

## Effect of Previous Benign Breast Biopsy on the Interpretive Performance of Subsequent Screening Mammography

Stephen H. Taplin, L. Abraham, B. M. Geller, B. C. Yankaskas, D. S. M. Buist, R. Smith-Bindman, C. Lehman, D. Weaver, P. A. Carney, W. E. Barlow

Manuscript received April 30, 2009; revised May 22, 2010; accepted May 26, 2010.

**Correspondence to:** Stephen H. Taplin, MD, MPH, 6130 Executive Blvd, MSC 7344, EPN 4005, Rockville, MD 20852-7344 (e-mail: taplins@mail.nih.gov).

**Background** Most breast biopsies will be negative for cancer. Benign breast biopsy can cause changes in the breast tissue, but whether such changes affect the interpretive performance of future screening mammography is not known.

**Methods** We prospectively evaluated whether self-reported benign breast biopsy was associated with reduced subsequent screening mammography performance using examination data from the mammography registries of the Breast Cancer Surveillance Consortium from January 2, 1996, through December 31, 2005. A positive interpretation was defined as a recommendation for any additional evaluation. Cancer was defined as any invasive breast cancer or ductal carcinoma in situ diagnosed within 1 year of mammography screening. Measures of mammography performance (sensitivity, specificity, and positive predictive value 1 [PPV1]) were compared both at woman level and breast level in the presence and absence of self-reported benign biopsy history. Referral to biopsy was considered a positive interpretation to calculate positive predictive value 2 (PPV2). Multivariable analysis of a correct interpretation on each performance measure was conducted after adjusting for registry, year of examination, patient characteristics, months since last mammogram, and availability of comparison film. Accuracy of the mammogram interpretation was measured using area under the receiver operating characteristic curve (AUC). All statistical tests were two-sided.

**Results** A total of 2007381 screening mammograms were identified among 799613 women, of which 14.6% mammograms were associated with self-reported previous breast biopsy. Multivariable adjusted models for mammography performance showed reduced specificity (odds ratio [OR] = 0.74, 95% confidence interval [CI] = 0.73 to 0.75,  $P < .001$ ), PPV2 (OR = 0.85, 95% CI = 0.79 to 0.92,  $P < .001$ ), and AUC (AUC 0.892 vs 0.925,  $P < .001$ ) among women with self-reported benign biopsy. There was no difference in sensitivity or PPV1 in the same adjusted models, although unadjusted differences in both were found. Specificity was lowest among women with documented fine needle aspiration—the least invasive biopsy technique (OR = 0.58, 95% CI = 0.55 to 0.61,  $P < .001$ ). Repeating the analysis among women with documented biopsy history, unilateral biopsy history, or restricted to invasive cancers did not change the results.

**Conclusions** Self-reported benign breast biopsy history was associated with statistically significantly reduced mammography performance. The difference in performance was likely because of tissue characteristics rather than the biopsy itself.

J Natl Cancer Inst 2010;102:1040–1051

Mammography is the only screening test known to reduce breast cancer mortality through the early detection of breast cancers (1,2). Breast biopsies are performed after 1%–2% of mammography screenings when suspect areas on the mammograms cannot be explained by additional imaging (3). However, cancer is not detected in 65%–75% of these biopsies (4). Total biopsy rates are two to three times higher in the United States than in the United Kingdom, despite similar cancer detection rates (5). Such high rates of biopsy demand that women and their health-care providers understand the adverse effects, if any, so that women are better informed about the effects and potential risks (6–8).

Although it is reported that breast biopsy can cause architectural changes in the breast, such as scarring and tissue distortion (9), it is unclear how these changes affect subsequent interpretive performance of screening mammography (10). One study reported that 3 years after a biopsy, 14% of the women had architectural distortion and 26% had skin distortion (11). Another study showed an association between previous breast surgery and reduced sensitivity of screening mammography (12). However, the reported reduction in the latter study (12) was not statistically significant, and the investigators did not adjust for risk factors such as breast density and months since last mammogram that can potentially

confound the association (13,14). A history of previous biopsy could also be associated with a higher mammography sensitivity because radiologists lower their threshold to call an examination abnormal. However, a prior benign biopsy may not only change subsequent interpretive sensitivity, it may also reduce specificity and therefore increase the likelihood of a subsequent false-positive test (6,15).

In this study, we evaluated the association between benign biopsies and future mammography interpretive performance using the large database of the Breast Cancer Surveillance Consortium (BCSC) (16). Our study also accounted for demographic and mammography characteristics associated with interpretive performance to determine the possible consequences of benign biopsy. We sought more precise estimates of screening mammography interpretive performance compared with previous reports (12) because we had a large number of cancer cases, data resources for prospective association between biopsy history and subsequent interpretation, and cancer occurrence.

## Methods

### Study Design

The purpose of our study was to evaluate the association between benign biopsies and future mammography interpretive performance. We used the radiologists' interpretations to calculate four measures that reflect the accuracy of mammography interpretive performance—two fundamental characteristics of a test (sensitivity and specificity), and two measures that are commonly monitored in clinical practice by the radiologists—positive predictive value 1 (PPV1) and positive predictive value 2 (PPV2). The measures are defined as follows: 1) sensitivity—the likelihood of a positive mammogram when cancer is present, 2) specificity—the likelihood of a negative mammogram when cancer is absent, 3) PPV1—the proportion of screening mammograms associated with cancer among all screening mammograms recommended for any additional evaluation, and 4) PPV2—the proportion of screening mammograms associated with cancer among all screening mammograms recommended for biopsy. We calculated a fifth characteristic (area under the receiver operating characteristic curve [AUC]) that reflects the overall accuracy of the radiologist's interpretations (17,18) and have defined it below.

For each performance measure that was used in the main analysis, we counted a woman as having cancer if ductal carcinoma in situ (DCIS) or invasive cancer was diagnosed within 1 year of the screening mammogram. If cancer was detected by a second screening mammogram performed between 9 and 12 months of the first mammogram, then the cancer was associated with the second mammogram. The study design is summarized in Figure 1.

Data on women's health history were collected by a survey at the time of each mammogram at the participating radiology facilities of the BCSC (16). We used breast biopsy history from this survey as our measure of biopsy exposure in the main analysis and evaluated the validity of our findings in secondary analyses of a subset of women who had documented biopsies. We hypothesized that changes in the breast tissue after a benign biopsy would reduce both sensitivity and specificity and thus reduce the overall accuracy. We further hypothesized that the quantity of breast tissue

---

## CONTEXT AND CAVEATS

### Prior knowledge

Breast biopsy is performed on women if additional imaging cannot explain a suspect finding detected on a mammogram, and about 65%–75% of the biopsies are negative for cancer. However, it is not known whether a benign breast biopsy affects future screening mammography interpretive performance.

### Study design

Multivariable analyses at the woman level and at the breast level included 2007381 screening mammograms to examine the association between biopsy history (self-reported and/or documented) and mammography interpretive performance by the radiologists. Data from mammography registries and the pathology database of the Breast Cancer Surveillance Consortium were used and linked with regional cancer registries for breast cancer occurrence.

### Contribution

Self-reported biopsy history was associated with reduced accuracy of mammography interpretive performance. The difference in performance was likely because of breast tissue characteristics and not the biopsy technique.

### Implications

The results may help clinicians to inform women about the potential risks of benign biopsy.

### Limitations

Mammography interpretive performance may be influenced by breast tissue characteristics that prompt a benign biopsy, as well as the biopsy itself, and their effects cannot be completely separated.

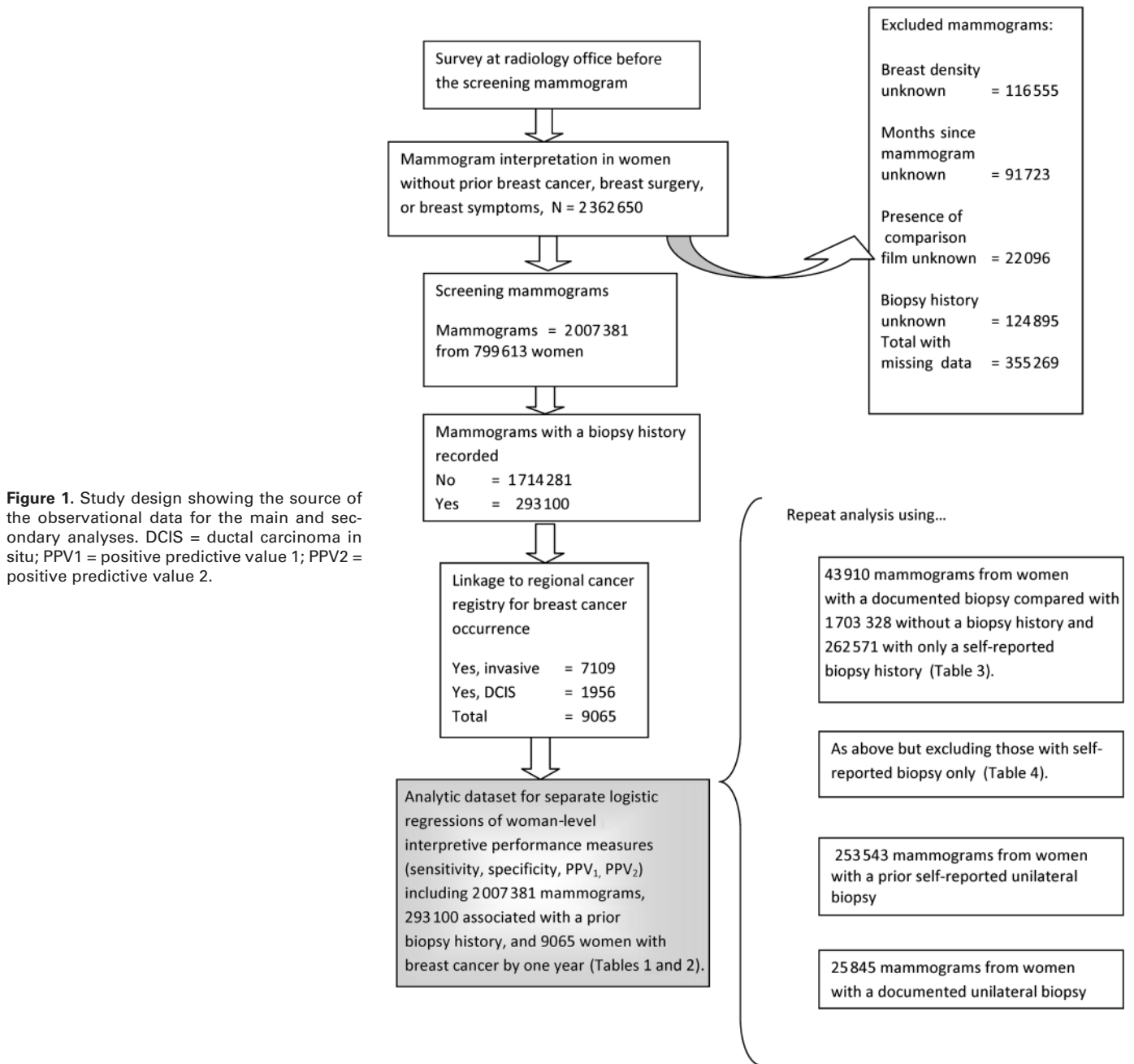
*From the Editors*

---

removed at the time of a biopsy would affect the screening mammography performance, so sensitivity would be lowest after an excisional biopsy and still reduced, but to a lesser extent, after a core biopsy.

### Setting

We used data from the mammography registries of the BCSC, which are funded by the National Cancer Institute (NCI), to link reports of radiologists' mammography interpretations with documentation of cancer occurrence in regional population-based cancer registries (<http://breastscreening.cancer.gov/>) (16). The mammography registries consist of the mammography facilities at the Group Health of Puget Sound in Washington, San Francisco Bay Area (California), Albuquerque (New Mexico), New Hampshire, Vermont, and 39 counties of North Carolina. Cancer occurrence data were collected from population-based cancer registries between January 2, 1996, and December 31, 2006. For analyses based on tissue, biopsy data were collected from the BCSC pathology database covering the populations served by the mammography registries of the BCSC (16). The BCSC pathology database contains information from the available pathology reports of the participating hospitals. The majority of the pathology reports are for biopsies done in 1995 or later; however, there are some reports also from earlier years. All mammography registries regularly sent common data elements and the cancer status of women to the Statistical Coordinating Center (SCC) of the BCSC



**Figure 1.** Study design showing the source of the observational data for the main and secondary analyses. DCIS = ductal carcinoma in situ; PPV1 = positive predictive value 1; PPV2 = positive predictive value 2.

for pooled analyses. Each mammography registry and the SCC received approval by the institutional review board for either active or passive consent process, waiver of consent to enroll participants, link data, and perform analytic studies. The SCC also received a federal Certificate of Confidentiality. All procedures were compliant with the Health Insurance Portability and Accountability Act. All registries and the SCC have protections for the identities of the women, physicians, and facilities who participated in this research (19).

### Subjects

The study included screening mammograms from January 2, 1996, through December 31, 2005, for asymptomatic women aged 40–89 years who reported their biopsy history at the time of the mammography examination. Screening mammograms were defined as bilateral examinations that the radiologist coded as “screening”

among women without any breast imaging within the previous 9 months. We included screening mammograms through December 31, 2005, to allow at least 1 year of follow-up to identify their cancer status (no breast cancer, invasive breast cancer, DCIS). We excluded mammograms from women with a history of breast cancer, previous breast surgery (mastectomy or lumpectomy for treatment of breast cancer, breast reconstruction, breast augmentation or reduction), or if the women reported breast symptoms at the time of screening mammography examination.

### Measurements

**Mammography Interpretations.** The analytic dataset included the radiologist’s initial interpretation based on two mammographic views and the final interpretation after completion of

additional breast imaging within 6 months. Tumors are most commonly missed at the initial interpretation, and scarring after a biopsy could affect the likelihood of the radiologist recommending additional imaging (recall) (20). Therefore, the initial interpretation was used to calculate sensitivity, specificity, PPV1, and overall accuracy (AUC). The final interpretation was used to calculate PPV2 because it is a commonly monitored measure (21). Radiologists recorded all interpretations using the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) assessment (0 = needs additional imaging, 1 = normal, 2 = benign finding, 3 = probably benign finding, 4 = suspicious abnormality, 5 = highly suggestive of malignancy) (17). Consistent with earlier work, BI-RADS 1, 2, or 3– (3 without a recommendation for immediate work-up or imaging) was considered a negative interpretation, and BI-RADS 0, 3+ (3 in association with a recommendation for immediate work-up or imaging), 4, or 5 was defined as a positive interpretation (P1) (14,22). After additional imaging, we used BI-RADS 4 or 5 as a positive interpretation (P2) in the calculation of PPV2 (23).

**Mammography Performance.** Based on the initial BI-RADS assessment, and whether women were diagnosed with DCIS or invasive cancer within 1 year, we classified each initial mammography interpretation as a true positive (TP1), false positive (FP1), true negative (TN1), or false negative (FN1) (14,18). To calculate the mammography performance based on the final BI-RADS assessment, we considered the interpretation after additional imaging was completed, and BI-RADS 4 and 5 were classified as positive interpretations. We calculated mammography performance measures based on the initial BI-RADS assessment as follows: sensitivity =  $TP1/(TP1 + FN1)$ , specificity =  $TN1/(FP1 + TN1)$ , and PPV1 =  $TP1/(FP1 + TP1)$ . We calculated PPV2 based on the final interpretation, and a recommendation for biopsy was considered to be a positive interpretation. We determined the number of women with cancer and a recommendation for biopsy (TP2) and the number of women without cancer among those recommended for biopsy (FP2). PPV2 was calculated as  $TP2/(FP2 + TP2)$ .

**Biopsy History.** At the time of each screening mammogram, women were asked whether they had one of the several breast procedures (fine needle or cyst aspiration, biopsy, lumpectomy, mastectomy, breast reduction, or breast implants) before screening. The type of biopsy was not always ascertained in self-reports, and in some years, at least two of the mammography facilities did not distinguish fine needle aspiration (FNA) from needle biopsy; therefore, women reporting FNA and cyst aspiration were included as biopsied women. In this study, women reporting a previous biopsy at the time of screening were designated as “biopsied” women, and those reporting no previous biopsy were designated as “unbiopsied” women. Biopsies were classified as benign if they were reported by women who did not have a breast cancer history and did not appear in the registry as having had breast cancer before the survey date. The main analysis of the association between biopsy history and sensitivity, specificity, PPV1, and AUC was performed using self-reported breast biopsy.

A secondary analysis of the association between biopsy history and the performance measures was performed using biopsies that

were recorded in the BCSC pathology database. If a woman had more than one biopsy recorded in the pathology database before a specific mammography examination, the type of biopsy was classified based on the most invasive procedure (FNA < core < excisional). A woman reporting all three biopsy types was classified under excisional biopsy.

## Statistical Analyses

**Overview.** The association between biopsy history and mammography interpretive performance was examined in two main multivariable analyses—one for the woman as a whole (woman level) and the other for the affected breast (breast level). In the main analysis, we classified women as having cancer if they had either invasive breast cancer or DCIS. In a secondary analysis, we excluded women with DCIS to see if it affected our results. We considered that the effect of biopsy could be different for the two cancer types because DCIS is more likely than invasive cancer to be calcified and therefore more easily visible despite tissue changes.

Mammograms missing one or more of the following variables were removed from the multivariable analyses because we thought they might confound the associations between biopsy and interpretive performance or were the exposure of interest—BI-RADS mammographic breast density, months since last mammogram, presence of comparison films at the time of interpretation, and biopsy history. Among the 2 362 650 eligible mammograms, 355 269 mammograms were excluded—116 555 mammograms without BI-RADS mammographic breast density, 91 723 mammograms without information on months since last mammogram, 22 096 mammograms where it was unknown if comparison films were available, and 124 895 mammograms from women who did not report their biopsy history. Among the mammograms excluded because of missing data, the age distribution was similar to those retained for the multivariable analyses. After removing all screening mammograms with missing values, 85.0% (2 007 381) of the eligible screening mammograms were included in the main multivariable analyses (Figure 1).

**Woman-Level Analysis.** The main multivariable analysis at the woman level was performed using 2 007 381 mammograms from 799 613 women (Figure 1) to evaluate the association between benign biopsy history and mammography interpretive performance. Radiologists record one BI-RADS assessment for each breast. To classify an interpretation in the woman-level analyses, we used the higher of the two BI-RADS assessments for each screening examination. We also repeated analyses of the associations between biopsy history and sensitivity, specificity, PPV1, and AUC using biopsies documented in the pathology database. As a result, we included 1 703 328 screening examinations without a previous biopsy, 43 910 screening examinations with a documented previous biopsy, and 262 571 screens with a self-reported previous biopsy that was not documented in the pathology database. Because we used both documented and self-reported biopsies, there were more screening mammograms (2 009 809) than the main analysis that relied on the self-reported biopsies alone. We further evaluated the association between the type of biopsy and mammography interpretive performance by also

including the 43 910 screens with a documented previous biopsy (Figure 1).

**Breast-Level Analysis.** Because biopsies could be done in either breast and at one or more locations within the breast, the associations between the interpretation and subsequent cancer occurrence could be inaccurate at the woman level if a self-reported biopsy occurred in a different breast than the subsequent cancer (misattribution). Furthermore, there could be adverse selection among women with a false-negative biopsy such that they are more likely to seek care from a new provider who might not be captured in the registry. Finally, we were concerned that the breast parenchymal pattern that led to a negative breast biopsy could also affect subsequent performance (parenchymal confounding). Therefore, we performed an analysis of the association between breast biopsy and interpretive performance on screening mammography at the breast level within women with a subsequent cancer.

The breast-level analysis evaluated interpretive performance based on the breast-specific BI-RADS interpretation, cancer outcome, and biopsy history for each breast (breast level) in 253 543 examinations with a self-reported history of a previous unilateral (one breast) benign breast biopsy. Each woman had two BI-RADS assessments at the time of each examination. Any cancer that occurred subsequently was associated with the specific breast in which it occurred. The sensitivity analysis compared interpretive performance in breasts with or without biopsy history in women who developed breast cancer. The specificity analysis compared interpretive performance in the biopsied and unbiopsied breasts without cancer. By conducting the breast-level analysis within women, we avoid adverse selection and mitigate the problem of misattribution because the effect of biopsy and cancer are associated in the same breast. We also mitigate the problem of parenchymal confounding to the extent that breast patterns are bilateral, and therefore, their effect on interpretive performance is similar in the biopsied and unbiopsied breasts.

The breast-level analysis was repeated using an additional 25 845 mammography examinations with previous unilateral biopsy documented in the pathology database. This independent analysis was performed to again evaluate whether the findings based on self-reported biopsy differed from the findings based on documented biopsy.

**Regression Models.** Logistic regression models were used to examine the association between self-reported biopsy history and each performance measure after accounting for differences across registries and patient characteristics influencing mammography interpretive performance. A separate logistic regression model was fit for each measure of performance to compute the odds of a correct interpretation among examinations in women reporting a biopsy history, compared with examinations among women who did not report such a history: 1) sensitivity—odds of a positive screen (BI-RADS assessment = 0, 4, 5, or 3 with immediate work-up) given a cancer diagnosis among women with a biopsy history compared with women without a history, 2) specificity—odds of a negative screen (BI-RADS assessment = 1, 2, or 3 without immediate work-up) given no cancer diagnosis among women with a biopsy history compared with women without a history, 3)

PPV1—odds of a cancer given a positive screen (BI-RADS assessment = 0, 4, 5, or 3 with immediate work-up) (PPV1) among women with a biopsy history compared with women without a history, and 4) PPV2—odds of a cancer given biopsy recommendation (BI-RADS assessment = 4, 5) (PPV2) among women with a biopsy history compared with women without a history. All analyses included year of mammogram and mammography registry as covariates and accounted for the following patient characteristics—age in 5-year intervals (14), BI-RADS breast density (1 = almost entirely fat, 2 = scattered fibroglandular, 3 = heterogeneously dense, 4 = extremely dense) (13,24), months since last mammography (no mammogram within past 59, 9–11, 12–35, 36–59) months (14), and the availability of a comparison film (no, yes) (25). To test the specific effects of race, we also performed an additional analysis of each performance measure testing for the effect of biopsy within race (white, non-Hispanic, black non-Hispanic, Asian/Pacific Islander, American Indian/Alaska Native, Other/mixed, and Hispanic) and then tested an interaction term between race and history of biopsy. We did not include family history because it has not been associated with differential mammography performance (25). The logistic regression models were fit using the SAS procedure GENMOD (26). Although multiple examinations can be included for a single woman in the analyses of specificity and PPV, we chose not to account for the inherent correlation between observations using a generalized estimating equations approach because such an approach did not change the conclusions in a previous study in a large sample of women (27), and a study by Njor et al. (28) also demonstrated the independence of mammographic examinations. All two-sided *P* values less than .05 were considered statistically significant.

**Receiver Operating Characteristic Analysis.** We conducted a receiver operating characteristic (ROC) analysis to examine the effect of “self-reported biopsy history,” “documented biopsy history,” and “no reported biopsy” on overall accuracy by estimating the area under the ROC curve (AUC). We used ordinal regression to fit an ROC model that adjusted for the following covariates—mammography registry, age at mammogram, breast density, months since last mammogram, presence of a comparison film, and year of mammogram. For this analysis, we used the initial interpretation and the following ordinal scale for the BI-RADS interpretations: 1, 2, 3 without additional evaluation; 3 with additional evaluation; and 0, 4, 5. The ROC analysis was fit using the SAS procedure NLMIXED (18,22). We compared the area under the ROC curve (AUC) for mammograms with a self-reported biopsy history and the AUC for those without such a history to determine whether self-reported biopsy history was associated with overall accuracy of performance. Likelihood ratio statistics were used to determine whether the AUCs were statistically significantly different. All two-sided *P* values less than .05 were considered statistically significant.

## Results

### Overview

We identified a total of 2 007 381 eligible screening mammograms among 799 613 women between January 2, 1996, and December 31, 2005. Among these mammograms, 9065 were associated with

breast cancer detected within 1 year of screening. The mammograms were obtained from women who were 75.9% white non-Hispanic, 6.4% black non-Hispanic, 2.7% Asian, 0.7% American Indian/Alaska Native, 1.1% Other/mixed (two or more races), 7% unknown race, and 6.3% of Hispanic ethnicity. Within 1 year of a screening mammogram, 7109 women were diagnosed with invasive breast cancer (5635 among previously unbiopsied women, 1474 among previously biopsied women), and 1956 with DCIS (1546 among previously unbiopsied women, 410 among previously biopsied women).

The characteristics of the women at the time of screening mammograms are shown in Table 1. The characteristics were compared among mammography examinations with and without a

prior benign breast biopsy and included age, breast density, months since last mammogram, presence of a comparison film, initial interpretation, and year at the time of the mammogram. Table 1 includes the screening mammograms from women who reported a prior benign biopsy at the time of the examination, and they accounted for 14.6% (n = 293 100) of all screening mammograms (N = 2 007 381). Women received between 1 and 10 screening mammograms in the analytic dataset; the mean and median numbers of mammograms were 2.5 and 2.0, respectively (data not in table). Compared with examinations in women who did not report a prior benign breast biopsy, examinations in women who reported a prior benign breast biopsy were more likely to be among those aged 50 years or older ( $P < .001$ ), with increased

**Table 1.** Description of women's characteristics by self-reported history of breast biopsy or aspiration at the time of 2007381 Breast Cancer Surveillance Consortium (BCSC) screening mammograms\*

Women and mammogram characteristics	No history of breast biopsy or aspiration (n = 1 714 281)	Self-reported history of breast biopsy or aspiration (n = 293 100)	Total screening examinations (n = 2 007 381)
	n (%)	n (%)	n (%)
<b>Age group, y†</b>			
40–44	263 154 (15.4)	23 352 (8.0)	286 506 (14.3)
45–49	284 425 (16.6)	37 293 (12.7)	321 718 (16.0)
50–54	299 837 (17.5)	49 928 (17.0)	349 765 (17.4)
55–59	236 287 (13.8)	47 751 (16.3)	284 038 (14.1)
60–64	181 755 (10.6)	39 225 (13.4)	220 980 (11.0)
65–69	154 151 (9.0)	33 588 (11.5)	187 739 (9.4)
70–74	130 877 (7.6)	28 332 (9.7)	159 209 (7.9)
75–89	163 795 (9.6)	33 631 (11.5)	197 426 (9.8)
<b>Density†</b>			
Almost entirely fat	155 682 (9.1)	16 823 (5.7)	172 505 (8.6)
Scattered fibroglandular	797 857 (46.5)	119 337 (40.7)	917 194 (45.7)
Heterogeneously dense	642 462 (37.5)	128 957 (44.0)	771 419 (38.4)
Extremely dense	118 280 (6.9)	27 983 (9.6)	146 263 (7.3)
<b>Months since last mammogram†</b>			
No mammogram within past 59 months	171 446 (10.0)	10 566 (3.6)	182 012 (9.1)
9–11	25 717 (1.5)	6 240 (2.1)	31 957 (1.6)
12–35	1 392 465 (81.2)	262 104 (89.4)	1 654 569 (82.4)
36–59	124 653 (7.3)	14 190 (4.8)	138 843 (6.9)
<b>Comparison film†</b>			
No	255 550 (14.9)	28 472 (9.7)	284 022 (14.1)
Yes	1 458 731 (85.1)	264 628 (90.3)	1 723 359 (85.9)
<b>Initial BI-RADS assessment†</b>			
1 (Normal)	1 255 593 (73.2)	182 938 (62.4)	1 438 531 (71.7)
2 (Benign finding)	291 913 (17.0)	74 147 (25.3)	366 060 (18.2)
3 (Probably benign finding)‡	21 833 (1.3)	4 023 (1.4)	25 856 (1.3)
3+ (Probably benign finding with immediate work-up)‡	6 891 (0.4)	1 359 (0.5)	8 250 (0.4)
0 (Needs additional imaging)	132 453 (7.7)	29 351 (10.0)	161 804 (8.1)
4 (Suspicious abnormality)	4 963 (0.3)	1 178 (0.4)	6 141 (0.3)
5 (Highly suggestive of malignancy)	635 (0.0)	104 (0.0)	739 (0.0)
<b>Year of mammogram</b>			
1996–1997	191 773 (11.2)	31 340 (10.7)	223 113 (11.1)
1998–1999	339 844 (19.8)	49 158 (16.8)	389 002 (19.4)
2000–2001	431 586 (25.2)	65 160 (22.2)	496 746 (24.7)
2002–2003	469 288 (27.4)	90 769 (31.0)	560 057 (27.9)
2004–2005	281 790 (16.4)	56 673 (19.3)	338 463§ (16.9)

\* Frequency distributions for each woman and mammogram characteristics computed for screening mammograms without a self-reported biopsy history or with a self-reported biopsy history and for the total of both. BI-RADS = Breast Imaging Reporting and Data System.

† All distributions differ statistically significantly between women with and without a biopsy history;  $P < .001$ , based on  $\chi^2$  test (two-sided).

‡ We separated the BI-RADS category 3 interpretations into those without and with a recommendation for additional evaluation (BI-RADS 3 and 3+, respectively).

§ The drop in number of screening examinations was because of a loss of data from a site that left the BCSC.

breast density ( $P < .001$ ), had a mammogram performed in the past 5 years ( $P < .001$ ), and had comparison films available (90.3% vs 85.1%,  $P < .001$ ). Approximately, 90.9% of the mammograms were preceded by a mammogram in the past 59 months, and approximately 84.0% of all mammograms were within the past 35 months (Table 1). Mammography recall for additional evaluation (all interpretations >BI-RADS 3 with immediate work-up) was more likely when a benign biopsy history was reported, compared with those without such a history (10.9% vs 8.4%,  $P < .001$ ). A total of 0.3% (6880 of 2 007 381) screening mammograms were associated with a recommendation for biopsy (BI-RADS 4 or 5) (Table 1). Among the documented biopsies performed before screening mammography, the proportion of core biopsies increased steadily from 34.8% in 1996 to 71.8% in 2003 (data not shown).

### Woman-Level Analysis

The radiologists' unadjusted performance measures for screening mammography in women with and without a benign biopsy history at the time of the examination are shown in Table 2. Also shown are the adjusted odds ratios (ORs) for a correct mammography interpretation in association with a benign biopsy history. A mammogram in a woman without a reported benign biopsy was always the referent. Unadjusted mammography sensitivity and specificity were lower for examinations in women reporting a biopsy at the time of a mammogram. After adjustment for age in 5-year intervals, BI-RADS breast density, presence of a comparison film, mammography registry, year of mammogram, and months since previous mammogram, mammography specificity and PPV2 were statistically significantly reduced ( $P < .001$ ) for

examinations associated with a history of a previous benign breast biopsy (Table 2).

The year of the mammogram was associated with mammography performance ( $P < .05$ ) in the multivariable models for all performance measures except PPV2, but the interaction between year of the mammogram and self-reported biopsy was only statistically significant for specificity (data not shown). We therefore kept year of the mammogram in the multivariable model but excluded the interaction term for year of mammogram and self-reported biopsy.

When we examined the effect of race on the association between previous benign biopsy and mammography interpretive performance, we found that biopsy history was not associated with differential effects across racial groups for sensitivity and PPV1 ( $P = .44$  and  $P = .56$ , respectively) but it was associated with specificity ( $P < .001$ ). Across all racial groups, the odds of a negative mammogram when no cancer was diagnosed were always lower in women with a biopsy history (OR = 0.67–0.87), compared with women without such a history, and race was associated with statistically significant differences in specificity ( $P < .001$ ) (data not shown in Table 2).

### Breast-Level Analysis

The breast-level analysis confirmed the findings of the woman-level analysis. The direction of association between biopsy history and the outcome of interest for each performance measure was the same. There were 253 543 screening mammograms among women reporting a unilateral biopsy history; 774 cancers subsequently occurred in an unbiopsied breast and 841 cancers in a biopsied

**Table 2.** Mammography performance measures and the adjusted odds ratios comparing screening mammograms with a self-reported biopsy history to screening mammograms without a self-reported biopsy history\*

Mammography performance measures	Unadjusted measures (95% CI)			Adjusted OR	
	No biopsy	Biopsy	Difference between no biopsy and biopsy (95% CI of difference)	OR* (95% CI)	P†
Sensitivity	82.3 (81.4 to 83.2)	79.6 (77.7 to 81.4)	2.70 (0.7 to 4.75)	0.88‡ (0.77 to 1.01)	.0597
Specificity	91.9 (91.8 to 91.9)	89.5 (89.4 to 89.6)	2.30 (2.21 to 2.45)	0.74§ (0.73 to 0.75)	<.0001
PPV1	4.1 (4 to 4.2)	4.7 (4.5 to 4.9)	−0.60 (−0.86 to −0.36)	1.01    (0.95 to 1.07)	.8213
PPV2	25.4 (24.8 to 26.1)	25.4 (24.2 to 26.7)	0.02 (−1.34 to 1.38)	0.85¶ (0.79 to 0.92)	<.0001

\* Odds ratios (ORs) and 95% confidence intervals (CIs) for a correct interpretation comparing screening mammograms with a self-reported biopsy history with screening mammograms without a self-reported biopsy history were computed from multivariable logistic regression models including age at mammogram (in 5-year intervals), months since last mammogram, breast density, availability of comparison film, mammography registry, and year of examination as covariates. Examinations not associated with a biopsy history (no biopsy) were the referent group and included 1 714 281 screens (7181 screens were associated with subsequent breast cancer). A total of 293 100 screening mammograms with a self-reported biopsy history were also included (1884 screens were associated with subsequent breast cancer). PPV1 = positive predictive value 1; PPV2 = positive predictive value 2.

† Two-sided  $P$  values for overall association of biopsy history with mammography performance measure, calculated using likelihood ratio tests.

‡ Odds of a positive screen (Breast Imaging Reporting and Data System [BI-RADS] = 0, 4, 5, or 3 with recommendation for additional evaluation) among screening mammograms from women with a subsequent cancer diagnosis (sensitivity), comparing screening mammograms from women with a biopsy history with screening mammograms from women without a biopsy history.

§ Odds of a negative screen (BI-RADS assessment = 1, 2, or 3 without recommendation for additional evaluation) among screening mammograms from women without a subsequent cancer diagnosis (specificity), comparing screening mammograms from women with a biopsy history with screening mammograms from women without a biopsy history.

|| Odds of a subsequent cancer among positive screening mammograms (BI-RADS = 0, 4, 5, or 3 with immediate work-up) (PPV1), comparing screening mammograms from women with a biopsy history with screening mammograms from women without a biopsy history.

¶ Odds of a subsequent cancer among screening mammograms with a recommendation for biopsy (BI-RADS = 4 or 5) (PPV2), comparing screening mammograms from women with a biopsy history with screening mammograms from women without a biopsy history.

breast. The adjusted odds of the appropriate interpretation for each performance measure were reduced for women with a previous biopsy history, compared with women with no previous biopsy: sensitivity (OR = 0.92, 95% confidence interval [CI] = 0.71 to 1.19), specificity (OR = 0.90, 95% CI = 0.88 to 0.92), PPV1 (OR = 0.97, 95% CI = 0.86 to 1.09), and PPV2 (OR = 0.84, 95% CI = 0.72 to 0.98). Similar to the woman-level analysis, the breast-level adjusted analysis showed a statistically significantly reduced association between biopsy history and performance for mammography specificity ( $P < .001$ ) and PPV2 ( $P = .025$ ) (data not shown in tables 2–4).

The analysis using documented benign biopsy also confirmed the woman-level analysis. There were 43 910 screening mammograms among women with a biopsy history documented in the BCSC pathology database, and 306 women developed cancer. Only a self-reported benign biopsy history was present in 262 571 women, including 1667 women who developed cancer (Figure 1 and Table 3). Screening mammography performance after documented benign biopsies was associated with statistically significantly lower specificity ( $P < .001$ ), PPV1 ( $P < .05$ ), and PPV2 ( $P < .001$ ) (Table 3). Cancer rates were lowest for women with no biopsy history (4.2 of 1000), compared with women with a self-reported biopsy history (6.3 of 1000) and documented biopsy history (7.0 of 1000) (data not shown in tables 2–4). PPV1 was slightly elevated for screening mammograms performed after documented biopsies, and this was also consistent with the analysis using self-reported biopsy history (Table 3).

After excluding 1962 women with DCIS from the analysis, our conclusions that only specificity and PPV2 were statistically significantly lower among women with self-reported (or documented)

previous benign biopsies remained unchanged (data not shown in tables 2–4). The odds ratios for a correct interpretation when no cancer was present (specificity) were 0.64 (95% CI = 0.62 to 0.66) for a documented previous biopsy and 0.75 (95% CI = 0.74 to 0.76) for a self-reported previous biopsy. The odds ratios for a recommendation for biopsy (PPV2) were 0.74 (95% CI = 0.61 to 0.89) for a documented previous biopsy and 0.85 (95% CI = 0.78 to 0.93) for a self-reported previous biopsy history.

When we repeated the logistic regression analysis using mammography examinations associated with documented biopsies and the types of biopsies performed, we found that compared with women with no previous biopsies, mammography specificity was statistically significantly reduced for all types of previous biopsies (Table 4). Specificity was lowest for women with previous FNAs or cyst aspirations. Women with previous core biopsies had a statistically significantly higher PPV1 but lower PPV2 compared with women without any previous biopsy (Table 4). The odds ratio of the adjusted sensitivity of a subsequent mammogram was higher among women with core and excisional biopsies but lower among those with FNA history, compared with women with no previous biopsies; however, these differences were not statistically significant for sensitivity as shown in the adjusted model (Table 4).

In the breast-level analysis, we observed that 25 845 mammography screenings among women with one previous documented unilateral benign breast biopsy were associated with 99 subsequent cancers in the biopsied breast and 90 cancers in the unbiopsied breast. Unadjusted mammography performance measures were lower in the breast with previous documented biopsy, and they remained statistically significantly reduced for specificity after

**Table 3.** Association between documented vs self-reported breast biopsy history and subsequent screening mammography performance measures\*

Mammography performance measures	Documented breast biopsy† (n = 43 910 screens),† OR (95% CI)	Self-reported breast biopsy only‡ (n = 262 571 screens),‡ OR (95% CI)	P§
Sensitivity	1.1 (0.82 to 1.53)	0.9 (0.75 to 0.99)	.0765
Specificity¶	0.64 (0.62 to 0.66)	0.75 (0.74 to 0.76)	<.001
PPV1#	1.18 (1.03 to 1.34)	0.99 (0.93 to 1.05)	<.05
PPV2**	0.78 (0.66 to 0.92)	0.87 (0.80 to 0.94)	<.001

\* Odds ratios (ORs) and 95% confidence intervals (CIs) for a correct interpretation when comparing screening mammograms with a biopsy history with screening mammograms without a biopsy history were computed from multivariable logistic regression models including age at mammogram (in 5-year intervals), months since last mammogram, breast density, availability of comparison film, mammography registry, and year of examination as covariates. The referent group included 1 703 328 mammography screenings among women without a breast biopsy, of which 7119 were associated with a subsequent breast cancer. PPV1 = positive predictive value 1; PPV2 = positive predictive value 2.

† Benign biopsies were documented in the cancer registries or pathology databases before mammograms performed from January 2, 1996, through December 31, 2005 (306 screens were associated with a subsequent cancer in a woman).

‡ Screens with self-reported previous biopsy history but have no previous biopsy documented in the pathology database (1667 screens were associated with subsequent breast cancer).

§ Two-sided  $P$  values for overall association of biopsy history with mammography performance measure were calculated using likelihood ratio tests.

|| Odds of a positive screen (Breast Imaging Reporting and Data System [BI-RADS] = 0, 4, 5, or 3 with recommendation for additional evaluation) among screening mammograms from women with a subsequent cancer diagnosis (sensitivity), comparing screening mammograms from women with a biopsy history (documented or only self-reported) with screening mammograms from women without a biopsy history.

¶ Odds of a negative screen (BI-RADS assessment = 1, 2, or 3 without recommendation for additional evaluations) among screening mammograms from women without a subsequent cancer diagnosis (specificity), comparing screening mammograms from women with a biopsy history (documented or only self-reported) with screening mammograms from women without a biopsy history.

# Odds of a subsequent cancer among positive screening examinations (BI-RADS = 0, 4, 5, or 3 with recommendation for additional evaluation) (PPV1), comparing screening mammograms from women with a biopsy history (documented or only self-reported) with screening mammograms from women without a biopsy history.

\*\* Odds of a subsequent cancer among screening mammograms with a recommendation for biopsy (BI-RADS assessment = 4,5) (PPV2), comparing screening mammograms from women with a biopsy history (documented or only self-reported) with screening mammograms from women without a biopsy history.



**Table 4.** Association between type of biopsy and subsequent mammography performance measures\*

Mammography performance measures	Documented FNA or cyst aspirations,† OR (95% CI)	Documented core biopsies,† OR (95% CI)	Documented excisional biopsies, surgical biopsy, or not otherwise specified,† OR (95% CI)	P‡
Sensitivity§	0.9 (0.55 to 1.47)	1.3 (0.87 to 2.16)	1.2 (0.56 to 3.07)	.5306
Specificity	0.58 (0.55 to 0.61)	0.69 (0.66 to 0.72)	0.66 (0.61 to 0.71)	<.001
PPV1¶	1.05 (0.82 to 1.31)	1.30 (1.09 to 1.54)	1.17 (0.8 to 1.64)	<.05
PPV2#	0.95 (0.70 to 1.28)	0.70 (0.57 to 0.86)	0.99 (0.61 to 1.55)	<.01

\* Odds ratios (ORs) and 95% confidence intervals (CIs) for a correct interpretation when comparing screening mammograms with prior documented biopsies (fine needle aspiration [FNA] or cyst aspiration, core biopsies, excisional biopsies, surgical biopsy, or not otherwise specified) with screening mammograms without a biopsy history were computed from multivariable logistic regression models including age at mammogram (in 5-year intervals), months since last mammogram, breast density, availability of comparison film, mammography registry, and year of examination as covariates. The model for sensitivity used fewer categories for age at mammogram and months since last mammogram and did not include the availability of comparison film because of small numbers. The analysis was based on 1 703 328 screens among women without a biopsy history (7119 associated with a subsequent cancer), 13 339 screens among women with a prior documented FNA or cyst aspiration (99 associated with a subsequent cancer), 22 889 screens with prior documented core biopsies (167 associated with a subsequent cancer), and 7682 screens among women with documented prior excisional biopsies, surgical biopsy, or not otherwise specified (40 associated with a subsequent breast cancer). PPV1 = positive predictive value 1; PPV2 = positive predictive value 2.

† Based on biopsies documented in the cancer registries or pathology databases before screening mammograms performed from January 2, 1996, through December 31, 2005.

‡ Two-sided *P* values for overall association of biopsy history with performance measure in the multivariate models were calculated using likelihood ratio tests.

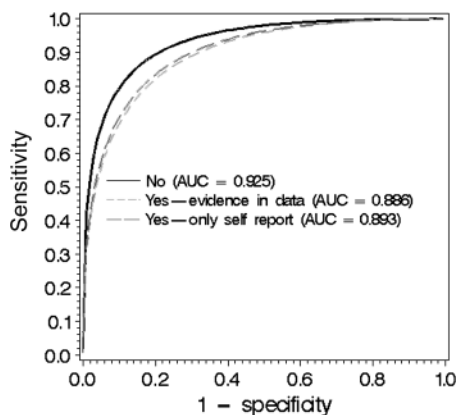
§ Odds of a positive screen (Breast Imaging Reporting and Data System [BI-RADS] = 0, 4, 5, or 3 with recommendation for additional evaluation) among screening mammograms from women with a subsequent cancer diagnosis (sensitivity), comparing examinations from women with a documented biopsy (FNA or cyst aspirations, core biopsies, excisional biopsies) with screening mammograms from women without a biopsy history.

|| Odds of a negative screen (BI-RADS assessment = 1, 2, or 3 without immediate follow-up) among screening mammograms from women without a subsequent cancer diagnosis (specificity), comparing screening mammograms from women with a documented biopsy (FNA or cyst aspirations, core biopsies, excisional biopsies) with screening mammograms from women without a biopsy history.

¶ Odds of a subsequent cancer among positive screening mammograms (BI-RADS = 0, 4, 5, or 3 with recommendation for additional evaluation) (PPV1), comparing screening mammograms from women with a documented biopsy (FNA or cyst aspirations, core biopsies, excisional biopsies) with screening mammograms from women without a biopsy history.

# Odds of a subsequent cancer among screening mammogram with a recommendation for biopsy (BI-RADS assessment = 4, 5) (PPV2), comparing screening mammograms from women with a documented biopsy (FNA or cyst aspirations, core biopsies, excisional biopsies) with screening mammograms from women without a biopsy history.

adjustment (OR = 0.82, 95% CI = 0.77 to 0.87, *P* < .001). The odds ratios of PPV1 (OR = 0.90, 95% CI = 0.65 to 1.23, *P* = .51) and PPV2 (OR = 0.75, 95% CI = 0.49 to 1.13, *P* = .18) were also reduced, but the confidence intervals were wide and the associations were not statistically significant (data not shown in tables 2–4). An adjusted logistic regression model was not run for the association between unilateral documented previous benign biopsy and mammogram sensitivity because there were too few women with documented unilateral biopsy who developed cancer.



**Figure 2.** Area under the receiver operating curve (AUC) for screening examinations associated with a biopsy history (yes vs no). Yes = self-reported biopsy history or biopsy tissue in the pathology database. No = no biopsy history.

### ROC Analysis

Overall mammography accuracy was examined using two ROC analyses to estimate the AUC of the radiologists' interpretations of screening mammograms. First, we found that accuracy was statistically significantly lower for mammograms with a self-reported biopsy history (AUC = 0.892) compared with no biopsy history (AUC = 0.925) (*P* < .001). Then, we reran the analysis including self-reported (AUC = 0.893) and documented biopsy history (AUC = 0.886), compared with no biopsy history (AUC = 0.925), and also showed a statistically significant reduction in accuracy (*P* < .001) (Figure 2).

### Discussion

To our knowledge, this is the largest analysis investigating the effect of a previous benign breast biopsy on the interpretive performance of screening mammography and accounting for other characteristics (age at the time of mammogram, time since last mammogram, breast density, availability of comparison film, mammography registry, and year of examination) that affect mammography interpretive sensitivity, specificity, PPV 1, PPV2, and overall accuracy (AUC). Consistent with our initial hypotheses, a self-reported benign breast biopsy history was associated with lower specificity, PPV2, and AUC. The findings for specificity and PPV2 were confirmed in the breast-level analysis, and the reduced odds of a correct interpretation were reinforced in analyses using a documented biopsy history and when only subsequent invasive

cancers were included. Sensitivity was not statistically significantly reduced based on either the self-reported or the documented biopsy history at the woman-level or breast-level analysis.

When documented biopsy history was examined, the association was contrary to our hypotheses—mammogram sensitivity was higher among women with core or excisional biopsies compared with less intrusive procedures like FNA. Therefore, it cannot be concluded that the biopsy itself affected sensitivity or was the cause for the observed difference in performance. Furthermore, when we evaluated specificity among women with known types of benign biopsy procedures, we found that the least intrusive biopsy technique was associated with the lowest mammogram specificity. This suggests that the intrinsic characteristics of the breast tissue, such as fibrocystic change or fibroglandular breast structure, may be responsible for interpretive differences and the biopsy history is an associated epiphenomenon. Therefore, we conclude that self-reported benign breast biopsy history is associated with lower mammography performance, although the difference is likely because of tissue characteristics rather than the biopsy itself. These findings can be used by clinicians to inform women that they are more likely to have a false-positive subsequent mammographic examination if they have a breast biopsy than if they do not have a biopsy. The clinicians can also inform the patients that it is unlikely that the biopsy itself will affect subsequent cancer detection.

Consistent with other work, we also demonstrate that women with a benign biopsy history are recalled for additional imaging more frequently (+2.5%) than women without such a history (29). Slanetz et al. (10) report a 1% difference in recall in their retrospective analysis of women with a biopsy history. The impact of a 1%–3% difference in recall can be estimated for women in the United States ( $n = 132\,489\,248$  women, aged >40 years, based on US Census estimates) assuming that 66% have been screened ( $n = 87\,442\,903$ ) and half of these women ( $n = 43\,721\,452$ ) were screened in each of the previous 2 years. If we use these numbers for the total screened population and assume that 14.6% will have a biopsy history and that a 1%–3% increase in the recall rate is associated with this biopsy history, then between 63 833 ( $0.146 \times 43\,721\,452 \times 0.01$ ) and 191 500 ( $0.146 \times 43\,721\,452 \times 0.03$ ) additional women will be recalled each year in the United States by virtue of having a previous benign biopsy.

The complexity of interpreting our findings lies in how to interpret the associations between biopsy history and subsequent mammography sensitivity. The odds ratio for detection is reduced after a negative biopsy, but the difference is small and not statistically significant. We found a 2.7% reduced unadjusted sensitivity among women with a biopsy vs no biopsy (79.6% vs 82.3%). Banks et al. (12) reported a 5.6% difference in unadjusted sensitivity among women with a history of benign biopsy compared with those with no prior benign biopsy (83.5% vs 89.1%), although the reduction was not statistically significant. Unlike the study by Banks et al. (12), we were able to control for additional factors expected to be associated with performance, such as breast density and months since last examination. The difference in sensitivity was marginally but not statistically significant for the overall association. Whereas a larger set of cancers might demonstrate an unadjusted difference, the dataset used in our study included a

large number of cancers and other data that allowed adjustment for factors confounding the association. The lack of statistical significance is important in the adjusted analysis, and the apparent reduced sensitivity among women with a previous benign biopsy reported in the previous unadjusted analysis is a limitation of observational studies.

Furthermore, the meaning of self-reported biopsy history has changed with time, so interpretation of its effect may have changed as well. For example, based on our pathology database, a much higher proportion of women had a core biopsy in 2003 (71.8%) than in 1996 (34.8%). In our main multivariable analysis, we tested for the differential effects of biopsy over time by testing an interaction term for calendar year and biopsy but did not see an association. In this analysis, we assumed a constant effect of each biopsy type over time. But since the distribution of biopsy types changed and we did not find differential effect of biopsy across calendar years, it was reasonable to conclude that the effect, if any, was small. Using the database of known biopsies, we were able to also show that core and excisional biopsies were associated with similar and not reduced sensitivity. Therefore, we concluded that the difference in mammography performance in women with previous breast biopsies was not because of biopsy techniques. An alternative explanation is that the differences in the mammography performance were because of intrinsic breast parenchymal patterns that resulted in the original negative biopsy. Because many factors cause changes in the breast tissue, it remains to be seen what factors contribute to changes that result in cancer and how they are associated with images that merit biopsy.

Because there can be a trade-off between sensitivity and specificity, we also used a single measure of accuracy—the area under the receiver operating curve (AUC). This analysis evaluates the ability of the radiologist to discriminate between healthy and diseased tissue by relating sensitivity to specificity and accounting for differences in the threshold for a positive interpretation (18). It is plausible that radiologists would lower their threshold for requesting additional evaluation or biopsy if women had a biopsy history but that should not change the AUC. Because we observed a reduced AUC in the presence of a documented or self-reported biopsy, our results demonstrate that interpretive performance in a biopsied woman appears to be different and less accurate than in an unbiopsied woman. Our findings regarding sensitivity and specificity suggest that specificity is the main cause of the difference in overall accuracy because it is persistently and significantly lower in examinations among previously biopsied women.

There are several potential limitations of this study including the use of self-reported biopsy information. Self-reported biopsy may be inaccurate, or information on the specific location of the biopsy within the breast may not be recorded and analyzed. In the main woman-level analysis, we do not know if the subsequent mammography performance was a result of changes at the location of the biopsy or a finding in another quadrant of the same breast or the other unbiopsied breast. We did not have information at the individual woman level regarding whether the mammogram was interpreted with the use of computer-assisted detection (CAD) or whether digital technology was used. However, because our analyses adjusted for the year of mammogram and we did not find

a biopsy effect that differed by year, we do not think the effects of digital mammography and CAD affected our conclusions.

Radiologists may also wonder about the use of BI-RADS 3+ interpretation because it does not appear in practice. It was used in our study to be consistent with the intent of BI-RADS to reflect an increasing level of concern, and as demonstrated elsewhere (22), the likelihood of cancer increases between BI-RADS 2, 3, and 3+. We included the 3+ category for analytic purposes to capture those mammography interpretations associated with a recommendation for additional imaging that was not recorded. Including all, the BI-RADS 3+ as a negative examination would underestimate the interpretive skills of the radiologist at the initial interpretation.

Also, there may be concerns that we identified the cancer status differentially among biopsied and unbiopsied women. However, our breast-level analysis found nearly identical results as the finding of reduced interpretive performance at the woman-level. However, in the breast-level analysis we used the biopsied and unbiopsied breasts of the same women. This within-woman analysis provides reassurance that the results are valid and not because of differential reporting or a spurious association in the woman-level analysis.

Throughout the time course of this study, many practices such as the use of CAD and the more common use of core biopsies have changed. We pointed out the increasing use of core biopsy and accounted for the year of mammogram in the analysis. The year of the mammogram was statistically significantly associated with performance measures, suggesting that those measures changed over time. We also tested the interaction between the year of mammogram and self-reported biopsy and did not find an association with the mammography performance measures. This result suggests that the reported effect of biopsy did not change over the period of the study. Factors that could have affected performance include the learning curve of the radiologist, the addition of CAD, improvements in technology, and the addition of double reading (two separate people reading the same mammogram) (30). None of these effects were accounted for in our analysis, but we do not think that they would alter our conclusions because we examined whether the association between mammography performance and history of biopsy changed with time and found no differential effects over time.

We believe that the reduced accuracy reported in this study is clinically important, regardless of whether the biopsy caused the differences, because this will help clinicians inform women about the potential adverse effects of benign biopsy. Our results showed that the unadjusted specificity was reduced by 2.3 percentage points. After adjustment for women's characteristics, evidence of reduced specificity and reduced PPV2 was statistically significant. These differences in specificity mean additional imaging evaluations and potentially more biopsies among women with a benign biopsy history, but our findings regarding sensitivity make it seem unlikely that more cancers were subsequently missed. Thus, our results could be used to prepare women with a history of a previous benign biopsy when they are being referred for their next mammogram. Although a woman with a previous benign biopsy is more likely to have cancer than someone without such a history, it can also be noted that she has a higher risk of a false-positive screening

mammogram. The message before the next mammogram for a woman with a benign biopsy history should be that a positive test must be taken seriously, but there is also a good chance of a false-positive test. Furthermore, it is not likely to affect subsequent cancer detection. Whether this mitigates anxiety at the time of a referral should be studied further because persistent anxiety is the principal long-term consequence of a false-positive mammogram (31,32).

## References

1. Humphrey LL, Helfand M, Chan BKS, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventative Services Task Force. *Ann Intern Med.* 2002;137(5):347–360.
2. Larsson LG, Andersson I, Bjurstam N, et al. Updated overview of the Swedish Randomized Trials on Breast Cancer Screening with Mammography: age group 40–49 at randomization. *J Natl Cancer Inst Monogr.* 1997;(22):57–61.
3. 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.
4. Kerlikowske K, Smith-Bindman R, Ljung BM, Grady D. Evaluation of abnormal mammography results and palpable breast abnormalities. *Ann Intern Med.* 2003;139(4):274–284.
5. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United Kingdom [Serial (Book, Monograph)]. *JAMA.* 2003;290(16):2129–2137.
6. Baines CJ. Mammography screening: are women really giving informed consent? *J Natl Cancer Inst.* 2003;95(20):1508–1511.
7. Schwartz LM, Woloshin S, Fowler FJ Jr, Welch HG. Enthusiasm for cancer screening in the United States. *JAMA.* 2004;291(1):71–78.
8. Thornton H, Edwards A, Baum M. Women need better information about routine mammography. *BMJ.* 2003;327(7406):101–103.
9. Brenner RJ, Pfaff JM. Mammographic changes after excisional breast biopsy for benign disease. *AJR Am J Roentgenol.* 1996;167(4):1047–1052.
10. Slanetz PJ, Giardino AA, McCarthy KA, et al. Previous breast biopsy for benign disease rarely complicates or alters interpretation on screening mammography. *AJR Am J Roentgenol.* 1998;170(6):1539–1541.
11. Sickles EA, Herzog KA. Mammography of the postsurgical breast. *Am J Radiol.* 1981;136(3):585–588.
12. Banks E, Reeves G, Beral V, et al. Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. *BMJ.* 2004;329(7464):477.
13. 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(3):168–175.
14. 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.
15. Elmore JG, Choe JH. Breast cancer screening for women in their 40s: moving from controversy about data to helping individual women. *Ann Intern Med.* 2007;146(7):529–531.
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. American College of Radiology. Bi-Rads®—Mammography: Assessment Categories [Electronic Citation]. Reston, VA: American College of Radiology; 2003. <http://www.birads.at/info.html>. Accessed June 30, 2008.
18. Tosteson ANA, Weinstein MC, Wittenberg J, Begg CB. ROC curve regression analysis: the use of ordinal regression models for diagnostic test assessment. *Environ Health Perspect.* 1994;102(suppl 8):73–78.
19. Carney PA, Geller BM, Moffett H, et al. Current medico-legal issues and confidentiality issues in a large multi-center research program: the National Cancer Institute's Breast Cancer Surveillance Consortium. *Am J Epidemiol.* 2000;154(4):371–378.

20. Geller BM, Vacek PM, Skelly J, Harvey SC. The use of additional imaging increased specificity and decreased sensitivity in screening mammography. *J Clin Epidemiol.* 2005;58(9):942–950.
21. Geller BM. ILBDSECPYBDMKKYKBWRR, Breast Cancer Surveillance Consortium. Improving the concordance of mammography assessment and management recommendations. *Radiology.* 2006;241(1):67–75.
22. 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.
23. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology.* 2006;241(1):55–66.
24. Laya MB, Larson EB, Taplin SH, White E. Effect of estrogen replacement therapy on the specificity and sensitivity of screening mammography. *J Natl Cancer Inst.* 1996;88(10):643–649.
25. 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.
26. SAS institute Inc. SAS/STAT User's Guide Version 8. Cary, NC.
27. Miglioretti DL, Heagerty PJ. Marginal modelling of multilevel binary data with time-varying covariates [generic]. *Biostatistics.* 2004;5(3):381–398.
28. Njor SH, Olsen AH, Schwartz W, Vejborg I, Lynge E. Predicting the risk of a false-positive test for women following a mammography screening programme. *J Med Screen.* 2007;14(2):94–97.
29. Cook AJ, Elmore JG, Miglioretti DL, et al. Decreased accuracy in interpretation of community-based screening mammography for women with multiple clinical risk factors. *J Clin Epidemiol.* 2010;63(4):441–451.
30. Miglioretti DL, Gard CC, Carney PA, et al. When radiologists perform best: the learning curve in screening mammogram interpretation. *Radiology.* 2009;253(3):632–640.
31. Barton MB, Morley DS, Moore S, et al. Decreasing women's anxieties after abnormal mammograms: a controlled trial. *J Natl Cancer Inst.* 2004; 96(7):529–538.
32. Brewer NT. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med.* 2007;146(7):502–510.

### Funding

This study was supported by a National Cancer Institute-funded Breast Cancer Surveillance Consortium (BCSC) co-operative agreement (U01CA63740 K Kerlikowske, U01CA86076 D Miglioretti, U01CA86082 T Onega, U01CA63736 G Cutter, U01CA70013 B Geller, U01CA69976 R Rosenberg, U01CA63731 D Buist, U01CA70040 B Yankaskas) and direct salary support for the first author through the National Cancer Institute.

### Notes

All opinions and findings are the sole responsibility of the authors. The views expressed do not necessarily represent those of the US Government. The collection of cancer occurrence data used in this study was provided in part by several state public health departments and cancer registries throughout the United States. For a full description of these sources and BCSC cancer registry acknowledgement, please visit <http://breastscreening.cancer.gov/work/acknowledgement.html>.

**Affiliations of authors:** Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD (SHT); Group Health Research Institute, Seattle, WA (LA, DSMB); Health Promotion Research, Vermont Cancer Center (BMG) and Department of Pathology (DW), University of Vermont, Burlington, VT; Department of Radiology, University of North Carolina, Chapel Hill, NC (BCY); Department of Radiology, Department of Epidemiology and Biostatistics, and Department of Obstetrics, Gynecology and Reproductive Medicine, University of California San Francisco, San Francisco, CA (RS-B); Seattle Cancer Care Alliance, Department of Radiology and Cancer Research and Biostatistics (CL) and Department of Biostatistics (WEB), University of Washington, Seattle, WA; Department of Family Medicine and Department of Public Health and Preventive Medicine, Oregon Health and Science University (PAC).