Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy[1]

Joann G. Elmore, MD, MPH
Sara L. Jackson, MD, MPH
Linn Abraham, MS
Diana L. Miglioretti, PhD
Patricia A. Carney, PhD
Berta M. Geller, EdD
Bonnie C. Yankaskas, PhD
Karla Kerlikowske, MD
Tracy Onega, PhD
Robert D. Rosenberg, MD
Edward A. Sickles, MD
Diana S. M. Buist, PhD

Purpose: To identify radiologists' characteristics associated with interpretive performance in screening mammography.

Materials and Methods: The study was approved by institutional review boards of University of Washington (Seattle, Wash) and institutions at seven Breast Cancer Surveillance Consortium sites, informed consent was obtained, and procedures were HIPAA compliant. Radiologists who interpreted mammograms in seven U.S. regions completed a self-administered mailed survey; information on demographics, practice type, and experience in and perceptions of general radiology and breast imaging was collected. Survey data were linked to data on screening mammograms the radiologists interpreted between January 1, 1998, and December 31, 2005, and included patient risk factors, Breast Imaging Reporting and Data System assessment, and follow-up breast cancer data. The survey was returned by 71% (257 of 364) of radiologists; in 56% (205 of 364) of the eligible radiologists, complete data on screening mammograms during the study period were provided; these data were used in the final analysis. An evaluation of whether the radiologists' characteristics were associated with recall rate, false-positive rate, sensitivity, or positive predictive value of recall (PPV_1) of the screening examinations was performed with logistic regression models that were adjusted for patients' characteristics and radiologist-specific random effects.

Results: Study radiologists interpreted 1 036 155 screening mammograms; 4961 breast cancers were detected. Median percentages and interquartile ranges, respectively, were as follows: recall rate, 9.3% and 6.3%–13.2%; false-positive rate, 8.9% and 5.9%–12.8%; sensitivity, 83.8% and 74.5%–92.3%; and PPV_1, 4.0% and 2.6%–5.9%. Wide variability in sensitivity was noted, even among radiologists with similar false-positive rates. In adjusted regression models, female radiologists or fellowship-trained radiologists had significantly higher recall and false-positive rates ($P < .05$, all). Fellowship training in breast imaging was the only characteristic significantly associated with improved sensitivity (odds ratio, 2.32; 95% confidence interval: 1.42, 3.80; $P < .001$) and the overall accuracy parameter (odds ratio, 1.61; 95% confidence interval: 1.05, 2.45; $P = .028$).

Conclusion: Fellowship training in breast imaging may lead to improved cancer detection, but it is associated with higher false-positive rates.

[1] From the Department of Medicine, University of Washington School of Medicine, Harborview Medical Center, 325 Ninth Ave, Box 359780, Seattle, WA 98104-2499 (J.G.E., S.L.J.), Department of Biostatistics, University of Washington School of Public Health, Seattle, Wash (D.L.M.); Group Health Research Institute, Seattle, Wash (L.A., D.L.M., D.S.M.B.); Department of Family Medicine, Oregon Health and Science University, Portland, Ore (P.A.C.); Department of Family Medicine and Radiology, University of Vermont, Burlington, Vt (B.M.G.); Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC (B.C.Y.); Departments of Medicine and Epidemiology and Biostatistics (K.K.), and Radiology (E.A.S.), University of California San Francisco, San Francisco, Calif; Department of Family Medicine, Dartmouth Medical School, Hanover, NH (T.O.); and Department of Radiology, University of New Mexico, Albuquerque, NM (R.D.R.). Received February 9, 2009; revision requested March 10; final revision received April 21; accepted May 8; final version accepted July 2. Supported by the National Cancer Institute (grants R01 CA107623, 1K05 CA104699), the Breast Cancer Surveillance Consortium (grants U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040), the Agency for Healthcare Research and Quality (grant 1R01 CA107623), the Breast Cancer Stamp Fund, and the American Cancer Society, made possible by a donation from the Longaberger Company's Horizon of Hope Campaign (grants SIRGS-07-271-01, SIRGS-07-272-01, SIRGS-07-273-01, SIRGS-07-274-01, SIRGS-07-275-01, SIRGS-06-281-01). **Address correspondence to** J.G.E. (e-mail: *jelmore@u.washington.edu*).

Mammography is the only breast cancer screening test shown in clinical trials to be associated with reduced breast cancer mortality (1,2). However, it is not a perfect test. As in other areas of medicine, variability among radiologists in the interpretive accuracy of mammograms is extensive (3–9). For example, the percentage of images from U.S. mammographic examinations interpreted as abnormal ranges from 0.6% for one radiologist to 28.7% for another (10), and the sensitivity (the percentage of cancers with abnormal findings from a screening examination) ranges from 31.6% to 96.2% (11). Ideally, screening performance should have a high sensitivity, increasing the likelihood that cancers are detected, and a low false-positive rate, reducing the monetary costs and adverse events associated with additional work-up among women without disease (12,13).

Studies of the effect of radiologists' training and clinical experience on interpretive performance have generated conflicting results, even when overlapping populations of radiologists were studied (6,14). Uncertainty about the effect of radiologists' characteristics is caused, in part, by the often limited number of radiologists and mammographic examinations in published studies and the varying measures used to define such characteristics as clinical experience. Radiologists' characteristics that have been studied in the United States and international cohorts include number of years of interpreting mammograms, receipt of specialized fellowship training in breast imaging, academic affiliation, and various measures of mammographic interpretive volume (6,14).

The purpose of our study was to identify radiologists' characteristics associated with interpretive performance in screening mammography.

Advances in Knowledge

- In adjusted regression models, radiologists who were women or who had fellowship training had higher false-positive rates.
- Wide variability in sensitivity was noted, even among radiologists with similar false-positive rates.
- The only characteristic significantly associated with improved sensitivity and the overall accuracy parameter was fellowship training in breast imaging.
- The false-positive rates noted among the fellowship-trained radiologists were higher than the desirable goals recommended for performance of U.S. radiologists.

Implications for Patient Care

- The highest sensitivity and overall accuracy parameter for screening mammography were noted when radiologists with fellowship training in breast imaging interpreted the images from mammographic examinations.
- Patients should be advised that these fellowship-trained radiologists also have higher false-positive rates.
- Although fellowship-trained radiologists may detect more cancers than non–fellowship-trained radiologists, they also may call back an additional 83 women for false-positive results for every additional breast cancer detected.

Materials and Methods

The authors had full responsibility in the design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

Overview and Institutional Review Board Approval

All radiologists who interpreted screening mammograms in 2005–2006 at the seven Breast Cancer Surveillance Consortium (BCSC) sites were invited to complete a self-administered mailed survey. These sites represent distinct patient populations and geographic regions in seven U.S. states (California, Colorado, North Carolina, New Mexico, New Hampshire, Vermont, Washington) (15,16). A previous, shorter survey was distributed to radiologists ($n = 139$) in just three BCSC sites 5 years earlier (6,17). The BCSC maintains large regional databases on mammographic examinations and patients' characteristics, and these databases are linked to Surveillance Epidemiology and End Results tumor registries, pathology databases, or both to ensure data completeness for breast cancer occurrence (16). Details of data collection by the BCSC have been reported previously (5,18–22). We linked survey results from participating radiologists to BCSC data on the screening mammograms that they interpreted.

The study was approved by the institutional review boards of the University of Washington (Seattle, Wash) and the institutions at all seven BCSC sites. All procedures were Health Insurance Portability and Accountability Act compliant, and all sites and the Statistical Coordinating Center (Seattle, Wash) received a Federal Certificate of Confidentiality and other protection for the identities of women and physicians who are subjects of this research and the facilities involved.

Radiologist Survey

The survey was developed by a diverse research group that included experts in breast imaging, clinical medicine,

Published online before print
10.1148/radiol.2533082308

Radiology 2009; 253:641–651

Abbreviations:
BCSC = Breast Cancer Surveillance Consortium
BI-RADS = Breast Imaging Reporting and Data System
PPV_1 = positive predictive value of recall

Author contributions:
Guarantors of integrity of entire study, J.G.E., P.A.C., K.K., T.O.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, J.G.E., S.L.J., P.A.C., B.M.G., T.O., E.A.S.; clinical studies, P.A.C., B.C.Y., K.K., T.O.; statistical analysis, S.L.J., L.A., D.L.M., D.S.M.B.; and manuscript editing, all authors

Funding:
This research was supported by the National Cancer Institute (grants 1R01 CA107623, 1K05 CA104699) and the Breast Cancer Surveillance Consortium (grants U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040).

See Materials and Methods for pertinent disclosures.

See also the article by Miglioretti et al and the editorial by Kopans in this issue.

health services research, biostatistics, epidemiology, behavioral sciences, and educational psychology. The survey was pilot tested among community radiologists working in breast imaging who were not associated with the BCSC. The survey was 10 pages long and required 10–15 minutes to complete. A copy of the survey is available on the BCSC Web site (23).

Of 364 eligible radiologists contacted, 257 (71%) responded to the survey and gave consent for linkage to performance measures collected by the BCSC. Of those who responded to the survey, we excluded 26 who did not have complete BCSC data on the interpretation of the screening mammograms and the patient outcomes associated with them during the study period and 26 in whom information was missing on patient breast density, as the density information was not forwarded to the BCSC registry from the facilities where they worked. Thus, our final population of radiologists was 205 (56%) of 364 who were eligible for inclusion.

Survey data included demographics (age, sex), practice type (full vs part time, primary or adjunct affiliation with an academic medical center), general clinical experience (years since graduation from residency), and breast imaging experience (fellowship training in breast imaging, years of mammographic interpretation, percentage of time spent in breast imaging and number of hours working in breast imaging per week, and self-reporting of annual interpretive volume for images from screening and diagnostic examinations).

Surveys were mailed to radiologists at different times (depending on the site) between January 1, 2006, and September 30, 2007, depending on each site's funding mechanism and institutional review board status. Survey mailing and collection were handled by individual BCSC sites to maintain radiologists' confidentiality. Study managers and principal investigators at each site attempted to contact radiologists by mail and/or telephone at least three times to maximize local study participation. Radiologists were informed that their survey responses would be linked to their actual BCSC interpretive performance data by an encrypted linkage variable and that their data would remain anonymous.

Incentives to complete the survey varied among the seven sites and included bookstore gift cards worth $25–$50 for radiologists (seven sites) and Breast Imaging Reporting and Data System (BI-RADS) manuals (12) for participating facilities (four sites).

Once each BCSC site obtained completed surveys with informed consent, the data were double entered and discrepancies were corrected. Encrypted data were sent to the BCSC Statistical Coordinating Center, where the survey data were linked to BCSC data on mammographic examinations.

Patient and Mammographic Outcome Data

The information obtained in each mammographic examination included the BI-RADS screening assessment results, recommendations for additional evaluations, and breast cancer diagnosis and outcomes. In addition, information was obtained on factors known to be associated with interpretive performance, including the patient's age (18), BI-RADS mammographic breast density (18), indication for mammography (screening vs diagnostic examination) (15), and time since last mammographic examination (24).

Included in the analysis were BCSC screening mammograms obtained in women aged 40 years or older interpreted by a participating radiologist at a BCSC site between January 1, 1998, and December 31, 2005. A screening mammogram was defined as a bilateral mammogram designated as a screening mammogram obtained in women without a history of breast cancer or breast augmentation (25). The mammogram had to be obtained at least 9 months after any other breast imaging examination to prevent misclassification of diagnostic examinations as screening examinations. Mammograms obtained in women with missing information about time since last mammographic examination was performed (n = 48 426), breast density (n = 26 431), or cancer outcome status (n = 1) were excluded. Our final study analysis included 1 036 155 screening mammograms obtained in 531 705 women. Radiologists self-reported their estimated number of annual mammograms that they interpreted in the radiologists' survey; however, the mammographic data reported here are the numbers of mammograms recorded in the BCSC database. Some radiologists had low numbers of mammograms recorded in the BCSC because they had practices outside of the consortium in addition to their practices at BCSC sites.

Definitions of Breast Cancer Cases and Interpretive Performance

Breast cancer cases included cases in patients with a diagnosis of either invasive carcinoma or ductal carcinoma in situ within 1 year of the screening mammographic examination and with a diagnosis date prior to the date of their next screening mammographic examination (26). Each screening mammogram was classified in one of six BI-RADS assessment categories by the interpreting radiologist (27). We defined mammograms that were classified in BI-RADS categories 1, 2, and 3 with no immediate follow-up recommended as mammograms with negative interpretations and mammograms that were classified in all other BI-RADS categories as mammograms with positive interpretations (28).

Performance measures included recall rate, sensitivity, false-positive rate, positive predictive value of recall (PPV_1), and cancer detection rate. Recall rate was defined as the percentage of screening examinations with findings interpreted as positive. Sensitivity was defined as the percentage of screening examinations with findings interpreted as positive among all women who received a diagnosis of breast cancer within the 1-year follow-up period. False-positive rate was defined as the percentage of screening examinations with findings interpreted as positive among all patients who did not receive a diagnosis of breast cancer within the follow-up period. PPV_1 was defined as the probability of cancer, given a screening examination with abnormal findings suggestive of cancer. Cancer detection

rate was defined as the number of true-positive mammograms for every 1000 screening mammograms.

Performance measures for the 205 radiologists in this study with screening mammographic data were compared with performance measures for the 951 radiologists with BCSC screening mammographic data at any time between January 1, 1998, to December 31, 2005: Recall rate was 9.1% and 9.6%, sensitivity was 82.5% and 82.0%, false-positive rate was 8.8 and 9.2%, and PPV_1 was 4.3% and 4.2%, respectively. Patients' characteristics (age, breast density, time since last mammographic examination), BI-RADS assessment distributions, and associated cancer detection rates noted in this study were similar to data overall from the BCSC.

Statistical Analysis

We examined frequency distributions of each of the following self-reported characteristics of the radiologists: sex, primary affiliation with an academic medical center, fellowship training in breast imaging, years of mammographic interpretation, percentage of time spent in breast imaging, hours working in breast imaging per week, average number of mammograms interpreted per year, and the percentage of images from all examinations interpreted that were screening mammograms (interpretive volumes were reported as averages per year over the past 5 years). We also looked at cross-tabulations between some of these characteristics to understand their associations with one another. Unadjusted medians and interquartile ranges for recall rate, sensitivity, false-positive rate, and PPV_1 were computed for each of the radiologists' characteristics. Sensitivity was plotted in comparison with false-positive rate to display the distribution and variability of these performance measures among radiologists, with the receiver operating characteristic curve overlaid. Recall rate was plotted in comparison with PPV_1 for each radiologist, with contour lines indicating the corresponding cancer detection rates.

Logistic regression models were fit to examine the association between each performance measure and each of the radiologists' characteristics, adjusting for the patients' characteristics (BCSC registry site, age at mammographic examination, breast density, and time since last mammographic examination) and other radiologists' characteristics. For recall rate, we modeled the probability of a positive mammogram and included a radiologist-specific random effect to account for correlation among mammograms interpreted by the same radiologist. For PPV_1, we modeled the probability of cancer, given a positive mammogram, and included a single radiologist-specific random effect.

To assess overall accuracy (differential changes in sensitivity and false-positive rate not caused by a threshold level effect), we jointly modeled sensitivity and false-positive rate and included separate radiologists' random effects for each performance measure. We defined the overall accuracy parameter as the coefficient corresponding to the interaction between cancer status and the radiologists' characteristic under study. The accuracy parameter is positive for a given characteristic if either sensitivity is increased more than the false-positive rate or if the false-positive rate is decreased more than the sensitivity for radiologists with that characteristic. A more detailed description of this approach for jointly modeling sensitivity and false-positive rate can be found in a previous article (29). This joint model is equivalent to the model proposed by Rutter and Gatsonis (30), with their scale parameter set to one. Separate logistic regression models (which included mammogram-level characteristics) were fit for each of the radiologists' characteristics prior to fitting the final multivariable models, which included all radiologists' characteristics. We first tested whether the group radiologists' characteristics significantly improved model fit before examining the significance of each characteristic individually ($P < .05$). We did not include the percentage of time spent in breast imaging in the multivariable models, as this variable might not be reflective of the amount of time spent working in breast imaging for radiologists working part time, and we included a similar variable in the model (hours working in breast imaging per week). Analyses were also performed, with restriction of the mammographic data to the period January 1, 2002, to December 31, 2005, a period close to that for the survey of the radiologists.

We estimated the effect of fellowship training on cancer detection and additional work-ups for a hypothetical population of 100 000 women, with the conditions noted in our study. Population estimates were calculated on the basis of the observed unadjusted false-positive rates and sensitivity values of fellowship-trained and non–fellowship-trained radiologists. The background cancer incidence of 4.8 women with breast cancer per 1000 examinations was used for both fellowship-trained and non–fellowship-trained radiologists.

Statistical analyses were performed by using software (SAS, version 9.1; SAS Institute, Cary, NC). A difference with $P < .05$ was considered significant. For the multiple regression analyses, a difference with $P < .016$ was considered significant if a Bonferroni multiple-comparisons adjustment was applied to adjust for fitting three regression models.

Results

The 205 radiologists in our study cohort interpreted a total of 1 036 155 screening mammograms in 531 705 women; 4961 women had breast cancer. The median number of screening mammographic examinations interpreted per radiologist was 3131 (range, 1–43 119), including a median of 13 (range, 0–265) examinations with mammograms interpreted in women who received a diagnosis of breast cancer. Screening mammograms were obtained at 111 facilities in seven U.S. states.

Table 1 presents radiologists' characteristics and the cross-tabulation of all characteristics with radiologists' sex, fellowship training in breast imaging, years of clinical experience interpreting mammograms, and percentage of time spent in breast imaging. Data on radiologists' age and years since graduation from radiology residency are not shown, as these were highly correlated with the number of years of mammographic interpretation

(Pearson correlation coefficients were 0.75 and 0.81, respectively). Most radiologists were men (71%), had no affiliation with an academic medical center (82%), and reported no fellowship training in breast imaging (92%). Thirty-nine percent reported interpreting mammograms for 20 years or longer, and almost half spent 40% or more of their time in breast imaging. Fifty-six percent reported interpreting more than 2000 mammograms per year, whereas 19% interpreted more

Table 1

Characteristics of 205 Radiologists Who Interpreted Screening Mammograms in Seven U.S. States

Characteristic	Total No.*	Percentage of Radiologists with Characteristic								
		Sex		Fellowship Training		Years of Experience			Time Working in Breast Imaging	
		Male ($n = 146$)	Female ($n = 59$)	No ($n = 189$)	Yes ($n = 16$)	<10 ($n = 50$)	10–19 ($n = 74$)	≥20 ($n = 80$)	<40% ($n = 105$)	≥40% ($n = 93$)
Sex										
Male	146 (71)	. . .	. . .	75†	31†	52†	70†	84†	83†	58†
Female	59 (29)	. . .	. . .	25†	69†	48†	30†	16†	17†	42†
Primary affiliation with academic medical center										
No	167 (82)	84†	77†	85†	53†	84	85	78	88	76
Yes, adjunct	14 (7)	9†	2†	6†	13†	2	7	10	6	8
Yes, primary	22 (11)	7†	21†	9†	33†	14	8	11	7	16
Breast imaging experience										
Fellowship training										
No‡	189 (92)	97†	81†	. . .	. . .	84†	91†	99†	99†	84†
Yes	16 (8)	3†	19†	. . .	. . .	16†	9†	1†	1†	16†
Years of mammographic interpretation										
<10	50 (25)	18†	41†	22†	50†	. . .	. . .	. . .	25	25
10–19	74 (36)	36†	37†	36†	44†	. . .	. . .	. . .	38	33
≥20	80 (39)	46†	22†	42†	6†	. . .	. . .	. . .	37	42
Percentage of time spent in breast imaging										
<20	51 (26)	33†	7†	27†	6†	14	30	30	49	. . .
20–39	54 (27)	28†	25†	30†	0†	39	27	19	51	. . .
40–79	33 (17)	11†	30†	16†	19†	18	14	18	. . .	35
80–100	60 (30)	27†	39†	26†	75†	29	30	32	. . .	65
Hours working in breast imaging per week										
0–8	47 (24)	31†	5†	26†	6†	11	24	32	45†	1†
>8 to 16	69 (35)	34†	40†	39†	0†	47	37	28	51†	17†
>16 to 32	26 (13)	11†	18†	11†	38†	15	11	14	4†	24†
>32	53 (27)	24†	36†	25†	56†	28	28	26	0†	58†
Volume										
Self-reported average no. of mammograms interpreted per year over the past 5 years										
≤1000	20 (10)	10	11	11†	0†	11	8	12	9†	12†
1001–2000	65 (34)	38	23	36†	7†	39	28	35	42†	23†
>2000	108 (56)	51	67	53†	93†	50	63	53	48†	65†
Percentage of images from all examinations interpreted that were screening mammograms§										
<83	83 (43)	39	53	40†	73†	41	37	51	31†	57†
≥83	110 (57)	61	47	60†	27†	59	63	49	69†	43†

Note.—The number of radiologists with data that were missing for specific characteristics were as follows: two radiologists, primary affiliation with an academic medical center; one radiologist, years of mammographic interpretation; seven radiologists, percentage of time spent in breast imaging; 10 radiologists, hours working in breast imaging per week; 12 radiologists, self-reported average number of mammograms interpreted per year; and 12 radiologists, percentage of images from all examinations interpreted that were screening mammograms.

* Numbers in parentheses are percentages that are based on the numbers of radiologists who responded to the survey question.

† Variable with significant association ($P < .05$).

‡ Unknowns were included in this category.

§ Average per year over past 5 years; ratio of five or more screening examinations to one diagnostic examination.

than 5000 mammograms per year (data not shown). For most of the radiologists, 83% or more of the images from mammographic examinations they interpreted were screening mammograms; 17% or less were diagnostic mammograms.

Radiologists with fellowship training in breast imaging were significantly more likely to be women, to be affiliated with academic centers, to have fewer than 10 years of experience interpreting mammograms, to spend 80% or more of their time in breast imaging, and to spend more than 32 hours per week working in breast imaging than those without specialized training ($P < .05$). Radiologists with fellowship training also were more likely to report annual interpretive volume of more than 2000 mammograms and to report that less than 83% of images from all examinations they interpreted were screening mammograms.

Thirty percent of radiologists reported interpreting images from mammographic examinations at facilities outside the BCSC, although this percentage varied widely across the seven sites. For self-reported volume measures, 30% of radiologists stated that their volume estimates were a "guess," 42% stated that their estimates were made with confidence, and 24% provided actual volume information obtained from their own audit reports.

The percentage of radiologists who reported using computer-aided detection when they interpreted screening mammograms was 77.2% for non–fellowship-trained and 75% for fellowship-trained radiologists. The non–fellowship-trained and fellowship-trained radiologists reported applying computer-aided detection to a similar mean percentage of screening examinations (84.6% and 85.4%, respectively).

Interpretive performance varied widely with the median and interquartile ranges for performance measures as follows: recall rate, 9.3% and 6.3%–13.2%; false-positive rate, 8.9% and 5.9%–12.8%; sensitivity, 83.8% and 74.5%–92.3%; and PPV_1, 4.0% and 2.6%–5.9%, respectively. Figure 1, which is based on data from 187 radiologists who interpreted one or more mammograms associated with a cancer diagnosis, shows that sensitivity ranged from 0% to 100%, whereas false-positive rates ranged from 1.7% to 24.7%. In Figure 1, the median number of screening mammographic examinations associated with breast cancer per radiologist was 16 (range, 1–265). Among the 119 radiologists who interpreted 10 or more mammograms associated with breast cancer, the median sensitivity was 82.8% (range, 40%–100%; interquartile range, 76.5%–88.2%). We found wide variability in sensitivity even among radiologists with similar false-positive rates. By assuming a constant accuracy among radiologists, and varying the threshold value for recall, the normalized partial area under the curve was 0.82, which corresponds to the average sensitivity over the range of observed false-positive rates.

As shown in Figure 1, 18 radiologists had both sensitivity and false-positive rates that were in the highest quartile of interpretive performance in the United States, on the basis of BCSC benchmarks for screening performance (11) (sensitivity, ≥86.4%; false-positive rate, ≤7.5%). Six radiologists had both sensitivity and false-positive rates in the lower 25th percentile (sensitivity, ≤75.7%; false-positive rate, ≥14%).

Higher recall and false-positive rates were noted among female radiologists, those with fellowship training in breast imaging, and those with fewer than 10 years of mammographic interpretation (Table 2). Female radiologists had higher

Figure 1

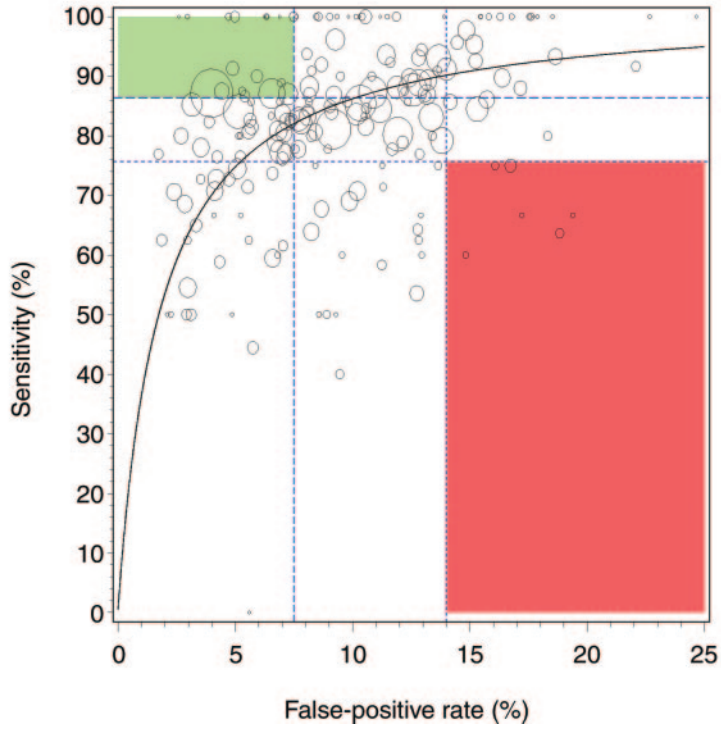

Figure 1: Performance of 187 U.S. radiologists who interpreted images from screening mammographic examinations (with images from one or more examinations associated with a cancer diagnosis). Sensitivity and false-positive rate are shown for each radiologist. Specificity was calculated by subtracting false-positive rate from one. Size of circle represents number of screening mammograms interpreted by that radiologist that were associated with a cancer diagnosis, with larger circles representing more cancers. Small circle may represent a radiologist with screening mammograms associated with one or two cancers. Normalized partial area under the curve is 0.82. Two areas are in color to highlight radiologists with both sensitivity and false-positive rate in highest (≥75th percentile in green) and lowest (≤25th percentile in red) rating for interpretive performance on basis of national BCSC benchmarks for screening performance (11).

Table 2

Characteristics of U.S. Radiologists Who Interpreted Screening Mammograms during 1998–2005 and Their Interpretive Performance

Performance	Recall Rate	False-Positive Rate	Sensitivity	PPV_1
Overall	9.3 (6.3–13.2)	8.9 (5.9–12.8)	83.8 (74.5–92.3)	4.0 (2.6–5.9)
Stratified according to radiologists' characteristics				
Sex				
Male	8.4 (5.6–12.5)	8.1 (5.3–12.1)	82.4 (70.7–91.3)	4.1 (2.6–6.1)
Female	11.4 (8.5–13.5)	11.2 (8.2–13.1)	87.0 (77.4–94.6)	3.9 (2.7–5.4)
Primary affiliation with academic medical center				
No	9.5 (6.3–13.5)	9.1 (6.0–13.2)	84.5 (75.0–93.0)	4.0 (2.7–5.8)
Yes, adjunct	7.4 (4.8–9.3)	7.1 (4.8–9.2)	81.7 (69.2–85.6)	4.5 (3.4–9.1)
Yes, primary	7.6 (6.3–13.0)	7.2 (6.3–12.6)	81.7 (77.5–89.3)	3.5 (2.6–5.1)
Breast imaging experience				
Fellowship training				
No	9.1 (6.0–13.1)	8.6 (5.7–12.8)	83.3 (72.7–91.3)	3.9 (2.6–5.8)
Yes	11.6 (9.3–14.7)	11.0 (8.8–14.3)	87.7 (80.9–100.0)	5.2 (3.1–7.4)
Years of mammographic interpretation				
<10	13.1 (7.8–16.6)	12.8 (7.4–16.3)	85.2 (75–100.0)	3.7 (2.3–5.5)
10–19	9.8 (7.4–13.0)	9.6 (6.9–12.6)	85.5 (77.3–91.7)	4.2 (3.1–5.6)
≥20	7.4 (4.9–10.7)	7.1 (4.8–10.2)	82.4 (72.7–88.9)	4.2 (2.7–6.0)
Percentage of time spent in breast imaging				
<20	9.7 (6.3–14.9)	9.6 (5.6–14.5)	85.7 (75.0–91.2)	3.7 (2.3–5.0)
20–39	8.6 (5.9–13.2)	8.2 (5.7–12.8)	82.3 (72.7–90.9)	4.4 (3.4–6.4)
40–79	8.6 (6.8–12.4)	8.4 (6.6–11.9)	82.6 (76.2, 90.0)	3.9 (2.6–5.1)
80–100	9.5 (6.7–12.9)	9.4 (6.4–12.6)	84.8 (71.4–100.0)	3.8 (2.7–6.5)
Hours working in breast imaging per week				
0–8	9.7 (6.3–14.5)	9.6 (5.3–14.0)	85.9 (76.5–93.3)	3.6 (2.3–5.5)
>8 to 16	9.0 (6.3–12.6)	8.4 (6.2–12.3)	84.0 (76.2–88.2)	4.3 (3.2–5.7)
>16 to 32	7.2 (5.9–13.8)	7.1 (5.6–13.4)	80.8 (66.7–93.0)	3.8 (2.6–5.1)
>32	9.9 (7.4–12.9)	9.5 (7.0–12.6)	83.9 (71.4–100.0)	3.8 (2.6–6.5)
Volume				
Self-reported average no. of mammograms interpreted per year over the past 5 years				
≤1000	9.7 (5.4–12.7)	9.2 (5.0–12.4)	80.0 (71.4–100.0)	3.9 (2.4–6.3)
1001–2000	9.6 (7.4–14.9)	9.3 (7.0–14.5)	85.7 (77.8–96.2)	4.0 (2.5–6.0)
>2000	9.2 (6.2–12.6)	9.0 (5.7–12.3)	83.9 (70.6–90.9)	3.9 (2.7–5.7)
Percentage of images from all examinations that were screening mammograms*				
<83	9.6 (6.8–13.6)	9.3 (6.5–13.2)	84.5 (73.7–95.7)	3.9 (3.1–6.0)
≥83	9.4 (6.3–13.1)	9.0 (6.0–12.8)	83.2 (73.6–92.0)	3.9 (2.4–5.8)

Note.—Data are median percentages, and numbers in parentheses are interquartile ranges. Recall and false-positive rates were based on 205 radiologists. Sensitivity was based on 187 radiologists with mammograms with cancer diagnosed in follow-up. PPV_1 was based on 201 radiologists with mammograms with positive results. The number of radiologists varies with covariates within each column because of missing values in the covariates.

* Average per year over past 5 years. Ratio of five or more screening examinations to one diagnostic examination.

sensitivity than did male radiologists and radiologists with fellowship training had higher sensitivity than did those without fellowship training.

Radiologists' characteristics significantly improved model fit for all outcomes (Table 3), with $P < .001$ for all three multiple regression models. Higher recall and false-positive rates were noted among female radiologists and radiologists with fellowship training, and lower recall and false-positive rates were noted among radiologists who had adjunct affiliations with an academic medical center and those with 10–19 years of experience interpreting mammograms (Table 3). Higher sensitivity was noted for fellowship-trained radiologists, and PPV_1 was lower among female radiologists. Fellowship training in breast imaging was the only characteristic of radiologists that was significantly associated with improved overall accuracy (odds ratio, 1.61; 95% confidence interval: 1.05, 2.45; $P = .028$) after adjusting for radiologists' random effects, patients' characteristics, and all radiologists' characteristics of interest, including sex and experience. Although most of the fellowship-trained radiologists were women, after adjusting for fellowship training, female sex was not significantly associated with overall accuracy. The results presented in Table 3 were similar when

analyses were restricted to BCSC data from 2002 to 2005, a period closer to that of the survey of radiologists.

Figure 2 shows the variability in recall rates and PPV_1 and their relationship to cancer detection rates for 203 radiologists with recall rates less than 30%. In general, PPV_1 was inversely associated with recall rate, but there was wide variability in PPV_1, recall rate, and cancer detection rate across radiologists. Very few radiologists had both a high recall rate and a high PPV_1. Radiologists with fellowship training tended to have higher PPV_1 and cancer detection rates than did those without fellowship training.

Table 4 shows the effect of the performance of fellowship-trained compared with non–fellowship-trained radiologists for a hypothetical screening population of 100 000 women. Fellowship-trained individuals had a sensitivity of 88% and a false-positive rate of 11%, but non–fellowship-trained radiologists had a sensitivity of 83% and a false-positive rate of 9% (Table 2). Al-

Table 3

Model Results of Radiologists' Characteristics Associated with Interpretive Performance after Adjustment

Radiologists' Characteristics	Recall Rate	False-Positive Rate	Sensitivity	PPV_1	Overall Accuracy Parameter
Sex					
P value	.047	.040	.414	.025	.629
Male	1.00	1.00	1.00	1.00	1.00
Female	1.20 (1.00, 1.43)	1.21 (1.01, 1.46)	1.14 (0.83, 1.56)	0.81 (0.67, 0.97)	0.94 (0.73, 1.21)
Primary affiliation with academic medical center					
P value	.011	.010	.178	.416	.830
No	1.00	1.00	1.00	1.00	1.00
Yes, adjunct	0.65 (0.48, 0.89)	0.64 (0.46, 0.88)	0.60 (0.35, 1.04)	1.25 (0.89, 1.77)	0.94 (0.60, 1.48)
Yes, primary	0.80 (0.61, 1.05)	0.80 (0.61, 1.05)	0.87 (0.57, 1.32)	1.05 (0.82, 1.36)	1.08 (0.78, 1.50)
Breast imaging experience					
Fellowship training					
P value	.004	.005	<.001	.456	.028
No	1.00	1.00	1.00	1.00	1.00
Yes	1.45 (1.13, 1.86)	1.45 (1.12, 1.87)	2.32 (1.42, 3.80)	1.12 (0.83, 1.53)	1.61 (1.05, 2.45)
Years of mammographic interpretation					
P value	<.001	<.001	.171	.263	.413
<10	1.00	1.00	1.00	1.00	1.00
10–19	0.90 (0.88, 0.93)	0.90 (0.87, 0.93)	1.02 (0.81, 1.28)	1.07 (0.96, 1.21)	1.13 (0.90, 1.42)
≥20	1.05 (0.99, 1.11)	1.06 (1.00, 1.13)	1.29 (0.95, 1.74)	0.97 (0.81, 1.15)	1.21 (0.90, 1.63)
Hours working in breast imaging per week					
P value	.253	.261	.228	.218	.379
0–8	1.00	1.00	1.00	1.00	1.00
>8 to 16	0.84 (0.69, 1.03)	0.84 (0.68, 1.03)	0.74 (0.52, 1.05)	1.24 (1.00, 1.52)	0.88 (0.66, 1.18)
>16 to 32	0.79 (0.61, 1.04)	0.79 (0.60, 1.04)	0.69 (0.44, 1.09)	1.10 (0.83, 1.45)	0.87 (0.60, 1.26)
>32	0.91 (0.73, 1.13)	0.91 (0.72, 1.14)	0.67 (0.44, 1.01)	1.09 (0.86, 1.38)	0.74 (0.52, 1.04)
Volume					
Self-reported average no. of mammograms interpreted per year over the past 5 years					
P value	.170	.210	.255	.910	.569
≤1000	1.00	1.00	1.00	1.00	1.00
1001–2000	1.19 (0.92, 1.55)	1.19 (0.90, 1.56)	1.12 (0.64, 1.97)	0.93 (0.68, 1.27)	0.94 (0.58, 1.55)
>2000	1.03 (0.79, 1.33)	1.03 (0.79, 1.35)	0.87 (0.51, 1.48)	0.94 (0.70, 1.28)	0.84 (0.53, 1.34)
Percentage of images from all examinations that were screening mammograms*					
P value	.812	.833	.766	.667	.832
<83	1.00	1.00	1.00	1.00	1.00
≥83	0.98 (0.84, 1.15)	0.98 (0.84, 1.15)	0.96 (0.73, 1.26)	0.96 (0.82, 1.14)	0.98 (0.78, 1.22)

Note.—Except as otherwise indicated, data are odds ratios, and numbers in parentheses are 95% confidence intervals. Adjustment was made for radiologists' random effects, all radiologists' characteristics listed in the table, and patients' characteristics (BCSC registry, age at mammography, breast density, and time since last mammographic examination). Overall *P* value for the model was less than .001 for recall rate, false-positive rate, sensitivity, PPV_1, and overall accuracy parameter.

* Average per year over past 5 years. Ratio of five or more screening examinations to one diagnostic examination.

though fellowship-trained radiologists detected breast cancer in an estimated additional 24 women per 100 000 women screened compared with the number of women in whom breast cancer was detected by non–fellowship-trained radiologists, they also called back an additional 1990 women with false-positive mammographic examination findings.

Discussion

Our study, which included 205 U.S. radiologists and mammograms from more than 1 million mammographic examinations, revealed significant variability in interpretive performance. Wide variability in sensitivity was noted even among radiologists with similar false-positive rates. In adjusted analyses, fellowship training in breast imaging was the only radiologists' characteristic that was significantly associated with greater sensitivity in cancer diagnosis and higher overall accuracy; however, radiologists with fellowship training also had significantly higher false-positive rates compared with those of radiologists without specialized training. In addition, the recall and false-positive rates noted among the fellowship-trained radiologists were higher than the desirable goal of a 10% recall rate recommended for U.S. radiologists' performance (12,13).

Although fellowship-trained radiologists detected more cancers than did non–fellowship-trained radiologists, they also called back an additional 83 women because of false-positive results of evaluations for every additional breast cancer detected. Can we do better? Although the content and duration of breast imaging fellowships in the United States are variable, the goal is always the same: to teach radiologists how to perform high-quality breast imaging. Fellowship training programs in breast imaging should emphasize decreasing radiologists' false-positive rates to within the recommended performance goals for U.S. radiologists while maintaining high sensitivity.

Most mammograms obtained in the United States are interpreted by general radiologists who have no fellowship training in breast imaging (31). The demand for radiologists to interpret findings from mammographic examinations is increasing as the U.S. population ages and greater numbers of women comply with screening guidelines (32,33). Fewer radiology residents apply for breast imaging fellowships than they do for fellowships in many other clinical areas of radiology (34). Attention to the performance of general radiologists is thus important.

It is thought that training of general

Figure 2

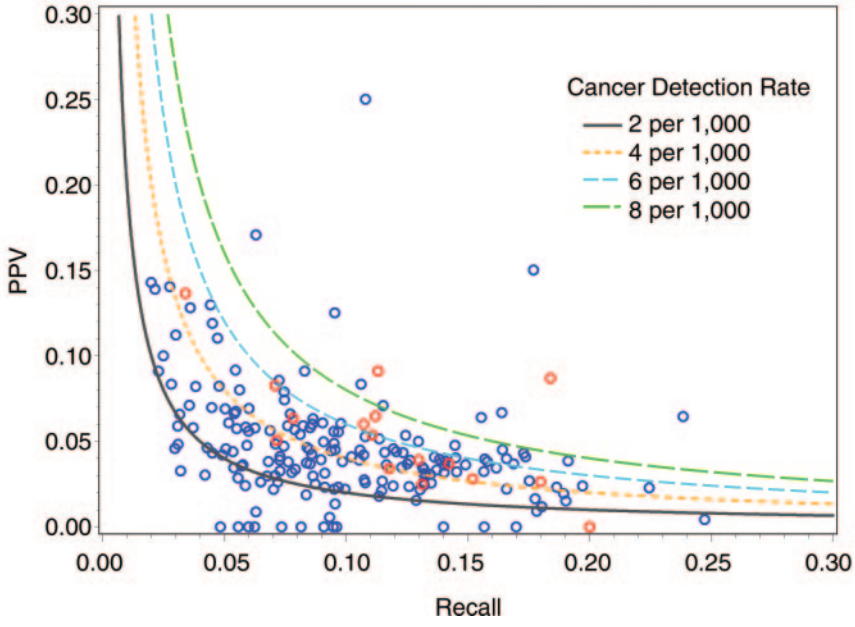

Figure 2: Unadjusted recall rate and PPV_1 for 203 U.S. radiologists with respect to theoretic cancer detection rates per 1000 screening mammographic examinations. Red circles designate fellowship-trained radiologists and blue circles designate non–fellowship-trained radiologists. Four curved lines represent theoretical cancer detection rates for a given PPV_1 and recall rate. Cancer detection rate is defined as the number of true-positive mammograms for every 1000 screening examinations. For example, a radiologist with PPV_1 of five cancers per 100 positive screening examinations (0.05) and a recall rate of five screening examinations with positive results per 100 screening examinations (0.05) would have a cancer detection rate of 2.5 per 1000 ($0.05 \cdot 0.05 = 0.0025$) screening examinations. This rate would be between the lines that represent cancer detection rates of two per 1000 screening examinations and four per 1000 screening examinations, the range for most of the radiologists in this study.

Table 4

Performance of Fellowship-trained versus Non–fellowship-trained Radiologists If 100 000 Women Were Screened with Mammography

	Fellowship-trained Radiologists		Non–fellowship-trained Radiologists	
Screening Mammographic Interpretation	Breast Cancer Present	Breast Cancer Absent	Breast Cancer Present	Breast Cancer Absent
Positive	422*	10 947†	398*	8957†
Negative	58‡	88 573	82‡	90 563
Total	480	99 520	480	99 520

* Number with cancer detected.

† Number with false-positive findings.

‡ Number with false-negative cancers.

radiologists in the interpretation of mammograms has been improving over time. For example, the Mammography Quality Standards Act (35) requires physicians to be board certified in diagnostic radiology or receive 3 months of formal training in mammography to qualify for independent interpretation of mammograms. In addition, physicians are required to interpret at least 240 mammograms with direct supervision before they can qualify as independent interpreting physicians. Since passage of the Mammography Quality Standards Act, a curriculum and dedicated time in breast imaging have been added to radiology residency training programs (34,36–40). Despite these improvements, we did not find any significant increase in overall accuracy among radiologists who had fewer years of clinical experience and, thus, were more likely to be recent graduates of a residency program.

The Mammography Quality Standards Act (35) requires radiologists to interpret a minimum of 960 mammograms every 2 years; other countries set much higher requirements, such as a minimum of 5000 per year (5). We found no association between self-reported annual volume of mammograms interpreted and interpretive performance. Researchers in two previous studies (5,6) in a smaller BCSC population of radiologists also found no association between greater interpretive volume and overall accuracy. Our study goes beyond the previous BCSC studies in that we included data from BCSC mammographic examinations that were performed more recently (up to December 31, 2005), whereas researchers in previous studies included data from screening mammographic examinations only through 2001. In addition, our study period was longer (8 years vs 4 years for Smith-Bindman et al [5] and 6 years for Barlow et al [6]). The longer study period increases the number of breast cancers detected per radiologist, yielding more reliable estimates of an individual radiologist's sensitivity.

One strength of this study was the inclusion of a diverse group of community-based radiologists who interpreted mammograms in women living in geographically disparate regions of the United States. Our findings are, thus, more generalizable than results of a survey of only academic physicians or specialists in breast imaging. In a smaller study, Sickles et al (7) compared interpretive performance reported on radiologists at the two ends of the spectrum of experience and expertise (n = 10) (41,42), whereas our study involved a far larger sample of 205 radiologists and a wider spectrum of experience and training among the study radiologists. Of note, the survey response rate of 71% was higher than that in most studies of physicians, who have a mean response rate of only 54% (43). Although it is possible that survey respondents were not representative of all community radiologists, the interpretive performance of respondents was similar to that of BCSC radiologists as a whole.

One limitation of our study was that, despite the added years of data, low numbers of examinations in women with cancer remained for some interpreting radiologists; this result added to the variability we found in sensitivity. However, the statistical modeling approach we used accounted for the numbers of mammograms interpreted by each radiologist and pooled information across radiologists to make more stable inferences about which characteristics were associated with performance. Other limitations of our study included the small number of fellowship-trained radiologists (n = 16) and the lack of data on the use of digital mammography. Although we used a 1-year standard definition of follow-up for breast cancer diagnosis, as recommended by the American College of Radiology (12), in some studies longer follow-up periods were used, and this discrepancy made comparisons difficult. Finally, 30% of the study radiologists interpreted mammograms at institutions outside of the BCSC; thus, their self-reported data on annual volume could not be verified. Last, many of the radiologists worked part time, and this factor made interpretation of the percentage of time spent in breast imaging challenging.

Because mammography is the only proven method to screen women for breast cancer, we must continue our efforts to maintain high quality and achieve high performance. Despite variability in interpretive performance, we found that radiologists with fellowship training in breast imaging had significantly higher sensitivity and higher overall accuracy in screening mammograms than did non–fellowship-trained radiologists. However, these fellowship-trained radiologists also had higher recall and false-positive rates. A new era of continuing medical education should target radiologists on the basis of their specific training and clinical practice and, ideally, link the educational programs to the individual radiologist's performance (44,45).

Acknowledgments: We thank the participating women, mammography facilities, and radiologists for the data they 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/*. The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the United States. A full description of these sources is at *http://breastscreening.cancer.gov/work/acknowledgement.html*.

References

1. U.S. Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. Ann Intern Med 2002;137(5 pt 1):344–346.
2. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137(5 pt 1):347–360.
3. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med 1994;331(22):1493–1499.
4. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample. Arch Intern Med 1996;156(2):209–213.
5. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst 2005;97(5):358–367.
6. 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.
7. Sickles EA, Wolverton DE, Dee KE. Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology 2002;224(3):861–869.
8. Elmore JG, Miglioretti DL, Reisch LM, et al. Screening mammograms by community radiologists: vari-

ability in false-positive rates. J Natl Cancer Inst 2002;94(18):1373–1380.

9. Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. J Natl Cancer Inst 2000;92(20):1657–1666.

10. Smoothed plots of frequency distributions of recall rates for 3,294,680 screening mammography examinations (among radiologists with 1000 or more examinations), 1996-2005 [figure 1]. Performance benchmarks for screening mammography. Breast Cancer Surveillance Consortium Web site. http://breastscreening.cancer.gov/data/benchmarks/screening/figure1.html. Updated November 29, 2007. Accessed December 26, 2008.

11. Smoothed plots of sensitivity for 16,324 cancers that were identified at screening mammography (among radiologists finding 30 or more cancers), 1996–2005 [figure 10]. Performance benchmarks for screening mammography. Breast Cancer Surveillance Consortium Web site. http://breastscreening.cancer.gov/data/benchmarks/screening/figure10.html. Updated November 29, 2007. Accessed December 26, 2008.

12. American College of Radiology. Breast imaging reporting and data system (BI-RADS). Reston, Va: American College of Radiology, 2003.

13. Bassett LW, Hendrick RE, Bassford TL, et al. Quality determinants of mammography. Clinical practice guideline no. 13. AHCPR publication no. 95-0632. Rockville, Md: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, October 1994.

14. 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.

15. Breast Cancer Surveillance Consortium. Breastscreening.cancer.gov Web site. http://breastscreening.cancer.gov/. Accessed May 29, 2008.

16. 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.

17. Elmore JG, Taplin SH, Barlow WE, et al. Does litigation influence medical practice? the influence of community radiologists' medical malpractice perceptions and experience on screening mammography. Radiology 2005;236(1):37–46.

18. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of breast density, age, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138(3):168–175.

19. 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.

20. Kerlikowske K, Carney PA, Geller B, et al. Performance of screening mammography among women with and without a first-degree relative with breast cancer. Ann Intern Med 2000;133(11):855–863.

21. Kerlikowske K, Miglioretti DL, Ballard-Barbash R, et al. Prognostic characteristics of breast cancer among postmenopausal hormone users in a screened population. J Clin Oncol 2003;21(23):4314–4321.

22. Miglioretti DL, Rutter CM, Geller BM, et al. Effect of breast augmentation on the accuracy of mammography and cancer characteristics. JAMA 2004;291(4):442–450.

23. Ongoing collaborations: FAVOR study (factors affecting variability of radiologists). National survey of mammography practices. FAVOR II mammography practice survey. Breast Cancer Surveillance Consortium Web site. http://breastscreening.cancer.gov/collaborations/favor_ii_mammography_practice_survey.pdf. Accessed June 27, 2008.

24. Yankaskas BC, Taplin SH, Ichikawa L, et al. Association between mammography timing and measures of screening performance in the United States. Radiology 2005;234(2):363–373.

25. BCSC Glossary of Terms. Breast Cancer Surveillance Consortium Web site. http://breastscreening.cancer.gov/data/bcsc_data_definitions.pdf. Last updated May 28, 2008. Accessed December 17, 2008.

26. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. Radiology 2006;241(1):55–66.

27. American College of Radiology. Illustrated Breast Imaging Reporting and Data System (BI-RADS). 3rd ed. Reston, Va: American College of Radiology, 1998.

28. Taplin SH, Ichikawa LE, Kerlikowske K, et al. Concordance of Breast Imaging Reporting and Data System assessments and management recommendations in screening mammography. Radiology 2002;222(2):529–535.

29. 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.

30. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001;20(19):2865–2884.

31. Leung JW, Margolin FR, Dee KE, Jacobs RP, Denny SR, Schrumpf JD. Performance parameters for screening and diagnostic mammography in a community practice: are there differences between specialists and general radiologists? AJR Am J Roentgenol 2007;188(1):236–241.

32. Meyer J. Census 2000 brief. http: //www.census.gov/prod/2001pubs/c2kbr01-12.pdf. Accessed September 1, 2008.

33. Centers for Disease Control and Prevention (CDC). Use of mammograms among women aged > or = 40 years: United States, 2000–2005. MMWR Morb Mortal Wkly Rep 2007;56(3):49–51.

34. Bassett LW, Monsees BS, Smith RA, et al. Survey of radiology residents: breast imaging training and attitudes. Radiology 2003;227(3):862–869.

35. Mammography Quality Standards Act and program. U.S. Food and Drug Administration Web site. http://www.fda.gov/Radiation-Emitting Products/MammographyQualityStandardsActandProgram/Regulations/ucm110906.htm#s90011. Accessed July 8, 2008.

36. Bassett LW, Cassady CI, Gold RH. Present status of residency training in mammography. AJR Am J Roentgenol 1991;156(1):59–62.

37. Bassett LW, Lubisich JP, Bresch JP, Jessop NW, Hendrick RE. Quality assurance in mammography: status of residency education. AJR Am J Roentgenol 1993;160(2):271–274.

38. Homer MJ. Mammography training in diagnostic radiology residency programs. Radiology 1980;135(2):529–531.

39. Jackson VP, Bresch JA, Bassett LW, Dershaw DD, Jessop NW. Status of mammography education and knowledge of radiology residents. Radiology 1996;201(3):773–776.

40. Lubisich JP, Bassett LW, Bresch JA, Jessop NW. Status of residency training in mammography. AJR Am J Roentgenol 1996;166(5):1189–1191.

41. Guenin MA. Generalists versus specialists in mammography [letter]. Radiology 2003;227(2):609.

42. Sickles EA, Wolverton DE, Dee KE. Reply. Radiology 2003;227(2):609–611.

43. Asch DA, Jedrziewski MK, Christakis NA. Response rates to mail surveys published in medical journals. J Clin Epidemiol 1997;50(10):1129–1136.

44. Continuing education in the health professions: improving healthcare through lifelong learning. New York, NY: Josiah Macy Jr Foundation, 2008.

45. Linver MN, Paster SB, Rosenberg RD, Key CR, Stidley CA, King WV. Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases. Radiology 1992;184(1):39–43.