Radiology

Effect of Background Parenchymal Enhancement on Breast MR Imaging Interpretive Performance in Community-based Practices[1]

Kimberly M. Ray, MD
Karla Kerlikowske, MD
Iryna V. Lobach, PhD
Michael B. Hofmann, MS
Heather I. Greenwood, MD
Vignesh A. Arasu, MD
Nola M. Hylton, PhD
Bonnie N. Joe, MD, PhD

Purpose: To evaluate the effect of background parenchymal enhancement (BPE) on breast magnetic resonance (MR) imaging interpretive performance in a large multi-institutional cohort with independent analysis of screening and diagnostic MR studies.

Materials and Methods: Analysis of 3770 breast MR studies was conducted. Examinations were performed in 2958 women at six participating facilities in the San Francisco Bay Area from January 2010 to October 2012. Findings were recorded prospectively in the San Francisco Mammography Registry. Performance measures were compared between studies with low BPE (mild or minimal) and those with high BPE (moderate or marked) by using binomial tests of proportions.

Results: Of 1726 MR imaging studies in the screening group, 1301 were classified as having low BPE and 425 were classified as having high BPE (75% vs 25%, respectively; $P < .001$). Of 2044 MR imaging studies in the diagnostic group, 1443 were classified as having low BPE and 601 were classified as having high BPE (71% vs 29%, respectively; $P < .001$). For low versus high BPE groups at screening, abnormal interpretation rate was 157 of 1301 versus 111 of 424 (12% vs 26%, $P < .001$); biopsy recommendation rate was 85 of 1301 versus 54 of 424 (7% vs 13%, $P < .001$); and specificity was 89% (95% confidence interval [CI]: 87, 91) versus 75% (95% CI: 71, 80) ($P = .01$). For the low versus high BPE groups at diagnostic MR imaging, biopsy recommendation rate was 325 of 1443 versus 195 of 601 (23% vs 32%, $P < .001$); and specificity was 86% (95% CI: 84, 88) versus 75% (95% CI: 74, 82) ($P < .001$). There were no significant differences between studies with low versus high BPE in sensitivity for screening (76% [95% CI: 55, 91] vs 83% [95% CI: 52, 98]; $P = .94$) or diagnostic (93% [95% CI: 87, 97] vs 96% [95% CI: 87, 99]; $P = .69$) MR imaging, nor were there significant differences in cancer detection rate per 1000 patients between the low BPE versus high BPE groups for screening (15 per 1000 vs 24 per 1000, $P = .30$) or diagnostic (78 per 1000 vs 85 per 1000, $P = .64$) MR imaging.

Conclusion: Relative to MR studies with minimal or mild BPE, those with moderate or marked BPE were associated with higher abnormal interpretation and biopsy rates and lower specificity, with no difference in cancer detection rate.

© RSNA, 2017

Online supplemental material is available for this article.

I n the 2013 edition of the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) Atlas, background parenchymal enhancement (BPE) was added to the breast magnetic resonance (MR) imaging lexicon to acknowledge that normal parenchymal enhancement may vary substantially in pattern and degree across MR imaging examinations in a manner analogous to the variation in breast density that is observed at mammography (1). BPE is affected by multiple variables, such as menopausal status, phase of the menstrual cycle,

lactation, antiestrogen therapies, and treatments for breast cancer, including chemotherapy and radiation therapy (2–8). Just as breast density reduces the sensitivity of mammography (9), it has been suggested that BPE may affect breast MR imaging interpretive performance (10–13).

Previous studies evaluated the effect of BPE on interpretive performance of breast MR imaging (11,14–16). However, these were single-institution studies with relatively small sample sizes, which limited the generalizability of results. In addition, these studies may have lacked the statistical power to detect differences in some performance metrics. Finally, in the largest prior study, combined analysis of screening and diagnostic MR imaging was performed, which may have obscured differences in outcomes across BPE categories (16). The purpose of our study was to evaluate the effect of BPE on breast MR imaging interpretive performance in a large multi-institutional cohort with independent analysis of screening and diagnostic MR imaging examinations.

Materials and Methods

Study Population

In accordance with an institutional review board–approved, Health Insurance Portability and Accountability Act–compliant protocol, we used data from the San Francisco Mammography Registry for this study. The registry participates in the National Cancer Institute Breast Cancer Surveillance Consortium (BCSC) *(https://www.bcsc-research.*

org), which standardizes and pools data across breast imaging facilities nationwide for analysis. At participating facilities, all women undergoing breast imaging are asked to complete a one-page breast health history questionnaire at the time of breast imaging. Information regarding breast imaging studies is provided by each facility. All registry and consortium data are kept confidential and are protected by a National Institutes of Health Federal Certificate of Confidentiality and Memorandum of Understanding filed with the Agency for Healthcare Research and Quality. A waiver for signed consent was obtained from the institutional review board for this research project.

The registry database was queried to identify all contrast material–enhanced breast MR imaging examinations performed at six facilities in the San Francisco Bay Area between

Author contributions:
Guarantors of integrity of entire study, K.M.R., K.K., B.N.J.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, K.M.R., K.K., H.I.G., V.A.A., B.N.J.; clinical studies, K.M.R., K.K., H.I.G., V.A.A., B.N.J.; statistical analysis, K.K., I.V.L., M.B.H.; and manuscript editing, K.M.R., K.K., I.V.L., H.I.G., V.A.A., N.M.H., B.N.J.

Conflicts of interest are listed at the end of this article.

Advances in Knowledge

- Relative to MR imaging examinations with minimal or mild background parenchymal enhancement (BPE), those with moderate or marked BPE were associated with higher abnormal interpretation rates (26% vs 12%, *P* < .001); higher biopsy recommendation rates (13% vs 6%; *P* < .001); and lower specificity (75% [95% confidence interval {CI}: 71, 80] vs 89% [95% CI: 87, 91], *P* = .01).

- Relative to diagnostic MR imaging examinations with minimal or mild BPE, those with moderate or marked BPE were associated with higher biopsy recommendation rates (32% vs 22%, *P* < .001) and lower specificity (75% [95% CI: 74, 82] vs 86% [95% CI: 84, 88], *P* < .001).

- There were no significant differences between examinations with minimal or mild BPE versus those with moderate to marked BPE in sensitivity for screening (76% [95% CI: 55, 91] vs 83% [95% CI: 52, 98], *P* = .94) or diagnostic (93% [95% CI: 87, 97] vs 96% [95% CI: 87, 99], *P* = .69) MR imaging, nor were there significant differences in cancer detection rate per 1000 patients for screening (15 per 1000 vs 24 per 1000, *P* = .30) or diagnostic (78 per 1000 vs 85 per 1000, *P* = .64) MR imaging.

Implications for Patient Care

- Moderate to marked BPE at breast MR imaging is associated with higher abnormal interpretation and biopsy rates, as well as with lowered specificity relative to minimal to mild BPE.

- Moderate to marked BPE does not reduce sensitivity of breast MR imaging in cancer detection relative to minimal to mild BPE.

January 2010 and December 2012. Facilities included one academic and five community practices; 4120 breast MR imaging studies were identified. We excluded 314 examinations in which BPE was not reported. An additional 36 examinations were excluded because the BI-RADS final assessment was not recorded, yielding our final study data set of 3770 examinations in 2958 women; 638 women had undergone two MR imaging examinations, 35 women had undergone three examinations, and 39 women had undergone more than three examinations. Because BPE may fluctuate across multiple examinations in the same woman owing to a variety of factors, such as timing of the examination relative to the menstrual cycle, menopausal status, and hormonal therapies, each examination was analyzed independently.

Indication for examination was determined by the interpreting radiologists. Examinations performed for all indications besides screening were considered diagnostic (Table E1 [online]).

Data on patient demographics, menopausal status, current use of postmenopausal hormone therapy or antiestrogen therapies, personal history of breast cancer, and first-degree family history of breast cancer were collected. Pathology results for percutaneous biopsies and surgical excisions performed within 180 days after MR imaging were obtained from radiology systems at participating facilities. In addition, the breast imaging registry links to the statewide California Cancer Registry, and studies included in this study had a minimum follow-up period of 12 months for cancer ascertainment. Cancer was defined as invasive cancer or ductal carcinoma in situ.

Statistical Analysis of Interpretive Performance

In general, imaging protocols for screening and diagnostic breast MR imaging are the same. However, screening and diagnostic MR imaging studies were analyzed separately because the probability of disease is higher among women who undergo diagnostic MR imaging.

In the ACR BI-RADS MR imaging lexicon (1), BPE refers to the physiologic enhancement of normal fibroglandular breast tissue as assessed on the initial contrast material–enhanced image of a dynamic contrast-enhanced MR study, which is the optimal phase for cancer detection. BPE may be described as minimal, mild, moderate, or marked. Because our study period preceded publication of the 2013 BI-RADS Atlas, there may have been variability in BPE assessment. Thus, we combined minimal and mild BPE into one low BPE category and moderate and marked BPE into one high BPE category to minimize the effect of interreader variability. Performance measures were compared for these two groups.

Performance measures were calculated according to the ACR BI-RADS Atlas Follow-up and Outcome Monitoring section (17). Abnormal interpretation rate was defined as the percentage of screening examinations with BI-RADS category 0, 3, 4, or 5 lesions. BI-RADS category 0 and 3 lesions were considered abnormal at screening because additional imaging evaluation was recommended. For diagnostic examinations, the abnormal interpretation rate was defined as the percentage of examinations with BI-RADS category 4 or 5 lesions and was therefore equivalent to the biopsy recommendation rate. For screening examinations, a negative examination was defined as one showing BI-RADS category 1 or 2 lesions; for diagnostic examinations, this was defined as one showing BI-RADS category 0, 1, 2, 3, or 6 lesions. The BI-RADS Atlas states only a small minority of diagnostic MR imaging examinations will warrant a final assessment of BI-RADS category 0; furthermore, the category 0 assessment should be replaced by a category 1 to category 5 assessment after additional imaging evaluation has been performed (17). In our study, only 62 of 2045 (3%) diagnostic MR imaging studies were issued a BI-RADS category 0 final overall assessment. In keeping with expert guidance from the BI-RADS committee, these were classified as negative examinations because

no biopsy recommendation was issued (E.A. Sickles, written communication, April 5, 2016).

Although final BI-RADS assessment categories were issued for each breast in the database, performance measures were calculated at the examination level. The following hierarchy was used to determine the overall BI-RADS category for each examination: BI-RADS 5 > 4 > 0 > 3 > 6 > 2 > 1.

MR imaging examinations with true-positive findings were assessed as abnormal with subsequent diagnosis of malignancy within the 12-month follow-up period. Examinations with false-positive findings were assessed as abnormal with no subsequent diagnosis of malignancy within the 12-month follow-up period. MR imaging examinations with true-negative findings were assessed as negative with no subsequent diagnosis of malignancy at 12-month follow-up. MR imaging examinations with false-negative findings were assessed as negative with subsequent malignant diagnosis at 12-month follow-up.

PPV_1 was defined as the number of true-positive screening results divided by the total number of positive examinations. PPV_2 was defined as the number of true-positive examinations divided by the total number of examinations with BI-RADS category 4 or 5 findings (biopsy recommended) at either screening or diagnostic MR imaging. PPV_3 was defined as the number of true-positive examinations divided by the total number of biopsies performed among examinations with a BI-RADS category 4 or 5 assessment.

Estimates of performance measures and their 95% confidence intervals (CIs) based on binomial distribution are reported. The measures were compared between BPE groups by using Fisher exact or χ^2 tests for comparison of proportions. Statistical comparisons for a subset of the performance measures were repeated after stratification by menopausal status, a potential confounding variable that is correlated with BPE. Comparisons were also repeated after stratification for the presence or absence of a personal history

of breast cancer and a family history of breast cancer to examine the effect of these potential confounding variables.

Post hoc analyses of sample size were conducted to examine minimum detectable differences in sensitivity, positive predictive value (PPV), and cancer detection rate (CDR) in the sample analyzed in this study. Analyses were based on binomial enumeration, 80% power, and two-sided 0.05 significance level.

All statistical analyses were performed with statistical software (R, version 3.2; R Foundation for Statistical Computing, Vienna, Austria). $P < .05$ was considered indicative of a significant difference.

Results

Characteristics of Patients

We included a total of 1726 screening breast MR imaging studies and 2044 diagnostic MR imaging studies. A total of 192 cancers were detected on the basis of these studies, excluding the index malignancies in women in the diagnostic group who had known breast cancer. Among both screening and diagnostic studies, higher BPE was associated with younger age and postmenopausal hormone therapy, whereas lower BPE was associated with postmenopausal status and antiestrogen therapy (tamoxifen or raloxifene) (Tables 1, 2). There was a greater proportion of women with a personal history of breast cancer in the high BPE group among both screening and diagnostic studies. Among the screening studies, there was a greater proportion of women with a strong family history of breast cancer in the high BPE group than in the low BPE group (Table 1).

Screening Performance

When comparing high versus low BPE groups among the screening examinations, both the abnormal interpretation rate (111 of 425 [26%] vs 157 of 1301 [12%], respectively; $P < .001$) and the biopsy recommendation rate (54 of 425 [13%] vs 85 of 1301 [7%], $P < .001$) were approximately two

times higher in the high BPE group (Table 2). There was no significant difference in CDR, sensitivity, or PPV_1. When comparing high versus low BPE groups, a trend was observed toward lower PPV_2 (11% [95% CI: 4, 23] vs 21% [95% CI: 13, 31]; $P = .25$) and PPV_3 (11% [95% CI: 4, 23] vs 26% [95% CI: 16, 38]; $P = .11$), though this difference was not significant. Specificity was lower in the high BPE group than in the low BPE group (75% [95% CI: 71, 80] vs 89% [95% CI: 87, 91]; $P = .01$).

Because there was a significant association between menopausal status

and BPE, examinations were stratified by menopausal status. Within the groups stratified by menopausal status, the rate of abnormal interpretation and biopsy recommendation remained elevated for examinations with high relative BPE to low BPE (Table 3), indicating that the effects of BPE on screening performance are independent of menopausal status. Similarly, when women were stratified by personal and family history of breast cancer—variables that were associated with low BPE and high BPE, respectively—the rate of abnormal interpretation and biopsy recommendation

---

Table 1

Characteristics of Women Undergoing Screening Breast MR Imaging

Characteristic	Minimal or Mild BPE	Moderate or Marked BPE	P Value
No. of patients	1301 (75)	425 (2)	<.001
Mean age (y)*	53 (26–83)	46 (21–76)	<.001
Postmenopausal†	792 (61)	106 (25)	<.001
Postmenopausal with hormone therapy‡	44 (3)	24 (6)	.05
Tamoxifen or raloxifene therapy§	166 (13)	27 (6)	<.001
Personal history of breast cancer	742 (57)	123 (29)	<.001
First-degree family history of breast cancer	594 (46)	250 (59)	<.001

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses.

* Data in parentheses are the range.

† Postmenopausal women were defined according to BCSC criteria as those whose periods had stopped naturally, those who had both ovaries removed, those who were on current hormone replacement therapy, or those aged at least 55 years.

‡ 535/1726 (31%) unknown.

§ 569/1726 (33%) unknown.

Table 2

Characteristics of Women Undergoing Diagnostic Breast MR Imaging

Characteristic	Minimal or Mild BPE	Moderate or Marked BPE	P Value
No. of patients	1443 (71)	601 (29)	<.001
Mean age (y)*	54 (21–91)	49 (22–79)	<.001
Postmenopausal†	805 (56)	193 (32)	<.001
Postmenopausal hormone therapy‡	56 (4)	30 (5)	.34
Tamoxifen or raloxifene therapy§	133 (9)	17 (3)	<.001
Personal history of breast cancer	1114 (77)	415 (69)	<.001
First-degree family history of breast cancer	367 (25)	162 (27)	.35

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses.

* Data in parentheses are the range.

† Postmenopausal women were defined according to BCSC criteria as those whose periods had stopped naturally, those who had both ovaries removed, those who were on current hormone replacement therapy, or those who were age 55 or older.

‡ 736/2044 (36%) unknown.

§ 798/2044 (39%) unknown.

Table 3

Performance Measures for Screening Breast MR Imaging

Performance Measure	Minimal or Mild BPE		Moderate or Marked BPE		P Value
	Quantity	95% CI	Quantity	95% CI	
No. of MR imaging examinations	1301 (75)	…	424 (25)	…	<.001
Abnormal interpretation					
No. of women	157 (12)	…	111 (26)	…	<.001
Premenopausal	84 (16)	13, 20	90 (28)	23, 34	.002
Postmenopausal	73 (9)	7, 11	21 (20)	13, 29	.006
First-degree family history of breast cancer	77 (13)	10, 15.9	78 (31)	26, 37	<.001
No first-degree family history of breast cancer	80 (11)	9, 14	33 (19)	13, 26	.03
Personal history of breast cancer	69 (9)	7, 11.7	20 (16)	10, 24	.04
No personal history of breast cancer	88 (16)	13, 19	91 (30)	25, 36	<.001
Biopsy recommended					
No. of women	85 (6)	…	54 (13)	…	<.001
Premenopausal	40 (8)	6, 11	47 (15)	11, 19	.007
Postmenopausal	45 (6)	4, 8	7 (7)	3, 14	.89
First-degree family history of breast cancer	37 (6)	4, 8	38 (15)	11, 20	<.001
No first-degree family history of breast cancer	48 (7)	5, 9	16 (9)	5, 14	.33
Personal history of breast cancer	45 (6)	4, 8	11 (9)	4, 15	.2
No personal history of breast cancer	40 (7)	5, 10	43 (14)	10, 19	.004
No. of cancers*	19 (15)	…	10 (24)	…	.30
PPV_1	0.12	0.07, 0.18	0.09	0.04, 0.16	.85
PPV_2	0.21	0.13, 0.31	0.11	0.04, 0.23	.25
PPV_3	0.26	0.16, 0.38	0.11	0.04, 0.23	.11
Sensitivity	0.76	0.55, 0.91	0.83	0.52, 0.98	.94
Specificity	0.89	0.87, 0.91	0.75	0.71, 0.80	.01

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses.

* Data in parentheses are number of cancers detected per 1000 patients.

remained elevated for examinations with high relative to low BPE (Table 3), indicating that the effects of BPE on screening performance are independent of personal and family history of breast cancer.

Diagnostic Performance

Among the diagnostic examinations, the rate of biopsy recommendation in the high BPE group was 1.4 times higher than that in the low BPE group (195 of 601 [32%] vs 325 of 1443 [23%]; $P < .001$) (Table 4). Specificity was significantly lower in the high relative to the low BPE group (75% [95% CI: 71, 78] vs 86% [95% CI: 84, 88], $P < .001$). There was no significant difference in the CDR or sensitivity between BPE groups. Relative to the low BPE group, there was a not significant trend in the high BPE group toward lower PPV_2 (37% [95% CI: 32, 43] vs 27%

[95% CI: 21, 34], $P = .09$) and PPV_3 (38% [95% CI: 29, 46] vs 49% [95% CI: 43, 56], $P = .20$).

Post hoc power analyses revealed the following: In the screening groups, our study had sufficient (80%) power to detect absolute differences in PPV_1, PPV_2, and PPV_3 of at least 10%, 15%, and 17%, respectively. In the diagnostic groups, the observed data provided sufficient power to detect an absolute difference in PPV_2 and PPV_3 between the BPE groups of at least 12% and 13%, respectively. Our data provided sufficient (80%) power to detect a difference in CDR of 27 per 1000 in the screening group and 41 per 1000 in the diagnostic group. These minimum detectable differences in PPV and CDR are larger than those observed in this study. Hence, for both PPV and CDR, larger samples are needed to provide sufficient power to test whether the

differences observed in our study are statistically significant. Although the differences we observed in sensitivity were small, the absolute values of sensitivity were high. Even if we assume perfect sensitivity of 100% for one of the study groups, there would be 7% power or lower to detect a difference from the lower diagnostic sensitivity value of 93% and 66% power or lower to detect a difference from the lower screening sensitivity value of 76%. Thus, for the small differences in sensitivity that we observed, larger sample sizes are needed to provide sufficient power to test whether these differences exceed what would be expected by chance alone.

When women were stratified by menopausal status, the rate of biopsy recommendation remained elevated in the high relative to the low BPE examinations (Table 4). When women were

Table 4

Performance Measures for Diagnostic Breast MR Imaging

Performance Measure	Minimal or Mild BPE		Moderate or Marked BPE		
	Quantity	95% CI	Quantity	95% CI	P Value
No. of MR imaging examinations	1443 (71)	…	601 (29)	…	<.001
Biopsy recommended					
No. of women	325 (23)	…	195 (32)	…	<.001
Premenopausal	175 (28)	24, 31	134 (32)	28, 37	.13
Postmenopausal	150 (18)	16, 21	61 (33)	26, 40	<.001
First-degree family history of breast cancer	70 (22)	18, 27	40 (26)	19, 34	.37
No first-degree family history of breast cancer	255 (23)	18, 22	155 (35)	30, 40	<.001
Personal history of breast cancer	263 (25)	22, 28	173 (35)	31, 39	<.001
No personal history of breast cancer	62 (16)	12, 20	22 (21)	14, 30	.27
No. of biopsies performed	227 (16)	…	136 (22)	…	<.001
No. of cancers detected*	112 (78)	…	51 (85)	…	.64
PPV_2	0.37	0.32, 0.43	0.27	0.21, 0.34	.09
PPV_3	0.49	0.43, 0.56	0.38	0.29, 0.46	.20
Sensitivity	0.93	0.87, 0.97	0.96	0.87, 0.99	.69
Specificity	0.86	0.84, 0.88	0.75	0.71, 0.78	<.001

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses.

* Data in parentheses are number of cancers detected per 1000 patients.

stratified by whether or not they had a personal history of breast cancer, the biopsy recommendation rate remained elevated in the women with diagnostic studies who had prior breast cancer (Table 4).

Discussion

In this study, we found that higher levels of BPE were associated with higher abnormal interpretation rates for screening breast MR imaging and higher rates of biopsy recommendation in both the screening and diagnostic settings. Specificity was also significantly lower among screening and diagnostic studies with higher BPE. There was a trend toward lower PPV of biopsy in screening and diagnostic studies with higher BPE, although this did not reach statistical significance.

Our results are largely consistent with those of prior studies. In a study of 250 baseline women at high risk who underwent screening MR imaging, Hambly et al (15) observed a higher abnormal interpretation rate in studies with greater than minimal BPE compared with those with minimal BPE but no significant difference in CDR. Similarly, DeMartini et al (16) evaluated 736 women who underwent screening or diagnostic MR imaging and found a higher abnormal interpretation rate among women with moderate or marked BPE compared with those with minimal BPE but no significant difference in CDR. Our results are consistent with the findings of these studies.

However, our findings of decreased specificity and a trend toward lower PPV of biopsy among screening and diagnostic MR imaging examinations with higher BPE differ from the findings of previous studies. Hambly et al (15) did not find a significant difference in the PPV of biopsy between BPE groups. DeMartini et al (16) found no significant difference in PPV or specificity between women with moderate or marked BPE and women with minimal or mild BPE. When compared with those two prior studies, our larger sample size likely increased our power to detect differences in specificity between BPE groups. Furthermore, in contrast to DeMartini et al, we audited

screening and diagnostic examinations separately because the prior probability of disease is different between these cohorts. This likely improved our ability to detect differences in performance between BPE groups. Nevertheless, the relatively small number of cancers detected in our study limited our ability to demonstrate a significant difference in PPV.

As documented in previous studies (8), we observed that younger age and premenopausal status were associated with higher BPE. To account for these variables, we stratified women by menopausal status. We found that the abnormal interpretation rate in the screening cohort and biopsy recommendation rates in both the screening cohort and the diagnostic cohort remained elevated in premenopausal women who had higher levels of BPE, indicating that BPE has an effect on interpretive performance that is independent of menopausal status.

Information regarding whether women were carriers of a breast cancer gene mutation was not available. Thus, we analyzed the available data regarding first-degree family history of breast cancer. Even in women with a first-degree family history of breast cancer, the abnormal interpretation and biopsy rates were significantly elevated in women with higher levels of BPE, indicating that BPE had an effect on screening performance that was independent of family history. Similarly, women with a personal history of breast cancer have an elevated cancer risk. However, in both the screening cohort and the diagnostic cohort, we found a higher rate of abnormal interpretation and biopsy in women with higher levels of BPE among women with a personal history of breast cancer, indicating that BPE had an effect on performance that was independent of personal breast cancer history.

In contrast to previous single-institution studies, our study used prospective data collected from a regional breast imaging registry, which represented a variety of practice settings. Consequently, our results may be more generalizable to most practices. A limitation

of this study and previous studies is the fact that clinical assessment of BPE is qualitative rather than quantitative and therefore inherently subjective. Definitions of BPE categories were added to the 2013 BI-RADS MR imaging lexicon (1), which serves as a reference standard. The literature shows that radiologist training in which standardized criteria are used improves interreader agreement in BPE assessment (18). Because our study period preceded the addition of BPE to the 2013 BI-RADS MR imaging lexicon (1), there may have been variation in BPE assessment due to lack of a published reference standard. The BPE classification criteria used by the radiologists in our study cannot ultimately be verified because this information was not recorded in the registry. There was likely greater interreader variability in BPE assessment in our study than in previous studies (15,16) because we had more radiologists at multiple facilities providing BPE assessments. To minimize the effect of interreader variability, we collapsed minimal and mild BPE into a single low BPE category and moderate and marked BPE into a single high BPE category. Despite the possibility of interreader variability, we were able to find a significant difference in specificity and biopsy rates between BPE groups.

Another limitation of the registry data is that information regarding whether prior MR imaging studies were available for comparison was missing for more than 95% of cases in our data set. Comparison with prior findings could potentially reduce the false-positive interpretation and biopsy rate for both screening and diagnostic MR imaging (19). Hambly et al (15) eliminated this confounding factor by limiting their analysis to baseline screening MR imaging examinations. To avoid including multiple examinations for the same woman, DeMartini et al (16) limited their data set to the most recent MR imaging examinations for women who had undergone multiple examination. In doing so, they excluded a subset of baseline MR imaging examinations and may have consequently overestimated MR imaging performance. Our study

consisted of a combination of baseline and subsequent MR imaging examinations, in accordance with ACR BI-RADS auditing guidelines (17). The performance measures we report are therefore more likely to reflect actual clinical performance. The abnormal interpretation rates reported in our study are either lower than or within the range of values reported previously (15,16), suggesting that the absence of prior comparisons did not significantly inflate this performance measure for either study group. Nevertheless, we cannot exclude the possibility that certain patient characteristics might be have been amplified by the inclusion of multiple examinations in a subset of the patients.

Our overall CDR of 17 per 1000 screening examinations is lower than the BI-RADS benchmark of 20–30 per 1000 screening examinations. Similarly, our overall screening sensitivity of 78% did not meet the BI-RADS benchmark threshold of greater than 80%. However, BI-RADS benchmarks are derived from expert practices involved in performance of screening MR imaging trials in women with a hereditary predisposition for breast cancer; in contrast, our data are derived from community practices with a more heterogeneous patient population. Recently, screening breast MR imaging outcome data were published by the BCSC (20), in which the San Francisco Mammography Registry is a participant. The BCSC includes practices that are geographically and ethnically representative of the U.S. population. Our CDR was identical to the BCSC-reported CDR of 17 per 1000 screening examinations, and our sensitivity was slightly lower than the BCSC-reported sensitivity of 81%. The relatively low sensitivities reported in this study and in the BCSC study also reflect the fact that a substantial number of women undergoing high-risk screening alternate mammography and MR imaging at 6-month intervals. Thus, a proportion of women who were initially screened with MR imaging also underwent a second round of screening with mammography during the 1-year follow-up interval. In accordance with BI-RADS auditing guidelines, if cancer

is diagnosed during the follow-up period, MR imaging examinations are deemed false negative. In this manner, alternating mammography with MR imaging every 6 months would be expected to lower the calculated sensitivity for either modality when the follow-up interval for the outcomes audit is 1 year.

Lastly, additional limitations of our study relate to the years of data collection and variability in MR imaging technique over time and across facilities. The MR imaging examinations were performed between 2010 and 2012, and we did not include more recent studies because 2 years are required for the state tumor registry to have complete cancer ascertainment. However, we are unaware of any substantial technical advances in clinical breast MR imaging that occurred after the study period that would render our findings less valid. Because we used registry data, we did not have access to information regarding MR imaging technique, including magnet field strength, breast coil specifications, contrast agents and injection rates, or scanning protocols. Variability in any of these factors could have affected our data.

In conclusion, moderate to marked BPE at breast MR imaging is associated with higher abnormal interpretation and biopsy rates, with no difference in cancer detection rate.

References

1. Morris EA, Comstock CE, Lee CH, et al. ACR BI-RADS Magnetic Resonance Imaging. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013.

2. Kuhl CK, Bieling HB, Gieseke J, et al. Healthy premenopausal breast parenchyma

in dynamic contrast-enhanced MR imaging of the breast: normal contrast medium enhancement and cyclical-phase dependency. Radiology 1997;203(1):137–144.

3. Müller-Schimpfle M, Ohmenhaüser K, Stoll P, Dietz K, Claussen CD. Menstrual cycle and age: influence on parenchymal contrast medium enhancement in MR imaging of the breast. Radiology 1997;203(1):145–149.

4. Delille JP, Slanetz PJ, Yeh ED, Kopans DB, Halpern EF, Garrido L. Hormone replacement therapy in postmenopausal women: breast tissue perfusion determined with MR imaging—initial observations. Radiology 2005;235(1):36–41.

5. Li J, Dershaw DD, Lee CH, Joo S, Morris EA. Breast MRI after conservation therapy: usual findings in routine follow-up examinations. AJR Am J Roentgenol 2010; 195(3):799–807.

6. King V, Kaplan J, Pike MC, et al. Impact of tamoxifen on amount of fibroglandular tissue, background parenchymal enhancement, and cysts on breast magnetic resonance imaging. Breast J 2012;18(6):527–534.

7. King V, Goldfarb SB, Brooks JD, et al. Effect of aromatase inhibitors on background parenchymal enhancement and amount of fibroglandular tissue at breast MR imaging. Radiology 2012;264(3):670–678.

8. King V, Gu Y, Kaplan JB, Brooks JD, Pike MC, Morris EA. Impact of menopausal status on background parenchymal enhancement and fibroglandular tissue on breast MRI. Eur Radiol 2012;22(12):2641–2647.

9. Kerlikowske K, Zhu W, Tosteson AN, et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. Ann Intern Med 2015;162(10):673–681.

10. Teifke A, Hlawatsch A, Beier T, et al. Undetected malignancies of the breast: dynamic contrast-enhanced MR imaging at 1.0 T. Radiology 2002;224(3):881–888.

11. Uematsu T, Kasami M, Watanabe J. Does the degree of background enhancement in breast MRI affect the detection and staging of breast cancer? Eur Radiol 2011;21(11):2261–2267.

12. Giess CS, Yeh ED, Raza S, Birdwell RL. Background parenchymal enhancement at breast MR imaging: normal patterns, diagnostic challenges, and potential for false-positive and false-negative interpretation. RadioGraphics 2014;34(1):234–247.

13. Shimauchi A, Jansen SA, Abe H, Jaskowiak N, Schmidt RA, Newstead GM. Breast cancers not detected at MRI: review of false-negative lesions. AJR Am J Roentgenol 2010;194(6):1674–1679.

14. Baltzer PA, Dietzel M, Vag T, et al. Clinical MR mammography: impact of hormonal status on background enhancement and diagnostic accuracy. Rofo 2011;183(5):441–447.

15. Hambly NM, Liberman L, Dershaw DD, Brennan S, Morris EA. Background parenchymal enhancement on baseline screening breast MRI: impact on biopsy rate and short-interval follow-up. AJR Am J Roentgenol 2011;196(1):218–224.

16. DeMartini WB, Liu F, Peacock S, Eby PR, Gutierrez RL, Lehman CD. Background parenchymal enhancement on breast MRI: impact on diagnostic performance. AJR Am J Roentgenol 2012;198(4):W373–W380.

17. Sickles EA, D'Orsi CJ. ACR BI-RADS Follow-up and Outcome Monitoring. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013.

18. Melsaether A, McDermott M, Gupta D, Pysarenko K, Shaylor SD, Moy L. Inter- and intra-reader agreement for categorization of background parenchymal enhancement at baseline and after training. AJR Am J Roentgenol 2014;203(1):209–215.

19. Abramovici G, Mainiero MB. Screening breast MR imaging: comparison of interpretation of baseline and annual follow-up studies. Radiology 2011;259(1):85–91.

20. Lee JM, Ichikawa L, Valencia E, et al. Performance benchmarks for screening breast MR imaging in community practice. Radiology 2017;285(1):44-52.