UC Davis

Title

Performance Benchmarks for Screening Breast MR Imaging in Community Practice

Authors

Lee, Janie M
Ichikawa, Laura
Valencia, Elizabeth
et al.

Radiology

Performance Benchmarks for Screening Breast MR Imaging in Community Practice[1]

Janie M. Lee, MD, MSc
Laura Ichikawa, MS
Elizabeth Valencia, MD, JD
Diana L. Miglioretti, PhD
Karen Wernli, PhD
Diana S. M. Buist, PhD, MPH
Karla Kerlikowske, MD
Louise M. Henderson, PhD, MSPH
Brian L. Sprague, PhD
Tracy Onega, PhD, MS, MA
Garth H. Rauscher, PhD
Constance D. Lehman, MD, PhD

Purpose: To compare screening magnetic resonance (MR) imaging performance in the Breast Cancer Surveillance Consortium (BCSC) with Breast Imaging Reporting and Data System (BI-RADS) benchmarks.

Materials and Methods: This study was approved by the institutional review board and compliant with HIPAA and included BCSC screening MR examinations collected between 2005 and 2013 from 5343 women (8387 MR examinations) linked to regional Surveillance, Epidemiology, and End Results program registries, state tumor registries, and pathologic information databases that identified breast cancer cases and tumor characteristics. Clinical, demographic, and imaging characteristics were assessed. Performance measures were calculated according to BI-RADS fifth edition and included cancer detection rate (CDR), positive predictive value of biopsy recommendation (PPV_2), sensitivity, and specificity.

Results: The median patient age was 52 years; 52% of MR examinations were performed in women with a first-degree family history of breast cancer, 46% in women with a personal history of breast cancer, and 15% in women with both risk factors. Screening MR imaging depicted 146 cancers, and 35 interval cancers were identified (181 total—54 in situ, 125 invasive, and two status unknown). The CDR was 17 per 1000 screening examinations (95% confidence interval [CI]: 15, 20 per 1000 screening examinations; BI-RADS benchmark, 20–30 per 1000 screening examinations). PPV_2 was 19% (95% CI: 16%, 22%; benchmark, 15%). Sensitivity was 81% (95% CI: 75%, 86%; benchmark, >80%), and specificity was 83% (95% CI: 82%, 84%; benchmark, 85%–90%). The median tumor size of invasive cancers was 10 mm; 88% were node negative.

Conclusion: The interpretative performance of screening MR imaging in the BCSC meets most BI-RADS benchmarks and approaches benchmark levels for remaining measures. Clinical practice performance data can inform ongoing benchmark development and help identify areas for quality improvement.

© RSNA, 2017

Radiology

International clinical trials of women at increased hereditary risk of developing breast cancer have shown that breast magnetic resonance (MR) imaging can allow detection of breast cancers that are not visible with mammography and are small and at an early stage (1). These screening MR imaging trials formed the evidence base for recommendations to use MR imaging as an adjunct to mammography for screening women at high risk of developing breast cancer, with a lifetime risk of 20% or higher as assessed with hereditary risk models per the American Cancer Society (2), the National Comprehensive Cancer Network (3), and the American College of Radiology (4).

Since the publication of these guidelines, the use of screening MR imaging in clinical practice has increased nationally (5,6). The use of breast MR imaging for the early detection of second breast cancers in asymptomatic women after treatment of primary breast cancer ("surveillance") has also increased, with reports from single-institution observational studies indicating incremental cancer detection beyond that of mammography alone (7–11). Currently, most breast MR examinations are being performed for screening and surveillance indications (5,6).

The American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) (12) provides a standardized lexicon to describe findings, assessments, and recommendations for mammography, ultrasonography (US), and MR examinations. It also provides guidance for conducting a medical outcomes audit of interpretive performance for quality assurance purposes, as well as performance benchmarks for screening and diagnostic mammography. The current (fifth) edition, published in 2013, added screening MR imaging performance benchmarks (13).

The BI-RADS performance benchmarks for screening and diagnostic mammography are based on data from the Breast Cancer Surveillance Consortium (BCSC) (14), a network of breast imaging registries throughout the United States that collect longitudinal breast imaging data linked to cancer outcomes. In contrast, screening MR imaging performance benchmarks are based on data from high-risk screening MR imaging trials, largely from academic medical centers, and have yet to include performance from community practice settings outside of clinical trials. The purpose of our study was to compare screening MR imaging performance in the BCSC with current BI-RADS performance benchmarks.

Materials and Methods

Data Sources

Six regional BCSC registries provided screening MR imaging data from 49 facilities in the following areas: Chicago, Western Washington (Kaiser Permanente), New Hampshire, North Carolina, the San Francisco Bay Area, and Vermont. Among BCSC facilities contributing data to this study, 12% were academic (six of 49 facilities) and contributed 33% of the MR examinations (2795 of 8387 examinations) included for analysis. Breast cancer diagnoses

Advances in Knowledge

- In a study of screening MR examinations collected between 2005 and 2013 from 5343 women (8387 MR examinations), the cancer detection rate was 17 per 1000 screening examinations (95% confidence interval [CI]: 15, 20 per 1000 screening examinations; Breast Imaging Reporting and Data System [BI-RADS] benchmark, 20–30 per 1000 screening examinations).

- The positive predictive value for biopsy recommendation was 19% (95% CI: 16%, 22%; BI-RADS benchmark, 15%).

- Sensitivity was 81% (95% CI: 75%, 86%; benchmark, >80%) and specificity was 83% (95% CI: 82%, 84%; BI-RADS benchmark, 85%–90%).

- Our results provide clinical practice–based values for measures currently noted as "To Be Determined" in the BI-RADS manual: percentage of stage 0 or 1 breast cancers (87%) and median size of invasive breast cancers (10 mm).

Implications for Patient Care

- The interpretive performance of screening breast MR imaging in clinical practices of the Breast Cancer Surveillance Consortium meets or approaches current BI-RADS benchmarks based on expert practice in clinical trials.

- Positive predictive values and MR imaging specificity (which approached but did not meet the benchmark of 85%–90%) suggest that a continued focus on reducing false-positive results while maintaining sensitivity and cancer detection remains important in ongoing quality improvement efforts.

- Clinical practice performance data can inform and supplement ongoing benchmark development.

and tumor characteristics were obtained by linkage of BCSC data to pathology databases and regional Surveillance, Epidemiology, and End Results Program or state tumor registries. Data were pooled at a central statistical coordinating center.

Registries and the statistical coordinating center received institutional review board approval for data collection and analysis. All procedures were compliant with the Health Insurance Portability and Accountability Act, and registries and the statistical coordinating center received a federal Certificate of Confidentiality for the identities of women, physicians, and facilities. Previous reports of BCSC registries and the statistical coordinating center are available at *http://www.bcsc-research.org/publications/index/html*.

Study Population

The cohort included women with and women without a personal history of breast cancer, because the BI-RADS manual recommends that breast MR examinations performed in asymptomatic women with a personal history of breast cancer be audited as screening examinations. We identified 13 312 MR examinations performed in women aged 18 years and older from 2005 to 2013 that were coded with an indication of screening. There were 6034 women with and 7278 women without a personal history of breast cancer. For women in the BCSC database who had undergone more than one screening MR examination, we included all available examinations for analysis.

Screening MR examinations in women with a personal history of breast cancer were excluded if MR imaging was performed within 6 months after cancer diagnosis ($n = 405$), if the examination was associated with previous bilateral mastectomy ($n = 143$), or if a previous MR examination had been performed within 60 days ($n = 34$). Screening MR examinations in women without a personal history of breast cancer were excluded if the examination was not bilateral ($n = 125$), if previous mastectomy was reported ($n = 134$), or if a previous MR examination

had been performed within 9 months ($n = 220$). MR examinations were also excluded if a BI-RADS category 6 (known malignancy, $n = 35$) had been assigned, if the BI-RADS assessment was missing ($n = 225$), or if less than 12 months of complete capture of cancer data were available ($n = 3523$). Because capture of data on breast cancer recurrence varies across Surveillance, Epidemiology, and End Results program and state tumor registries, we excluded MR examinations ($n = 81$) from women with a personal history of breast cancer from facilities with biopsy and pathologic data capture rates of less than 65% after BI-RADS category 4 or 5 assessments.

Demographic, clinical, and imaging characteristics examined included age, race and/or ethnicity, first-degree family history or personal history of breast cancer, previous breast biopsy, mammography within the previous 12 months, and BI-RADS breast density. A first-degree family history of breast cancer was defined as breast cancer in a mother, sister, or daughter. A personal history of breast cancer was defined as previous breast cancer, either self-reported or documented in a pathologic information database, a state tumor registry, or the Surveillance, Epidemiology, and End Results program registry. Previous mammographic examination was defined on the basis of the most recent date from either self-report or the BCSC database. Breast density was classified according to the BI-RADS manual (13) and was obtained from the most recent mammogram in the BCSC database.

Measures, Definitions, and Statistical Analysis

The MR examination was the unit of analysis, as recommended in the BI-RADS manual, to derive performance and outcomes to be used as clinically relevant benchmarks. The BI-RADS assessment categories are defined as follows (13): 0 (incomplete, need additional imaging evaluation), 1 (negative), 2 (benign), 3 (probably benign), 4 (suspicious for malignancy), and 5 (highly suggestive of malignancy). Positive

screening results were defined as those classified as BI-RADS category 0, 3, 4, or 5; negative results were defined as those classified as BI-RADS category 1 or 2. If each breast was given an assessment, we applied a hierarchical ranking of assessment categories to determine the overall assessment for the examination, as follows: $5 > 4 > 0 > 3 > 2 > 1$.

True-positive findings (screening-detected cancers) were defined as breast cancer diagnosis within 12 months of a positive screening MR examination. False-positive findings were defined as positive screening MR imaging results and no breast cancer diagnosis within 12 months. True-negative findings were defined as no breast cancer within 12 months of a negative screening MR examination. False-negative findings (interval breast cancers) were defined as diagnosis within 12 months of a negative screening MR imaging examination. The mode of interval cancer detection (screening mammography, diagnostic mammography or US, or other) and the time between negative MR imaging results and cancer diagnosis were recorded.

The following performance metrics were calculated according to the BI-RADS manual, fifth edition: cancer detection rate (CDR); positive predictive value (PPV) of positive screening result (PPV_1), PPV of biopsy recommendation (PPV_2), PPV of biopsies performed (PPV_3), sensitivity, and specificity. The CDR was calculated as the number of true-positive cancers per 1000 examinations. The current (fifth) edition of BI-RADS no longer recommends calculation of PPV after positive screening MR imaging (PPV_1). However, because data were collected while the fourth edition of BI-RADS (15) was used in clinical practice and PPV_1 calculation was recommended, we calculated PPV_1 as the percentage of category 0, 3, 4, and 5 assessments with a tissue diagnosis of cancer within the follow-up period. PPV_2 was calculated as the percentage of category 4 and 5 assessments with a tissue diagnosis of cancer within the follow-up period. PPV_3 was calculated as the percentage of biopsies performed with a tissue diagnosis of

Table 1

Characteristics of 8387 Screening MR Examinations

Characteristic	No. of Examinations
Year of examination	
2005	151 (2)
2006	537 (6)
2007	855 (10)
2008	997 (12)
2009	1097 (13)
2010	1539 (18)
2011	1587 (19)
2012	1551 (18)
2013*	73 (1)
Patient age (y)	
18–39	1013 (12)
40–49	2348 (28)
50–59	2993 (36)
60–69	1624 (19)
70+	409 (5)
Race or ethnicity[†]	
White, non-Hispanic	6559 (85)
Black, non-Hispanic	190 (3)
Hispanic	242 (3)
Asian	466 (6)
Other	261 (3)
Any mammography in previous 12 months	
No	2063 (25)
Yes	6324 (75)
First-degree family history[‡]	
No	3704 (48)
Yes	4020 (52)
Personal history of breast cancer	
No	4509 (54)
Yes	3878 (46)
Previous breast biopsy[§]	
No	2299 (29)
Yes	5600 (71)
BI-RADS breast density[||]	
Fatty	327 (4)
Scattered	2149 (29)
Heterogeneously dense	3512 (47)
Extremely dense	1420 (19)

Note.—Numbers in parentheses are percentages, which may not add up to 100% owing to rounding.

* Partial year of examinations with complete cancer follow-up.

[†] Data are missing for 669 examinations.

[‡] Data are missing for 663 examinations.

[§] Data are missing for 488 examinations.

[||] Data are missing for 979 examinations.

cancer within the follow-up period, including core-needle biopsy, fine-needle aspiration, and excisional biopsy. For auditing at the facility level as recommended in the BI-RADS manual (13), PPV_2 and PPV_3 typically have the same numerator, and the denominator for PPV_3 is that of PPV_2 after subtraction of the cases with no recorded biopsy performed. When PPV_3 was calculated within the BCSC, the denominator included examinations with confirmed biopsies preceded by a final assignment of category 4 or 5 at screening MR examination. Within this group, the subset of true-positive findings formed the numerator. Sensitivity was calculated as the percentage of true-positive results among women with cancer within the follow-up period. Specificity was calculated as the percentage of true-negative results among women without cancer within the follow-up period.

PPV_2 and PPV_3 calculations were based on final assessments after imaging follow-up. BI-RADS category 0 assessments ($n = 306$) were resolved to final assessments in the following order: (a) We used the earliest BI-RADS category 1–5 assessment identified within a 90-day follow-up period ($n = 188$); (b) we imputed the BI-RADS assessment on the basis of the recommendation in the follow-up period (as BI-RADS category 2 if recommendation was for routine screening, BI-RADS category 3 if recommendation was for short-interval follow-up, and BI-RADS category 4 if recommendation was for biopsy [$n = 6$]); (c) we imputed the BI-RADS assessment as a category 4 assessment if a biopsy was performed in the follow-up period ($n = 35$); and (d) we assigned final assessment on the basis of recommendations from the original examination ($n = 9$) or (e) we left as unresolved BI-RADS category 0 and excluded these examinations from PPV_2 and PPV_3 calculations ($n = 68$). The false-positive biopsy recommendation rate was defined as MR examinations with no known tissue diagnosis of cancer within 12 months after biopsy recommendation (BI-RADS category 4 or

5) per 1000 examinations. All other performance measures were based on the initial assessment from screening MR imaging.

The following cancer outcomes were determined overall and stratified according to screening detection and interval presentation, as follows: percentage of minimal cancer (defined as ductal carcinoma in situ [DCIS] or invasive carcinoma ≤10 mm) (13), percentage of node-negative invasive cancers, percentage of stage 0 and 1 cancers, and median size of invasive cancer. BCSC performance measures and outcomes were compared with BI-RADS fifth edition benchmarks (13).

Descriptive mean performance measures and confidence intervals (CIs) were calculated on the basis of a binomial distribution by using α of .05. All statistical analyses were performed by using software (SAS, version 9.3; SAS Institute, Cary, NC).

Results

Data were collected from 13 312 MR examinations with a screening indication. After application of exclusion criteria, the final sample was 8387 screening MR examinations performed in 5343 women from 2005 to 2013. Of the 5343 women, 3630 (68%) underwent one breast MR examination, 967 (18%) underwent two examinations, 405 (8%) underwent three examinations, and 341 (6%) underwent four or more examinations.

Table 1 summarizes the characteristics of screening MR examinations. During the study period, screening MR imaging use increased steadily between 2005 and 2010 and then remained stable through 2012. The median age of women undergoing screening MR imaging was 52 years (range, 19–92 years). MR examinations were performed in women primarily between the ages of 40 and 59 years (5341 of 8387 examinations [64%]). Examination-level race and ethnicity associated with screening MR imaging showed that 85% of examinations were performed in non-Hispanic white women (6559 of 7718 examinations) and 12% in black, Asian, or Hispanic women (898 of
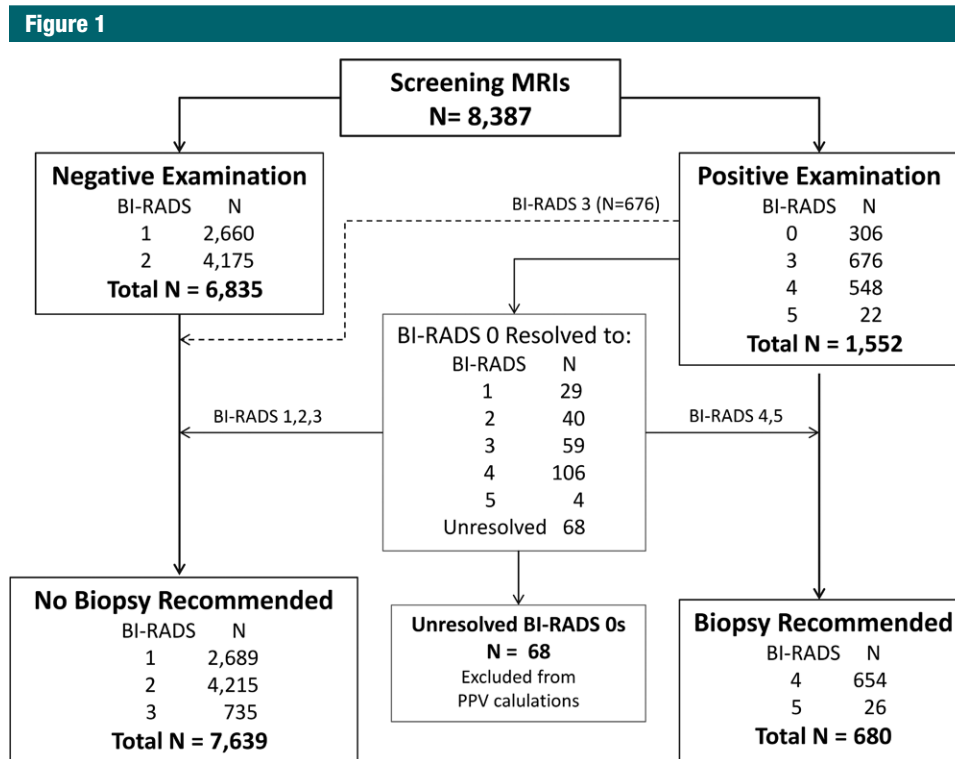
Figure 1: BI-RADS audit flowchart. CDR, sensitivity, and specificity were calculated on the basis of positive or negative examination results. PPV$_2$ and PPV$_3$ were calculated on the basis of whether biopsy was recommended. Unresolved BI-RADS category 0 examinations were excluded from PPV calculations.

7718 examinations). Forty-six percent of the MR examinations (3878 of 8387 examinations) were performed in women with a personal history of breast cancer and 52% were performed in women with a first-degree family history (4020 of 7724 examinations). Both of these risk factors were reported in 15% of MR examinations (1134 of 7724 examinations). Most MR examinations (6324 of 8387 examinations [75%]) were preceded by screening or diagnostic mammography within the previous 12 months. Breast density was classified as either heterogeneously dense or extremely dense in 67% of previous mammograms (4932 of 7408 examinations).

Screening MR Imaging Performance

Of the 8387 screening MR examinations, 6835 (81%) were negative and 1552 (19%) were positive (Fig 1). Of 306 BI-RADS category 0 assessments, 238 (78%) were resolved to final assessment of positive (BI-RADS category

4 or 5; n = 110) or negative (BI-RADS category 1, 2, or 3; n = 128). The remaining 68 examinations with unresolved assessments (0.8% of total examinations) were excluded from PPV$_2$ and PPV$_3$ calculations. Biopsy was ultimately recommended for lesions in 680 examinations—8% of all MR imaging examinations and 44% of examinations with positive screening results (initial BI-RADS assessment 0, 3, 4, or 5).

Screening MR imaging performance in the BCSC and BI-RADS benchmarks are shown in Table 2. The CDR was 17 per 1000 examinations (146 of 8387 examinations; 95% CI: 15, 20 per 1000). PPV$_1$ was 9% (146 of 1552 examinations; 95% CI: 8%, 11%), PPV$_2$ was 19% (132 of 680 examinations; 95% CI: 16%, 22), and PPV$_3$ was 21% (115 of 558 examinations; 95% CI: 17%, 24%). Sensitivity and specificity were 81% (146 of 181 examinations; 95% CI: 75%, 86%) and 83% (6800 of 8206 examinations; 95% CI: 82%, 84%), respectively. The

false-positive biopsy recommendation rate was 66 per 1000 examinations (548 of 8319 examinations; 95% CI: 62, 71).

Breast Cancer Characteristics

Breast cancers diagnosed within 1 year of MR imaging screening tended to be small and at an early stage (Table 2). Minimal cancer accounted for 69% of breast cancers (110 of 160 cancers), and 88% of invasive cancers were node negative (95 of 108 examinations). Among breast cancers, 87% (134 of 154 cancers) were stage 0 or 1; the median size of invasive cancers was 10 mm.

Cancer characteristics stratified according to mode of detection (ie, screening detected vs interval cancers) are shown in Table 3. Both screening-detected and interval cancers had comparable, favorable prognostic characteristics. DCIS represented a similar proportion of screening-detected cancers (44 of 145 cancers [30%]) and interval cancers (10 of 34 cancers

Table 2

Performance Benchmarks for Breast MR Imaging Screening

Performance Measure	BCSC Performance	BI-RADS Benchmark
CDR (per 1000 examinations)	17 (146/8387)	20–30
PPV_1 (%)*	9 (146/1552)	…
PPV_2 (%)	19 (132/680)	15
PPV_3 (%)	21 (115/558)	20–50
False-positive biopsy recommendation rate (per 1000 examinations)†	66 (548/8319)	NA
Sensitivity (%)	81 (146/181)	>80
Specificity (%)	83 (6800/8206)	85–90
Percentage minimal cancer‡	69 (110/160)	>50
Percentage stage 0 or 1 cancer	87 (134/154)	TBD
Percentage node-negative of invasive cancers	88 (95/108)	>80
Median size of invasive cancers (mm)	10	TBD

Note.—Numbers in parentheses are raw data. NA = not applicable, TBD = to be determined.

* Not included in fifth edition of BI-RADS but included in earlier editions.

† MR examinations with no known tissue diagnosis of cancer within 1 year after biopsy recommendation ([BI-RADS category 4 or 5/all examinations) after exclusion of unresolved BI-RADS 0 assessments, standardized per 1000 examinations.

‡ Minimal cancer is invasive cancer 10 mm or smaller or DCIS.

Table 3

Breast Cancer Characteristics

Characteristic	Screening-detected Cancer (n = 146)	Interval Cancer (n = 35)	Total (n = 181)
Histologic type*			
DCIS	44 (30)	10 (29)	54 (30)
Invasive	101 (70)	24 (71)	125 (70)
Invasive cancer size†			
1–5 mm	11 (13)	5 (29)	16 (16)
6–10 mm	32 (37)	6 (35)	38 (37)
11–15 mm	25 (29)	3 (18)	28 (27)
16–20 mm	8 (9)	0 (0)	8 (8)
>20 mm	10 (12)	3 (18)	13 (13)
Minimal cancer‡			
Yes	89 (67)	21 (78)	110 (69)
No	44 (33)	6 (22)	50 (31)
Axillary lymph node status§			
Negative	78 (88)	17 (89)	95 (88)
Positive	11 (12)	2 (11)	13 (12)
American Joint Committee on Cancer stage‖			
0	44 (35)	10 (37)	54 (35)
I	66 (52)	14 (52)	80 (52)
II	13 (10)	3 (11)	16 (10)
III	4 (3)	0 (0)	4 (3)

Note.—Data are numbers of cancers, with percentages in parentheses.

* Data are missing for two cancers.

† Data are missing for 22 cancers.

‡ Minimal cancer is invasive cancer 10 mm or smaller or DCIS. Data are missing for 21 cancers.

§ Among invasive cancers only. Data are missing for 17 cancers.

‖ Data are missing for 27 cancers.

[29%]). Invasive screening-detected and interval cancers had comparable proportions of cancers that were larger than 20 mm (12% and 18%, respectively) or node positive (12% and 11%). Sensitivity according to histologic type was 82% (95% CI: 71%, 92%) for DCIS and 81% (95% CI: 74%, 88%) for invasive cancers. Median time to diagnosis for interval breast cancers was 227 days (interquartile range, 45–305 days).

Figure 2 displays the frequency of screening mammography relative to screening MR examinations. Of the 8387 MR imaging examinations, 4557 (54%) were performed in women who had undergone screening mammography within the period of 60 days before MR imaging through 365 days after a screening MR imaging examination. Screening mammography occurred most frequently within 60 days prior to a screening MR examination or on the same day (1811 of 4557 examinations [40%]), followed by another frequency peak at around 6 months after a screening MR examination (range, 151–210 days; 1081 of 4557 examinations [24%]).

To examine whether the interval cancers after negative screening breast MR examinations might have been detected in asymptomatic women choosing MR imaging and mammography screening that alternated every 6 months, we searched the BCSC database for any additional imaging examinations performed within 60 days prior to interval cancer diagnosis. Eight of 35 interval cancers (23%) were identified with screening mammography and 12 (34%) were identified with diagnostic mammography, breast US, or both. Three of 35 interval cancers (9%) were identified by means of MR imaging; two of these examinations were performed for screening within the 365-day follow-up period. For the remaining 12 interval cancers (34%), either imaging records did not specify modality or indication (n = 5) or no imaging was performed before diagnosis within the BCSC catchment area (n = 7).

Discussion

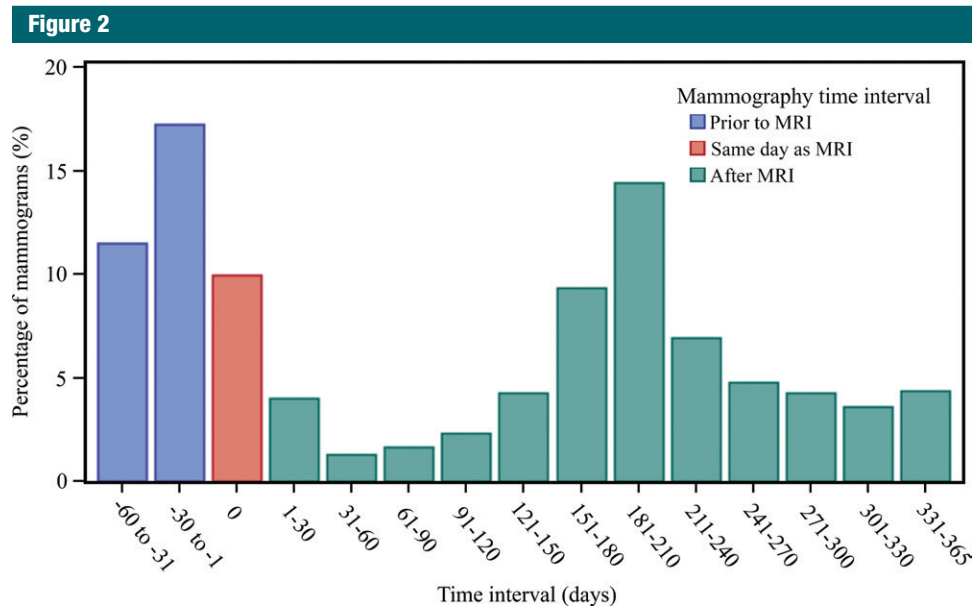Our findings summarize the current range of performance and outcomes for

Figure 2: Bar chart shows screening mammography frequency relative to screening MR imaging. Distribution of time (in 30-day increments) between screening MR examination and associated screening mammography is shown. Of 8387 MR examinations, 4557 (54%) involved women who had undergone screening mammography within 60 days of MR imaging up to 365 days after screening MR examination.

screening MR imaging in clinical practice within the BCSC and place these results in the context of clinical benchmarks recommended by the American College of Radiology. Overall, screening MR imaging in the BCSC met BI-RADS benchmarks for most performance measures and approached benchmark levels for the remaining measures.

The CDR of screening MR imaging approached the performance benchmark, consistent with the application of screening MR imaging in women at increased risk of developing breast cancer. Compared with women receiving screening digital mammography in the BCSC (16), women receiving screening MR imaging were younger, with 40% of screening MR imaging versus 29% of screening mammography examinations performed in women younger than 50 years and in women who were more likely to have a family history of breast cancer (52% vs 17%, respectively) or a personal history of breast cancer (46% vs 5%). Accordingly, the observed CDR of screening MR imaging is higher than that of screening mammography (17.4 vs 5.1 per 1000 screening procedures, respectively).

When the cancer yield of biopsy is examined, both PPV_2 (PPV of biopsies recommended) and PPV_3 (PPV of biopsies performed) met the benchmark at the lower end of the benchmark range. The PPVs, along with MR imaging specificity, which approached but did not meet the benchmark of 85%–90%, suggest that a continued focus on reducing false-positive results while maintaining sensitivity and cancer detection remains important for ongoing quality improvement efforts at population, facility, and individual radiologist levels. We also calculated the false-positive biopsy recommendation rate (66 per 1000 examinations). Although this is not a current American College of Radiology performance benchmark, its standardization per 1000 examinations enables more direct comparison of information along with CDR (17 per 1000 examinations) to describe the benefits and potential harms of screening MR imaging.

In terms of characteristics, breast cancers diagnosed after screening MR imaging are small and at an early stage. Benchmarks for percentage of minimal cancers and percentage of node-negative

invasive breast cancers were exceeded in this study. In addition, our results provide clinical practice–based values for measures currently noted as "To Be Determined" in the BI-RADS manual (13): percentage of stage 0 or 1 breast cancers (87%) and median size of invasive breast cancers (10 mm). Previous analyses of BCSC data informed BI-RADS benchmark values for performance and outcomes for screening and diagnostic mammography (16–19) in the third (20), fourth (15), and fifth (21) editions of the manual. The results of our study provide additional data to inform and guide continued revision of screening MR imaging performance benchmarks.

Further examination of breast cancer characteristics indicated that both screening-detected and interval cancers had favorable prognostic characteristics. The proportion of cancers that were larger than 20 mm or node positive was comparable across the two groups. This suggests that screening MR imaging is effective in detecting most breast cancers within the detectable preclinical phase (22,23). It may be that both screening-detected and

interval cancers are being identified before the critical point of disease, beyond which treatment becomes less effective.

When we examined the mode of detection for interval breast cancers, 23% (eight of 35 cancers) were detected by means of subsequent screening mammography during the follow-up period. This proportion may be an underestimate, because designation of a mammographic examination as diagnostic after breast conservation therapy, even if the patient is asymptomatic, is considered appropriate by the American College of Radiology (24). Although breast cancers detected with screening mammography were counted as a false-negative finding for screening MR imaging, the cancers were still identified when patients were asymptomatic and were detected with a multimodality regimen concordant with guidelines for screening women at high risk of developing breast cancer (2–4). Two additional interval cancers were identified by means of screening MR imaging in women who returned for screening before the end of the 365-day follow-up period (on days 362 and 365, respectively).

In this study, the observed sensitivity of screening MR imaging in community practice met the benchmark value of 80%. Multimodality screening with both mammography and MR imaging is recommended for women at high risk of developing breast cancer, and many women choose screening regimens in which MR imaging and mammography are alternated at 6-month intervals. Although concordant with clinical guidelines, multimodality screening with alternating tests at intervals shorter than the standard follow-up period poses an auditing challenge. The single-modality auditing approach recommended by the American College of Radiology defines false-negative (interval) cancers as any that are identified during the follow-up period and includes both breast cancers that manifest clinically and those that are asymptomatic and detected with a second screening modality such as mammography. Asymptomatic breast cancers detected with a second screening modality (mammography) during the follow-up period after an initial screening modality

(MR imaging) contribute a higher proportion of breast cancers classified as false-negative to the calculation of sensitivity for the initial screening test, compared with regimens in which both screening tests are performed concurrently, with no additional screening tests performed during the follow-up period. As we move to implement risk-based screening and surveillance, revision of auditing methods to account for multimodality regimens in addition to single-modality assessments may be needed to improve accurate assessment of screening programs and outcomes.

With regard to screening MR imaging use at the population level, evaluation of appropriate utilization includes examination of whether MR imaging utilization in women at low risk of breast cancer and lack of MR imaging use in women at high risk occurs. Use of advanced imaging techniques such as breast MR imaging has the potential to exacerbate existing disparities in access to breast cancer screening and diagnostic services. Previous BCSC analyses showed that, among women at high risk, those with lower educational attainment (high school graduate, General Educational Development certificate of high school equivalency, or lower) were 60% less likely to use screening MR imaging compared with women with at least a college degree (25). At the same time, among women at average risk, those with at least a college education were almost 2.5 times more likely to use screening MR imaging. These results suggest that improved risk communication may improve appropriate use of screening MR imaging, further increasing achievable CDRs into the range observed in clinical trials.

Strengths of this study include the large, diverse sample of breast imaging facilities in the BCSC linked to pathology databases as well as state and regional tumor registries to provide comprehensive capture of cancer outcomes for accurate assessment of performance and comparison with benchmarks. Data were collected from community and academic practices that serve a geographically and racially representative sample of the U.S. population, and

our results are likely to reflect clinical radiology practice. Our results indicate that clinical practice performance can meet or approach benchmarks based on expert practice in clinical trials. In addition, our results provide additional data that could be used to define new benchmarks where a value of "To Be Determined" exists in the current (fifth) edition of the BI-RADS manual, to inform and guide continued revision of screening MR imaging performance benchmarks.

A limitation of this study is the lack of inclusion of genetic mutation data, which would improve characterization of the underlying risk distribution in women undergoing screening MR imaging. In addition, 46% of screening MR examinations in this study were performed in women with a personal history of breast cancer, for whom risk models for second breast cancer events are not currently available. We did not stratify the performance of screening MR imaging between women with and women without a personal history of breast cancer because the BI-RADS manual recommends that breast MR examinations in asymptomatic women with a personal history of breast cancer be audited as screening examinations. For women in the BCSC database who had undergone more than one screening examination, we included all available examinations in our analysis. Following the BI-RADS audit guidelines, we report overall performance, combining first and subsequent screening MR examinations, and did not separately analyze prevalence versus incidence screening performance.

When determining breast cancer status in this study, we adhered to BI-RADS auditing guidance that an examination be classified as "true-negative" when there is no known tissue diagnosis of breast cancer within 1 year of a negative examination. Linkage of MR examinations in this study to state and regional tumor registries enables broad capture of breast cancer status, even when diagnosis occurs at a different facility from the one that conducted the screening examination. However, it is possible for an MR examination to be misclassified as true-negative

in the event that a woman moved out of a state or regional cancer registry catchment area and was diagnosed with breast cancer elsewhere within 12 months. Following BI-RADS guidelines, we classified negative examinations as true-negative even if a breast cancer was diagnosed 13 months after an MR examination with negative results.

In conclusion, the interpretive performance of screening breast MR imaging in U.S. community practice meets or approaches current BI-RADS benchmarks. Individual practices can compare their performance with BCSC performance as well as BI-RADS benchmarks to better understand their performance. Clinical practice performance data can inform and supplement ongoing benchmark development.

References

1. Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. Ann Intern Med 2008;148(9):671–679.

2. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin 2007;57(2):75–89. [Published correction appears in CA Cancer J Clin 2007;57(3):185.]

3. National Comprehensive Cancer Network. Breast cancer, v.1.2015. In: NCCN clinical practice guidelines in oncology (NCCN guidelines). Fort Washington, Pa: National Comprehensive Cancer Network, 2015.

4. Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. J Am Coll Radiol 2010;7(1):18–27.

5. Wernli KJ, DeMartini WB, Ichikawa L, et al. Patterns of breast magnetic resonance imaging use in community practice. JAMA Intern Med 2014;174(1):125–132.

6. Stout NK, Nekhlyudov L, Li L, et al. Rapid increase in breast magnetic resonance imaging use: trends from 2000 to 2011. JAMA Intern Med 2014;174(1):114–121.

7. Brennan S, Liberman L, Dershaw DD, Morris E. Breast MRI screening of women with a personal history of breast cancer. AJR Am J Roentgenol 2010;195(2):510–516.

8. Schacht DV, Yamaguchi K, Lai J, Kulkarni K, Sennett CA, Abe H. Importance of a personal history of breast cancer as a risk factor for the development of subsequent breast cancer: results from screening breast MRI. AJR Am J Roentgenol 2014;202(2):289–292.

9. Gweon HM, Cho N, Han W, et al. Breast MR imaging screening in women with a history of breast conservation therapy. Radiology 2014;272(2):366–373.

10. Giess CS, Poole PS, Chikarmane SA, Sippo DA, Birdwell RL. Screening breast MRI in patients previously treated for breast cancer: diagnostic yield for cancer and abnormal interpretation rate. Acad Radiol 2015;22(11):1331–1337.

11. Lehman CD, Lee JM, DeMartini WB, et al. Screening MRI in women with a personal history of breast cancer. J Natl Cancer Inst 2016;108(3):djv349.

12. American College of Radiology. Breast imaging reporting and data system (BI-RADS). Reston, Va: American College of Radiology, 1992.

13. Sickles EA, D'Orsi CJ. ACR BI-RADS follow-up and outcomes monitoring. In: ACR BI-RADS atlas, breast imaging reporting and data system. 5th ed. Reston, Va: American College of Radiology, 2013.

14. 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.

15. D'Orsi CJ, Mendelson EB, Ikeda DM, et al. Breast imaging reporting and data system: ACR BI-RADS—breast imaging atlas. 4th ed. Reston, Va: American College of Radiology, 2003.

16. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology 2017;283(1):49–58.

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

18. Sickles EA, Miglioretti DL, Ballard-Barbash R, et al. Performance benchmarks for diagnostic mammography. Radiology 2005;235(3):775–790.

19. Sprague BL, Arao RF, Miglioretti DL, et al. National performance benchmarks for modern diagnostic digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology 2017;283(1):59–69.

20. American College of Radiology. Breast imaging reporting and data system: ACR BI-RADS—breast imaging atlas. 3rd ed. Reston, Va: American College of Radiology, 1998.

21. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS atlas, breast imaging reporting and data system. 5th ed. Reston, Va: American College of Radiology, 2013.

22. Black WC, Welch HG. Screening for disease. AJR Am J Roentgenol 1997;168(1):3–11.

23. Obuchowski NA, Graham RJ, Baker ME, Powell KA. Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. AJR Am J Roentgenol 2001;176(6):1357–1362.

24. Moy L, Newell MS, Mahoney MC, et al. ACR appropriateness criteria stage I breast cancer: initial workup and surveillance for local recurrence and distant metastases in asymptomatic women. J Am Coll Radiol 2014;11(12 Pt A):1160–1168.

25. Haas JS, Hill DA, Wellman RD, et al. Disparities in the use of screening magnetic resonance imaging of the breast in community practice by race, ethnicity, and socioeconomic status. Cancer 2016;122(4):611–617.