

UCSF

UC San Francisco Previously Published Works

Title

Accuracy and interpretation time of computer-aided detection among novice and experienced breast MRI readers.

Permalink

<https://escholarship.org/uc/item/2x6274wj>

Journal

American Journal of Roentgenology, 200(6)

ISSN

0361-803X

Authors

Lehman, Constance D
Blume, Jeffrey D
DeMartini, Wendy B
[et al.](#)

Publication Date

2013-06-01

DOI

10.2214/ajr.11.8394

Peer reviewed



Published in final edited form as:

AJR Am J Roentgenol. 2013 June ; 200(6): W683–W689. doi:10.2214/AJR.11.8394.

Accuracy and Interpretation Time of Computer-Aided Detection Among Novice and Experienced Breast MRI Readers

Constance D. Lehman^{1,2}, Jeffrey D. Blume³, Wendy B. DeMartini^{1,2}, Nola M. Hylton⁴, Benjamin Herman⁵, and Mitchell D. Schnall⁶

¹Department of Radiology, Section of Breast Imaging, School of Medicine, University of Washington, Seattle, WA

²Breast Imaging, Seattle Cancer Care Alliance, 825 Eastlake Ave E, G2-600, Seattle WA 98109

³Biostatistics Collaboration Center, Vanderbilt University Medical Center North, Nashville, TN

⁴Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA

⁵Biostatistics Department, Brown University, Providence, RI

⁶Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA

Abstract

OBJECTIVE—The purpose of this study was to compare the diagnostic accuracy and interpretation times of breast MRI with and without use of a computer-aided detection (CAD) system by novice and experienced readers.

SUBJECTS AND METHODS—A reader study was undertaken with 20 radiologists, nine experienced and 11 novice. Each radiologist participated in two reading sessions spaced 6 months apart that consisted of 70 cases (27 benign, 43 malignant), read with and without CAD assistance. Sensitivity, specificity, negative predictive value, positive predictive value, and overall accuracy as measured by the area under the receiver operating characteristic curve (AUC) were reported for each radiologist. Accuracy comparisons across use of CAD and experience level were examined. Time to interpret and report on each case was recorded.

RESULTS—CAD improved sensitivity for both experienced (AUC, 0.91 vs 0.84; 95% CI on the difference, 0.04, 0.11) and novice readers (AUC, 0.83 vs 0.77; 95% CI on the difference, 0.01, 0.10). The increase in sensitivity was statistically higher for experienced readers ($p = 0.01$). Diagnostic accuracy, measured by AUC, for novices without CAD was 0.77, for novices with CAD was 0.79, for experienced readers without CAD was 0.80, and for experienced readers with CAD was 0.83. An upward trend was noticed, but the differences were not statistically significant. There were no significant differences in interpretation times.

Address correspondence to C. D. Lehman (lehman@u.washington.edu).

Presented at the 2009 annual meeting of the RSNA, Chicago, IL.

Supported by an Avon-NCI Progress for Patients (PPF) Award, supplement to FHCRC Cancer Center support grant P30 CA15704.

COCLUSIONS—MRI sensitivity improved with CAD for both experienced readers and novices with no overall increase in time to evaluate cases. However, overall accuracy was not significantly improved. As the use of breast MRI with CAD increases, more attention to the potential contributions of CAD to the diagnostic accuracy of MRI is needed.

Keywords

accuracy; breast MRI; computer-aided diagnosis; reader study

MRI is currently considered a key tool in detection and management of breast cancer. Several medical organizations that establish policy for breast cancer screening and management recommend the use of MRI for specific applications in defined patient populations [1–3]. Consideration of breast MRI is recommended to assess extent of disease in patients with a recent diagnosis of breast cancer [4–9] and to screen women who are at high risk of breast cancer, such as those who have a *BRCA1* or *BRCA2* mutation or those with a lifetime risk greater than 20% according to models that depend on family history [2].

The assessment of breast MRI examinations is based on evaluation of morphologic and kinetic features of enhancing lesions. These assessments depend on both high-spatial- and high-temporal-resolution imaging. A typical breast MRI examination may include as many as 1000 images as thin slices obtained through the breasts at multiple time points before and after contrast material is injected. To address the challenges of reviewing these large datasets, software programs have been developed to assist radiologists in image review. These tools focus primarily on providing visual cues to present kinetic information to the radiologist in a more accessible manner. Color overlays are provided that direct the radiologist to regions with more suspicious patterns of enhancement, such as regions with rapid uptake and delayed-phase washout of contrast medium. Although these computer-aided detection (CAD) systems are commercially available and widely used in breast MRI interpretation, little is known regarding the effect of use of CAD systems on diagnostic accuracy and interpretation time for breast MRI.

The purpose of this study was to compare the diagnostic accuracy of breast MRI interpretations by novice and experienced readers using and not using a CAD system. A secondary aim was to compare times for breast MRI interpretation by novice and experienced readers using and not using a CAD system.

Subjects and Methods

Study Design

This study was exempt from institutional review board review. It was a reader study involving 20 radiologists, each reading the same set of cases with and without CAD. Nine of the radiologists were determined to be experienced readers who had performed at least 100 breast MRI examinations before study participation, and 11 were novice readers. Each participant completed two reading sessions spaced 6 months apart to minimize recall bias. For each session, before interpreting images the participants participated in a short review of BI-RADS breast MRI terminology (e.g., finding types and kinetics) led by one of the study

investigators. For CAD-assisted readings, color overlays and CAD tools were available and used at the discretion of the radiologist. Non-CAD readings included visual assessment of the dynamic series.

Each session involved reading one half of the cases with CAD assistance and one half without. The selection and order of cases were randomized for each reader to further minimize the influence of temporal bias, recall bias, and learning effects (i.e., improvement in reading accuracy due to the extra practice of being in this study). The case series was composed of 70 breast MRI cases, 27 of which had benign outcome and 43 of which had malignant outcome. The reference standard was obtained in the International Breast MR Image Consortium (IBMC) study 6883, in which a combined biopsy-proven histologic and clinical follow-up protocol was used for obtaining final outcomes of the lesions.

The primary endpoint was the reader's evaluation of the presence of malignant lesions in the breast. For each case, the reader was asked to identify any suspicious lesions (detection) and to rate the likelihood of malignancy or benignancy (classification) of each case. Readers were instructed to provide an overall assessment of the case as a whole based on the most suspicious lesion identified. Results for each reader allowed calculations of sensitivity, specificity, and overall accuracy as measured by the area under the receiver operating characteristic (ROC) curve (AUC).

Primary Endpoint Capture

Three different scales were used to capture each reader's interpretation: modified BI-RADS (BI-RADS scale without category 0), probability of malignancy scale (1, definitely absent; 2, probably absent; 3, equivocal; 4, probably present; 5, definitely present), and percentage probability of malignancy scale (0–100%, i.e., each reader's subjective assessment of the probability that the breast had malignant foci). Three interpretation scales were used because each plays a different role in the assessment of reader performance and because they have differing clinical utility. Moreover, the domain of the scales affects the statistical performance of the fitting algorithms used to derive ROC curves. The plurality of these measures can count for practical issues like reader preference and familiarity and clarity and precision. Finally, consistent performance among the various rating scales has strong internal validity, excellent external generalizability, and robustness of results compared with statistical fitting algorithms, which have a tendency to result in overestimation of the AUC when the number of categories in an ordinal scale is small.

A sample size of 70 MRI examinations with approximately 40% abnormal and 60% normal cases was determined to provide excellent statistical precision and power for a two-sided 5%-level test to detect a difference in CAD performance of at least 10%. For this determination a correlation between the same reader interpretations was assumed to be 0.4, the correlation reader-specific differences in the areas was assumed to be 0.1, and the range of AUC estimates was as follows: novice not using CAD, 0.65; experienced not using CAD, 0.8; novice using CAD, 0.75; and experienced using CAD, 0.85 [10–14]. This was also projected to provide a CI margin of error (i.e., half width) of less than 0.06 for experienced readers and less than 0.075 for novice readers [15]. These assumptions were supported by observed trends in the National Cancer Institute–sponsored IBMC study 6883.

Computer-Aided Detection System

In this study, we used the CADstream system developed by Confirma. The CADstream software is designed to enhance data analysis of contrast-enhanced images. CADstream is a validated computer-aided method that addresses the numerous issues unique to MRI with contrast enhancement, including registration technology and temporal analysis. Segmentation of enhancing lesions based on characteristic intensity signatures is designed to improve tumor detection, classification, and measurement (Figs. 1 and 2). At the time this study was proposed, CADstream was the only device cleared by the U.S. Food and Drug Administration for computer-aided analysis of MR images for the identification of similar tissue types. Other CAD systems that generate similar kinetic assessment summaries have since become commercially available.

Images

Images used in this study were selected by the study investigators from the image data archive established at the American College of Radiology in conjunction with IBMC study 6883. This dataset was chosen because it includes images obtained with a variety of acquisition techniques at various institutions. This provides a representative real-world set of images on which to test the utility of the CAD system. Images in this data archive were provided by facilities that adhered to a rigorous image acquisition protocol [10].

The image acquisition protocol and patient and lesion characteristics are described in detail in a previous publication [16]. In brief, the protocol included a series of 3D T1-weighted gradient-echo sequences performed before and after IV administration of 0.1 mmol (0.2 mL) gadolinium chelate per kilogram of body weight. Slice thickness was 3 mm or less, and the acquisition time for each T1-weighted sequence was less than 4 minutes. T1-weighted dynamic series, T2-weighted unenhanced, subtraction, and maximum-intensity-projection images were provided for review. All images in the data archive were reviewed by members of the IBMC quality control committee to assure proper image acquisition. Each image was assigned a quality score based on this review. Only images that met current clinical standards for image quality were considered for use in our CAD reader study. Lesions in the IBMC dataset included the following histologic findings: 43.5% benign, 42.4% invasive carcinoma, 7.7% ductal carcinoma in situ, and 6.3% atypical ductal or lobular hyperplasia.

Readers

The readers came from diverse backgrounds, approximately one half practicing in academic settings and one half in private practice. Eight of the nine experienced readers had completed a fellowship (one half completed fellowships in MRI), and 5 of the 11 (45%) novices were fellowship trained (one with a fellowship in MRI). Most of the experienced readers had been out of residency at least 6 years, and slightly more than one half of the novice readers had been out of residency for 6 years or more. Forty percent of the readers were breast imagers or mammographers; 25% were MRI or abdominal imagers; 36% were general radiologists; and 27% were residents.

Data Analysis

Data were cleaned and analyzed with the SAS 9.0 (SAS Institute), Stata 10.0 (StataCorp), and R (R Project) software packages. Reader interpretation scales were dichotomized to compute standard operating metrics (e.g., sensitivity, positive predictive value [PPV], and negative predictive value [NPV]). We calculated 95% CIs for these metrics, AUCs, and comparisons between them (e.g., differences of sensitivities or of AUCs). The method of Leisenring et al. [17] was used to compare PPVs and NPVs. Reader interpretations were analyzed with a standard ROC method [18, 19]. Empirical AUCs were compared with fitted AUCs determined with Rokit software (version 1.1b2, Metz CE, University of Chicago) [20, 21]. Differences in observed AUCs were assessed with standard *U* statistics and standard regression approaches (both allow correlation between AUCs) [19, 22].

A random effects analysis of variance (ANOVA) model was used to compare average AUCs across CAD groups [22]. The interaction between reader expertise and CAD covariates was examined, and regression models were fit with and without reader expertise.

A separate regression analysis was used to model interpretation time. Time to complete an examination included both image review and reporting of results. The regression model was adjusted for reading session (first or second), reader demographics and expertise, case difficulty, and assistance of CAD. The effect of reading session was essentially a learning effect. Generalized estimating equations with robust standard errors were used to account for the correlation in interpretation times due to repeat reads of an image within and across readers [23].

Results

Results across interpretative scales were remarkably consistent. Table 1 summarizes the AUCs by scale and expertise. For the BI-RADS scale, the AUC for novices without CAD was 0.77 and with CAD was 0.79; the AUC for experienced readers with CAD was 0.80 and without CAD was 0.83. A general upward trend moving from novice to experienced readers and from without to with CAD was noticed, but the difference was not statistically significant even in a flexible model that included an interaction for this effect. The results were suggestive, but an even larger study is needed to detect this improvement if the magnitude of improvement is as observed here. We found no evidence to support an improvement due to CAD of more than the 0.1 AUC units assumed in our initial projections (which is admittedly a large difference, but also the smallest that could practically be assessed at the time).

Table 2 summarizes the reader-specific AUCs for the probability of malignancy scale, and Figures 3A and 3B show these curves. Figures 3A and 3B show that the reader-to-reader variability of the ROC of novice and experienced readers is substantial and is the major contributor to the statistical variance. Almost identical results were observed for the BI-RADS and percentage probability of malignancy scales (not shown). These findings indicate that attempts to expand the interpretive scale do little to reduce the statistical variability of readings despite increasing the number of categories used to interpret images.

Summary measures of accuracy are presented in Table 3 for the BI-RADS scale (dichotomized at category 3 or less vs 4 or greater). Use of CAD improved sensitivity for both experienced (AUC, 0.91 versus 0.84; 95% CI on the difference, 0.04, 0.11) and novice readers (AUC, 0.83 versus 0.77; 95% CI on the difference, 0.01, 0.10). There was a significant difference for interaction in the comparison of experienced readers using CAD to novices without CAD assistance (AUC, 0.91 versus 0.77; $p = 0.01$). This difference was also significant on the percentage probability of malignancy scale ($p = 0.047$) but not on the probability of malignancy scale. No differences were seen in specificity or PPV. A difference in NPV was seen within the experienced group when reading with and without CAD (NPV, 0.81 versus 0.70 for BI-RADS scale; 95% CI, 0.03, 0.20; $p < 0.05$ for both scales). These results were remarkably consistent across rating scales.

To compare the efficiency of breast MRI interpretation with and without use of a CAD system between novice and experienced readers, we measured the amount of time it took for readers to complete each examination within a session (Tables 4 and 5). No significant differences were found overall in time to assess cases with or without CAD. However, there was a small session effect, likely due to a learning effect. The generalized estimating equation model used to adjust for session effect yielded a significant interaction between reference standard and CAD reading after adjustment for session ($p = 0.021$). This result implied the images with abnormal findings read with CAD took slightly longer to assess (1.18 minutes; 95% CI, 0.175, 2.2 minutes).

Discussion

To our knowledge, this multicenter reader study of the effect of CAD on diagnostic accuracy of breast MRI is the first of its kind and the only one designed to minimize the influence of temporal bias, recall bias, and learning effects. We found that CAD does improve the sensitivity of breast MRI interpretation. Previous studies showed that CAD applied to breast MRI can improve the distinction between malignant and benign lesions [24–27], although one study did not find improved accuracy when CAD was used [28]. We found the sensitivity of MRI improved with CAD for both experienced and novice readers. We also found an interaction effect whereby the highest sensitivity was achieved when experienced readers were using CAD. A nonsignificant trend was also seen toward improved diagnostic accuracy among novices not using CAD to experienced readers using CAD.

Our findings are similar to those reported by Shimauchi et al. [29]. In that study, 50 benign and 50 malignant lesions were reviewed with a specialized CAD program that provided six radiologists with a malignancy probability score. The authors reported an increase in sensitivity when CAD was applied to conventional methods of MRI interpretation. They also reported improvement in overall diagnostic accuracy. In that study, readings with a specialized CAD program that provided a probability of malignancy were compared with readings that entailed conventional interpretations with commercially available software for kinetic image data processing and visualization. As CAD products are developed that incorporate malignancy probability scales and kinetic information, further study will be important to guide clinical use and practice.

This study highlights important issues regarding the design of reader studies for evaluating CAD performance of radiologists. The consistency of our results across rating scales is an important finding. First, it implies that in similar settings attempts to increase statistical power by developing more refined assessments are not likely to be successful. It also implies that accuracy estimates from interpretations in similar settings that are assessed with 5-point rating scales can be as precise as continuous scales. Because radiologists often find continuous scales difficult to use in practice and because in our experience they do not necessarily use them uniformly, there does not appear to be a strong case for continuing to use them in settings in which they will be burdensome. Specific to our breast imaging results, the 5-point BI-RADS scale is the one with which radiologists are most adept and is consistently used in clinical practice across all breast imaging examination assessments.

One noteworthy exception is that the interaction for comparing experienced readers using CAD to novices not using CAD was significant on the BI-RADS and percentage probability of malignancy scales but not on the probability of malignancy scale. Experienced readers using CAD had higher performance than novices not using CAD. This finding, however, appears to be due to a lack of statistical precision on the probability of malignancy scale rather than a differential finding.

It is also important that by documenting the reader-specific AUCs, we identified significant reader-to-reader variability in the effect of CAD use on interpretive accuracy. As the use of breast MRI with CAD increases, more attention to the potential contributions of CAD to the diagnostic accuracy of MRI is needed, and additional reader studies are indicated. Even larger reader studies will be needed to detect changes and improvement in reader accuracy. As imaging methods continue to rapidly evolve, carefully designed reader studies are needed to assess new techniques and their potential effect on diagnostic accuracy.

There were limitations to this study. The designation of experienced was subjective and based on the requirements of the American College of Radiology guidelines for radiologists to be accredited in interpretation of breast MR images. More stringent requirements for designating experienced readers might have yielded different results. Radiologists did not undergo specific training in the use of CAD applied to MRI before participation in the study. This training also might have influenced the results.

This study provides valuable information on reader variation for future research. CAD appears to improve the sensitivity of breast MRI interpretations. Education programs on CAD should attend to methods of improving both overall diagnostic accuracy and the efficiency of MRI interpretation.

Acknowledgments

We thank the 20 radiologists who participated in the reader study. Participating radiologists were recruited from Seattle; Philadelphia and the University of Pennsylvania Department of Radiology; and the San Francisco Bay Area and the University of California, San Francisco, Department of Radiology and Biomedical Imaging.

References

1. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin.* 2007; 57:75–89. [PubMed: 17392385]
2. Lehman CD, Smith RA. The role of MRI in breast cancer screening. *J Natl Compr Canc Netw.* 2009; 7:1109–1115. [PubMed: 19930977]
3. Lehman CD, DeMartini W, Anderson BO, Edge SB. Indications for breast MRI in the patient with newly diagnosed breast cancer. *J Natl Compr Canc Netw.* 2009; 7:193–201. [PubMed: 19200417]
4. Lehman CD, Gatsonis C, Kuhl CK, et al. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Engl J Med.* 2007; 356:1295–1303. [PubMed: 17392300]
5. Lehman CD, Blume JD, Weatherall P, et al. Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. *Cancer.* 2005; 103:1898–1905. [PubMed: 15800894]
6. Fischer U, Kopka L, Brinck U, Korabiowska M, Schauer A, Grabbe E. Prognostic value of contrast-enhanced MR mammography in patients with breast cancer. *Eur Radiol.* 1997; 7:1002–1005. [PubMed: 9265662]
7. Lee SG, Orel SG, Woo JJ, et al. MR imaging screening of the contralateral breast in patients with newly diagnosed breast cancer: preliminary results. *Radiology.* 2003; 226:773–778. [PubMed: 12601182]
8. Liberman L, Morris EA, Dershaw DD, Abramson AF, Tan LK. MR imaging of the ipsilateral breast in women with percutaneously proven breast cancer. *AJR.* 2003; 180:901–910. [PubMed: 12646427]
9. Liberman L, Morris EA, Kim CM, et al. MR imaging findings in the contralateral breast of women with recently diagnosed breast cancer. *AJR.* 2003; 180:333–341. [PubMed: 12540428]
10. Schnall MD, Blume J, Bluemke DA, et al. MRI detection of distinct incidental cancer in women with primary breast cancer studied in IBMC 6883. *J Surg Oncol.* 2005; 92:32–38. [PubMed: 16180227]
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. [PubMed: 9419702]
12. Obuchowski NA. Computing sample size for receiver operating characteristic studies. *Invest Radiol.* 1994; 29:238–243. [PubMed: 8169102]
13. Obuchowski NA. Sample size calculations in studies of test accuracy. *Stat Methods Med Res.* 1998; 7:371–392. [PubMed: 9871953]
14. Obuchowski NA, McClish DK. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Stat Med.* 1997; 16:1529–1542. [PubMed: 9249923]
15. Blume JD. Bounding sample size projections for the area under a ROC curve. *J Stat Plan Inference.* 2009; 139:711–721. [PubMed: 20160839]
16. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. *Radiology.* 2006; 238:42–53. [PubMed: 16373758]
17. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics.* 2000; 56:345–351. [PubMed: 10877288]
18. Pepe, MS. *The statistical evaluation of medical tests for classification and prediction.* New York, NY: Oxford University Press; 2003.
19. Zhou, XH.; Obuchowski, NA.; McClish, DK. *Statistical methods in diagnostic medicine.* New York, NY: Wiley; 2002.
20. Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. *Med Decis Making.* 1998; 18:110–121. [PubMed: 9456215]
21. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Stat Med.* 1998; 17:1033–1053. [PubMed: 9612889]

22. Obuchowski NA, Rockette HE. Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests: an ANOVA approach with dependent observations. *Commun Stat Simul Comput.* 1995; 24:285–308.
23. Diggle, PJ.; Liang, KY.; Zeger, SL. *Analysis of longitudinal data.* New York, NY: Oxford University Press; 1994.
24. Lehman CD, Peacock S, DeMartini WB, Chen X. A new automated software system to evaluate breast MR examinations: improved specificity without decreased sensitivity. *AJR.* 2006; 187:51–56. [PubMed: 16794155]
25. Williams TC, DeMartini WB, Partridge SC, Peacock S, Lehman CD. Breast MR imaging: computer-aided evaluation program for discriminating benign from malignant lesions. *Radiology.* 2007; 244:94–103. [PubMed: 17507720]
26. Wang LC, DeMartini WB, Partridge SC, Peacock S, Lehman CD. MRI-detected suspicious breast lesions: predictive values of kinetic features measured by computer-aided evaluation. *AJR.* 2009; 193:826–831. [PubMed: 19696298]
27. Meeuwis C, van de Ven SM, Stapper G, et al. Computer-aided detection (CAD) for breast MRI: evaluation of efficacy at 3.0 T. *Eur Radiol.* 2010; 20:522–528. [PubMed: 19727750]
28. Arazi-Kleinman T, Causer PA, Jong RA, Hill K, Warner E. Can breast MRI computer-aided detection (CAD) improve radiologist accuracy for lesions detected at MRI screening and recommended for biopsy in a high-risk population? *Clin Radiol.* 2009; 64:1166–1174. [PubMed: 19913125]
29. Shimauchi A, Giger ML, Bhooshan N, et al. Evaluation of clinical breast MR imaging performed with prototype computer-aided diagnosis breast MR imaging workstation: reader study. *Radiology.* 2011; 258:696–704. [PubMed: 21212365]

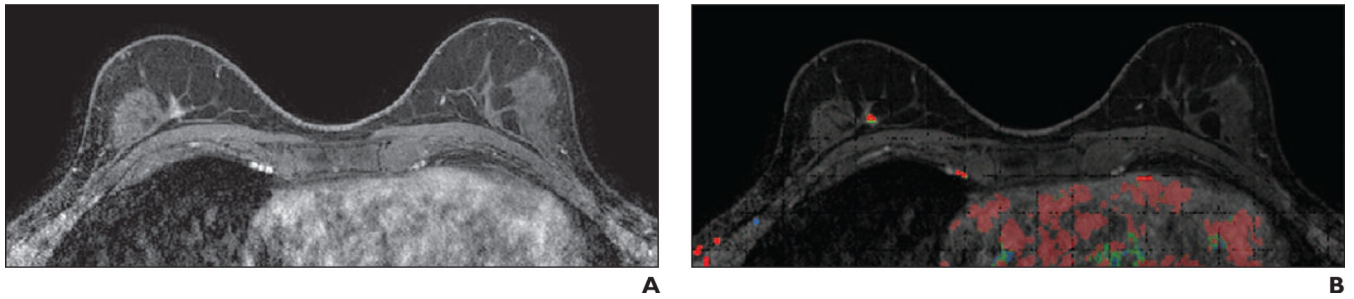


Fig. 1. 58-year-old woman with ductal carcinoma in situ

A and **B**, Initial axial contrast-enhanced MR images without (**A**) and with (**B**) computer-aided detection (CAD) color overlay show enhancing mass in right posterior medial breast that is more evident with color overlay. CAD detailed synopsis showed rapid initial phase (96% of lesion) uptake of contrast material with mixed delayed phase kinetics (40% washout in lesion) in this intermediate grade ductal carcinoma in situ.

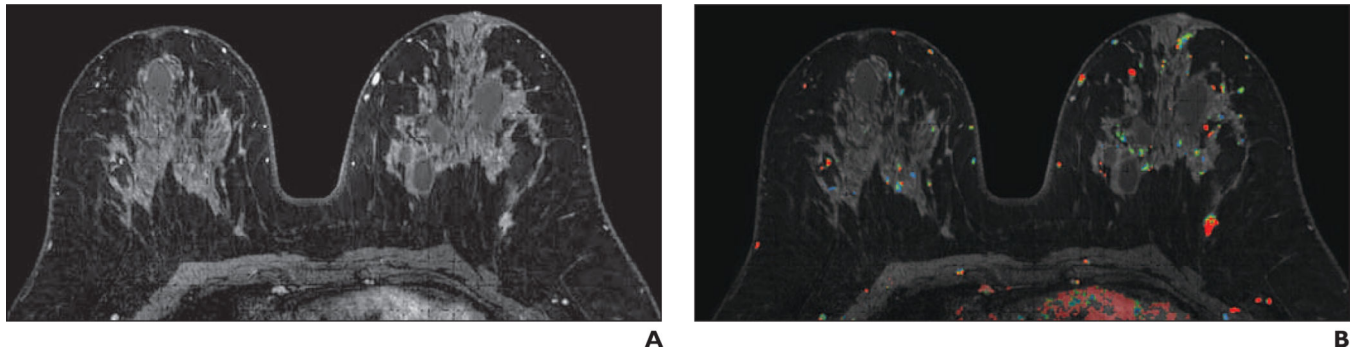


Fig. 2. 47-year-old woman with invasive ductal carcinoma

A and B, Initial axial contrast-enhanced MR images without (**A**) and with (**B**) computer-aided detection (CAD) color overlay show enhancing mass in left posterior lateral breast that is more evident with color overlay. CAD detailed synopsis showed rapid initial phase (56% of lesion) uptake of contrast material with mixed delayed phase kinetics (27% washout in lesion) in this invasive ductal carcinoma.

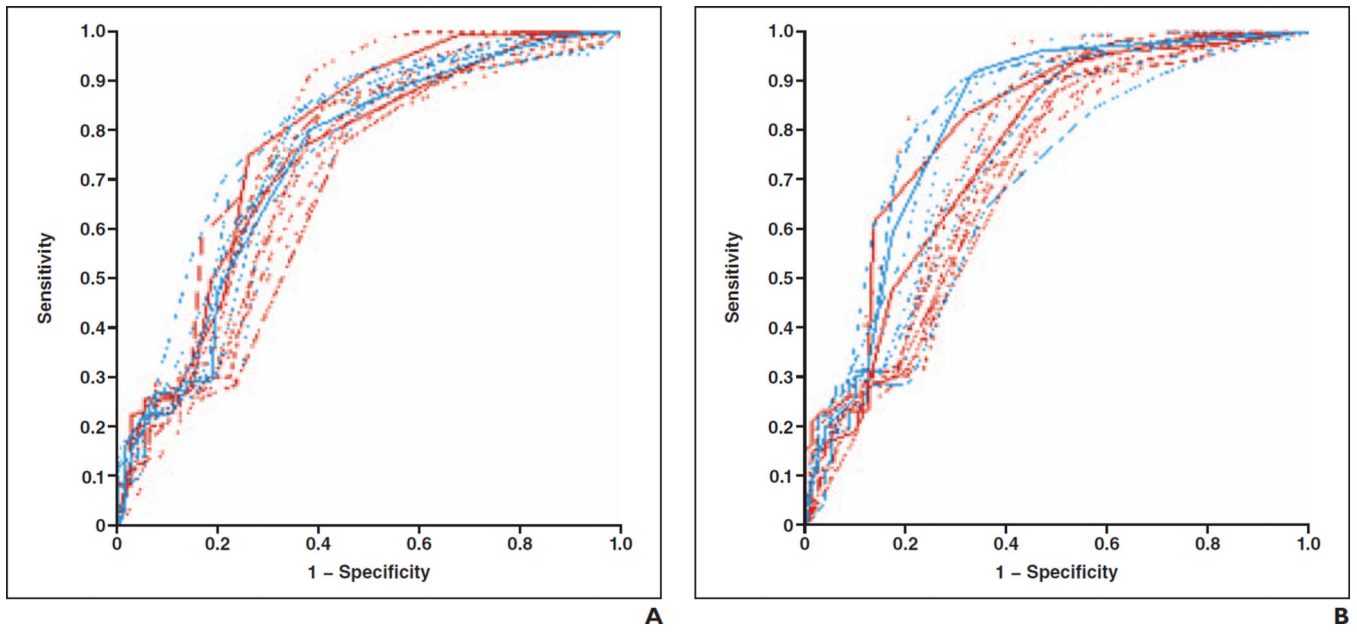


Fig. 3. Receiver operating characteristic curves
A, Without computer-aided detection. Blue indicates experienced reader; red, novice reader.
B, With computer-aided detection. Blue indicates experienced reader; red, novice reader.

TABLE I

Area Under the Curve (AUC) for Experienced and Novice Readers With and Without Computer-Aided Detection (CAD)

Experience Level	No. of Readers	Average AUC		<i>p</i> ^a
		Without CAD	With CAD	
BI-RADS				
All	20	0.7841	0.8091	0.0890
Experienced	9	0.7972	0.8266	0.1308
Novice	11	0.7734	0.7949	0.2665
Probability of malignancy scale				
All	20	0.7917	0.8191	0.0865
Experienced	9	0.8058	0.8383	0.0752
Novice	11	0.7801	0.8033	0.2390
Percentage probability of malignancy scale				
All	20	0.8036	0.8238	0.2529
Experienced	9	0.8231	0.8431	0.1941
Novice	11	0.7876	0.8080	0.3881

Note—AUCs compared by use of empirical AUCs and a random effects model [20–22].

^aH₀ = no difference.

TABLE 2
Area Under the Curve (AUC) for Each Reader With and Without Computer-Aided Detection (CAD)

Reader No.	No CAD (Modality 2)		CAD (Modality 1)		<i>p</i> ^a
	AUC	Standard Error	AUC	Standard Error	
Novice					
1	0.8039	0.0500	0.8586	0.0411	0.21
2	0.7967	0.0489	0.8771	0.0475	0.12
3	0.8274	0.0468	0.8405	0.0445	0.79
4	0.7471	0.0575	0.7761	0.0567	0.64
5	0.7963	0.0557	0.8577	0.0444	0.26
6	0.6763	0.0656	0.7765	0.0559	0.12
7	0.8396	0.0513	0.8119	0.0558	0.45
8	0.7984	0.0561	0.7757	0.0577	0.66
9	0.7862	0.0537	0.7281	0.0588	0.27
10	0.7458	0.0591	0.7866	0.0568	0.47
11	0.7639	0.0567	0.7479	0.0606	0.83
Experienced					
12	0.8077	0.0486	0.8493	0.0501	0.27
13	0.7921	0.0532	0.7416	0.0601	0.41
14	0.8245	0.0460	0.8809	0.0434	0.21
15	0.8068	0.0512	0.8434	0.0502	0.50
16	0.7942	0.0511	0.8636	0.0412	0.15
17	0.8253	0.0505	0.8346	0.0534	0.84
18	0.7997	0.0553	0.8733	0.0433	0.04
19	0.8136	0.0529	0.8632	0.0449	0.10
20	0.7879	0.0599	0.7946	0.0587	0.83
Average	0.7917		0.8191		0.09

Note—Probability of malignancy scale.

^aResults of overall test that average AUCs are equivalent with random-reader effect model yields *F** = 3.268 with 1 and 19 *df* and corresponding *p* = 0.0865. The 95% CIs for difference in mean accuracy between modalities are -0.0043 and 0.0591 for CAD versus no CAD.

TABLE 3

Performance in Reading Study Images for Experienced and Novice Readers With and Without Computer-Aided Detection (CAD) When Using BI-RADS Scale Dichotomized to Category 3 or Less Versus 4 or Greater

Performance Characteristic	No. of Readers	Mean AUC	Standard Error	Confidence Limits for Mean AUC	
				Lower	Upper
Sensitivity ^a					
Experienced ^a					
CAD	9	0.91	0.0149	0.88	0.95
No CAD	9	0.84	0.0247	0.78	0.90
Difference		0.07		0.04	0.11
Novice					
CAD	11	0.83	0.0293	0.77	0.90
No CAD	11	0.77	0.0284	0.71	0.84
Difference		0.06		0.01	0.10
Specificity					
Experienced					
CAD	9	0.62	0.0448	0.52	0.72
No CAD	9	0.61	0.0405	0.52	0.71
Difference		0.01		-0.06	0.07
Novice					
CAD	11	0.63	0.0340	0.56	0.71
No CAD	11	0.67	0.0388	0.58	0.76
Difference		-0.04		-0.10	0.02
Positive predictive value					
Experienced					
CAD	9	0.80	0.0189	0.76	0.83
No CAD	9	0.78	0.0201	0.74	0.82
Difference		0.018		-0.04	0.07
Novice					
CAD	11	0.79	0.0181	0.75	0.82

Performance Characteristic	No. of Readers	Mean AUC	Standard Error	Confidence Limits for Mean AUC	
				Lower	Upper
No CAD	11	0.79	0.0187	0.76	0.83
Difference		-0.01		-0.06	0.05
Negative predictive value Experienced	9	0.81	0.0287	0.76	0.87
CAD	9	0.70	0.0315	0.64	0.76
Difference		0.11		0.03	0.20
Novice	11	0.70	0.0280	0.64	0.75
CAD	11	0.64	0.0273	0.59	0.70
Difference		0.05		-0.02	0.13

^a Significant interaction experienced by CAD; $p = 0.013$.

TABLE 4

Mean Time (min) to Read One Examination

Reading Session	No. of Observations	Mean	SD	Minimum	Maximum
First					
CAD	293	5.83	5.58	0.2833	77.97
No CAD	289	4.66	4.14	0.0333	56.13
Second					
CAD	283	4.45	2.97	0.1167	21.42
No CAD	290	3.83	3.91	0.0667	59.30

Note—Reading sessions were at least 6 months apart in an effort to avoid strong recall bias.

TABLE 5

Mean Time (min) to Read One Examination

Reference Standard	First Session		Second Session	
	CAD	No CAD	CAD	No CAD
Negative	5.08 (4.45, 5.71)	4.92 (4.32, 5.52)	4.02 (3.41, 4.63)	3.86 (3.24, 4.48)
Positive	6.03 (5.54, 6.53)	4.69 (4.17, 5.20)	4.97 (4.46, 5.48)	3.62 (3.13, 4.12)

Note—Values in parentheses are confidence intervals. Model is based on a generalized estimating equation model for time with adjustment for reader correlation. The second session is likely to be more reflective of current practice because it accounts for learning effects.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript