

Variability of Interpretive Accuracy Among Diagnostic Mammography Facilities

Sara L. Jackson, Stephen H. Taplin, Edward A. Sickles, Linn Abraham, William E. Barlow, Patricia A. Carney, Berta Geller, Eric A. Berns, Gary R. Cutter, Joann G. Elmore

- Background** Interpretive performance of screening mammography varies substantially by facility, but performance of diagnostic interpretation has not been studied.
- Methods** Facilities performing diagnostic mammography within three registries of the Breast Cancer Surveillance Consortium were surveyed about their structure, organization, and interpretive processes. Performance measurements (false-positive rate, sensitivity, and likelihood of cancer among women referred for biopsy [positive predictive value of biopsy recommendation (PPV2)]) from January 1, 1998, through December 31, 2005, were prospectively measured. Logistic regression and receiver operating characteristic (ROC) curve analyses, adjusted for patient and radiologist characteristics, were used to assess the association between facility characteristics and interpretive performance. All statistical tests were two-sided.
- Results** Forty-five of the 53 facilities completed a facility survey (85% response rate), and 32 of the 45 facilities performed diagnostic mammography. The analyses included 28100 diagnostic mammograms performed as an evaluation of a breast problem, and data were available for 118 radiologists who interpreted diagnostic mammograms at the facilities. Performance measurements demonstrated statistically significant interpretive variability among facilities (sensitivity, $P = .006$; false-positive rate, $P < .001$; and PPV2, $P < .001$) in unadjusted analyses. However, after adjustment for patient and radiologist characteristics, only false-positive rate variation remained statistically significant and facility traits associated with performance measures changed (false-positive rate = 6.5%, 95% confidence interval [CI] = 5.5% to 7.4%; sensitivity = 73.5%, 95% CI = 67.1% to 79.9%; and PPV2 = 33.8%, 95% CI = 29.1% to 38.5%). Facilities reporting that concern about malpractice had moderately or greatly increased diagnostic examination recommendations at the facility had a higher false-positive rate (odds ratio [OR] = 1.48, 95% CI = 1.09 to 2.01) and a non-statistically significantly higher sensitivity (OR = 1.74, 95% CI = 0.94 to 3.23). Facilities offering specialized interventional services had a non-statistically significantly higher false-positive rate (OR = 1.97, 95% CI = 0.94 to 4.1). No characteristics were associated with overall accuracy by ROC curve analyses.
- Conclusions** Variation in diagnostic mammography interpretation exists across facilities. Failure to adjust for patient characteristics when comparing facility performance could lead to erroneous conclusions. Malpractice concerns are associated with interpretive performance.

J Natl Cancer Inst 2009;101:814–827

Women undergoing diagnostic mammography for a clinical breast symptom are at a 10-fold higher risk for breast cancer compared with asymptomatic women who are screened for cancer (1). Facilities performing screening mammography vary in the accuracy of their interpretation, and those facilities with more frequent audit feedback and a breast specialist on staff may provide more accurate screening tests (2). However, little is known about the effect of facility characteristics on the accuracy of diagnostic mammography.

Previous studies (3–9) described patient and radiologist characteristics associated with the accuracy of diagnostic mammography, such as a patient's breast density and radiologist's time spent in breast imaging. At the mammography facility level, one study in Denmark (6) reported greater accuracy at facilities that have at least one high-volume radiologist. This study, however, did not

Affiliations of authors: Department of Internal Medicine, University of Washington School of Medicine, Seattle, WA (SLJ, JGE); Division of Cancer Control and Population Sciences, Applied Research Program, National Cancer Institute, Bethesda, MD (SHT); Department of Radiology, University of California San Francisco Medical Center, San Francisco, CA (EAS); Center for Health Studies, Group Health, Seattle, WA (LA); Cancer Research and Biostatistics, Seattle, WA (WEB); Department of Family Medicine, Oregon Health and Science University, Portland, OR (PAC); College of Medicine, University of Vermont, Burlington, VT (BG); Department of Radiology, University of Colorado Hospital, Denver, CO (EAB); School of Public Health, University of Alabama at Birmingham, Birmingham, AL (GRC).

Correspondence to: Sara L. Jackson, MD, MPH, Department of Internal Medicine, University of Washington School of Medicine, Box 359854, Seattle, WA 98104 (e-mail: sljack@u.washington.edu).

See "Funding" and "Notes" following "References."

DOI: 10.1093/jnci/djp105

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.

adjust for patient or radiologist factors that could have affected their findings, and the health-care system and screening program in Denmark are quite different from those in other countries, such as the United States (6). Identifying facility characteristics that are associated with greater diagnostic accuracy for women at high risk of breast cancer would be useful for patients and their primary care providers in choosing a facility and to radiologists and institutions in the planning and quality assessment of mammography services.

We had the unique opportunity in a multicenter study to implement two cross-sectional surveys, one of mammography facilities and another of radiologists interpreting mammograms at these facilities, and to link interpretive performance data from the Breast Cancer Surveillance Consortium (BCSC) (10). Our study goals were to evaluate facility performance associated with interpreting diagnostic mammography examinations to see whether variability exists among facilities after adjustment for patient and radiologist characteristics and to assess facility characteristics associated with better accuracy.

We used a blended conceptual framework (11–13) that considered facility organization and structure, including volume, clinical services offered, financial and malpractice characteristics, and scheduling traits, as well as facility processes related to interpretation and auditing. We hypothesized that greater accuracy in diagnostic mammography (higher sensitivity and/or lower false-positive rates) would be associated with academic facilities, higher volume facilities, those with financial incentives for providing more care (eg, higher charges or fiscal market competition), and those that use audit feedback mechanisms designed to improve cancer detection. Conversely, we hypothesized that higher false-positive rates would be noted in for-profit settings and at facilities where malpractice concerns were perceived to affect recommendations, with increased incentive to obtain additional diagnostic evaluations.

Participants and Methods

Three geographically dispersed mammography registries within the BCSC (<http://breastcancerscreening.cancer.gov>) (10) contributed outcome data and participated in the two cross-sectional surveys: one of mammography facilities and the other of radiologists.

All data were analyzed at a central statistical coordinating center (SCC). Each registry and the SCC have received institutional review board approval for either active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytic studies. All procedures are Health Insurance Portability and Accountability Act compliant, and all registries and the SCC have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities who are subjects of this research.

Data Sources and Study Population

The three BCSC mammogram registries included Group Health Breast Cancer Surveillance System, a nonprofit integrated health plan in the Pacific Northwest that includes more than 100 000 women who are 40 years or older (14); the New Hampshire Mammography Network that captures approximately 90% of mammograms performed in New Hampshire (15); and the

CONTEXT AND CAVEATS

Prior knowledge

It is known that interpretive performance of mammography screening varies by facility; whether performance of diagnostic interpretation varies by facility has not been investigated.

Study design

Survey of 45 diagnostic mammography facilities in the Breast Cancer Surveillance Consortium to compare structure, organization, and interpretive processes. Performance measurements were compared and adjusted for patient and radiologist characteristics.

Contribution

Variations in mammography interpretation occurred across facilities, but after adjustment for patient and radiologist characteristics, only variation in false-positive rates remained.

Implications

When comparing the performance of mammography interpretation between facilities, patient and radiologist characteristics should be considered.

Limitations

These results may not be generalizable to other regions of the United States or to other countries where mammography programs, screening guidelines, and systems and requirements for interpretation differ from those of facilities included in this study.

From the Editors

Colorado Mammography Program that provides mammograms to approximately 50% of the women in the six-county metropolitan area of Denver, Colorado (10). The registries collect standardized data on all mammograms including date of examination interpretation, patient characteristics (age, breast density, date of most recent prior mammogram), interpreting radiologist, and facility. The BCSC mammography data were linked to cancer outcomes from regional cancer registries and pathology databases to obtain both benign and malignant outcomes. These data were pooled at a single SCC for analysis. Cancer ascertainment is estimated to be better than 94.3% at BCSC sites (16).

Diagnostic mammograms were defined by the interpreting radiologist as examinations indicated for the evaluation of a breast problem (clinical sign or symptom of breast cancer). Examinations performed between January 1, 1998 and December 31, 2005 were included in the study. We did not include diagnostic mammograms performed for additional diagnostic evaluation after a screening examination or for short-interval follow-up of a probably benign finding because these examinations are performed on asymptomatic women and the outcomes of these examinations are known to be different from those performed on primarily symptomatic women (1,17,18). Mammograms of women who had breast augmentation, reduction, or reconstruction and of women younger than 18 years were excluded.

Mammography Facilities Survey

Facilities providing screening mammography within the three BCSC registries were eligible for participation. The facility survey, which was developed by a panel of experts in multiple disciplines,

was based on a conceptual framework of factors known or suspected to influence mammography interpretation (2). The survey assessed facility organizational and structural processes, including academic affiliation, volume, clinical services, financial malpractice, and scheduling characteristics, and interpretive and audit processes, such as double reading, and method of audit feedback. Detailed description of the survey was previously published (2), and the survey is available online at <http://breastscreening.cancer.gov/collaborations/favor.html>.

Surveys were mailed to designated contact persons at each eligible facility and completed by one or more persons depending on the facility administrator and specific survey question; respondents included a lead technologist, a radiologist, the radiology department and/or the facility business manager. If no response was obtained after the second mailing, then telephone or in-person contact was attempted, with data collection completed in September 2002, after a 10-month period.

Radiologist Survey

A survey addressing a wide variety of radiologist characteristics and perceptions was mailed during the same period to radiologists who interpreted mammograms at any of the participating registries. This survey has been previously described (19) and is available online at <http://breastscreening.cancer.gov/collaborations/favor.html>.

Specification of Study Variables

Outcome measures included false-positive rate, sensitivity, positive predictive value of biopsy recommendation (PPV2), and area under the receiver operating characteristic curve (AUC) of diagnostic mammography. To calculate these variables, we defined a negative assessment as breast imaging-reporting and data system (BI-RADS) 1 (negative), 2 (benign), 3 (probably benign), 0 (incomplete—needs additional imaging evaluation), when the latter two were not associated with a recommendation for a biopsy (ie, fine-needle aspiration [FNA], biopsy, or surgical consultation). We defined a positive assessment as BI-RADS 4 (suspicious) or 5 (highly suggestive of malignancy). Assessments of 3 or 0 with a recommendation for FNA, surgical consultation, or biopsy were also considered positive (20). These assessments were based on the final, not initial, interpretations. For each diagnostic mammogram, we looked for breast cancer (ie, invasive carcinoma or ductal carcinoma in situ but not lobular carcinoma in situ) diagnosed during the subsequent year.

The false-positive rate was defined as the proportion of diagnostic examinations interpreted as positive among all patients who did not have a diagnosis of breast cancer within 1 year. Sensitivity (also called the true-positive rate) was defined as the proportion of diagnostic examinations interpreted as positive among all patients who had a diagnosis of breast cancer within 1 year. PPV2 was defined as the number of mammograms that were associated with a breast cancer diagnosis within 1 year among those with a positive assessment.

Statistical Analysis

Initially, a logistic regression model for each performance measure was fit, which included a facility random effect variable only. We

then built logistic regression models adjusted for mammography registry, patient characteristics (patient age [<40 , 40–49, 50–59, 60–69, ≥ 70 years], breast density [almost entirely fat, scattered fibroglandular densities, heterogeneously dense, extremely dense], time since last mammogram [<1 , 1 to <3 , ≥ 3 years, no previous mammography], self-reported breast lump [no, yes]), radiologist characteristics (years of mammography interpretation [<10 , 10–19, ≥ 20 years], percent time working in breast imaging [<20 , 20–39, ≥ 40], number of mammograms interpreted in previous year [<1000 , 1000–2000, ≥ 2001]), and a facility random effect. By comparing the models with a facility random effect only to those with a facility random effect and patient and radiologist characteristics, we sought to determine whether statistically significant variation remained across facilities after adjusting for patient and radiologist factors.

The association between characteristics of patients and radiologists included in the study and false-positive rate and sensitivity was assessed. Then, means and 95% confidence intervals (CIs) were computed for false-positive rate, sensitivity, and PPV2 for each facility, both at the facility level (each facility contributes equally) and at the mammogram level.

Next, we constructed multivariable logistic regression models for each facility characteristic of interest, adjusted for patient and radiologist characteristics, mammography registry, and a facility random effect. Patient and radiologist characteristic variables were categorical (*see* Table 1) and based on previous literature and adequate spread of the data (4,19). Throughout the analyses, the patient and radiologist characteristics that were adjusted for remained consistent, as listed above.

We also built fully adjusted multivariable logistic regression models including all facility characteristics that were univariately associated with any outcome at the level of P less than or equal to .1 with any performance outcome, to allow borderline statistically significant factors to be included, and adjusting for patient and radiologist characteristics, mammography registry, and a facility random effect. The variable “imaging services offered” was excluded from the models because it was highly related to “interventional services offered.” For all models, mammograms missing data on patient, radiologist, or facility characteristics were excluded (Figure 1).

Receiver operating characteristic (ROC) curve analysis was performed to summarize the association between facility characteristics and overall diagnostic mammography accuracy. AUC compares sensitivity to 1 minus the specificity over a range of values for these two measurements, thus incorporating both sensitivity and specificity information in one analytic assessment. AUC values can range from 0 to 1, although only values greater than 0.5 reflect accuracy greater than chance alone, with higher values indicating better ability to discriminate between mammograms with and without breast cancer (eg, better accuracy).

We used the SAS procedure NLMIXED (SAS version 9.1; SAS Institute, Inc., Cary, NC), with the BI-RADS assessment codes as an ordinal response: 1, 2, 3, or 0 without a recommendation for biopsy; 3 or 0 with a recommendation for biopsy; or 4 or 5. We then fit ordinal regression models including all facility characteristics that were statistically significant at the P less than or equal to .1 level in the models described above and adjusted for patient and

Table 1. Characteristics of patients obtaining diagnostic mammography for evaluation of a breast problem

Patient characteristic*	Total diagnostic mammograms, No.	Diagnostic mammograms	
		Without cancer, No. (%)	With cancer, No. (%)
Patient's age, y			
<40	5467	5351 (20.0)	116 (8.7)
40–49	9827	9502 (35.5)	325 (24.5)
50–59	6201	5872 (21.9)	329 (24.8)
60–69	3120	2923 (10.9)	197 (14.8)
≥70	3485	3123 (11.7)	362 (27.2)
Mammographic breast density			
Almost entirely fat	1504	1435 (5.4)	69 (5.2)
Scattered fibroglandular densities	9266	8838 (33.0)	428 (32.2)
Heterogeneously dense	12 582	11 935 (44.6)	647 (48.7)
Extremely dense	4748	4563 (17.0)	185 (13.9)
Time since last screening or diagnostic mammogram			
<1 y	7822	7460 (27.9)	362 (27.2)
1 to <3 y	12 255	11 726 (43.8)	529 (39.8)
≥3 y	3521	3290 (12.3)	231 (17.4)
No previous mammography	4502	4295 (16.0)	207 (15.6)
Reported presence of a breast lump			
No	10 189	9943 (37.1)	246 (18.5)
Yes	17 911	16 828 (62.9)	1083 (81.5)

* Based on main study cohort of 28 100 diagnostic mammograms (26 771 without cancer and 1329 with cancer) from 32 facilities.

radiologist characteristics. The multivariable models enabled us to estimate ROC curves and the AUC associated with specific facility characteristics while adjusting for patient and radiologist characteristics. The actual AUC value was computed from the estimates associated with covariates. Additional detail regarding the method of fitting the ROC curves is available in an earlier publication (19). All model score equations converged to 0 when tested with the default SAS convergence criterion. Likelihood ratio statistics were used to determine whether each facility factor was statistically significantly associated with accuracy ($P < .05$). All P values were two-sided.

Results

The study populations for the main analysis and the subsequent models that adjust for patient and radiologist characteristics are described in Figure 1. Forty-five of the 53 eligible facilities (85% response rate) completed the facility survey. Thirty-two of the 45 facilities performed diagnostic mammograms in the study period and were therefore included in the main descriptive analyses. A total of 28 100 diagnostic mammograms were performed as an evaluation of a breast problem during the 8-year study period at these 32 facilities, with 1329 mammograms associated with breast cancer. Data were available for 118 radiologists who interpreted diagnostic mammograms at the 32 facilities during the study period (Figure 1).

In the final model, which was fully adjusted for patient, radiologist, and statistically significant facility characteristics, we excluded five facilities because of missing data on facility characteristics in the model and one radiologist who responded but interpreted diagnostic mammograms at a facility that was missing data on facility characteristics of interest. Thus, 27 facilities with 117 radiologists interpreting 20 019 evaluations of breast problems (including

942 examinations associated with cancer) were included in the final multivariable models (Figure 1).

We assessed characteristics of patients and radiologists included in the study and their relationship to the number of diagnostic mammograms with and without cancer (Tables 1 and 2). The relationships between the patient and the radiologist characteristics and false-positive rates and sensitivity are consistent with published results within the same study population assessing radiologist characteristics and accuracy of diagnostic examinations (4) (data not shown).

Individuals who responded to the facility survey included 20 lead mammography technologists, seven radiologists, and two radiology department or facility business managers (respondent was not identified in three facilities). Among the 32 facilities, 28 reported offering breast ultrasound, five reported offering breast magnetic resonance imaging, and two reported offering breast computed tomography. Interventional services performed at the facilities included FNA ($n = 10$), core biopsy ($n = 16$), vacuum-assisted biopsy ($n = 9$), cyst aspirations ($n = 22$), and needle localization ($n = 20$). A median of seven radiologists per year were working in breast imaging at a facility, with 14 facilities reporting 10 or more radiologists working in breast imaging, and four facilities having at least one radiologist with fellowship training in breast imaging. A median of 81 diagnostic mammograms were performed to evaluate a breast problem per year per facility (range = 1–883). Facility practice characteristics were assessed in relation to unadjusted performance measurements (Table 3).

Variability of unadjusted diagnostic mammography performance measurements for the 32 facilities, ordered by number of diagnostic evaluations (highest to lowest volume), is shown in Figure 2. Overall mean false-positive rate (at the facility level) was 6.5% (95% CI = 5.5% to 7.4% and range = 1%–14%), mean sensitivity was 73.5% (95% CI = 67.1% to 79.9% and range = 33%–100%), and

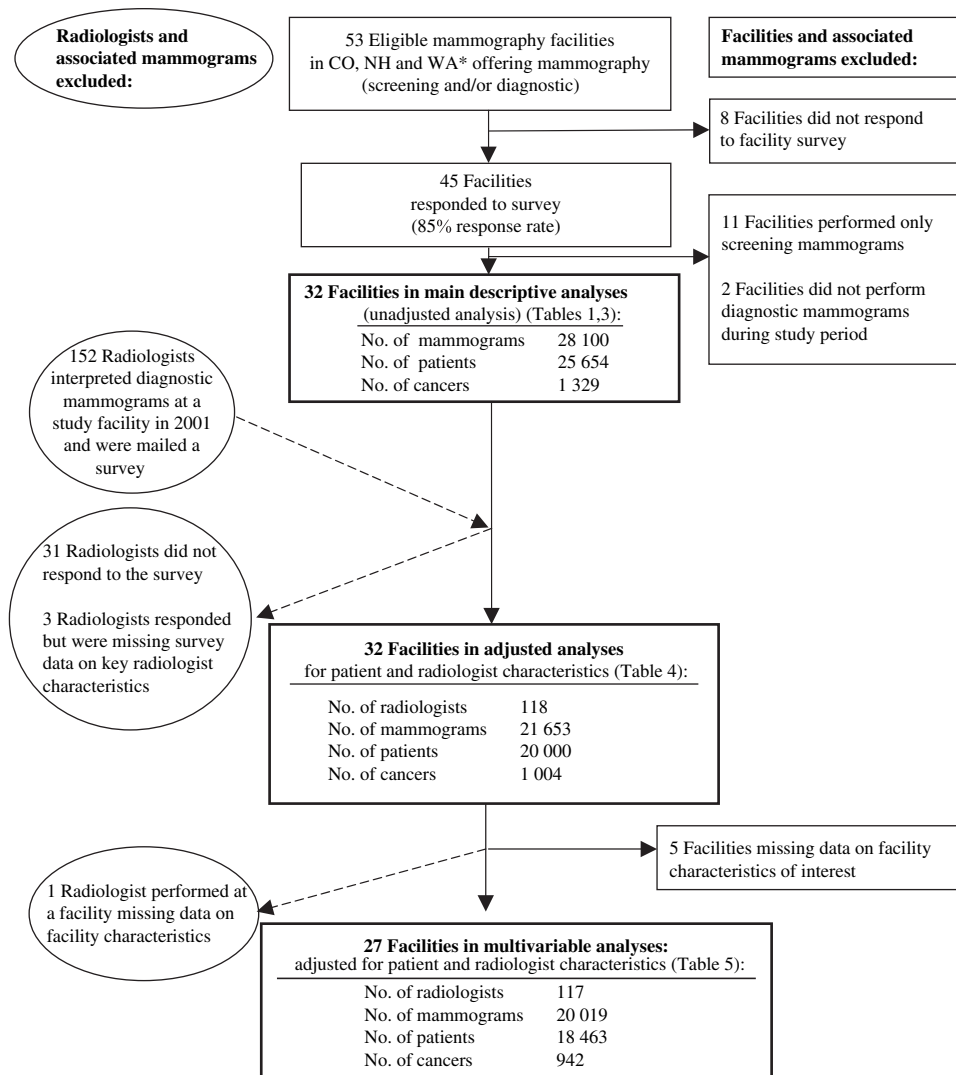


Figure 1. Study populations included in the analyses, including number of facilities, radiologists, patients, mammograms, and cancers. *CO = Colorado; NH = New Hampshire; WA = Washington.

mean PPV2 was 33.8% (95% CI = 29.1% to 38.5% and range = 0%–61%). In unadjusted logistic regression analyses, statistically significant variability was found for all three performance measures across facilities (sensitivity, $P = .006$; false-positive rate, $P < .001$; and PPV2, $P < .001$) (Figure 2). This statistical variability in performance persisted after adjustment for patient and radiologist characteristics only for false-positive rate ($P < .001$).

Statistically significant associations were noted between frequencies of the facility characteristics and their unadjusted measures of performance (false-positive rate, sensitivity, and PPV2) (Table 3). In unadjusted analyses, no facility characteristics were associated with false-positive rate at the level of P less than .05, and multiple facility characteristics were associated with sensitivity and PPV2 (Table 3).

The facility characteristics and performance measures are shown adjusted for patient and radiologist characteristics, mammography registry, and a facility random effect (Table 4). The variables that were associated with performance after adjustment are markedly different compared with the unadjusted results

shown in Table 2. Those that are statistically significantly associated are footnoted (Table 4).

Many facility characteristics associated with performance measurements changed between unadjusted models (Table 3) and those adjusted for patient and radiologist characteristics (Table 4). To isolate the effect of adjustment for patient characteristics alone, we performed logistic regression models adjusted for patient characteristics (patient age, breast density, time since last mammogram, self-reported breast lump), mammography registry, and a facility random effect variable, but not for radiologist characteristics. The variables identified in Table 4 were also at least borderline statistically significantly associated with the same performance measures, when adjusted for patient characteristics only, with similar odds ratios (ORs) and 95% confidence intervals (data not shown).

Multivariable model results were adjusted for patient characteristics, radiologist characteristics, mammography registry, a facility random effect, and all facility characteristics noted in Table 3 to be associated with false-positive rate, sensitivity, or PPV2 at the P less than or equal to .1 level (Table 5). Facilities reporting that

Table 2. Characteristics of radiologists interpreting examinations included in the study

Radiologist characteristic*	Total radiologists, No.†	Diagnostic mammograms	
		Without cancer, No. (%)	With cancer, No. (%)
Experience			
Years of mammography interpretation			
<10	28	3715 (17.8)	187 (18.4)
10–19	54	12 028 (57.6)	609 (59.8)
≥20	37	5124 (24.6)	222 (21.8)
Percentage of time working in breast imaging			
<20	57	6308 (30.2)	278 (27.3)
20–39	48	11 663 (55.9)	570 (56.0)
≥40	14	2896 (13.9)	170 (16.7)
No. of mammograms interpreted in the previous year			
<1000	33	2246 (10.8)	114 (11.1)
1001–2000	45	7110 (34.1)	354 (34.5)
≥2001	42	11 508 (55.2)	557 (54.3)
Practice characteristics			
Primary affiliation with an academic medical center			
No	115	20 215 (95.9)	998 (96.1)
Yes	6	867 (4.1)*	41 (3.9)*
Percentage of mammograms interpreted that were diagnostic			
0–24	62	10 619 (50.4)	505 (48.6)
25–49	49	9084 (43.1)	481 (46.3)
50–100	10	1379 (6.5)*	53 (5.1)*
Performed breast biopsy examinations in the previous year			
No	35	3895 (18.9)	172 (16.9)
Yes	83	16 727 (81.1)	847 (83.1)

* Based on 22 121 diagnostic mammograms (21 082 without cancer and 1039 with cancer) from 32 facilities and 121 radiologists.

† Number of radiologists may not add up to 121 because of radiologists missing responses to certain questions on the radiologist survey, and subsequently the number of diagnostic mammograms may be less than 22 121.

malpractice concerns had moderately to greatly increased recommendations for additional tests at the facility were associated with a higher false-positive rate (OR = 1.48, 95% CI = 1.09 to 2.01) and a non-statistically significantly greater sensitivity (OR = 1.74, 95% CI = 0.94 to 3.23). Offering specialized interventional services was associated with a non-statistically significantly higher false-positive rate (OR = 1.97, 95% CI = 0.94 to 4.10). Because of small numbers (n = 3), interpreting all diagnostic mammograms on-site was not statistically significantly associated with higher sensitivity compared with no on-site reading (OR = 3.12, 95% CI = 0.54 to 18.09).

ROC curve analyses were adjusted for patient and radiologist characteristics. The analyses did not identify any facility characteristics that were associated with statistically significantly improved or diminished overall diagnostic accuracy.

Discussion

Mammography facilities vary in their size, organization, services, and processes, but it is not known whether any of these facility differences affect the interpretive performance of diagnostic mammography. In this multisite cross-sectional study, we found that US facilities have statistically significant variability in the false-positive rates of diagnostic mammography after adjusting for patient and radiologist factors. Reporting that concerns about malpractice increased additional diagnostic testing at the facility was associated with a statistically significantly higher false-positive rate and a non-statistically significantly higher sensitivity. A non-statistically significant association was also noted between a greater

false-positive rate and offering specialized interventional services. Off-site interpretation of diagnostic mammograms, which occurred in three facilities in this study, was non-statistically significantly associated with lower sensitivity than facilities with on-site reading. None of the facility characteristics we assessed were associated with changes in overall accuracy.

The extensive variability in interpretive performance that we noted was markedly reduced after adjustment for patient and radiologist characteristics. Statistically significant variability was present for sensitivity and PPV2 in unadjusted analyses, but after accounting for differences in patient and radiologist characteristics, the variation was reduced. Our sample was based on 20 019 diagnostic mammograms, but some low-volume facilities had few cancers detected, as demonstrated by the wide confidence intervals around the individual facility sensitivity values and PPV2 values demonstrated in Figure 2. Because sensitivity and PPV2 depend on the number of breast cancers detected, which is a relatively infrequent outcome, low statistical power may have affected our analyses.

Detailed patient and radiologist data allowed us to discriminate facility-level associations from those due to patient and radiologist factors. We sequentially assessed interpretive performance adjusted for patient factors; then for patient and radiologist factors; and finally for patient, radiologist, and multiple facility-level characteristics. The facility traits associated with performance changed dramatically after adjustment for patient attributes, then were consistent with additional adjustment for radiologist factors, and diminished when multiple facility factors

Table 3. Facility-reported practice procedures and unadjusted performance measurements at the level of the diagnostic mammogram (based on 32 facilities with 28 100 evaluations of a breast problem)*

Procedures and performance measurements	No.† (%)	False-positive rate,‡ % (95% CI)	Sensitivity,§ % (95% CI)	PPV2, % (95% CI)
Overall mean at mammogram level¶	32	6.4 (5.5 to 7.4)	82.1 (78.4 to 85.3)	39.1 (35.3 to 43.0)
Facility structure and organization				
Associated with an academic medical center				
No	27 (84)	5.9 (5.0 to 7.0)	81.0 (77.6 to 84.0)	39.3 (35.1 to 43.7)
Yes	5 (16)	7.6 (6.2 to 9.2)	84.8 (74.7 to 91.3)	38.5 (30.9 to 46.7)
Volume				
Facility volume of all screening and diagnostic mammograms (average per year) from BCSC data#				
≤1500	5 (16)	8.5 (5.7 to 12.4)	73.7 (58.5 to 84.7)*	31.8 (20.9 to 45.2)
1501–2500	7 (23)	6.0 (5.1 to 7.0)	70.0 (55.8 to 81.2)*	32.6 (26.6 to 39.2)
2501–6000	9 (29)	6.2 (5.2 to 7.3)	77.5 (72.3 to 81.9)*	38.0 (32.2 to 44.1)
>6000	10 (32)	6.3 (5.0 to 8.0)	86.1 (83.4 to 88.4)*	41.0 (36.0 to 46.1)
Facility volume of diagnostic evaluations for breast problems (average per year) from BCSC data#				
≤100	13 (42)	7.7 (6.3 to 9.3)	71.8 (63.6 to 78.8)*	31.5 (26.6 to 36.8)*
101–400	9 (29)	7.0 (5.2 to 9.5)	83.6 (76.1 to 89.1)*	36.3 (29.9 to 43.1)*
>400	9 (29)	5.6 (4.6 to 6.8)	83.8 (79.6 to 87.3)*	43.2 (40.3 to 46.1)*
Clinical services				
Interventional services offered (FNA to core or vacuum-assisted biopsy, cyst aspirations, needle localization, or other procedures)				
No	8 (25)	4.1 (2.7 to 6.1)	68.7 (51.5 to 81.9)	34.7 (27.2 to 43.0)
Yes	24 (75)	6.7 (5.7 to 7.7)	83.2 (79.7 to 86.2)	39.4 (35.5 to 43.5)
Specialized imaging services offered (breast CT, breast MRI, breast nuclear medicine scans)**				
No	13 (57)	5.8 (4.1 to 8.2)	81.3 (74.1 to 86.9)	37.8 (29.3 to 47.2)
Yes	10 (43)	7.3 (6.2 to 8.6)	84.6 (79.1 to 88.8)	38.8 (34.0 to 43.8)
Facility is currently short staffed (ie, not enough radiologists)				
Strongly disagree/disagree/neutral	17 (53)	6.2 (5.2 to 7.5)	79.0 (70.6 to 85.6)	39.5 (33.9 to 45.4)
Agree/strongly agree	15 (47)	6.4 (5.1 to 8.0)	84.5 (81.4 to 87.2)	38.8 (33.9 to 43.9)
Financial and malpractice				
Profit status				
Nonprofit	19 (59)	6.4 (5.4 to 7.5)	85.0 (82.1 to 87.6)*	41.4 (37.9 to 45.0)
For profit	13 (41)	6.3 (4.5 to 8.8)	73.4 (66.4 to 79.3)*	32.8 (26.7 to 39.5)
What does your facility charge self-pay patients (uninsured) for diagnostic mammograms (facility and radiologist fees)?				
<\$200 per examination	17 (63)	7.0 (5.4 to 8.9)	78.9 (72.7 to 84.0)	34.4 (29.5 to 39.7)*
≥\$200 per examination	10 (37)	6.0 (5.0 to 7.2)	86.1 (83.3 to 88.5)	43.3 (39.8 to 46.9)*
How have medical malpractice concerns influenced recommendations of diagnostic mammograms, ultrasounds, or breast biopsies at your facility following screening mammograms?				
Not changed	11 (41)	5.2 (4.2 to 6.5)	74.8 (70.2 to 78.9)*	39.0 (33.5 to 44.9)
Moderately increased/greatly increased	16 (59)	6.9 (5.7 to 8.4)	85.0 (81.6 to 87.8)*	39.5 (34.5 to 44.7)
Do you feel that your facility has fiscal market competition from other mammography facilities in the area?				
No competition	7 (23)	5.5 (4.8 to 6.4)	85.5 (82.3 to 88.2)	42.4 (36.5 to 48.6)
Some competition	11 (37)	6.9 (5.2 to 9.2)	84.9 (80.1 to 88.7)	39.1 (32.8 to 45.8)
Moderate to extreme competition	12 (40)	6.2 (5.3 to 7.2)	76.3 (69.7 to 81.9)	37.3 (32.0 to 42.9)

(Table continues)

Table 3 (continued).

Procedures and performance measurements	No.† (%)	False-positive rate,‡ % (95% CI)	Sensitivity,§ % (95% CI)	PPV2, % (95% CI)
Scheduling process				
Average wait time to schedule diagnostic mammogram				
≤3 d	16 (52)	5.4 (4.4 to 6.7)	74.6 (67.2 to 80.7)*	36.4 (31.5 to 41.7)
>3 d	15 (48)	7.0 (5.8 to 8.3)	85.3 (82.3 to 87.9)*	40.0 (35.1 to 45.1)
Do women wait for interpretation of diagnostic mammogram?				
No, never/yes, some of the time	14 (44)	6.1 (5.2 to 7.1)	76.7 (70.2 to 82.1)*	38.0 (34.1 to 42.0)
Yes, all of the time	18 (56)	6.5 (5.3 to 7.9)	84.6 (81.2 to 87.5)*	39.6 (34.5 to 44.8)
Interpretation and audit processes				
Interpretive processes				
Are clinical breast examinations done routinely for women getting a screening mammogram?				
No	25 (78)	6.5 (5.3 to 7.9)	78.3 (73.6 to 82.3)*	35.1 (30.9 to 39.5)*
Yes	7 (22)	6.1 (5.0 to 7.5)	87.2 (85.7 to 88.6)*	45.3 (43.5 to 47.1)*
What percentage of diagnostic mammograms are interpreted on-site?				
0	3 (9)	5.2 (4.2 to 6.5)	53.6 (35.6 to 70.7)	27.8 (22 to 34.4)*
100	29 (91)	6.4 (5.5 to 7.4)	82.7 (79.2 to 85.7)	39.3 (35.5 to 43.2)
How many radiologists interpret mammograms full time (ie, 40 h per week)?				
0	26 (84)	6.5 (5.5 to 7.6)	82.5 (78.5 to 85.9)	39.6 (35.5 to 43.9)
≥1	5 (16)	5.4 (3.8 to 7.8)	80.4 (69.4 to 88.1)	35.8 (28.1 to 44.3)
Are any diagnostic mammograms interpreted by more than one radiologist?				
No	15 (50)	5.3 (4.7 to 5.9)	82.2 (77.8 to 85.9)	40.1 (35.0 to 45.4)
Yes, 1%–5% of the time	6 (20)	5.7 (3.4 to 9.5)	76.0 (59.4 to 87.2)	33.3 (24.7 to 43.1)
Yes, 10%–25% or less of the time	6 (20)	7.2 (5.8 to 9.0)	86.0 (82.8 to 88.7)	42.3 (36.7 to 48.2)
Yes, ≥80% of the time	3 (10)	7.8 (6.3 to 9.6)	73.3 (68.7 to 77.4)	32.8 (29.2 to 36.6)
Audit processes				
Individual performance data reported back to radiologists				
Once a year	13 (41)	5.6 (4.8 to 6.5)	80.3 (75.3 to 84.6)	40.7 (37.1 to 44.5)
Twice or more per year	15 (47)	7.3 (5.9 to 9.1)	83.2 (77.7 to 87.6)	37.3 (30.8 to 44.3)
Unknown	4 (13)	4.8 (4.0 to 5.8)	83.5 (77.9 to 87.9)	42.8 (35.9 to 50.0)
Method of performance feedback information review				
Reviewed together in meeting	18 (56)	6.6 (5.5 to 8.0)	81.2 (76.9 to 84.8)	37.7 (33.2 to 42.4)
Reviewed by facility or department manager or lead radiologist alone	5 (16)	4.4 (3.1 to 6.2)	80.6 (68.6 to 88.8)	41.9 (31.3 to 53.4)
Reviewed by each radiologist alone	4 (13)	7.0 (6.0 to 8.0)	73.7 (50.8 to 88.4)	28.3 (15.4 to 46.1)
Unknown	5 (16)	6.9 (5.5 to 8.6)	86.9 (82.8 to 90.2)	44.3 (41.8 to 46.7)

* Values are statistically significantly different, that is, $P < .05$ (two-sided χ^2 tests). PPV2 = positive predictive value of biopsy; CI = confidence interval; BCSC = Breast Cancer Surveillance Consortium; FNA = fine-needle aspiration; CT = computed tomography; MRI = magnetic resonance imaging.

† Columns may not total 32 because of missing values.

‡ False-positive rate is defined as the percentage of mammograms without a cancer diagnoses within 1 year that also have a BI-RADS assessment 4, 5, 0, or 3 with a recommendation for biopsy, surgical consultation, or FNA.

§ Sensitivity is defined as the percentage of mammograms with a cancer diagnoses within 1 year that also have a BI-RADS assessment 4, 5, 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA.

|| PPV2 is defined as the percentage of mammograms with either a BI-RADS assessment 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA or a BI-RADS assessment 4 or 5 that result in a cancer diagnosis within 1 year.

¶ The overall mean at the mammogram level was computed by weighting each facility by the number of mammography screening examinations interpreted at that facility and included in this analysis (facilities with more mammography screening examinations would be given a higher weight in this calculation compared with the overall mean at the facility level shown in Figure 2).

Based on mammography registry data.

** Does not include ultrasound or ductography.

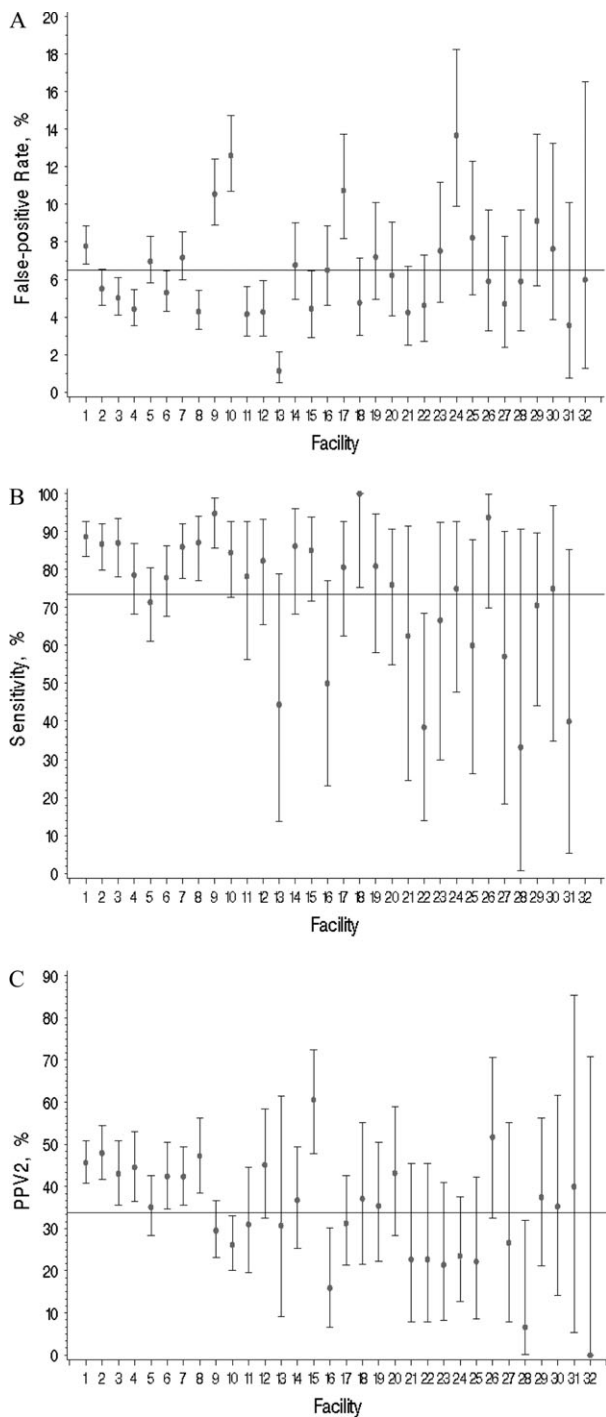


Figure 2. Unadjusted diagnostic mammography performance measures for the 32 facilities. Overall mean at the facility level indicated by a line. **A)** False-positive rate. The unadjusted facility-level mean was 6.5% (95% CI = 5.5 to 7.4). **B)** Sensitivity. The unadjusted facility-level mean was 73.5% (95% CI = 67.1 to 79.9); facility no. 32 was without any cancers during follow-up. **C)** PPV2. The unadjusted facility-level mean was 33.8% (95% CI = 29.1 to 38.5). Overall mean at the facility level was computed by calculating the performance measures for each facility and taking the average across facilities. Each facility was given the same weight regardless of the number of screens interpreted at the facility. **Diamonds** indicate mean values; **error bars** correspond to 95% CIs. Facilities are ordered by number of diagnostic evaluations (highest to lowest) included in the analysis. Statistically significant variability was found across the facilities in unadjusted analyses for all performance measures ($P < .01$, calculated using two-sided F tests). CI = confidence intervals; PPV2 = positive predictive value of biopsy.

were also included. These analyses suggest that the factors most strongly associated with interpretive performance for diagnostic mammograms are characteristics of a patient, such as age, breast density, time since last mammogram, and self-reported breast lump. These patient factors may differ substantially across facilities. Many health and insurance agencies encourage or publish “report cards” and quality rating systems for physicians and hospitals (21). We caution that analyses comparing differences among mammography facilities that do not adjust for important characteristics of patients using the facility may falsely conclude that there is more facility variation than actually exists or that one facility is above or below average.

Reported concerns about malpractice were the predominant factors associated with interpretive performance at the facility level, above all the other facility characteristics measured. The influence of malpractice on US radiologists is postulated to play a role in higher recall rates in the United States than in other countries (22,23). In a previous report of 124 US radiologists, 118 of whom were also participants in this study (13), the majority (58.5% or 72 of 123) indicated that malpractice concerns moderately or greatly increased their recommendations for biopsy examinations after screening mammograms. However, the data did not demonstrate that they had higher recall and false-positive rates for screening mammograms than their colleagues without such perceptions (13). In this facility study, belief that malpractice concerns at the facility level moderately to greatly increased recommendations for diagnostic evaluations was associated with a higher false-positive rate and non-statistically significantly higher sensitivity. It is possible that when the clinical level of perceived patient risk of breast cancer is higher, as in diagnostic mammograms compared with screening examinations, the effect of malpractice concern is more pronounced and thus measurable.

We hypothesized a priori that malpractice concerns would be associated with lower overall accuracy by increasing the false-positive rate; however, we did not find any change in overall accuracy, using ROC curve analysis. Malpractice concern was associated with both an increase in false-positive rate and an increase in cancer detection. This finding suggests a shift in the threshold for calling examinations abnormal, not a true increase in the ability to discriminate between cancer-free and cancer-associated mammograms. Although false-positive examinations are associated with expense and anxiety, any attempt to lower the false-positive rate must not be at the expense of decreasing the cancer detection rate.

Facilities that offered interventional services had a non-statistically significantly higher false-positive rate (OR = 1.97, 95% CI = 0.94 to 4.1). It is plausible that if diagnostic modalities are readily available on-site, radiologists might have a lower threshold for recommending these procedures, resulting in more false-positive and true-positive examinations. It may also be that radiologists in diagnostic facilities perceive a higher probability of cancer in general and are therefore more likely than other radiologists to make a positive interpretation. Financial incentive to perform well-reimbursed interventional services may also play a role. Sensitivity was also increased in facilities offering interventional services (OR = 1.75, 95% CI = 0.47 to 6.45), although not statistically significantly, suggesting a shift in threshold for calling

Table 4. Associations between individual facility characteristics and false-positive rate, sensitivity, and positive predictive value of biopsy, adjusted for patient characteristics, radiologist characteristics, and a facility random effect (based on 32 facilities with 21653 evaluations of a breast problem)*

Facility characteristic	Odds of having a positive mammogram given no cancer diagnosis (false-positive rate),† OR (95% CI)	Odds of having a positive mammogram given a cancer diagnosis (sensitivity),‡ OR (95% CI)	Odds of having a cancer diagnosis given a positive mammogram (PPV2),§ OR (95% CI)
Facility structure and organization			
Associated with an academic medical center			
No	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes	1.1 (0.75 to 1.68)	1.1 (0.67 to 1.65)	1 (0.77 to 1.32)
Volume			
Facility volume of all screening and diagnostic mammograms (average per year) from BCSC data			
≤1500	1.0 (referent)	1.0 (referent)	1.0 (referent)
1501–2500	0.6 (0.36 to 1.14)	0.5 (0.19 to 1.37)	0.9 (0.46 to 1.65)
2501–6000	0.7 (0.39 to 1.15)	1.1 (0.43 to 2.6)	1.0 (0.58 to 1.69)
>6000	0.7 (0.42 to 1.23)	1.3 (0.56 to 3.2)	0.9 (0.56 to 1.56)
Facility volume of diagnostic evaluations for breast problems (average per year) from BCSC data			
≤100	1.0 (referent)	1.0 (referent)	1.0 (referent)
101–400	0.9 (0.61 to 1.41)	1.6 (0.90 to 2.82)	0.9 (0.60 to 1.21)
>400	0.8 (0.46 to 1.23)	1.3 (0.72 to 2.51)	1.1 (0.75 to 1.63)
Clinical services			
Interventional services offered (FNA, core or vacuum-assisted biopsy, cyst aspirations, needle localization, or other procedures)			
No	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes	1.6 (1.08 to 2.28)*	2.4 (1.30 to 4.31)*	1.3 (0.86 to 2.00)
Specialized imaging services offered (breast CT, breast MRI, breast nuclear medicine scans)			
No	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes	1.6 (1.11 to 2.18)*	1.1 (0.69 to 1.76)	1.1 (0.84 to 1.42)
Facility is currently short staffed (ie, not enough radiologists)			
Strongly disagree/disagree/neutral	1.0 (referent)	1.0 (referent)	1.0 (referent)
Agree/strongly agree	1.3 (0.96 to 1.81)	1.6 (1.05 to 2.50)*	0.8 (0.61 to 0.97)*
Financial and malpractice			
Profit status			
Nonprofit	1	1	1
For profit	0.7 (0.51 to 1.02)	0.7 (0.45 to 1.20)	0.9 (0.69 to 1.22)
What does your facility charge self-pay patients (uninsured) for diagnostic mammograms (facility and radiologist fees)?			
<\$200 per examination	1.0 (referent)	1.0 (referent)	1.0 (referent)
≥\$200 per examination	1.1 (0.68 to 1.77)	1.5 (0.52 to 4.13)	0.8 (0.45 to 1.33)
How have medical malpractice concerns influenced recommendations of diagnostic mammograms, ultrasounds, or breast biopsies at your facility following screening mammograms?			
Not changed	1.0 (referent)	1.0 (referent)	1.0 (referent)
Moderately increased/greatly increased	1.6 (1.13 to 2.15)*	1.5 (0.92 to 2.45)	0.7 (0.55 to 0.98)*
Do you feel that your facility has fiscal market competition from other mammography facilities in the area?			
No competition	1.0 (referent)	1.0 (referent)	1.0 (referent)

(Table continues)

Table 4 (continued).

Facility characteristic	Odds of having a positive mammogram given no cancer diagnosis (false-positive rate),† OR (95% CI)	Odds of having a positive mammogram given a cancer diagnosis (sensitivity),‡ OR (95% CI)	Odds of having a cancer diagnosis given a positive mammogram (PPV2),§ OR (95% CI)
Some competition	1.1 (0.70 to 1.78)	1.0 (0.54 to 1.73)	1.1 (0.79 to 1.52)
Moderate to extreme competition	1.0 (0.61 to 1.59)	0.8 (0.43 to 1.52)	1.0 (0.73 to 1.48)
Scheduling process			
Average wait time to schedule diagnostic mammogram			
≤3 d	1.0 (referent)	1.0 (referent)	1.0 (referent)
>3 d	1.4 (1.00 to 2.08)*	1.6 (0.93 to 2.69)	0.8 (0.59 to 1.06)
Do women wait for interpretation of diagnostic mammogram?			
No, never/yes, some of the time	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes, all of the time	1.2 (0.85 to 1.74)	1.7 (1.07 to 2.80)*	0.8 (0.64 to 1.08)
Interpretation and audit processes			
Interpretive process			
Are clinical breast examinations done routinely for women getting a screening mammogram?			
No	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes	2.3 (0.67 to 7.91)	0.5 (0.04 to 5.77)	2.0 (0.36 to 10.53)
What percentage of diagnostic mammograms are interpreted on-site?			
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
100	1.2 (0.65 to 2.31)	4.3 (1.66 to 11.10)*	1.1 (0.50 to 2.40)
How many radiologists interpret mammograms full time (ie, 40 h per week)?			
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
≥1	1.1 (0.66 to 1.87)	1.3 (0.60 to 2.64)	0.8 (0.49 to 1.20)
Are any diagnostic mammograms interpreted by more than one radiologist?			
No	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes, 1%–5% of the time	0.7 (0.48 to 1.10)	0.6 (0.33 to 1.26)	0.9 (0.58 to 1.52)
Yes, 10%–25% or less of the time	1.2 (0.85 to 1.81)	1.1 (0.65 to 1.84)	1.0 (0.71 to 1.33)
Yes, ≥80% of the time	1.2 (0.68 to 1.99)	0.8 (0.37 to 1.78)	0.9 (0.53 to 1.59)
Audit processes			
Individual performance data reported back to radiologists			
Once a year	1.0 (referent)	1.0 (referent)	1.0 (referent)
Twice or more per year	1.0 (0.72 to 1.42)	1.1 (0.74 to 1.72)	0.9 (0.71 to 1.16)
Unknown	1.1 (0.63 to 1.87)	1.0 (0.48 to 2.12)	0.8 (0.57 to 1.27)
Method of performance feedback information review			
Reviewed together in meeting	1.0 (referent)	1.0 (referent)	1.0 (referent)
Reviewed by facility/department manager or lead radiologist alone	0.7 (0.40 to 1.31)	1.4 (0.65 to 3.01)	1.1 (0.66 to 1.68)
Reviewed by each radiologist alone	1.1 (0.62 to 2.12)	1.9 (0.49 to 7.56)	0.9 (0.44 to 1.70)
Unknown	1.2 (0.73 to 1.99)	0.7 (0.40 to 1.24)	1 (0.73 to 1.42)

* Values are statistically significantly different, that is, $P < .05$ (two-sided F tests). PPV2 = positive predictive value of biopsy; OR = odds ratio; CI = confidence interval; BCSC = Breast Cancer Surveillance Consortium; FNA = fine-needle aspiration; CT = computed tomography; MRI = magnetic resonance imaging.

† False-positive rate is defined as the percentage of mammograms without a cancer diagnoses within 1 year that also have a BI-RADS assessment 4, 5, 0, or 3 with a recommendation for biopsy, surgical consultation, or FNA.

‡ Sensitivity is defined as the percentage of mammograms with a cancer diagnoses within 1 year that also have a BI-RADS assessment 4, 5, 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA.

§ PPV2 is defined as the percentage of mammograms with either a BI-RADS assessment 4, 5, 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA that result in a cancer diagnosis within 1 year.

|| Based on mammography registry data.

¶ Does not include ultrasound or ductography.

Table 5. Multivariable models adjusting for patient characteristics, radiologist characteristics, facility characteristics, and a facility random effect (2019 mammograms, 27 facilities)*

Characteristic	Odds of having a positive mammogram given no cancer diagnosis (false-positive rate),† OR (95% CI)	Odds of having a positive mammogram given a cancer diagnosis (sensitivity),‡ OR (95% CI)	Odds of having a cancer diagnosis given a positive mammogram (PPV2),§ OR (95% CI)
Facility structure and organization			
Clinical services			
Interventional services offered (FNA, core or vacuum-assisted biopsy, cyst aspirations, needle localization, or other procedures)			
No	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes	1.97 (0.94 to 4.10)	1.75 (0.47 to 6.45)	1.41 (0.50 to 4.00)
Facility is currently short staffed (ie, not enough radiologists)			
Strongly disagree/disagree/neutral	1.0 (referent)	1.0 (referent)	1.0 (referent)
Agree/strongly agree	1.21 (0.92 to 1.58)	1.36 (0.79 to 2.33)	0.83 (0.62 to 1.11)
Financial and malpractice			
Profit status			
Nonprofit	1.0 (referent)	1.0 (referent)	1.0 (referent)
For profit	0.97 (0.72 to 1.32)	0.91 (0.52 to 1.59)	0.87 (0.63 to 1.20)
How have medical malpractice concerns influenced recommendations of diagnostic mammograms, ultrasounds, or breast biopsies at your facility following screening mammograms?			
Not changed	1.0 (referent)	1.0 (referent)	1.0 (referent)
Moderately increased/greatly increased	1.48 (1.09 to 2.01)	1.74 (0.94 to 3.23)	0.81 (0.57 to 1.15)
Scheduling process			
Average wait time to schedule diagnostic mammogram			
≤3 d	1.0 (referent)	1.0 (referent)	1.0 (referent)
>3 d	1.09 (0.76 to 1.57)	1.17 (0.57 to 2.41)	0.99 (0.66 to 1.47)
Do women wait for interpretation of diagnostic mammogram?			
No, never/yes, some of the time	1.0 (referent)	1.0 (referent)	1.0 (referent)
Yes, all of the time	1.02 (0.72 to 1.45)	1.17 (0.58 to 2.35)	0.88 (0.62 to 1.26)
Interpretation and audit processes			
Interpretive process			
What percentage of diagnostic mammograms are interpreted on-site?			
0	1.0 (referent)	1.0 (referent)	1.0 (referent)
100	0.74 (0.30 to 1.85)	3.12 (0.54 to 18.09)	0.74 (0.19 to 2.86)

* PPV2 = positive predictive value of biopsy; OR = odds ratio; CI = confidence interval; FNA = fine-needle aspiration.

† False-positive rate is defined as the percentage of mammograms without a cancer diagnoses within 1 year that also have a BI-RADS assessment 4, 5, 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA.

‡ Sensitivity is defined as the percentage of mammograms with a cancer diagnoses within 1 year that also have a BI-RADS assessment 4, 5, 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA.

§ PPV2 is defined as the percentage of mammograms with either a BI-RADS assessment 4, 5, 0 or 3 with a recommendation for biopsy, surgical consultation, or FNA that result in a cancer diagnosis within 1 year.

|| Value is statistically significant, $P = .01$ (two-sided F test).

examinations abnormal at facilities where more interventional services are readily available.

Malpractice concerns and the availability of interventional services were associated with high false-positive rates and sensitivities, without changing overall accuracy. There is much speculation that malpractice and the ready access to technology in the United States drive the rapidly increasing costs of medical care in this

country, potentially even resulting in poorer medical outcomes (24). Our study quantitatively demonstrates associations of diagnostic performance with malpractice concerns and access to interventional services.

Three of the 32 facilities sent their diagnostic mammograms to other locations for interpretation. These facilities had a much lower sensitivity than facilities where examinations were

interpreted on-site. It is possible that these three facilities were in remote locations and/or had limited staffing and thus required off-site interpretation. When facilities send examinations out for interpretation, the interpreting radiologist may have less potential for direct patient contact and could have less feedback about the outcome of positive examinations. Although these results are based on very small numbers, the accuracy of off-site interpretation of diagnostic mammograms merits further study.

A strength of this multisite study of diagnostic mammography facilities is that it represents diverse geographic regions of the United States. Another is the ability to adjust for important patient and radiologist characteristics and link to mammography performance data using BCSC information. Both facility and radiologist surveys had high response rates and assessed a broad range of characteristics.

The study also has several limitations. Concern about malpractice is a characteristic that is closely connected to the radiologist as the party with the most personal risk; however, in our study, it was assessed as a facility trait. Because most of the respondents to this survey were technologists who might have less personal fear of malpractice lawsuits than radiologists, the finding that malpractice concerns are associated with performance suggests that perhaps within facilities, there is a culture that influences radiologist's decision making. More work, specifically assessing diagnostic interpretation and radiologist-level malpractice concerns, may help clarify these findings. Another limitation involves defining diagnostic mammography. We limited our diagnostic mammograms to examinations that were designated by the radiologist as performed for evaluation of a breast problem. However, radiologist reporting of this designation includes two groups: women with a palpable lump and women with no or unknown lump status. The latter group includes a wide variety of indications—from women with breast pain and nipple discharge to women with a palpable lump that was not self-reported. It is possible that designation of diagnostic mammograms could vary between facilities (1,17). Last, our results may not be generalizable to other regions of the United States or to other countries where mammography programs may have different screening guidelines and different systems and requirements for interpreting diagnostic mammograms.

In summary, false-positive rates vary statistically significantly between facilities performing diagnostic mammography and are higher at facilities where the concern about malpractice is believed to increase recommendations for diagnostic evaluations at the facility. Analyses comparing differences among mammography facilities that do not adjust for important patient characteristics may falsely conclude that there is more facility variation in overall accuracy than what actually exists.

References

- Sickles EA, Miglioretti DL, Ballard-Barbash RD, et al. Performance benchmarks for diagnostic mammography. *Radiology*. 2005;235(3):775–790.
- Taplin S, Abraham L, Barlow WE, et al. Mammography facility characteristics associated with interpretive accuracy of screening mammography. *J Natl Cancer Inst*. 2008;100(12):876–887.
- Barlow WE, Lehman CD, Zheng Y, et al. Performance of diagnostic mammography in women with breast signs or symptoms. *J Natl Cancer Inst*. 2002;94(15):1151–1159.
- Miglioretti DL, Smith-Bindman R, Abraham L, et al. Radiologist characteristics associated with interpretive performance of diagnostic mammography. *J Natl Cancer Inst*. 2007;99(24):1854–1863.
- Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. *Radiology*. 2002;224(3):861–869.
- Jensen A, Vejborg I, Severinsen N, et al. Performance of clinical mammography: a nationwide study from Denmark. *Int J Cancer*. 2006;119(1):183–191.
- Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The influence of clinical information on the accuracy of diagnostic mammography. *Breast Cancer Res Treat*. 2004;85(3):223–228.
- Houssami N, Irwig L, Simpson JM, McKessar M, Blome S, Noakes J. The contribution of work-up or additional views to the accuracy of diagnostic mammography. *Breast*. 2003;12(4):270–275.
- Burnside ES, Sickles EA, Sohlich RE, Dee KE. Differential value of comparison with previous examinations in diagnostic versus screening mammography. *AJR Am J Roentgenol*. 2002;179(5):1173–1177.
- Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol*. 1997;169(4):1001–1008.
- Fishbein M. Factors influencing health behaviors: an analysis based on a theory of reasoned action. In: Landry F, ed. *Health Risk Estimation, Risk Reduction and Health Promotion*. Ottawa, ON, Canada: Canadian Public Health Association; 1983:203–214.
- Green LW, Kreuter MW. Health promotion planning: an educational and environmental approach. In: *Health Promotion Today and a Framework for Planning*. Mountain View, CA: Mayfield Publishing Co; 1991:24.
- Elmore JG, Carney PA, Taplin S, D'Orsi CJ, Cutter G, Hendrick E. Does litigation influence medical practice? The influence of community radiologists, medical malpractice perceptions and experience on screening mammography. *Radiology*. 2005;236(1):37–46.
- Taplin SH, Ichikawa L, Buist DS, Seger D, White E. Evaluating organized breast cancer screening implementation: the prevention of late-stage disease? *Cancer Epidemiol Biomarkers Prev*. 2004;13(2):225–234.
- Poplack SP, Tosteson AN, Grove M, Wells WA, Carney PA. The practice of mammography in 53,803 women from the New Hampshire Mammography Network. *Radiology*. 2000;217(3):832–840.
- Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst*. 2002;94(20):1546–1554.
- Dee KE, Sickles EA. Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently. *AJR Am J Roentgenol*. 2001;176(3):729–733.
- Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology*. 2006;241(1):55–66.
- Barlow WE, Chi C, Carney PA, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst*. 2004;96(24):1840–1850.
- D'Orsi C, Bassett L, Berg W, et al. *Breast Imaging Reporting and Data System—Mammography*. 4th ed. Reston, VA: American College of Radiology; 2003.
- Green J, Wintfeld N. Report cards on cardiac surgeons. Assessing New York State's approach. *N Engl J Med*. 1995;332(18):1229–1232.
- Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA*. 2003;290(16):2129–2137.
- Elmore JG, Nakano CY, Koepsell TD, Desnick LM, D'Orsi CJ, Ransohoff DF. International variation in screening mammography interpretations in community-based programs. *J Natl Cancer Inst*. 2003;95(18):1384–1393.
- Deyo RA, Patrick DL. *Hope or Hype: The Obsession With Medical Advances and the High Cost of False Promises*. 1st ed. New York, NY: AMACOM; 2005.

Funding

Agency for Healthcare Research and Quality (public health service grant R01 HS-010591 to J.G.E.) and National Cancer Institute (NCI) (grants R01 CA-107623 and K05 CA-104699 to J.G.E.). NCI-funded Breast Cancer Surveillance Consortium cooperative agreement (grants U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, and U01CA70040 for data collection). Dr Jackson is supported by Dr Elmore's R01 grant. The collection of cancer incidence data used in this study was supported in part by several state public health departments and cancer

registries throughout the United States. For a full description of these sources, please see <http://breastscreening.cancer.gov/work/acknowledgement.html>.

Notes

The sponsors had no role in the study design, the data collection and analysis, the interpretation of the results, the preparation of the manuscript, or the decision to submit the manuscript for publication.

We thank the Breast Cancer Surveillance Consortium (BCSC) investigators, participating mammography facilities, radiologists, and women undergoing mammography for the data they have provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at <http://breastscreening.cancer.gov/>.

Manuscript received October 27, 2008; revised February 27, 2009; accepted March 31, 2009.