83 (0.76–0.89)	122/150	0.81 (0.74–0.87)	0.250
9	135/150	0.90 (0.84–0.94)	136/150	0.91 (0.85–0.95)	1.000
10	125/150	0.83 (0.76–0.89)	115/150	0.77 (0.69–0.83)	0.006
11	130/150	0.87 (0.80–0.92)	130/150	0.87 (0.80–0.92)	1.000
12	136/150	0.91 (0.85–0.95)	136/150	0.91 (0.85–0.95)	1.000
13	132/150	0.88 (0.82–0.93)	133/150	0.89 (0.82–0.93)	1.000
14	143/149	0.96 (0.91–0.99)	143/149	0.96 (0.91–0.99)	1.000
15	122/149	0.82 (0.75–0.88)	120/149	0.81 (0.73–0.87)	0.500
Average		0.87 (0.83–0.91)		0.86 (0.82–0.90)	0.238

Note—Except where otherwise indicated, data shown are specificities with 95% CIs in parentheses. iCAD = SecondLook CAD system (version 1.4, iCAD), R2 = ImageChecker Cenova CAD system (version 1.0, Hologic).