Accuracy of Short-Interval Follow-Up Mammograms by Patient and Radiologist Characteristics

Erin J. Aiello Bowles[1], **Diana L. Miglioretti**[1], **Edward A. Sickles**[2], **Linn Abraham**[1], **Patricia A. Carney**[3], **Bonnie C. Yankaskas**[4], and **Joann G. Elmore**[5]

[1]Group Health Center for Health Studies, Group Health Cooperative, 1730 Minor Ave., Ste. 1600, Seattle, WA 98101

[2]Department of Radiology, University of California, San Francisco, San Francisco, CA

[3]Department of Family Medicine, Oregon Health Sciences University, Portland, OR

[4]Department of Radiology, University of North Carolina, Chapel Hill, NC

[5]Department of Internal Medicine, University of Washington School of Medicine, Seattle, WA

Abstract

OBJECTIVE—The purpose of our study was to examine the accuracy of short-interval follow-up mammograms and evaluate patient and radiologist characteristics associated with accuracy.

MATERIALS AND METHODS—We evaluated 45,007 initial short-interval follow-up mammograms from the Breast Cancer Surveillance Consortium interpreted 3–9 months after a probably benign assessment on a screening or diagnostic examination between 1994 and 2004. We linked these mammograms with patient characteristics and breast cancer diagnoses within 12 months. A subset of short-interval follow-up examinations ($n = 13,907$) was merged with radiologist characteristics collected from survey data from 130 interpreting radiologists. Using logistic regression, we fit generalized estimating equations to model sensitivity and specificity of short-interval follow-up mammograms by patient and radiologist characteristics.

RESULTS—For every 1,000 women, 8.0 women (0.8%) were diagnosed with breast cancer within 6 months and 11.3 (1.1%) within 12 months. Sensitivity was 83.3% (95% CI, 79.4–87.3%) for cancers diagnosed within 6 months and 60.5% (56.2–64.7%) for those diagnosed within 12 months. Specificity was 97.2% (96.9–97.6%) at 6 months and 97.3% (96.9–97.6%) at 12 months. Sensitivity at 12 months increased among women with unilateral short-interval follow-up mammograms (odds ratio, 1.56 [95% CI, 1.06–2.29]) and when the interpreting radiologist spent more than 10 hours a week in breast imaging (odds ratio, 3.25 [1.00–10.52]).

CONCLUSION—Initial short-interval follow-up mammography examinations had a lower sensitivity for detecting breast cancer within 12 months than other diagnostic mammograms (61% for short-interval follow-up vs 80% for diagnostic mammograms reported in the literature). However, sensitivity within the 6-month interval that is usually recommended for subsequent follow-up was 83%. Accuracy of short-interval follow-up mammograms was influenced by few patient and radiologist characteristics.

Keywords

breast cancer; diagnostic mammography; sensitivity; short-interval follow-up mammography; specificity

Address correspondence to E. J. Aiello Bowles (aiello.e@ghc.org).

Women with probably benign abnormalities noted on screening or diagnostic mammograms often receive recommendations to obtain a diagnostic mammogram in 6 months. These diagnostic mammograms, or short-interval follow-up examinations, are surprisingly common [1]. Short-interval follow-up examinations are most often recommended for women with a BI-RADS category 3, probably benign, mammogram assessment after a full diagnostic workup [2, 3]. Because probably benign lesions have a low probability of malignancy (< 2%) [1–8], they are recommended for periodic short-interval follow-up surveillance rather than immediate biopsy to avoid unnecessary invasive procedures, patient anxiety, and medical costs [9–12]. Numerous previous studies have shown that periodic short-interval follow-up surveillance appropriately detects visible changes in probably benign lesions that represent small malignant tumors at an early stage and have a favorable prognosis [1, 2, 5–8].

Despite the frequency of short-interval follow-up examinations in clinical practice, few data such as sensitivity and specificity are available on the performance characteristics of these examinations. Our study was stimulated by data available on the Website of the National Cancer Institute provided by the Breast Cancer Surveillance Consortium (BCSC) [13]. One table on this Website shows performance characteristics of diagnostic mammography examinations in the United States. The reported sensitivity of short-interval follow-up examinations was substantially lower than that of other diagnostic examinations (55.8% vs 79.8%) [1, 13]. We found this lower sensitivity of concern because it suggests that nearly one half of breast cancers are missed on these short-interval follow-up examinations.

The purpose of this study was to examine the accuracy of short-interval follow-up examinations and factors that might affect interpretive performance relative to other diagnostic examinations. We used detailed patient information from the BCSC linked to radiologist information from a self-administered survey to describe the sensitivity and specificity of short-interval follow-up mammograms. We also evaluated the association between the performance of short-interval follow-up mammograms and patient and radiologist characteristics in clinical practice. Patient characteristics, such as age and breast density, have been associated with the interpretive performance of both screening and diagnostic mammography examinations [14–16]. We do not know whether patient characteristics are associated with the accuracy of short-interval follow-up examinations, and this information could be used to determine which women would be best suited for periodic short-interval follow-up surveillance. Previous studies have also shown that radiologist characteristics, such as number of years in practice, workload, and type of training, are associated with interpretive performance of screening examinations [17–20]. Therefore, we explored whether any radiologists factors predict better performance when interpreting short-interval follow-up examinations.

Materials and Methods

Study Population

This study used data from seven mammography registries that are part of the BCSC [21]: Group Health Cooperative, Western Washington; Colorado Mammography Project, Colorado; New Hampshire Mammography Network, New Hampshire; Carolina Mammography Registry, North Carolina; New Mexico Mammography Project, New Mexico; San Francisco Mammography Registry, California; and Vermont Breast Cancer Surveillance System, Vermont. We also collected radiologist information via a survey of three of the registries: Group Health Cooperative, New Hampshire Mammography Network, and Colorado Mammography Project. A statistical coordinating center pooled data for analysis. Each registry and the statistical coordinating center received institutional review

board approval to enroll participants, link data, and perform analytic studies. All procedures were HIPAA-compliant, and all registries and the statistical coordinating center have received a federal certificate of confidentiality that protects the identities of research subjects.

Data Collection

Short-interval follow-up mammograms—We included mammograms from 1994 to 2004 in this study if the interpreting radiologist indicated that the examination was a short-interval follow-up. We included each woman's first unilateral or bilateral diagnostic mammography examination that was classified as short-interval follow-up ($n = 110,942$ women). We excluded women with a history of breast cancer ($n = 7,271$) because the recommendation for and interpretation of short-interval follow-up mammograms may be different for these women. We made several additional exclusions to ensure that our analysis included only initial short-interval follow-up mammograms. BI-RADS guidelines recommend short-interval follow-up mammograms only after a probably benign assessment [3]; therefore, we excluded women with examinations without a previous assessment of probably benign in the BCSC database ($n = 25,490$). Because initial short-interval follow-up examinations normally occur about 6 months after the previous mammogram or sonogram, we also excluded women who had an examination less than 3 months or more than 9 months before ($n = 31,487$). Finally, we excluded women with short-interval follow-up examinations that were missing a final assessment ($n = 445$) and those with unknown laterality of the mammogram ($n = 1,242$), for a total sample size of 45,007 short-interval follow-up mammograms.

We linked each short-interval follow-up mammogram to the radiologist's final BI-RADS assessment and recommendation within 180 days after the completion of all imaging workup. We also linked each mammogram to a Surveillance Epidemiology and End Results (SEER) registry, state cancer registry, or local benign and malignant pathology database to determine whether a cancer diagnosis had been made within 365 days of the short-interval follow-up examination.

Patient Characteristics

We obtained patient information by linking to standardized questionnaires completed by women at the time of the short-interval follow-up examination [21]. The questionnaires collected information on demographics (e.g., age and race), breast cancer risk factors (e.g., family history of breast cancer, breast symptoms, menopausal status, and current use of hormone therapy), and clinical history of previous breast procedures (e.g., biopsy).

Radiologist Characteristics

Radiologists were eligible to complete a self-administered survey if they interpreted mammograms in the year 2001 at one of three participating sites: Group Health Cooperative, New Hampshire Mammography Network, or Colorado Mammography Project. Of the 181 surveys mailed in 2002, 139 surveys were returned, for a response rate of 77%; 130 of these radiologists interpreted one or more short-interval follow-up mammograms from 1994 to 2004, making them eligible for the present study. A detailed description of the radiologist survey has been published elsewhere [19]. Briefly, we collected demographic and clinical characteristics of radiologists, including age, sex, academic affiliation, number of years spent working in breast imaging, workload, percentage of time spent in diagnostic imaging, number of breast procedures conducted, and malpractice history. We double-entered all survey data at each site and transferred the data file via secure file transfer protocol to the statistical coordinating center. We linked these survey responses to a subgroup of the short-interval follow-up mammograms described previously that were interpreted by radiologists

with survey data (*n* = 13,907). This subanalysis did not link to all 45,007 short-interval follow-up mammograms because we included mammograms in the main analyses that were interpreted by radiologists who did not complete the survey.

Statistical Analyses

We evaluated rates of abnormal interpretation and breast cancer diagnosis among all short-interval follow-up examinations and for each category of patient and radiologist characteristics. We then evaluated the sensitivity and specificity among all examinations and for each category of patient and radiologist characteristics using the standard 12- month follow-up interval for detecting breast cancer [3]. Because guidelines for short-interval follow-up examinations suggest that women should return for a second short-interval follow-up examination after 6 months, we also calculated sensitivity and specificity using a 6-month outcome interval. Positive examinations were defined as those with a final BI-RADS assessment of 4, 5, or 0 with a recommendation for biopsy after the completion of all diagnostic workup. Negative examinations were defined as those with a final BI-RADS assessment of 1, 2, 3, or 0 without recommendation for biopsy. We determined final BI-RADS assessments by looking for the first non-0 assessment from additional imaging within 180 days of the short-interval follow-up mammogram; however, some 0 assessments remained unresolved. We defined the abnormal interpretation rate as the number of positive examinations divided by the total number of examinations. The definitions we used for sensitivity and specificity are as follows:

- 12-month sensitivity = number of positive examinations/all examinations in women diagnosed with cancer within 12 months of the short-interval follow-up mammogram,

- 12-month specificity = number of negative examinations/all examinations in women without cancer within 12 months of the short-interval follow-up mammogram,

- 6-month sensitivity = number of positive examinations/all examinations in women diagnosed with cancer within 6 months of the short-interval follow-up mammogram, and

- 6-month specificity = number of negative examinations/all examinations in women without cancer within 6 months of the short-interval follow-up mammogram.

The 6- and 12-month definitions were not mutually exclusive (i.e., cancers included in the 6-month definition were also included in the 12-month definition). We matched cancer laterality among unilateral short-interval follow-up mammograms so that cancers diagnosed only in the same breast as the unilateral examination were counted as a cancer diagnosis. If a cancer was diagnosed in the opposite breast, it was not counted as a cancer diagnosis for the calculation of sensitivity, and instead was included as a negative examination in the calculation of specificity. We calculated 95% CIs for sensitivity and specificity using the robust variance estimates from generalized estimating equations (GEE) with an exchangeable correlation structure to account for correlation within radiologists who interpreted more than one short-interval follow-up mammogram [22].

Using logistic regression fit with GEE to account for correlation within radiologists, we modeled the odds of a positive examination given a cancer diagnosis (sensitivity) and the odds of a negative examination given no cancer diagnosis (specificity). We conducted univariate analyses to determine which covariates should be included in the multivariable models; those that were statistically signifi cant (*p* < 0.05) in one or more models (either sensitivity or specificity) were included. The final models for the patient characteristics were adjusted for patient age (continuous), menopausal status and hormone therapy use

(premenopausal, post-menopausal with no hormone therapy, post-menopausal with hormone therapy), mammogram laterality (bilateral vs unilateral), and breast density (categorized as BI-RADS density categories 3 and 4 [dense breasts] versus BI-RADS density categories 1 and 2 [fatty breasts]). The final models for the radiologist characteristics were adjusted for all patient characteristics just described and for self-reported radiologist characteristics, including age group (35–44, 45–54, ≥ 55 years old), academic affiliation (primary appointment vs affiliate or no appointment), hours per week spent in breast imaging (< 10 vs ≥ 10), number of mammograms interpreted in 2001 (≤ 1,000, 1,001–2,000, ≥ 2,001), percentage of mammograms interpreted that were diagnostic (≤ 25% vs 25–100%), and work full-time (yes vs no). We conducted all statistical analyses using Stata (StataCorp).

Results

Patient Characteristics

The characteristics of the women who underwent 45,007 short-interval follow-up examinations from 1994 to 2004 and associated performance characteristics are shown in Table 1. Most short-interval follow-up mammograms occurred among women 40–59 years old (61%) and among women with no breast symptoms (76%). In addition, most women were postmenopausal, not taking hormone therapy, and had no family history of breast cancer. Approximately 63% of examinations were unilateral short-interval follow-up mammograms. The examinations were ordered to follow up on abnormalities noted on a previous screening mammogram (42%) or as follow-up after a diagnostic evaluation was completed (50% of the previous diagnostic evaluations were after screening examinations and 8% were diagnostic examinations to evaluate a breast problem). The average abnormal interpretation rate of short-interval follow-up examinations was 3.4%. Breast cancer (matched on laterality to the breast that underwent the short-interval follow-up examination) was diagnosed in 11.3 per 1,000 women (1.1%) within 12 months of their short-interval follow-up examinations; more than half of these cases (8.0 per 1,000 examinations, 0.8%) were diagnosed within 6 months.

The 12-month sensitivity of all short-interval follow-up examinations was 60.5% (95% CI, 56.2–64.7%); at 6 months this increased to 83.3% (79.4–87.3%). We noted trends in accuracy of the short-interval follow-up examinations by patient characteristics; however, none was statistically significant in unadjusted models. Crude sensitivity increased with patient age and among post-menopausal women not receiving hormone therapy. Sensitivity was also higher among unilateral short-interval follow-up examinations compared with bilateral examinations. The sensitivity was lower for women with extremely dense breasts and whose examination before the short-interval follow-up mammogram was obtained to evaluate a specific breast problem. The average specificity was 97.3% (97.0–97.6%) at 12 months and was unchanged when calculated at 6 months; specificity increased only among women with almost entirely fatty breasts (98.4%; 95% CI, 97.8–99.0%).

Adjusted odds ratios (ORs) for sensitivity and specificity by patient characteristics are shown in Table 2. Because trends in unadjusted sensitivity and specificity at 6 months were similar to those at 12 months, and the number of cancers at 6 months was small, we presented adjusted rates for 12 months only. Although sensitivity increased among postmenopausal women who did not use hormone therapy (compared with premenopausal women), confidence intervals were wide. Women who had unilateral mammograms had significantly increased sensitivity compared with those with bilateral examinations (OR, 1.56 [95% CI, 1.06–2.29]). Postmenopausal women (regardless of hormone therapy use) and women with breast symptoms had slightly lower specificity (and thus more false-positive examinations) than premenopausal women and women without symptoms, respectively.

Women with dense breasts had significantly lower specificity than women with fatty breasts (OR, 0.78 [0.69–0.97]).

Radiologist Characteristics

We present interpretive performance by radiologists' characteristics in Table 3. The overall rates of abnormal interpretation, cancer diagnoses, sensitivity, and specificity for this subgroup of short-interval follow-up examinations were similar to those in the larger patient population obtaining these examinations described in Table 1. We noted trends in which unadjusted sensitivities increased among radiologists who had less than 10 years of experience interpreting mammograms (compared with ≥ 10), spent 10 or more hours per week in breast imaging (compared with < 10 hours), or were female (compared with male).

A few radiologist characteristics were associated with sensitivity and specificity (Table 4). Radiologists who spent 10 or more hours per week in breast imaging had increased sensitivity and specificity compared with those who spent less than 10 hours per week (sensitivity OR, 3.25 [1.00–10.52]; specificity OR, 1.35 [0.88–2.07]). Radiologists with an affiliate or no academic appointment had increased specificity compared with those with a primary academic appointment (OR, 2.51 [1.20–5.22]).

Discussion

Diagnostic mammograms obtained as short-interval follow-up examinations to follow up a probably benign abnormality have a low sensitivity for detecting cancers diagnosed within the following 12 months (61%) compared with what previous studies have shown for other types of diagnostic mammograms in the same population (usually ≈ 80%) [1, 14]. Few patient or radiologist characteristics were statistically significantly associated with sensitivity.

To our knowledge, this is the first article to describe the accuracy of initial short-interval follow-up mammograms and to evaluate the accuracy by patient and radiologist characteristics. Sickles et al. [1, 13] showed a similarly low sensitivity at 12 months (for all short-interval follow-up examinations) in results that were posted on the public BCSC Website; these results were not published or discussed in the original articles. Our study population was also from the BCSC, but included one new study site and additional years of data beyond those included in Sickles' articles. We also examined the influence of both patient and radiologist characteristics on sensitivity and specificity. These differences and that we matched laterality of cancer diagnoses and examinations might have accounted for the slightly higher sensitivity noted in our results (60.5% vs 55.8% in Sickles' earlier article [1, 13]). Other previous studies have evaluated the accuracy of the examination producing the initial short-interval follow-up recommendation (i.e., the examination given the probably benign assessment), but these studies did not evaluate the sensitivity or specificity of the initial short-interval follow-up examinations themselves [4, 7].

The reason for the low 12-month sensitivity of initial short-interval follow-up examinations is unclear, but there are several possible explanations. First, cancers assessed as probably benign (BI-RADS category 3) may not grow as rapidly as cancers that appear more suspicious for malignancy (BI-RADS category 4 or 5). Therefore, it may be more difficult to identify interval change (hence, recommend biopsy) at initial short-interval follow-up examinations because these examinations usually are performed 6 months rather than 1 year after the index mammogram that prompted the short-interval follow-up. The rationale for recommending an initial short-interval follow-up examination is to identify 6 months earlier those poorer-prognosis "probably benign" cancers that do grow sufficiently rapidly to be detected early [10–12].

A second possible reason for the low sensitivity of initial short-interval follow-up examinations is that radiologists interpreting them might be reassured by the previous radiologist's probably benign interpretation and thus have a higher threshold for calling the initial examination suspicious compared with other diagnostic examinations. This would be false reassurance if it is resulting in low sensitivity for diagnosing breast cancer. It would be interesting to evaluate short-interval follow-up sensitivity among radiologists who interpreted both the examination that resulted in a probably benign assessment and the initial short-interval follow-up examination; however, we were unable to do this in our study.

A third possible reason for the low sensitivity of initial short-interval follow-up examinations may be that the standard BI-RADS definition to use a 12-month follow-up period for evaluating sensitivity and specificity does not match the 6-month follow-up period recommended after an initial short-interval follow-up examination. Our data showed that using a 6-month follow-up interval for the definition of sensitivity increased the unadjusted sensitivity of short-interval follow-up examinations to 83%, which is similar to the 12-month sensitivity for other diagnostic examinations ($\approx$ 80%) [14]. However, a trade-off in defining the follow-up interval for sensitivity always exists—if you shorten the follow-up interval, sensitivity increases, because there are fewer false-negative examinations that appear as interval cancers during this shorter time. It has been recommended that the follow-up interval for defining sensitivity and specificity of a screening or diagnostic test should match the follow-up interval recommended for that test, so long as that is what occurs in clinical practice [23]. Future research should evaluate the specific follow-up intervals that radiologists are actually recommending after initial short-interval follow-up examinations and whether women are complying with those recommendations.

Our study had several limitations. We were unable to retrospectively evaluate whether cancers did or did not show interval progression on short-interval follow-up mammograms, which might have helped to determine more effective thresholds for recommending biopsy rather than continued surveillance. In addition, we were unable to evaluate performance of initial short-interval follow-up examinations by the types of lesions requiring follow-up (e.g., mass, focal asymmetry, calcifications) or by the size and stage of the cancers that were diagnosed. These analyses were beyond the scope of our project.

Although our sample included 130 radiologists from three geographic areas in the United States, we had limited ability to evaluate the importance of individual radiologist characteristics because of a small sample size in some categories. However, all sensitivity and specificity analyses were based on multiple short-interval follow-up mammograms interpreted by the radiologists, increasing the statistical power of these calculations. Despite the large size of our cohort, we were also limited by the small number of women with breast cancer used to evaluate sensitivity, especially within the 6-month follow-up period. Overall, the rate of cancer diagnoses in this study was smaller than has been reported for other short-interval follow-up studies. This may be because a large proportion of short-interval follow-up examinations in our study directly followed screening mammograms, which have a lower cancer rate compared with other diagnostic examinations. Given the low cancer rate and differences between this and previous studies, we caution the reader in interpreting our results.

Our study had several unique strengths. We were able to evaluate the interpretive performance of short-interval follow-up examinations in a large, geographically diverse population. The large population size allowed us to make several exclusions (such as women with a history of breast cancer or women without outcome information) and to analyze a population of women eligible for short-interval follow-up mammograms. Had we not made these exclusions, we likely would have increased the variability in our sensitivity and

specificity estimates, thus decreasing the ease of interpretation of our results. Our population included unique detailed information on patient and radiologist characteristics available from linking several study databases. We also had the ability to link to cancer registry data, which enabled us to calculate sensitivity and specificity.

In conclusion, the sensitivity of diagnostic mammograms obtained as initial short-interval follow-up examinations is low when using the standard 12-month auditing definition for follow-up period. The reasons for this low sensitivity should be elucidated. We noted increases in sensitivity among women who underwent unilateral short-interval follow-up examinations and among radiologists who spent 10 or more hours per week in breast imaging; but overall, few patient or radiologist characteristics were associated with accuracy. The value of using a 6-month (rather than a 12-month) follow-up period for defining sensitivity also should be examined in future studies.

Acknowledgments

References

1. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. Radiology. 2005; 235:775–790. [PubMed: 15914475]

2. Kerlikowske K, Smith-Bindman R, Abraham LA, et al. Breast cancer yield for screening mammographic examinations with recommendation for short-interval follow-up. Radiology. 2005; 234:684–692. [PubMed: 15734926]

3. American College of Radiology. ACR Breast Imaging Reporting and Data System: breast imaging atlas. 4. Reston, VA: American College of Radiology; 2003. ACR BI-RADS: mammography.

4. Yasmeen S, Romano PS, Pettinger M, et al. Frequency and predictive value of a mammographic recommendation for short-interval follow-up. J Natl Cancer Inst. 2003; 95:429–436. [PubMed: 12644536]

5. Varas X, Leborgne JH, Leborgne F, Mezzera J, Jaumandreu S, Leborgne F. Revisiting the mammographic follow-up of BI-RADS category 3 lesions. AJR. 2002; 179:691–695. [PubMed: 12185047]

6. Sickles EA. Periodic mammographic follow-up of probably benign lesions: results in 3,184 consecutive cases. Radiology. 1991; 179:463–468. [PubMed: 2014293]

7. Vizcaino I, Gadea L, Andreo L, et al. Short-term follow-up results in 795 nonpalpable probably benign lesions detected at screening mammography. Radiology. 2001; 219:475–483. [PubMed: 11323475]

8. Varas X, Leborgne F, Leborgne JH. Nonpalpable, probably benign lesions: role of follow-up mammography. Radiology. 1992; 184:409–414. [PubMed: 1620838]

9. Geller BM, Barlow WE, Ballard-Barbash R, et al. Use of the American College of Radiology BI-RADS to report on the mammographic evaluation of women with signs and symptoms of breast disease. Radiology. 2002; 222:536–542. [PubMed: 11818625]

10. 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; 222:529–535. [PubMed: 11818624]

11. Sickles EA. Probably benign breast lesions: when should follow-up be recommended and what is the optimal follow-up protocol? Radiology. 1999; 213:11–14. [PubMed: 10540635]

12. Sickles EA. Management of probably benign breast lesions. Radiol Clin North Am. 1995; 33:1123–1130. [PubMed: 7480660]

13. Sickles, EA.; Miglioretti, DL.; Ballard-Barbash, R., et al. Performance benchmarks for diagnostic mammography. National Cancer Institute, Breast Cancer Surveillance Consortium Website; http://breast-screening.cancer.gov/data/benchmarks/diagnostic/tableSensSpec.html

14. Barlow WE, Lehman CD, Zheng Y, et al. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. J Natl Cancer Inst. 2002; 94:1151–1159. [PubMed: 12165640]

15. Kavanagh AM, Cawson J, Byrnes GB, et al. Hormone replacement therapy, percent mammographic density, and sensitivity of mammography. Cancer Epidemiol Biomarkers Prev. 2005; 14:1060–1064. [PubMed: 15894653]

16. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med. 2003; 138:168–175. [PubMed: 12558355]

17. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst. 2005; 97:358–367. [PubMed: 15741572]

18. Beam CA, Conant EF, Sickles EA. Association of volume and volume-independent factors with accuracy in screening mammogram interpretation. J Natl Cancer Inst. 2003; 95:282–290. [PubMed: 12591984]

19. Barlow WE, Chi C, Carney PA, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst. 2004; 96:1840–1850. [PubMed: 15601640]

20. Moss SM, Blanks RG, Bennett RL. Is radiologists' volume of mammography reading related to accuracy? A critical review of the literature. Clin Radiol. 2005; 60:623–666. [PubMed: 16038688]

21. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR. 1997; 169:1001–1008. [PubMed: 9308451]

22. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. Biometrics. 1986; 42:121–130. [PubMed: 3719049]

23. Rosenberg RD, Yankaskas BC, Hunt WC, et al. Effect of variations in operational definitions on performance estimates for screening mammography. Acad Radiol. 2000; 7:1058–1068. [PubMed: 11131050]

TABLE 1

Characteristics of 45,007 Women Who Underwent Short-Interval Follow-Up Mammography, 1994 2004

Characteristic	No.	%	Abnormal Interpretation Rate	Cancers Diagnosed per 1,000 Examinations		Sensitivity (95% CI)[a] (%)		Specificity (95% CI)[a] (%)	
				Within 12 Months (n = 506)	Within 6 Months (n = 360)	12 Months	6 Months	12 Months	6 Months
Overall	45,007	100.0	3.4	11.3	8.0	60.5 (56.2–64.7)	83.3 (79.4–87.3)	97.3 (96.9–97.6)	97.2 (96.9–97.6)
Patient age group (y)									
< 40	3,179	7.1	2.5	3.1	1.6	40.0 (9.4–70.6)	90.0 (44.6–115.4)	97.6 (97.1–98.2)	97.6 (97.1–98.2)
40–49	14,797	32.9	3.0	5.9	4.1	55.2 (44.7–65.7)	77.0 (66.4–87.7)	97.3 (97.0–97.7)	97.3 (97.0–97.7)
50–59	12,843	28.5	3.5	10.6	7.8	62.5 (54.7–70.3)	83.0 (75.3–90.7)	97.2 (96.7–97.6)	97.2 (96.7–97.6)
60–69	7,742	17.2	4.0	19.5	13.4	59.6 (51.7–67.5)	84.6 (77.8–91.4)	97.1 (96.6–97.5)	97.1 (96.6–97.5)
≥ 70	6,446	14.3	3.9	18.9	14.0	64.8 (55.2–74.3)	86.7 (78.7–94.6)	97.3 (96.8–97.7)	97.3 (96.8–97.7)
Breast symptoms									
None	26,209	76.3	3.4	10.5	6.8	58.4 (52.6–64.2)	86.6 (81.5–91.7)	97.2 (96.8–97.6)	97.2 (96.8–97.6)
Lump	1,484	4.3	3.8	17.5	12.8	57.7 (40.8–74.6)	78.9 (60.1–97.8)	97.2 (96.3–98.1)	97.2 (96.3–98.1)
Other	6,646	19.4	3.6	12.0	9.2	51.3 (40.9–61.6)	67.2 (53.5–80.9)	96.9 (96.4–97.4)	96.9 (96.4–97.4)
Menopausal status and current hormone therapy use									
Premenopausal	16,502	38.9	2.8	6.5	4.5	54.2 (45.0–63.5)	74.7 (64.9–84.4)	97.5 (97.2–97.9)	97.5 (97.2–97.9)
Postmenopausal, no hormone therapy	16,890	39.9	3.8	15.7	11.7	65.7 (5.6–71.7)	87.8 (82.9–92.7)	97.2 (96.8–97.5)	97.2 (96.8–97.5)
Postmenopausal, hormone therapy	8,985	21.2	3.5	12.8	8.3	56.5 (47.3–65.7)	82.7 (74.4–90.9)	97.2 (96.7–97.6)	97.1 (96.7–97.6)
First degree family history of breast cancer									
No	28,036	85.5	3.4	10.4	7.1	58.9 (53.3–64.5)	83.0 (77.7–88.3)	97.2 (96.9–97.5)	97.2 (96.9–97.5)
Yes	4,768	14.5	4.2	13.6	10.1	64.6 (52.2–77.1)	87.5 (78.4–96.6)	96.6 (96.0–97.2)	96.7 (96.0–97.3)
Breast density									
Almost entirely fat	1,697	4.5	1.9	7.7	3.5	38.5 (10.5–66.4)	83.3 (53.2–113.5)	98.4 (97.8–99.0)	98.4 (97.8–99.0)
Scattered fibroglandular	17,065	45.5	2.9	9.9	7.0	62.1 (55.3–69.0)	84.9 (78.7–91.0)	97.7 (97.4–97.9)	97.6 (97.4–97.9)
Heterogeneously dense	16,215	43.2	3.4	11.8	8.9	62.3 (56.4–68.2)	81.3 (75.6–86.9)	97.3 (96.9–97.6)	97.2 (96.9–97.6)

Characteristic	No.	%	Abnormal Interpretation Rate	Cancers Diagnosed per 1,000 Examinations		Sensitivity (95% CI)[a] (%)		Specificity (95% CI)[a] (%)	
				Within 12 Months (n = 506)	Within 6 Months (n = 360)	12 Months	6 Months	12 Months	6 Months
Extremely dense	2,553	6.8	3.3	12.5	6.7	40.6 (26.0–55.2)	76.5 (59.6–93.3)	97.2 (96.6–97.8)	97.2 (96.6–97.8)
Mammogram laterality									
Bilateral	16,608	36.9	3.0	11.1	7.6	54.9 (47.9–61.9)	77.8 (70.5–85.0)	97.6 (97.3–97.9)	97.6 (97.3–97.9)
Unilateral	28,399	63.1	3.6	11.3	8.2	63.7 (58.5–68.8)	86.3 (82.0–90.7)	97.0 (96.7–97.4)	97.0 (96.6–97.4)
Type of previous mammogram[b]									
Screening	18,411	42.0	3.2	10.4	7.5	62.0 (55.7–68.2)	84.2 (78.2–90.2)	97.4 (97.0–97.7)	97.0 (97.0–97.7)
Diagnostic evaluation after screening	21,873	50.0	3.5	11.7	8.2	60.9 (55.2–66.7)	85.0 (79.6–90.4)	97.2 (96.8–97.6)	97.2 (96.8–97.6)
Diagnostic evaluation of a breast problem	3,502	8.0	3.4	11.1	7.4	46.2 (29.6–62.7)	69.2 (51.0–87.5)	97.1 (96.3–97.8)	97.0 (96.3–97.8)

[a] CIs for sensitivity and specificity estimates account for clustering within radiologists.

[b] As indicated by radiologist.

TABLE 2

Adjusted[a] Odds Ratios (ORs) and 95% CIs for Sensitivity and Specificity by Patient Characteristics

Characteristic	Sensitivity OR (n=500)	95% CI	Specificity OR (n=44,501)	95% CI
Patient age group (y)				
<50	1.00		1.00	
50–59	1.07	0.51–2.25	0.98	0.79–1.23
60–69	0.93	0.39–2.31	0.92	0.75–1.21
≥70	1.12	0.45–2.78	1.00	0.76–1.36
Menopausal status and hormone use				
Premenopausal	1.00		1.00	
Postmenopausal, no hormone therapy	1.74	0.89–3.43	0.80	0.62–1.03
Postmenopausal, hormone therapy	1.12	0.55–2.25	0.82	0.67–1.02
Breast symptoms				
None	1.00		1.00	
Lump or other symptoms	0.84	0.51–1.38	0.85	0.70–1.03
First-degree family history of breast cancer				
No	1.00		1.00	
Yes	1.44	0.71–2.93	0.98	0.79–1.05
Breast density				
Low	1.00		1.00	
High	1.08	0.69–1.64	0.78	0.66–0.90
Mammogram laterality				
Bilateral	1.00		1.00	
Unilateral	1.58	1.08–2.29	0.82	0.69–0.97
Type of previous mammogram[b]				
Screening	1.00		1.00	
Diagnostic evaluation after screening	0.87	0.57–1.31	1.01	0.85–1.19
Diagnostic evaluation of breast problem	0.49	0.20–1.22	1.09	0.91–1.47

[a] Adjusted for patient age (continuous), menopausal status and hormone use, breast density, mammogram laterality, and clustering within radiologists.

[b] As indicated by radiologist.

TABLE 3

Characteristics of 130 Radiologists and Their Interpretive Performances on 13,907 Short-Interval Follow-Up Mammograms, 1994 – 2004

Characteristic	No.	%	Abnormal Interpretation Rate	Cancers Diagnosed per 1,000 Examinations		Sensitivity (95% CI)[a] (%)		Specificity (95% CI)[a] (%)	
				Within 12 Months (n = 137)	Within 6 Months (n = 99)	12 Months	6 Months	12 Months	6 Months
Overall	13,907	100.0	3.0	9.9	7.1	65.7 (57.7–73.7)	89.9 (83.2–96.6)	97.6 (97.2–98.1)	97.6 (97.2–98.1)
Radiologist age group (y)									
34–44	4,157	29.9	3.7	10.1	7.5	69.0 (55.9–82.2)	90.3 (78.2–102.4)	96.9 (96.0–97.9)	96.9 (95.9–97.9)
45–54	6,647	47.8	2.7	8.3	5.9	63.6 (50.8–76.4)	89.7 (79.3–100.1)	97.8 (97.3–98.3)	97.8 (97.3–98.3)
≥55	3,103	22.3	2.6	12.9	9.3	65.0 (49.0–81.0)	89.7 (76.9–102.4)	98.2 (97.7–98.8)	98.2 (97.7–98.8)
Affiliated with academic medical center									
Yes, primary appointment	403	2.9	5.2	14.9	9.9	66.7 (35.0–98.4)	100.0 (—)	95.7 (93.5–97.9)	95.7 (93.5–97.9)
Yes, adjunct or affiliate	2,370	17.0	2.6	12.2	9.7	69.0 (54.0–83.9)	87.0 (74.1–99.8)	98.2 (97.5–98.9)	98.2 (97.5–98.9)
No	11,134	80.1	3.0	9.2	6.5	64.7 (55.1–74.3)	90.3 (82.0–98.6)	97.6 (97.1–98.1)	97.6 (97.1–98.1)
Years of mammogram interpretation									
< 10	2,709	19.7	3.3	8.5	6.3	73.9 (60.5–87.3)	100.0 (—)	97.3 (96.4–98.3)	97.3 (96.4–98.3)
10–19	7,865	57.2	2.8	9.4	6.9	64.9 (53.7–76.1)	87.0 (77.1–96.9)	97.8 (97.3–98.4)	97.8 (97.2–98.4)
≥ 20	3,176	23.1	3.0	11.3	7.9	61.1 (43.5–78.7)	88.0 (73.6–102.4)	97.7 (96.9–98.5)	97.7 (96.9–98.5)
Hours per week spent in breast imaging									
< 10	4,443	32.5	2.9	6.5	4.1	48.3 (25.5–71.1)	77.8 (54.8–100.7)	97.4 (96.7–98.1)	97.4 (96.7–98.1)
≥ 10	9,212	67.5	2.9	11.1	8.3	69.6 (61.8–77.4)	92.1 (85.6–98.6)	97.8 (97.4–98.3)	97.8 (97.3–98.4)
No. of mammograms interpreted in 2001									
≤ 1,000	1,244	9.0	2.0	4.8	3.2	66.7 (27.8–105.5)	100.0 (—)	98.3 (97.2–99.4)	98.3 (97.2–99.4)
1,001–2,000	6,140	44.5	2.8	10.3	7.7	65.1 (50.8–79.3)	87.2 (75.3–99.2)	97.9 (97.3–98.4)	97.9 (97.4–98.4)
> 2,000	6,401	46.4	3.3	10.5	7.3	65.7 (57.0–74.3)	91.5 (84.2–98.8)	97.3 (96.6–98.0)	97.3 (96.6–98.0)
% of mammograms that were diagnostic									
0–24	7,265	52.6	3.3	9.6	6.7	68.6 (56.6–80.6)	95.9 (90.0–101.8)	97.4 (96.7–98.1)	97.4 (96.6–98.1)

Characteristic	No.	%	Abnormal Interpretation Rate	Cancers Diagnosed per 1,000 Examinations		Sensitivity (95% CI)[a] (%)		Specificity (95% CI)[a] (%)	
				Within 12 Months (n = 137)	Within 6 Months (n = 99)	12 Months	6 Months	12 Months	6 Months
25–49	6,121	44.3	2.4	9.5	7.2	62.1 (52.2–72.0)	81.8 (70.1–93.6)	98.2 (97.8–98.5)	98.2 (97.8–98.5)
50–100	417	3.0	6.0	16.8	9.6	57.0 (6.3–107.9)	100.0 (—)	94.9 (92.7–97.1)	94.9 (92.7–97.1)
Work full-time									
Yes	9,134	66.4	3.0	8.1	5.5	63.5 (50.7–76.3)	92.0 (84.0–100.0)	97.5 (96.9–98.0)	97.5 (96.9–98.0)
No	4,616	33.6	2.8	12.8	10.0	67.8 (58.6–77.0)	87.0 (76.0–97.9)	98.1 (97.6–98.6)	98.1 (97.6–98.6)
Sex									
Men	10,141	72.9	3.1	9.8	6.7	62.6 (52.3–73.0)	89.7 (81.7–97.8)	97.5 (97.0–98.1)	97.5 (97.0–98.1)
Women	3,766	27.1	2.7	10.1	8.2	73.7 (63.2–84.2)	90.3 (78.2–102.5)	98.0 (97.5–98.5)	98.0 (97.5–98.5)

[a] 95% CIs for sensitivity and specificity estimates account for clustering within radiologists.

TABLE 4

Adjusted[a] Odds Ratios (ORs) and 95% CIs for 12-Month Sensitivity and Specificity by Radiologist Characteristics

Characteristic	Sensitivity OR (n = 137 examinations)	95% CI	Specificity OR (n = 13,770 examinations)	95% CI	Adjusted Sensitivity OR	Adjusted Specificity OR
Radiologist age group (y)						
34–44	1.00		1.00			
45–54	0.83	0.20–3.50	1.14	0.73–1.79		
≥ 55	0.86	0.19–2.30	1.50	0.88–2.58		
Affiliated with an academic medical center						
Yes, primary appointment	1.00		1.00			
Adjunct, affiliate, or none	2.67	0.35–20.08	2.51	1.20–5.22		
Years of mammogram interpretation						
1–9	1.00		1.00			
10–19	0.24	0.09–1.31	0.90	0.54–1.51		
≥ 20	0.30	0.06–1.40	0.90	0.47–1.70		
Hours per week spent in breast imaging						
< 10	1.00		1.00			
≥ 10	3.25	1.00–10.52	1.35	0.88–2.07		
No. of mammograms interpreted in 2001						
≤ 1,000	1.00		1.00			
1,001–2,000	1.15	0.18–7.26	0.85	0.27–1.97		
> 2,000	0.73	0.13–4.11	0.62	0.28–1.37		
% of mammograms that were diagnostic						
0–24	1.00		1.00			
25–100	0.70	0.26–1.90	1.23	0.85–1.79		
Work full-time						
Yes	1.00		1.00			
No	1.18	0.43–3.28	1.21	0.83–1.75		
Sex						
Men	1.00		1.00			
Women	2.48	0.66–9.28	1.05	0.69–1.60		

[a] Adjusted for patient age (continuous), menopausal status and hormone use, breast density, mammogram laterality, radiologist age group, academic affiliation, hours spent in breast imaging, number of mammograms interpreted in 2001, percentage of mammograms that were diagnostic, work full-time, and multiple observations per radiologist.