Effect of Observing Change from Comparison Mammograms on Performance of Screening Mammography in a Large Community-based Population[1]

Bonnie C. Yankaskas, PhD
Ryan C. May, MS
Jeanine Matuszewski, BS
J. Michael Bowling, PhD
Molly P. Jarman, MPH
Bruce F. Schroeder, MD

Purpose: To evaluate the effect of comparison mammograms on accuracy, sensitivity, specificity, positive predictive value (PPV_1), and cancer detection rate (CDR) of screening mammography to determine the role played by identification of change on comparison mammograms.

Materials and Methods: This HIPAA-compliant and institutional review board–approved prospective study was performed with waiver of patient informed consent. A total of 1 157 980 screening mammograms obtained between 1994 and 2008 in 435 183 women aged at least 40 years were included. Radiologists recorded presence of comparison mammograms and change, if seen. Women were followed for 1 year to monitor cancer occurrence. Performance measurements were calculated for screening with comparison mammograms versus screening without comparison mammograms and for screening with comparison mammograms that showed a change versus screening with comparison mammograms that did not show a change while controlling for age, breast density, and data clustering.

Results: Comparison mammograms were available in 93% of examinations. For screening with comparison mammograms versus screening without comparison mammograms, CDR per 1000 women was 3.7 versus 7.1; recall rate, 6.9% versus 14.9%; sensitivity, 78.9% versus 87.4%; specificity, 93.5% versus 85.7%; and PPV_1, 5.4% versus 4.8%. For screening with comparison mammograms that showed a change versus screening with comparison mammograms that did not show a change, CDR per 1000 women was 25.4 versus 0.8; recall rate, 41.4% versus 2.0%; sensitivity, 96.6% versus 43.5%; specificity, 60.4% versus 98.1%; and PPV_1, 6.0% versus 3.9%. Detected cancers with change were 21.1% ductal carcinoma in situ and 78.9% invasive carcinoma. Detected cancers with no change were 19.3% ductal carcinoma in situ and 80.7% invasive carcinoma.

Conclusion: Performance is affected when change from comparison mammograms is noted. Without change, sensitivity is low and specificity is high. With change, sensitivity is high, with a high false-positive rate (low specificity). Further work is needed to appreciate changes that might indicate cancer and to identify changes that are likely not indicative of cancer.

© RSNA, 2011

Supplemental material: *http://radiology.rsna.org/lookup /suppl/doi:10.1148/radiol.11110653/-/DC1*

The medical community generally accepts the notion that the availability of prior mammograms improves interpretation of screening mammograms and clinical outcome. The American College of Radiology states in their practice guidelines, "Comparison with available prior breast imaging studies is an important part of mammography," and "If previous breast imaging studies are needed for assessing mammographic findings, an attempt should be made to obtain them" (1). This recommendation is based on the findings of Burnside et al (2), who used consecutive screening mammograms obtained at fixed sites within one institution. They reported that false-positive results were significantly reduced with use of comparison mammograms; however, there

Advances in Knowledge

- Use of comparison mammograms at screening mammography results in lower recall rates (6.9% with comparison mammograms vs 14.9% without comparison mammograms) and higher specificity (93.5% with comparison mammograms vs 85.7% without comparison mammograms).

- When change from comparison mammograms is noted at screening mammography, recall rate, sensitivity, and cancer detection rate are higher than when no change is noted; however, specificity is lower, indicating a high false-positive rate.

- On the majority of screening mammograms on which change from comparison mammograms is not noted, recall rate, sensitivity, and cancer detection rate are low (2.0%, 43.5%, 0.8%, respectively, without comparison), while specificity is high.

- Excess cancers detected are mostly ductal carcinoma in situ (DCIS) when there is a change from comparison mammograms (detected cancers are 21.1% DCIS with change and 19.3% DCIS without change).

was no difference in the number of true-positive findings (2). When comparison mammograms were used, cancers were detected at an earlier stage. In another study, researchers evaluated abnormal screening mammograms obtained as part of a screening program at one institution. They compared initial screening mammograms for which no comparison mammograms were available with mammograms obtained at subsequent screening for which comparison mammograms were available. They found the abnormal rate (recall rate) was two times higher when no comparison mammograms were available, but they did not control for age. Cancers detected with comparison mammograms available had more favorable characteristics (3). There are four other studies reported in the literature in which researchers used test sets read by multiple radiologists with varying methods. All studies report that the use of comparison mammograms improves specificity, with fewer false-positive results (4–7). In none of these studies did researchers find any difference in cancer detection rates. In one report, there was no difference in sensitivity (7).

Radiologists feel more confident reporting an abnormality on a current mammogram when previous images are available for comparison (8). However, two studies in which researchers evaluated

Implications for Patient Care

- Cancers detected with or without change from comparison mammograms are similar.

- Careful review is indicated when change is noted to reduce the false-positive rate but not the sensitivity.

- The percentage of women in whom biopsy is recommended in the presence of change on comparison mammograms is close to the percentage of women in whom biopsy is recommended in the presence of no change; however, a large percentage of women with no change undergo additional imaging and do not have cancer.

the cost-effectiveness of obtaining comparison mammograms showed that the process of retrieving previous mammograms to compare with current mammograms has a high cost and yields small benefits (3,8); meanwhile, it does not provide any clinical benefit to the majority of the patients.

The objective of this study was to evaluate the effect of comparison mammograms on the accuracy, sensitivity, specificity, positive predictive value (PPV_1), and cancer detection rate of screening mammography to determine the role played by identification of change on mammograms.

Materials and Methods

This study was reviewed and approved by the biomedical institutional review board at the University of North Carolina at Chapel Hill and was compliant with the Health Insurance Portability and Accountability Act. The study received a waiver of patient informed consent.

Study Population

We identified 1 342 662 screening mammograms obtained in 504 589 women aged 40 or more years who were seen at facilities participating in the Carolina Mammography Registry between 1994

and 2008. Screening mammography was defined as a mammographic examination performed in an asymptomatic woman, as reported by the radiologist or technologist at the time of the visit. To reduce the possibility of having diagnostic mammograms included as screening mammograms, mammograms were excluded if there were not at least 9 months between a previous examination and the current examination. We excluded a total of 184682 (13.8%) mammograms for the following reasons: A total of 12333 (0.9%) mammograms were obtained in women who had a personal history of breast cancer. A total of 19136 (1.4%) mammograms were obtained in women who had breast implants. For 161148 (12.0%) mammograms, it was unknown if a comparison mammogram had been used. Of the excluded mammograms, 2492 (1.3%) were excluded for two or more reasons. The final study population included 1157980 screening mammograms obtained in 435183 women (Fig 1).

The Carolina Mammography Registry is a population-based study of screening in community practice. Data are prospectively collected from practices located in 39 counties in North Carolina. Patient demographics and imaging information, results, and recommendations are recorded for every screening examination in all women. Demographic and risk factor data are recorded directly from women's self-reports. Demographic data include date of birth, race or ethnicity, zip code, and education level. Risk factor data include personal and family history of breast cancer, previous mammography, breast symptoms, and history of breast procedures, including surgery. Women also report hormone use and menopausal status. The radiologist or technologist records specifics about the examinations, including the indication for the examination, what type of examination was performed, breast density as seen on the mammogram, mammographic findings, and recommendation for follow-up. By definition, all screening examinations were bilateral (craniocaudal and mediolateral oblique). The coding system of the Breast Imaging Reporting and Data System (BI-RADS) of the American College of Radiology was used
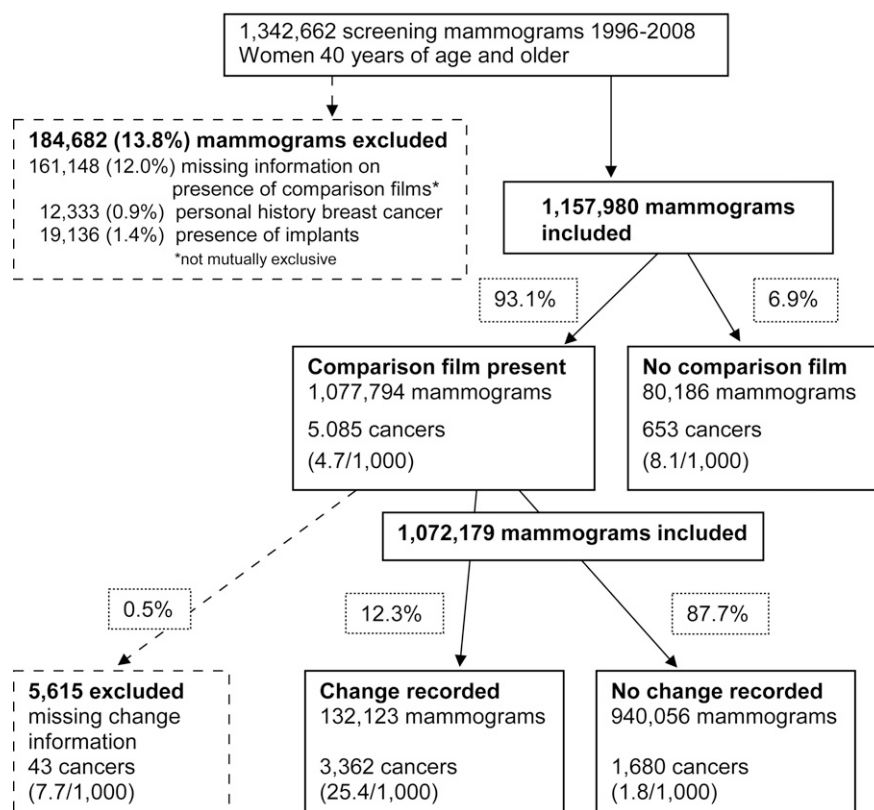
Figure 1

Figure 1: Flowchart provides an overview of study inclusion criteria, mammograms, and cancers.

to classify breast density and to assess the screening images (9). Mammographic findings were classified as negative if the BI-RADS score was 1 (normal), 2 (benign), or 3 (probably benign), unless the latter classification was associated with a recommendation for immediate work-up. Mammographic findings were considered positive if the BI-RADS score was 3 with recommendation for immediate work-up, 4 (suspicious abnormality), 5 (highly suggestive of cancer), or 0 (needs additional work-up). Recommendation for biopsy was based on the final assessment at the end of the imaging work-up. If the BI-RADS score was 4 or 5 or if biopsy or surgical consultation was recommended, recommendation for biopsy was coded *yes*; otherwise, it was coded *no*. Cancer diagnosis was determined by linking to the North Carolina Central Cancer Registry to indentify cancers diagnosed within 12 months after screening mammography and to a pathology database from

rapid case ascertainment from participating hospitals.

The radiologist was asked about the comparison mammogram and could respond as follows: *no change*, *change*, *no comparison film*, or *film pending*. There was a comparison mammogram if the radiologists responded no change or change, and there was no comparison mammogram if the radiologists responded no comparison film or film pending. Change was coded *yes* if the radiologist responded *change* and *no* if he or she responded *no change*.

Data Analysis

We will first describe the characteristics of the women whose mammograms were included in the study, then we will describe the characteristics of the women whose mammograms were excluded. Sensitivity was defined as the proportion of patients in whom breast cancer was diagnosed who had positive

mammographic findings within the previous 12 months. Specificity was defined as the proportion of patients without breast cancer who had negative mammographic findings within the previous 12 months. Sensitivity, specificity, and PPV_1 were calculated by using standard definitions (10) for mammographic interpretations with and without a comparison mammogram and then for the subgroup with comparison mammograms on which either change or no change was recorded.

We calculated unadjusted empirical estimates of cancer detection rates (per 1000 mammograms), sensitivity, specificity, and PPV_1 separately for mammograms with comparison mammograms and for those without comparison mammograms. In addition, we calculated those measures separately for mammograms that showed a change from the comparison mammogram and those that showed no change from the comparison mammogram. These measures were also calculated by using breast density, which was dichotomized (entirely fat and scattered fibrodensities vs heterogeneously dense and extremely dense). Bivariate associations between availability of comparison mammograms and age, race, family history of breast cancer, use of hormone therapy, breast density, and history of breast surgery were also examined. Similar bivariate associations were examined between changes in the comparison image and the variables of interest. χ^2 tests were used to determine if any of the bivariate associations were significant. Logistic regression analysis was used to determine whether performance measures depend on the availability of comparison images after adjusting for other variables, including age (10-year age groups) and breast density. We assumed an independent working correlation matrix and used generalized estimating equations to account for data clustering. In addition, we accounted for the nonnested structure of clustering by fitting three models with different clustering levels. The nonnested structure–adjusted standard error of the parameter estimates is calculated by using the fitted models (11). The same approach is then used to determine if performance

measures depend on whether there is change in the comparison image after adjusting for the same variables. Since the standard errors were calculated by using generalized estimating equations (11) to calculate statistical significance, a z test was performed with the parameter estimates from the logistic regression models by using generalized estimating equations and the corresponding standard errors, which were adjusted for the nonnested structure of the data. A P value of less than .05 was considered to indicate a significant difference. All statistical analysis was conducted by using SAS software (version 9.2; SAS Institute, Cary, NC).

Results

Descriptive Results

The excluded mammograms were obtained in women who were slightly younger (30% were obtained in women aged 40–49 years, 31% were obtained in women aged 50–59 years, 22% were obtained in women aged 60–69 years, and 17% were obtained in women aged ≥70 years), had a slightly different racial background (black, 19.9%; white, 75.4%; other, 4.7%); had a slightly higher family history of breast cancer (12.9%), reported less use of hormone therapy (12.8%), had a slightly different breast density distribution (extremely dense, 4.4%; heterogeneously dense, 31.6%; scattered fibrodensity, 29.8%; and entirely fat, 6.2%), and had a slightly lower history of breast surgery (20.8%) than did the overall study population.

For the descriptive data, all of the bivariate associations with whether a comparison image was available were significant ($P < .001$). Similarly, all of the bivariate associations with whether there was a change were significant ($P < .001$).

Comparison mammogram versus no comparison mammogram.—In the youngest age group (40–49 years), there were more women who had no comparison mammograms (46.3%) than there were women who had comparison mammograms (25.7%).

We adjusted the remaining characteristics by age for comparison and found the results were similar. Thus, we present the crude unadjusted distributions. When no comparison mammogram was available, the women were more likely to be black (26.6% with no comparison mammogram vs 16.7% with a comparison mammogram), less likely to have a family history of breast cancer (5.9% with no comparison mammogram vs 10.6% with comparison mammogram), less likely to be using hormone therapy (12.9% with no comparison mammogram vs 22.2% with comparison mammogram), and less likely to have a history of breast surgery (9.8% with no comparison mammogram vs 22.9% with comparison mammogram). There was little difference in the distribution of breast density (Table 1).

Change versus no change.—There were 1 077 794 (93.1%) mammograms for which a comparison mammogram was available. Of these, 132 123 (12.3%) showed a change, while 940 056 (87.7%) showed no change. The mammograms that showed a change were obtained in women who were similar in age to those with no change. For other characteristics, the percentage with change compared to the percentage with no change was: *(a)* higher in mammograms obtained in white women than in those obtained in black women, *(b)* higher in women with no family history of breast cancer than in women who had a family history of breast cancer, *(c)* higher in women who used hormone therapy than in women who did not use hormone therapy, *(d)* higher in women with less-dense breasts than in women with extremely or heterogeneously dense breasts, and *(e)* higher in women with a history of breast biopsy or surgery (Table 1). Because of the large number of images, all comparisons were significant ($P < .001$).

Cancer Outcomes and Screening Performance
Cancer rate.—A total of 5738 cancers were diagnosed within 1 year of screening mammography in the study population. Of these, 5085 were diagnosed with the aid of a comparison mammogram,

Table 1

Age-specific and Age-adjusted Characteristics Associated with Screening Mammograms for Presence and Absence of Comparison Mammograms and Change and No Change at Comparison

Characteristic	All Mammograms (n = 1 157 980)	Comparison Mammogram		Change	
		No (n = 80 186)	Yes (n = 1 077 794)	No (n = 940 056)	Yes (n = 132 123)
Age (y)					
40–49	313 881 (27.1)	37 160 (46.3)	276 721 (25.7)	239 894 (25.5)	34 607 (26.2)
50–59	355 580 (30.7)	20 411 (25.5)	335 169 (31.1)	290 550 (30.9)	42 957 (32.5)
60–69	260 509 (22.5)	12 193 (15.2)	248 316 (23.0)	216 681 (23.1)	30 651 (23.2)
≥70	228 010 (19.7)	10 422 (13.0)	217 588 (20.2)	192 931 (20.5)	23 908 (18.1)
Race					
White	859 597 (81.0)	48 104 (70.5)	811 493 (81.7)	729 869 (81.8)	77 289 (78.5)
Black	183 536 (17.3)	18 099 (26.6)	165 437 (16.7)	144 450 (16.5)	20 099 (20.4)
Other	17 933 (1.7)	1944 (2.9)	15 989 (1.6)	14 785 (1.7)	1051 (1.1)
Missing	96 914 (8.4)	12 039 (15.0)	84 875 (7.9)	50 952 (5.4)	33 684 (25.5)
Family history					
Yes	118 807 (10.3)	4694 (5.9)	114 113 (10.6)	96 316 (10.0)	17 429 (13.2)
No	1 039 173 (89.7)	75 492 (94.1)	963 681 (89.4)	843 740 (90.0)	114 694 (86.8)
Hormone therapy use					
Yes	214 902 (20.9)	9596 (12.9)	205 306 (21.5)	185 196 (22.8)	19 578 (17.6)
No	814 304 (79.1)	64 659 (87.1)	749 645 (78.5)	655 053 (77.2)	91 639 (82.4)
Missing	128 774 (11.1)	5931 (7.4)	122 843 (11.4)	99 807 (10.6)	20 906 (15.8)
Breast density					
Extremely dense	72 286 (6.3)	5122 (6.6)	67 164 (6.3)	61 086 (6.5)	5825 (4.5)
Heterogeneously dense	448 380 (39.2)	30 053 (38.5)	418 327 (39.3)	360 199 (38.7)	56 290 (43.5)
Scattered fibrodensities	557 935 (48.8)	37 381 (47.9)	520 554 (48.8)	452 981 (48.7)	64 351 (49.8)
Fatty	65 175 (5.7)	5449 (7.0)	59 726 (5.6)	56 679 (6.1)	2807 (2.2)
Missing	14 204 (1.2)	2181 (2.7)	12 023 (1.1)	9111 (1.0)	2850 (2.2)
History of breast surgery					
Yes	254 874 (22.0)	7839 (9.8)	247 035 (22.9)	219 976 (23.4)	26 048 (19.7)
No	903 106 (78.0)	72 347 (90.2)	830 759 (77.1)	720 080 (76.6)	106 075 (80.3)

Note.—Data are numbers of mammograms, and data in parentheses are percentages. A total of 435 183 women were included. Some of these women contributed multiple mammograms to the analysis. All χ^2 tests were significant for all characteristics stratified by both comparison mammogram and change ($P < .001$) due to the large numbers. Change versus no change was based on comparison mammograms.

and 653 were diagnosed without a comparison mammogram. The cancer rate per 1000 mammograms was higher when there was no comparison image than when there was a comparison image (8.1 per 1000 mammograms vs 4.7 per 1000 mammograms). The cancer rate was lower when no change was reported than when a change was reported (1.8 per 1000 mammograms vs 25.4 per 1000 mammograms) (Fig 1).

Recall rate.—The overall recall rate for screening was 7.4% (85 926 of 1 157 980 mammograms). When cancer was present, the recall rate was 79.9% (4585 of 5738 cancers). When cancer was not present, the recall rate was 7.1% (81 348 per 1 152 242 mammograms).

When no comparison mammogram was present, the recall rate of 14.9% (11 945 per 80 186 mammograms) was higher than that when a comparison mammogram was present (6.9%, 73 981 per 1 077 794 mammograms). The recall rate without a comparison mammogram was 87.4% (571 per 653 cancers) when cancer was present, which was higher than that when cancer was not present (14.3%, 11 374 per 79 533 mammograms). The recall rate with a comparison mammogram was 78.9% (4014 per 5085 mammograms) when cancer was present, which was higher than that when cancer was not present (6.5%, 69 967 per 1 072 709 mammmograms) (Table 2).

When no change was present, the recall rate was 2.0% (18 965 per 940 056 mammograms), which was lower than that when change was present (41.1%, 54 292 per 132 123 mammograms). The recall rate without change when cancer was present was 43.5% (731 of 1680 mammograms), which was higher than that when cancer was not present (1.9%, 18 234 of 938 376 mammograms). The recall rate with change when cancer was present was 96.6% (3248 of 3362 mammograms), which was higher than that when cancer was not present 39.6% (51 044 of 128 761 mammograms) (Table 2).

Sensitivity, specificity, and PPV_1.—Overall unadjusted sensitivity was 79.9%

Radiology

(4585 of 5738 cancers). Sensitivity was higher (87.4%, 571 of 653 cancers) when there was no comparison mammogram than when there was a comparison mammogram (78.9%, 4014 of 5085 cancers). Sensitivity was lower (43.5%, 731 of 1680 cancers) when no change was reported than when change was reported (96.6%, 3248 of 3362 cancers) (Fig 2).

Overall unadjusted specificity was 92.9% (1 070 901 of 1 152 242 mammograms). Specificity was lower when there was no comparison mammogram (85.7%, 68 159 of 79 533 mammograms) than when there was a comparison mammogram (93.5%, 1 002 742 of 1 072 709 mammograms); it was high when no change was reported (98.1%, 920 142 of 938 376 mammograms) and low when a change was reported (60.4%, 77 717 of 128 761 mammograms).

Overall PPV_1 was 5.3% (4585 of 85 926 mammograms). PPV_1 was 4.8% (571 of 11 945 mammograms) when there was no comparison mammogram and 5.4% (4014 of 73 981 mammograms) when there was a comparison mammogram. PPV_1 was 3.9% (731 of 1965 mammograms) when no change was reported and 6.0% (3248 of 54 292 mammograms) when a change was reported.

Logistic regression analysis, which was used to control for multiple mammograms per patient and per facility and for age and breast density, did not alter the results.

Performance stratified by breast density.—We calculated performance measures stratified by breast density with a bivariate classification of density. Comparison of *(a)* mammograms for which comparison mammograms were available with mammograms for which comparison mammograms were not available and *(b)* comparison of mammograms with a change with mammograms with no change yielded results similar to those presented. Across all measures, performance is better in fatty breasts than in dense breasts. Data are shown in Table E1 (online).

Cancer detection rates.—The overall cancer detection rate was 4.0 per 1000 screening mammograms. The cancer detection rate was higher when no

comparison mammogram was present (7.1 per 1000 screening mammograms, 571 cancers on 80 186 mammograms) than when a comparison mammogram was present (3.7 per 1000 screening mammograms, 4014 cancers on 1 077 794 mammograms). When no change was noted, with a sensitivity of 44% and a low cancer rate, the cancer detection rate was 0.8 cancers per 1000 screening mammograms (731 of 940 056 mammograms). In the 12% of comparisons in which a change was noted, higher sensitivity and cancer rate led to a cancer detection rate of 24.5 cancers per 1000 mammograms (3248 of 132 123 mammograms) (Fig 2).

Recommendation for biopsy.—We looked to see whether mammograms that revealed cancer in the presence of change and those that revealed cancer with no change led to different rates of recommendation for biopsy. With change, the biopsy recommendation rate was 6.9%; without change, the rate was 5.6%. When there was a change, 21.1% of cancers detected were ductal carcinoma in situ, and 79.0% were invasive carcinoma. When no change was noted, 19.3% of cancers detected were ductal carcinoma in situ, and 80.7% were invasive carcinoma (Table 3).

Logistic regression analysis results.—The unadjusted empirical estimates of

sensitivity and specificity did not vary greatly from the model-adjusted estimates. The model-adjusted sensitivity was 87.2% for mammography without comparison mammograms, 78.8% for mammography with comparison mammograms, 43.6% for mammograms without change, and 96.2% for mammograms with change. The model-adjusted specificity was 85.8% for mammography without comparison mammograms, 93.6% for mammography with comparison mammograms, 97.6% for mammograms without change, and 66.9% for mammograms with change. The first set of models was run to determine whether sensitivity and specificity depended on the presence of comparison mammograms. The Z statistics for the indicator of comparison mammograms for both sensitivity and specificity were significant ($P < .01$). The next set of models was run to determine whether the sensitivity and specificity depended on whether change was seen on the mammogram. The Z statistics for the indicator of change for both sensitivity and specificity were significant ($P < .01$).

Discussion

Our results enable us to confirm that having comparisons mammograms in a large community-based population leads

---

Table 2

Recall Rates for Comparison and Change by Cancer Status

Mammogram Group	All Mammograms	Comparison Mammogram		Change	
		No	Yes	No	Yes
All Mammograms					
Total no. of mammograms	1 157 980	80 186	1 077 794	940 056	132 123
No. of mammograms recalled	85 933	11 945	73 981	18 965	54 292
Recall rate (%)	7.4	14.9	6.9	2.0	41.1
Mammograms that show cancer					
Total no. of mammograms	5738	653	5085	1680	3362
No. of mammograms recalled	4585	571	4014	731	3248
Recall rate (%)	79.9	87.4	78.9	43.5	96.6
Mammograms that do not show cancer					
Total no. of mammograms	1 152 242	79 533	1 072 709	938 376	128 761
No. of mammograms recalled	81 348	11 374	69 967	18 234	51 044
Recall rate (%)	7.1	14.3	6.5	1.9	39.6

Note.—Recalled mammograms are those that were positive at screening.

---

to lower recall rates and higher overall specificity (2–4,6,7). We also found that comparison mammograms lead to lower sensitivity. For first mammograms without comparison mammograms, it is known that sensitivity will be higher and specificity will be lower because the women in whom these images were obtained are younger on average and a higher proportion of them have dense breasts (12). Comparison mammograms are reviewed to look for change, and whether or not change is noted has a large effect on the recall rates and performance measures. Recall rates were 2.2 times higher when change was noted compared with when no change was noted. When change was noted, the radiologist was more likely to recommend further work-up. This occurred about 12% of the time and resulted in high sensitivity and low specificity. When change was not noted, the finding was less likely to be recommended for work-up; this resulted in high specificity and low sensitivity.

It is important to note that comparison mammograms were available in 93% of the study mammograms and that change was noted in 12.3% of these. As others have reported, use of comparison mammograms increases specificity (4–7). This can be explained by the preponderance of comparisons in which there is no change that lead to decreased recall rates and lower false-positive rates. However, the price for this is low sensitivity in the absence of change. The group of women in whom no comparison mammograms were available (6.9% of subjects) had a larger proportion of younger women than did the group of women in whom comparison mammograms were available; thus, the former group had a higher proportion of prevalent screening mammograms. As a result, they had more cancers, a higher sensitivity, and a higher cancer detection rate.

The 12% of mammograms on which change was noted had less effect on overall accuracy when we looked at mammography with comparison mammograms versus mammography with no comparison mammograms. The cancers for which change was noted had a detec-

Figure 2

Figure 2: Graphs show performance stratified by availability of comparison mammograms and change.

tion rate of 24.5 per 1000, while the cancers for which no change was noted had a cancer detection rate of less than 1 per 1000. Thus, looking for change had the benefit of enabling us to identify cancer, as few cancers were missed. The price paid for identification of cancers was low specificity and an associated false-positive rate of 40%. If one evaluated the performance only on the basis of comparison, he or she would believe that sensitivity was lowered and specificity was increased by comparison mammograms. This is because our results are weighted by the mammograms on which there was no change (88% of mammograms). When there was no change, sensitivity was less than 50% and the cancer detection rate was less than one per 1000 mammograms. Clearly, the challenge is to identify findings that may be indicative

of cancer even though no change can be seen. With the low rate of cancer in the no-change group, this is a difficult task. The challenge is to improve discrimination, have less confidence to ignore findings when there is no change, and not immediately perform work-up when the mammogram is changed. In spite of the higher sensitivity with change, PPV_1 is still low at 7%.

Burnside et al (2) found no difference in cancer detection rates when they used comparison mammograms, but they reported that the cancers were detected at an earlier stage and had better prognostic characteristics. We found that the cancers detected when we used comparison mammograms had a much higher proportion of ductal carcinoma in situ than when no comparison mammogram was present. The result of identification
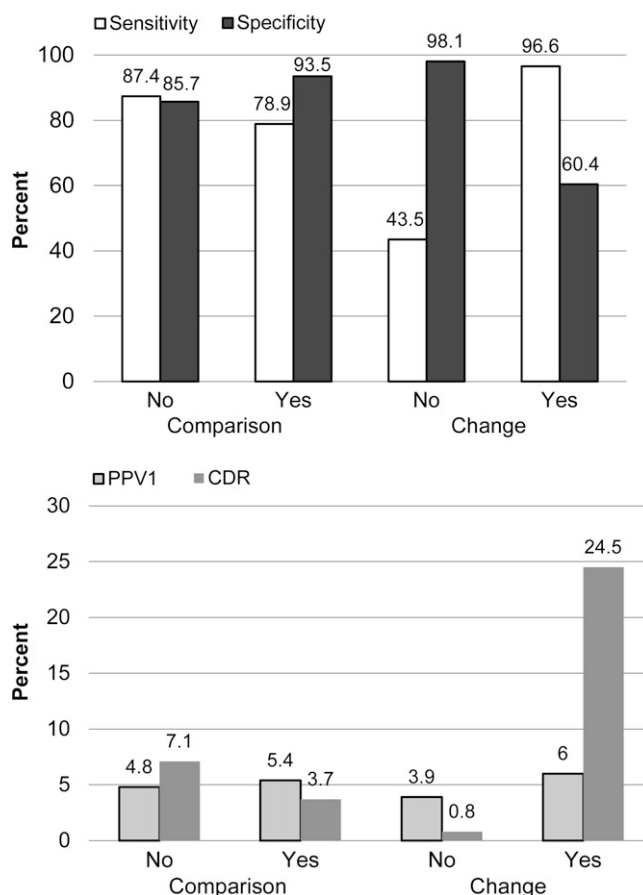
Radiology

Table 3

Positive Mammograms Recommended for Biopsy and Cancer Type by Change

A: Positive Mammograms

Mammogram Type	Comparison Mammogram		Change*	
	No	Yes	No	Yes
Positive mammogram	73 981 (100)	73 981 (100)	54 292 (100)	18 965 (100)
Recommended for biopsy				
No	69 114 (93.4)	69 114 (93.4)	50 520 (93.1)	17 904 (94.4)
Yes	4867 (6.6)	4867 (6.6)	3772 (6.9)	1061 (5.6)

B: Cancer Type

Cancer Type	Total No. of Cancers	Comparison Mammogram		Total No. of Cancers	Change*	
		No	Yes		No	Yes
Screening-detected cancer						
Ductal carcinoma in situ	939	832 (18.9)	107 (18.7)	826	141 (19.3)	685 (21.1)
Invasive	3646	3182 (81.1)	464 (81.3)	3153	590 (80.7)	2563 (78.9)
Total	4585	4401 (100)	571 (100)	3979	731 (100)	3248 (100)
Interval cancer						
Ductal carcinoma in situ	145	131 (12.2)	14 (17.1)	130	108 (11.4)	22 (19.3)
Invasive	1008	940 (87.8)	68 (82.9)	933	841 (88.6)	92 (80.7)
Total	1153	1071 (100)	82 (100)	1063	949 (100)	114 (100)

Note.—Unless otherwise indicated, data are numbers of mammograms. Data in parentheses are percentages. All χ^2 tests were significant for all characteristics stratified by both comparison mammograms and change.

* Change versus no change was based on presence of comparison mammograms = yes.

of change is a high false-positive rate (no cancer) and potentially a substantial addition to over-diagnosis of breast cancer. This should be of concern to radiologists and calls for further research.

In a population of 1 million women who undergo screening, a change will be seen on mammograms in approximately 123 000 women. Of these women, approximately 50 553 will undergo further examination. Cancer will be detected in approximately 3024 women, and approximately 47 519 women will undergo additional examinations when no cancer is present. Further examination will be performed in approximately 17 women for each cancer diagnosed. No change will be seen on approximately 877 000 mammograms. Approximately 17 693 women will undergo further examination, and approximately 885 cancers will be missed.

The strengths of this study are the population base and the number of screening mammograms. The results were obtained in actual community practice in a wide variety of settings and among a diverse population of women and radiologists. We excluded 0.5% of

the mammograms because there was no information on whether the comparison mammogram was used. The cancer rate in this group of mammograms was close to that in the group of mammograms for which there were no comparison mammograms. It is our belief that when there was no information, the data entry clerks skipped this question rather than enter *no mammogram*. Regardless, the number of these cases is small and did not alter our results. In 12% of cases, information on whether there was a comparison mammogram was missing. Unfortunately, the comparison mammogram question asks whether there is a change from the previous mammogram. The responses are *yes*, *no*, *no mammogram*, and *unknown*. We did not ask what the change was. Another study will explore in more detail whether changes are new findings, a change in density, or something else entirely, along with the actual findings associated with accuracy and change. We did not use information on the time between the screening mammogram and the comparison mammogram. We will evaluate this in our continuing work.

We conclude that use of comparison mammograms when interpreting screening mammograms is helpful but should be viewed in terms of whether change is noted. When change is noted, the false-positive rate is high. Attention needs to be paid to those recommended for further work-up to reduce the high rate of false-positive findings without lowering the high sensitivity. The reverse is true when no change is noted; the specificity is high but at the expense of low sensitivity, and a large proportion of cancers are missed in this subgroup. Future research should focus on improving sensitivity when there is no change and specificity where there is a change.

who underwent mammography at these facilities. Without these women, this work would not have been possible.

References

1. American College of Radiology. ACR practice guideline for the performance of screening and diagnostic mammography. In: Practice guidelines and technical standards. Reston, Va: American College of Radiology, 2008; 1–10.

2. Burnside ES, Sickles EA, Sohlich RE, Dee KE. Differential value of comparison with previous examinations in diagnostic versus screening mammography. AJR Am J Roentgenol 2002;179(5):1173–1177.

3. Frankel SD, Sickles EA, Curpen BN, Sollitto RA, Ominsky SH, Galvin HB. Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. AJR Am J Roentgenol 1995;164(5):1107–1109.

4. Thurfjell MG, Vitak B, Azavedo E, Svane G, Thurfjell E. Effect on sensitivity and specificity of mammography screening with or without comparison of old mammograms. Acta Radiol 2000;41(1):52–56.

5. Callaway MP, Boggis CR, Astley SA, Hutt I. The influence of previous films on screening mammographic interpretation and detection of breast carcinoma. Clin Radiol 1997;52(7):527–529.

6. Roelofs AA, Karssemeijer N, Wedekind N, et al. Importance of comparison of current and prior mammograms in breast cancer screening. Radiology 2007;242(1):70–77.

7. Sumkin JH, Holbert BL, Herrmann JS, et al. Optimal reference mammography: a comparison of mammograms obtained 1 and 2 years before the present examination. AJR Am J Roentgenol 2003;180(2):343–346.

8. Wilson TE, Nijhawan VK, Helvie MA. Normal mammograms and the practice of obtaining previous mammograms: usefulness and costs. Radiology 1996;198(3):661–663.

9. American College of Radiology. ACR BI-RADS: mammography. Reston, Va: American College of Radiology, 2003.

10. Rosenberg RD, Yankaskas BC, Hunt WC, et al. Effect of variations in operational definitions on performance estimates for screening mammography. Acad Radiol 2000;7(12): 1058–1068.

11. Miglioretti DL, Heagerty PJ. Marginal modeling of nonnested multilevel data using standard software. Am J Epidemiol 2007;165(4): 453–463.

12. Yankaskas BC, Haneuse S, Kapp JM, et al. Performance of first mammography examination in women younger than 40 years. J Natl Cancer Inst 2010;102(10):692–701.