Title

Can Radiologists Predict the Presence of Ductal Carcinoma In Situ and Invasive Breast Cancer?

Journal

American Journal of Roentgenology, 208(4)

ISSN

0361-803X

Authors

Aminololama-Shakeri, Shadi

Flowers, Chris I

McLaren, Christine E

et al.

Publication Date

2017-04-01

DOI

10.2214/ajr.16.16073

Peer reviewed

HHS Public Access

Author manuscript

AJR Am J Roentgenol. Author manuscript; available in PMC 2021 January 02.

Published in final edited form as:

AJR Am J Roentgenol. 2017 April ; 208(4): 933–939. doi:10.2214/AJR.16.16073.

Can Radiologists Predict the Presence of Ductal Carcinoma In Situ and Invasive Breast Cancer?

Shadi Aminololama-Shakeri[1], **Chris I. Flowers**[2,3], **Christine E. McLaren**[4], **Dorota J. Wisner**[5,6], **Jade de Guzman**[7], **Joan E. Campbell**[8], **Lawrence W. Bassett**[9], **Haydee Ojeda-Fournier**[7], **Karen Gerlach**[8,10], **Jonathan Hargreaves**[1], **Sarah L. Elson**[5,11], **Hanna Retallack**[5], **Bonnie N. Joe**[5], **Stephen A. Feig**[8], **Colin J. Wells**[9], **ATHENA Breast Health Initiative**

[1]Department of Radiology, University of California Davis, 4860 Y St, Ste 3100, Sacramento, CA 95817.

[2]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA.

[3]Present address: Moffitt Cancer Center, Tampa, FL.

[4]Department of Epidemiology, University of California, Irvine, CA.

[5]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA.

[6]Present address: Kaiser Permanente, San Rafael, CA.

[7]Department of Radiology, University of California, San Diego, CA.

[8]Department of Radiological Sciences, University of California, Irvine, CA.

[9]Breast Imaging Section, Department of Radiology, David Geffen School of Medicine, University of California, Los Angeles, CA.

[10]Present address: The University of Texas MD Anderson Cancer Center, Houston, TX.

[11]Present address: 23andMe, Inc., Mountain View, CA.

Abstract

OBJECTIVE.—We hypothesize that radiologists' estimated percentage likelihood assessments for the presence of ductal carcinoma in situ (DCIS) and invasive cancer may predict histologic outcomes.

MATERIALS AND METHODS.—Two hundred fifty cases categorized as BI-RADS category 4 or 5 at four University of California Medical Centers were retrospectively reviewed by 10 academic radiologists with a range of 1–39 years in practice. Readers assigned BI-RADS category (1, 2, 3, 4a, 4b, 4c, or 5), estimated percentage likelihood of DCIS or invasive cancer (0–100%), and confidence rating (1 = low, 5 = high) after reviewing screening and diagnostic mammograms and ultrasound images. ROC curves were generated.

Address correspondence to S. Aminololama-Shakeri (sshakeri@ucdavis.edu).

Based on a presentation at the Cancer Therapy and Research Center–American Association for Cancer Research 2013 San Antonio Breast Cancer Symposium, San Antonio, TX.

RESULTS.—Sixty-two percent (156/250) of lesions were benign and 38% (94/250) were malignant. There were 26 (10%) DCIS, 20 (8%) invasive cancers, and 48 (19%) cases of DCIS and invasive cancer. AUC values were 0.830–0.907 for invasive cancer and 0.731–0.837 for DCIS alone. Sensitivity of 82% (56/68), specificity of 84% (153/182), positive predictive value (PPV) of 66% (56/85), negative predictive value (NPV) of 93% (153/165), and accuracy of 84% ([56 + 153]/250) were calculated using an estimated percentage likelihood of 20% or higher as the prediction threshold for invasive cancer for the radiologist with the highest AUC (0.907; 95% CI, 0.864–0.951). Every 20% increase in the estimated percentage likelihood of invasive cancer increased the odds of invasive cancer by approximately two times (odds ratio, 2.4). For DCIS, using a threshold of 40% or higher, sensitivity of 81% (21/26), specificity of 79% (178/224), PPV of 31% (21/67), NPV of 97% (178/183), and accuracy of 80% ([21 + 178]/250) were calculated. Similarly, these values were calculated at thresholds of 2% or higher (BI-RADS category 4) and 95% or higher (BI-RADS category 5) to predict the presence of malignancy.

CONCLUSION.—Using likelihood estimates, radiologists may predict the presence of invasive cancer with fairly high accuracy. Radiologist-assigned estimated percentage likelihood can predict the presence of DCIS, albeit with lower accuracy than that for invasive cancer.

Keywords

BI-RADS; breast cancer; digital mammography; ductal carcinoma in situ; invasive breast cancer; kappa coefficients; ROC curves

The American College of Radiology's BI-RADS, which was developed for standardization of breast imaging reporting, is widely used across the world and has now reached its 5th edition [1]. A final BI-RADS assessment is given for each patient at the conclusion of a screening or diagnostic study. BI-RADS categories 1, 2, and 3 represent the smallest likelihood of malignancy, between 0% and 2%. BI-RADS categories 4 and 5 represent the remaining greater than 2% to 100% likelihood of malignancy. BI-RADS category 4 (suspicious for malignancy) confers a greater than 2% to 95% likelihood of malignancy and encompasses a heterogeneous group of findings, all of which are referred for further evaluation by tissue sampling. The 4th (2003) edition of the BI-RADS manual [2] provided optional stratification of BI-RADS category 4 lesions into low, intermediate, and moderate suspicion for malignancy (4a, 4b, and 4c) in an effort to provide more guidance regarding likelihood of malignancy in the radiology report. The lack of percentage ranges for the BI-RADS categories 4a, 4b, and 4c left room for variation in interpretation. The 5th (2013) edition of the BI-RADS manual [1] attempts to standardize the subcategories to make them more clinically meaningful by setting likelihood of malignancy ranges of 2–10% for category 4a, greater than 10% to 50% for category 4b, and greater than 50% to 95% for category 4c.

Current BI-RADS assessment categories estimate the overall likelihood for malignancy without attempting to distinguish between the likelihood of ductal carcinoma in situ (DCIS) and that of invasive cancer. Separate likelihood estimates of DCIS versus invasive cancer could inform radiologic-pathologic correlation, mediating underestimation rates of invasive cancer when DCIS only is identified at core biopsy. Currently, such upgrade rates of DCIS to invasive cancer at the time of surgery are 16–35% using a 14-gauge automated needle

versus 0–19% using a vacuum-assisted biopsy device [3]. Also, treatment options for DCIS are evolving. A trial of neoadjuvant treatment for postmenopausal women with DCIS (ClinicalTrials.gov identifier NCT01439711) is currently evaluating the feasibility of hormone treatment for DCIS only, which may eventually lead to nonsurgical options for DCIS. For these populations, it would be important to accurately exclude the possibility of invasive disease; hence, the separation of likelihood of malignancy estimates into DCIS versus invasive cancer would be particularly impactful.

In this study, we address the feasibility of separately assessing DCIS and invasive cancer likelihood on the basis of academic breast imagers' interpretation of imaging features. Prior studies have found subcategorization of BI-RADS category 4 to be useful in predicting the likelihood of malignancy [4] and found that radiologists could successfully stratify microcalcifications by malignant potential using BI-RADS assessment categories [5]. A pilot study of a single breast imaging radiologist showed the feasibility of assigning percentage likelihood estimates of DCIS and invasive cancer [6].

The purpose of this study was to determine whether radiologist-assigned likelihood for DCIS and invasive cancer may be predictive of histologic outcomes for the presence of DCIS and invasive cancer.

Materials and Methods

Reader Study Database

One hundred fifty consecutive patients with a final BI-RADS category 4 or 5 lesion were retrospectively identified at each of four University of California Medical Centers, totaling 600 patient studies. From the initial set of 600 cases, 250 (62–63 per center) were randomly chosen using a random number generator for inclusion in the study. Cases with multiple mammographic findings were excluded. Included studies were given new sequential identification numbers.

Images and data for this study were obtained under a HIPAA-compliant protocol. Institutional review board approval was obtained from the University of California Los Angeles with an intent-to-rely agreement by the other involved medical centers. All subjects had final outcomes determined by core biopsy or surgical excision. The BI-RADS category 4 or 5 findings included both palpable and nonpalpable lesions, as well as those with suspicious features on either diagnostic mammogram or targeted ultrasound or both. Clinical and histopathologic outcomes for all cases were provided in an online database.

Image Acquisition

Each site submitted up to 10 DICOM 3.0 images per patient that included the screening examination and diagnostic workup for which BI-RADS category 4 or 5 was assigned. Up to four diagnostic mammogram images and no more than two orthogonal ultrasound images were chosen per case. No prior examinations were provided. Protected health information was removed using a DICOM anonymization tool. All cases were individually reviewed to ensure appropriate deidentification and the presence of the complete set of images. A designated site radiologist was responsible for submitting the most representative images

available for making the BI-RADS category 4 or 5 assessments. A central radiologist performed additional review for quality control. If these criteria were not met, the cases were excluded. Each site provided available final histopathologic diagnoses for every case.

Radiologists

Ten academic radiologists subspecialized in breast imaging from five University of California Medical Centers reviewed the 250 cases. Years of practice after training ranged from 1 to 39 (mean, 14.8 years), and all but one radiologist (who had been in practice for over 30 years) had completed a breast imaging fellowship. In their daily practice, seven of the 10 radiologists exclusively read breast studies; three dedicated at least 60% of their clinical time to reading breast studies. The numbers of screening and diagnostic studies read over the past year were 1450–5849 (mean, 3282 studies) and 121–8123 (mean, 1837 studies), respectively; the number of procedures performed was 38–433 (mean, 254 procedures; two radiologists reported that their daily practice volumes were outside of ranges of the other eight readers). When a participating radiologist had also contributed to identification and acquisition of cases for the study, the washout period was at least 3 months between case review and evaluation.

Image Interpretation

All images pertaining to each individual case were made available simultaneously to the readers, who were blinded to clinical data and final outcomes. The images were reviewed on U.S. Food and Drug Administration–approved workstations (SecureView, Hologic) at a dedicated non–University of California site over a period of 1–2 days. The readers were instructed to evaluate cases as they would in routine clinical practice. For each case, readers assigned an overall BI-RADS category. Cases assigned BI-RADS categories of 1–3 required no further data evaluation. For cases given BI-RADS categories of 4a–4c or 5, readers assigned independent likelihood estimates (0–100%) for DCIS and invasive cancer and provided their level of confidence in each score (1–5, with 5 being the highest). Readers also reported any mammogram findings (calcifications, mass, focal asymmetry, asymmetry, or architectural distortion) and presence of a mass on ultrasound. When calcifications were reported, readers indicated the most suspicious morphologic features (punctate, amorphous, coarse heterogeneous, fine pleomorphic, or fine linear branching) and distribution (diffuse, regional, clustered, linear, or segmental). All data were entered into an online form; quality control was performed to confirm that there was a unique complete record for each case from each reader.

Statistical Analysis

Histologic outcomes——The frequency and percentage of cases by histologic classification recorded on the pathology index were calculated and reported. We analyzed the data for the 10 radiologists. First, for each radiologist, we used logistic regression modeling to predict the outcome of the presence of a DCIS lesion, as determined from histopathologic analysis, from two predictors: the radiologist’s estimated percentage risk of DCIS and an indicator variable representing modality, mammogram plus ultrasound versus mammogram only, with mammogram only considered the reference type of case. By use of

the same strategy, for each radiologist we used logistic regression modeling to predict the outcome of presence of an invasive cancer lesion as determined from histopathologic analysis, from two predictors: the radiologist's estimated percentage risk of invasive cancer and an indicator variable representing the two types of modalities, mammogram plus ultrasound versus mammogram only, with mammogram only considered the reference modality.

ROC curves——Logistic regression was used to model and estimate the odds of the presence of invasive cancer as recorded on the pathology index, for every 20% increase in the estimated percentage likelihood of invasive cancer, which was determined by a reader. ROC curves were determined from the sensitivity and specificity of thresholds for each reader. The AUC value and its 95% CI was calculated. A graph of the 10 ROC curves was constructed, and the threshold that was farthest from the diagonal line in the ROC graph was identified and used to calculate the sensitivity, specificity, PPV, NPV, and accuracy. In addition, prespecified thresholds for estimated percentage likelihood of invasive cancer of 2% and 95% were also used to report the distributions of the sensitivity, specificity, PPV, and NPV for predicting the presence of invasive cancer. The same analytic strategy was followed to model the outcomes of the presence of DCIS and invasive cancer and the presence of DCIS as the sole histopathologic finding recorded on the pathology index, from the estimated percentage likelihood of DCIS as a predictor.

In addition, a logistic regression model was formed using the data from an individual reader, with the presence of malignancy as the outcome and estimated percentage likelihood of invasive cancer as the predictor. This procedure was repeated using data from each reader. The mean and SD of values for sensitivity, specificity, PPV, and NPV were calculated at prespecified thresholds for estimated percentage likelihood of invasive cancer of 2% and 95%. The same analytic strategy was followed to model the presence of malignancy from the estimated percentage likelihood of DCIS for each reader. The mean and SD of values for sensitivity, specificity, PPV, and NPV were calculated at prespecified thresholds for the estimated percentage likelihood of DCIS of 2% and 95%.

Interobserver variability——The overall and per BI-RADS category kappa statistics were calculated for examining the agreement in given BI-RADS categories among 10 readers. On the basis of a normal distribution approximation, the 95% confidence limits of the kappa statistics were also obtained. The SAS macro, %MAGREE (version 1.3, SAS Institute), with its associated method was used [7, 8]. Kappa coefficients were interpreted using the method of Landis and Koch [9].

Results

Ten academic radiologists retrospectively reviewed image datasets from 250 female patients with BI-RADS category 4 or 5 lesions as determined by routine clinical diagnostic workup from four University of California Medical Centers. Of these 250 cases, 94 (38%) were malignant and 156 (62%) were benign. The malignant lesions consisted of 26 pure DCIS, 20 invasive cancers, and 48 invasive cancers with DCIS. Final histopathologic diagnoses of the

study lesions are outlined in Table 1. Findings were classified into mass, calcifications, asymmetry, architectural distortion, and other categories by each reader.

Likelihood estimates for the presence of invasive cancer and DCIS assigned by each reader were used to generate ROC curves for predicting the presence of invasive cancer (Fig. 1) and DCIS (Fig. 2). The distribution of cases used in ROC curve generation is shown in Table 2. The AUCs for the 10 radiologists ranged from 0.830 (95% CI, 0.767–0.893) to 0.907 (95% CI, 0.864–0.951) for the prediction of the presence of an invasive cancer. The AUC values for predicting the presence of DCIS as a concomitant finding with invasive cancer (Fig. 2A) ranged from 0.588 (95% CI, 0.504–0.671) to 0.778 (95% CI, 0.714–0.841). The AUC for prediction of the presence of DCIS as the sole histopathologic finding (Fig. 2B) was lower than that of invasive cancer but higher than for DCIS as a concomitant finding with invasive cancer, ranging from 0.731 (95% CI, 0.641–0.820) to 0.837 (95% CI, 0.748–0.926) for the 10 readers. Years of experience and radiologists' practice characteristics are shown with their associated AUC estimates for DCIS and invasive cancer in Table 3. Regression analysis confirmed a lack of association between radiologists' years of experience and estimated AUC for invasive cancer ($p = 0.29$, F statistic) and for DCIS only ($p = 0.26$, F statistic).

For every increase of 20% in the estimated percentage likelihood of invasive cancer, the odds of the presence of invasive cancer increased by 2.2–3.2 times across all readers, and similarly, for every increase of 20% in the estimated percentage likelihood of DCIS alone, the odds of the presence of DCIS increased by 1.4–2.5 times across all readers. A threshold of 20% likelihood was noted to be the farthest from the diagonal line in the ROC graph for the reader (reader 8) with the highest AUC (0.907; 95% CI, 0.864–0.951) for predicting the presence of invasive disease. Using this level of likelihood threshold, a sensitivity of 82% (56/68), specificity of 84% (153/182), PPV of 66% (56/85), NPV of 93% (153/165), and accuracy of 84% ([56 + 153]/250) were calculated. For the reader (reader 7) with the lowest AUC (0.830) in predicting invasive cancer, using a likelihood threshold of 30%, a sensitivity of 75% (51/68), specificity of 79% (144/182), PPV of 57% (51/89), NPV of 89% (144/161), and accuracy of 78% ([51 + 144]/250) were calculated. In predicting the presence of DCIS alone, reader 7 had the highest AUC (0.837; 95% CI, 0.748–0.926). For this reader, the threshold farthest from the diagonal line in the ROC graph for predicting the likelihood of DCIS was at the 40% likelihood level, producing a sensitivity of 81% (21/26), specificity of 79% (178/224), PPV of 31% (21/67), NPV of 97% (178/183), and accuracy of 80% ([21 + 178]/250). For the reader (reader 1) with the lowest AUC (0.731) in predicting DCIS alone, a likelihood threshold of 5% yielded a sensitivity of 58% (15/26), specificity of 72% (161/224), PPV of 19% (15/78), NPV of 94% (161/172), and accuracy of 70% ([15 + 161]/250).

In addition, we formed a logistic regression model using the data from an individual reader, with the presence of malignancy as the outcome and the estimated percentage likelihood of invasive cancer as the predictor. This procedure was repeated using data from each reader. The mean and SD of values for sensitivity, specificity, PPV, and NPV were calculated. At a threshold for estimated percentage likelihood of invasive cancer of 2% or greater (BI-RADS category 4), the mean values were as follows: sensitivity, 87% (SD, 9%); specificity, 41% (SD, 25%); PPV, 50% (SD, 9%); and NPV, 85% (SD, 7%). At an estimated likelihood of

invasive cancer of 95% or greater (BI-RADS category 5), the mean values were as follows: sensitivity, 33% (SD, 12%); specificity, 98% (SD, 2%); PPV, 92% (SD, 5%), and NPV, 70% (SD, 4%).

Similarly, we formed a logistic regression model using the data from each individual reader with the presence of malignancy as the outcome and the estimated percentage likelihood of DCIS as the predictor. At a threshold for estimated percentage likelihood of DCIS of 2% or greater (BI-RADS category 4), the mean values were as follows: sensitivity, 75% (SD, 18%); specificity, 33% (SD, 17%); PPV, 41% (SD, 3%); and NPV, 72% (SD, 11%). At an estimated likelihood of DCIS of 95% or greater (BI-RADS category 5), the mean values were as follows: sensitivity, 14% (SD, 6%); specificity, 99% (SD, 1%); PPV, 90% (SD, 8%); and NPV, 65% (SD, 2%).

Table 4 shows the number of lesions in the DCIS or invasive cancer categories detected by mammogram plus ultrasound versus detection by mammogram alone. The results from nine of 10 models indicated that there was a statistically significant relationship between the outcome, presence of a DCIS lesion, and the modality (all $p < 0.05$, Wald test), adjusted for the radiologist's estimated percentage risk of DCIS. The result based on the information for radiologist 7 showed no statistically significant relationship between the presence of a DCIS lesion and the type of case ($p = 0.07$, Wald test). The estimated odds ratio of the presence of a DCIS lesion ranged from 0.25 to 0.38, indicating that detection of DCIS lesions was less likely for mammogram-plus-ultrasound than for mammogram-only cases. The results from four of 10 models (radiologists 1, 4, 7, and 10) indicated that there was a statistically significant relationship between the outcome, presence of an invasive cancer, and the modality (all $p < 0.04$, Wald test) adjusted for the radiologist's estimated percentage risk of invasive cancer. For these four radiologists, the estimated odds ratios of the presence of invasive cancer ranged from 2.6 to 2.8, indicating that detection of invasive cancer was more likely for mammogram-plus-ultrasound cases than for mammogram-only cases. The results based on the information from the remainder of the radiologists showed no statistically significant relationship between the presence of invasive cancer alone and the modality ($p = 0.06$ to $p = 0.92$, Wald test).

Interobserver variability among the 10 readers was measured using kappa statistics. Agreement among the readers was scored as low for kappa values below 0.00, slight for kappa of 0.00–0.20, fair for kappa of 0.21–0.40, moderate for kappa of 0.41–0.60, substantial for kappa of 0.61–0.80, and almost perfect agreement for kappa of 0.81–1.00.

Kappa statistics according to designated BI-RADS categorization are shown in Table 5. There was overall fair agreement ($\kappa = 0.21$; 95% CI, 0.20–0.22) in BI-RADS category assignment among the 10 readers (Table 5). There was greater agreement for BI-RADS category 5 determination ($\kappa = 0.48$; 95% CI, 0.47–0.50; $p < 0.001$) and only slight agreement for BI-RADS category 4 determination ($\kappa = 0.20$; 95% CI, 0.18–0.22; $p < 0.001$). When subcategorizing BI-RADS category 4 lesions, the readers were in slight agreement for category 4a ($\kappa = 0.13$; 95% CI, 0.12–0.15; $p < 0.001$) and category 4b ($\kappa = 0.14$; 95% CI, 0.13–0.16; $p < 0.001$) but had increased agreement for category 4c ($\kappa = 0.26$; 95% CI, 0.24–0.28; $p < 0.001$) lesions. In assigning lesions to categories 4a and 4b, the radiologists

showed fair agreement ($\kappa = 0.31$; 95% CI, 0.29–0.33; $p < 0.001$). As shown in Table 5, readers had slight agreement ($\kappa = 0.11$; 95% CI, 0.1–0.13; $p < 0.001$) in categorizing BI-RADS categories 1 and 2 lesions. The least agreement was noted for BI-RADS category 3 ($\kappa = 0.07$; 95% CI, 0.05–0.09; $p < 0.001$). When only the subgroup of 94 malignant lesions was considered, the overall agreement among radiologists remained fair ($\kappa = 0.24$; 95% CI, 0.22–0.26; $p < 0.001$), increasing to moderate agreement when only categories 4c and 5 were combined ($\kappa = 0.48$; 95% CI, 0.45–0.51; $p < 0.001$).

Discussion

A recent pilot study showed the feasibility for a dedicated breast imager to accurately provide likelihood estimates for DCIS and invasive cancer [6]. In our study using 250 image datasets of BI-RADS categories 4 and 5 lesions, we also found that radiologist readers subspecializing in breast imaging were able to accurately assign likelihood estimates for predicting invasive cancer. The accuracy of readers in estimating the presence of DCIS when present concomitantly with invasive cancer was lower. This may partly be a result inherent in the daily practice of breast imagers. The presence of findings characteristic of invasive cancer supersede concomitant presence of microcalcifications or findings of DCIS in the same lesion, thereby possibly biasing the likelihood estimate toward the invasive component and minimizing the DCIS likelihood estimate. This would have affected the distribution of cases for ROC analysis in which, if the reader did not record DCIS, it was assumed that no DCIS was present. The accuracy of readers in estimating the presence of DCIS was lower compared with invasive cancer when DCIS was the only histopathologic abnormality present. The number of these cases, however, was small. Additionally, these data may reflect the heterogeneity of DCIS as a disease entity, with low-grade DCIS acting as an indolent process and high-grade DCIS frequently behaving similarly to invasive disease. Our data may also be a result of the diversity of mammographic findings attributable to DCIS. Although most cases of DCIS are represented as microcalcifications on mammography, approximately 10% are due to masses. The spectrum of microcalcification features that could represent DCIS as reflected in the BI-RADS lexicon covers a rather large associated likelihood of malignancy, ranging from 13% for coarse heterogeneous to nearly 70% for fine linear or fine-linear branching microcalcifications [1]. Moreover, microcalcifications are the only indicator of invasive disease 20% of the time [10].

Radiologists’ years of experience did not correlate with the estimated AUC for invasive cancer and for DCIS in our study. The point estimates of AUC for both invasive cancer and DCIS only were similar for the 10 radiologists, and the 95% CIs all overlapped (Table 3). Regression analysis confirmed this finding, which may be a reflection of the benefit of using the BI-RADS descriptors to guide assessment of the likelihood of malignancy regardless of level of experience.

Our study has several limitations. One limitation is a priori reader knowledge of study design, including the use of cases referred for biopsy. This may have biased the assignment of likelihood estimation by knowing that another interpreting radiologist had previously deemed the lesion suspicious and referred the case for biopsy. Our study is also limited by the lack of prior examinations. In the daily practice of breast imaging, prior examinations are

an important contributing factor in assigning BI-RADS categories. Another limitation is the small number of DCIS-only cases. As mentioned, the likelihood of assignment to DCIS may have been minimized when present with invasive cancer in the same lesion. This is an area that may be further studied in future reader studies with inclusion of larger numbers of pure DCIS lesions from our larger cohort of nearly 750 cases collected from the five University of California Medical Centers.

Others have previously reported interobserver variability in mammographic interpretation [11, 12]. Interobserver agreement was lowest for BI-RADS categories 1, 2, and 3. This may be due, in part, to differences in interpretation of prestudy instructions. Readers who selected the BI-RADS category 4 and 5 cases from each of the University of California Medical Centers may not have considered assigning the other BI-RADS category designations as frequently as others on the basis of this prior knowledge about the study design. The readers in our study had the most agreement in BI-RADS category 5 lesions followed by BI-RADS category 4c lesions. The highest agreement was recorded when the most suspicious categories for malignancy, BI-RADS categories 4c and 5, were considered together. There was only slight agreement on categories 4a and 4b findings. Lack of comparison examinations to evaluate whether lesions were new, stable, or increased may have contributed to reader variability. Moreover, this study was performed before the publication of the BI-RADS 5th edition [1]. Readers did not have access to this edition's expanded use of specific examples of lesion types appropriate for categories 4a, 4b, and 4c for assignment on the basis of appearance. The cases were a combination of findings including masses as well as microcalcifications. Follow-up reader studies may include larger cohorts of similar category findings such as microcalcifications.

Although the 10 readers are dedicated breast imagers, there were differences among them in terms of years of experience. Training, including review of microcalcification characteristics, before future reader studies may be undertaken to evaluate its utility in increasing interobserver agreement. Nakayama and colleagues [13] have suggested that presentation of similar images may improve radiologists' performance of microcalcification interpretation. In future studies, appropriate descriptors could be assigned to each finding: mass and calcification descriptors with comparison with prior imaging studies. Once consensus is reached among readers for appropriate descriptors for each case, a multivariate analysis may be performed to measure how well malignancy may be predicted. Such a study would also yield a training set of images and pathology results, which could reduce interpretive variability among other radiologists.

We did not adjust for patient age in our analyses because this information was removed during the anonymization process. Prior work using logistic regression models to predict the probability of invasive cancer versus DCIS on the basis of clinical and mammographic features found a dependence on patient age such that the models performed best in the older (≥ 65 years) age group [14]. This work shows the potential for radiologists to predict the likelihood of invasive disease at mammography. Additional studies are required to look at the predictive value of specific imaging features and patient characteristics that may indicate the presence of invasive disease.

In this study, we have shown that radiologists are able to predict the presence of invasive cancer with fairly high accuracy in assigning likelihood assessments. Similarly, radiologists are able to predict the presence of DCIS, albeit with lower accuracy than for invasive cancer.

Acknowledgments

We thank Wen-Pin Chen (Biostatistical Shared Resource, Chao Family Comprehensive Cancer Center, University of California Irvine) for assistance with statistical analyses and graphics, Hologic for providing the workstations and technical help compiling the cases, and Hannah Lui Park (Department of Epidemiology, University of California, Irvine) and Carlie K. Thompson (University of California, San Francisco) for assistance with manuscript preparation.

Supported by the University of California Office of the President, the Safeway Foundation, and award number P30CA062203 from the National Cancer Institute, National Institutes of Health.

References

1. D'Orsi C, Sickles EA, Mendelson EB, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology, 2013
2. D'Orsi CJ, Mendelson EB, Ikeda DM, et al. Breast Imaging Reporting and Data System: ACR BI-RADS—breast imaging atlas. Reston, VA: American College of Radiology, 2003
3. Liberman L Percutaneous image-guided core breast biopsy. Radiol Clin North Am 2002; 40:483–500 [PubMed: 12117188]
4. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. Radiology 2006; 239:385–391 [PubMed: 16569780]
5. Bent CK, Bassett LW, D'Orsi CJ, Sayre JW. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. AJR 2010; 194:1378–1383 [PubMed: 20410428]
6. Flowers CI, O'Donoghue C, Moore D, et al. Reducing false-positive biopsies: a pilot study to reduce benign biopsy rates for BI-RADS 4A/B assessments through testing risk stratification and new thresholds for intervention. Breast Cancer Res Treat 2013; 139:769–777 [PubMed: 23764994]
7. Fleiss JL, Levin B, Paik MC, Statistical methods for rates and proportions, 3rd ed. New York, NY: Wiley, 2003
8. Kendall MG. Rank correlation methods. London, UK: Charles Griffin & Co. Ltd, 1955
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159–174 [PubMed: 843571]
10. Venkatesan A, Chu P, Kerlikowske K, Sickles EA, Smith-Bindman R. Positive predictive value of specific mammographic findings according to reader and patient variables. Radiology 2009; 250:648–657 [PubMed: 19164116]
11. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994; 331:1493–1499 [PubMed: 7969300]
12. Skaane P, Diekmann F, Balleyquier C, et al. Observer variability in screen-film mammography versus full-field digital mammography with softcopy reading. Eur Radiol 2008; 18:1134–1143 [PubMed: 18301902]
13. Nakayama R, Abe H, Shiraishi J, Doi K. Potential usefulness of similar images in the differential diagnosis of clustered microcalcifications on mammograms. Radiology 2009; 253:625–631 [PubMed: 19789245]
14. Ayvaci MU, Alagoz O, Chhatwal J, et al. Predicting invasive breast cancer versus DCIS in different age groups. BMC Cancer 2014; 14:584 [PubMed: 25112586]

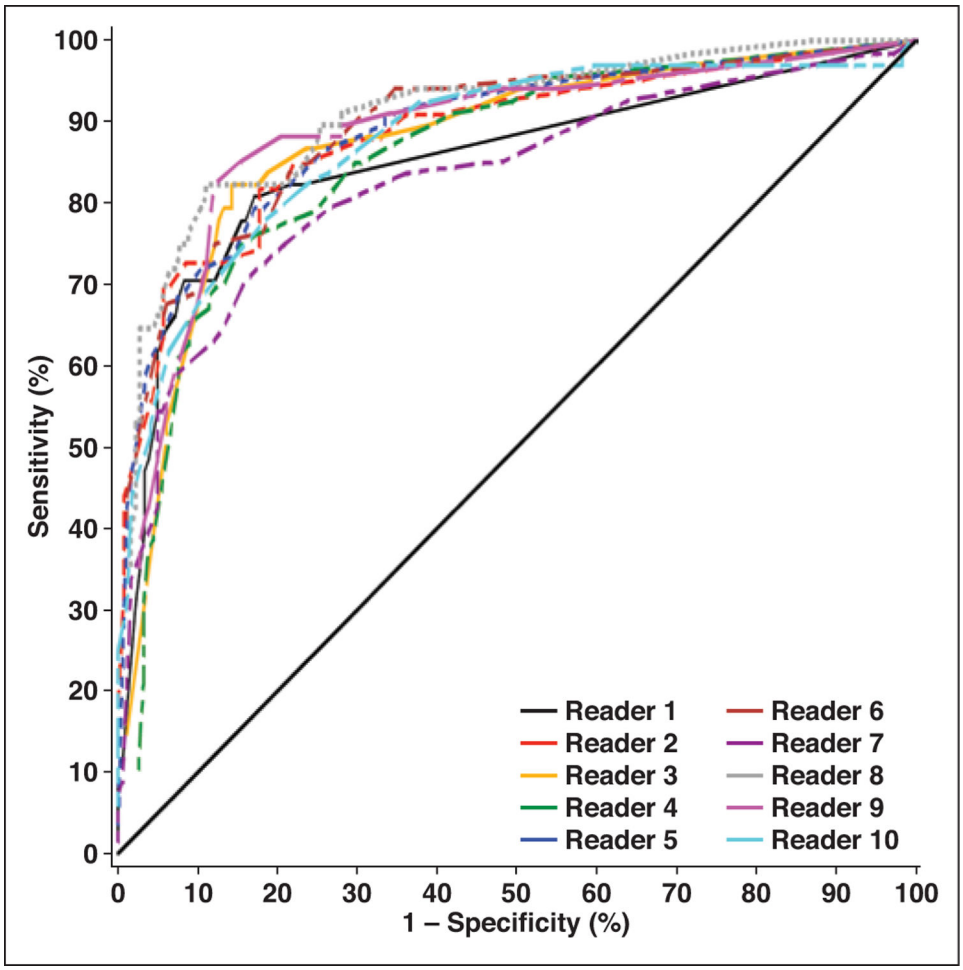

Fig. 1—.

ROC curves for prediction of presence of invasive cancer by estimated percentage likelihood of invasive cancer assigned by each of 10 readers. ROC curve for reader 8 had highest AUC of 0.907 (95% CI, 0.864–0.951). Diagonal line represents line of no discrimination.

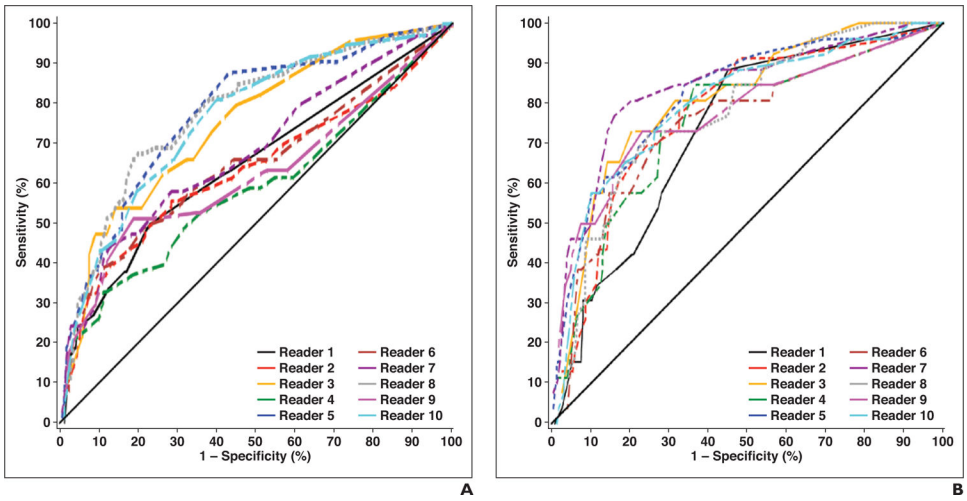

Fig. 2—. ROC curves for prediction of presence of ductal carcinoma in situ (DCIS).

A, ROC curves for prediction of presence of DCIS as sole finding or associated with invasive cancer by estimated percentage likelihood of DCIS. ROC curve for reader 5 had highest AUC of 0.778 (95% CI, 0.714–0.841). Diagonal line represents line of no discrimination.

B, ROC curves for prediction of presence of DCIS alone by estimated percentage likelihood of DCIS. ROC curve for reader 7 had highest AUC of 0.837 (95% CI, 0.748–0.926). Diagonal line represents line of no discrimination.

TABLE 1:

Final Histopathologic Profile of Study Lesions

Histopathologic Profile	No. of Lesions
Malignant lesions	94
Invasive breast cancer, no. (%) of lesions	68 (72)
Invasive ductal carcinoma	56
Grade 1	9
Grade 2	17
Grade 3	15
Grade, not otherwise specified	15
Invasive cancer, not otherwise specified	12
Invasive cancer plus DCIS	48
DCIS alone, no. (%) of lesions	26 (28)
Grade 1	2
Grade 2	11
Grade 3	13
Benign lesions[a]	156

Note—DCIS = ductal carcinoma in situ.

[a]Benign lesions include acellular amorphous debris, atypical ductal hyperplasia, atypical lobular hyperplasia, apocrine metaplasia, benign (not otherwise specified), benign microcalcifications, cyst, duct ectasia, fibroadenoma, fibroadipose tissue, fibrocystic changes, fibrosis, lobular carcinoma in situ, lobular hyperplasia, proteinaceous debris and neutrophils, sclerosing adenosis, stromal hyperplasia, and usual ductal hyperplasia.

TABLE 2:

Distribution of Lesion Classifications (Outcome Variable) for ROC Analyses

Lesion Classification	No. (%) of Lesions (*n* = 250)
Presence of lesion	
Benign	156 (62)
Malignant	94 (38)
Presence of DCIS alone	
DCIS alone	26 (10)
No DCIS alone	224 (90)
Presence of invasive cancer	
Invasive cancer	68 (27)
No invasive cancer	182 (73)

Note—DCIS = ductal carcinoma in situ.

TABLE 3:

Summary of Radiologists' Characteristics With AUC Estimates

Reader	Years in Practice	Type of Practice	AUC (95% CI)	
			Invasive Cancer	Ductal Carcinoma In Situ Only
1	39	Exclusively (only breast)	0.853 (0.796–0.910)	0.731 (0.641–0.820)
4	38	Exclusively (only breast)	0.857 (0.803–0.911)	0.748 (0.645–0.852)
5	29	Exclusively (only breast)	0.890 (0.841–0.938)	0.816 (0.727–0.906)
2	14	Exclusively (only breast)	0.884 (0.831–0.938)	0.775 (0.675–0.874)
8	8	Exclusively (only breast)	0.907 (0.864–0.951)	0.776 (0.687–0.866)
9	5	Exclusively (only breast)	0.885 (0.834–0.935)	0.775 (0.665–0.884)
10	3	Exclusively (only breast)	0.879 (0.827–0.931)	0.798 (0.703–0.892)
7	10	Mostly (≥ 60% interpreting breast studies)	0.830 (0.767–0.893)	0.837 (0.748–0.926)
3	1	Mostly (≥ 60% interpreting breast studies)	0.874 (0.822–0.925)	0.806 (0.722–0.891)
6	1	Mostly (≥ 60% interpreting breast studies)	0.893 (0.845–0.941)	0.749 (0.643–0.855)

Note—Reader numbers correspond to reader numbers shown in the figures.

AJR Am J Roentgenol. Author manuscript; available in PMC 2021 January 02.

TABLE 4:

Numbers of Cases of Ductal Carcinoma In Situ (DCIS) and Invasive Cancer Lesions as Detected by Mammogram Plus Ultrasound Versus Mammogram Alone

Type of Lesion	Mammogram Plus Ultrasound (*n* = 144)	Mammogram Alone (*n* = 106)
DCIS (*n* = 250)		
Present (*n* = 26)	7	19
Absent (*n* = 224)	137	87
Invasive cancer (*n* = 250)		
Present (*n* = 68)	60	8
Absent (*n* = 182)	84	98

Note—Data are number of lesions.

TABLE 5:

Kappa Statistics for 10 Readers by BI-RADS Category

BI-RADS Category	κ (95% CI)
Single category	
1	0.058 (0.04–0.077)
2	0.093 (0.075–0.111)
3	0.068 (0.05–0.087)
4a	0.134 (0.116–0.153)
4b	0.143 (0.125–0.162)
4c	0.260 (0.241–0.278)
5	0.480 (0.465–0.501)
Overall	0.206 (0.197–0.215)
Combined categories	
1 and 2	0.113 (0.095–0.132)
3	0.068 (0.05–0.087)
4a and 4b	0.312 (0.294–0.331)
4c and 5	0.593 (0.575–0.612)
Overall	0.370 (0.356–0.383)

Note—$p < 0.0001$ for all comparisons.