The Value of Patient and Tumor Factors in Predicting Preoperative Breast MRI Outcomes

Habib Rabbar, MD • Daniel S. Hippe, MS • Ahmed Alaam, MS • Safia H. Cheeney, MD • Mihadea van der Schaar, PhD • Savannah C. Partridge, PhD • Christoph I. Lee, MD, MS, MBA

From the Department of Radiology, University of Washington School of Medicine, Seattle Cancer Care Alliance, 1144 Eastlake Ave E, Room LG2-211, Seattle, WA 98109 (H.R., D.S.H., S.H.C., S.C.P., C.I.L.); Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, Calif (A.A., M.v.d.S.); and Departments of Applied Mathematics and Theoretical Physics and Public Health, University of Cambridge, Cambridge, England (M.v.d.S.). Received December 17, 2019; revision requested February 1, 2020; revision received February 12; accepted March 16. Address correspondence to H.R. (e-mail: hrabbar@uw.edu).

H.R. supported by the National Institutes of Health (grant 1R01 CA203883).

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

See also the commentary by Grimm in this issue.

Purpose: To identify patient and tumor features that predict true-positive, false-positive, and negative breast preoperative MRI outcomes.

Materials and Methods: Using a breast MRI database from a large regional cancer center, the authors retrospectively identified all women with unilateral breast cancer who underwent preoperative MRI from January 2005 to February 2015. A total of 1396 women with complete data were included. Patient features (ie, age, breast density) and index tumor features (ie, type, grade, hormone receptor, human epidermal growth factor receptor type 2/neu, Ki-67) were extracted and compared with preoperative MRI outcomes (ie, true positive, false positive, negative) using univariate (ie, Fisher exact) and multivariate machine learning approaches (ie, least absolute shrinkage and selection operator, AutoPrognosis). Overall prediction performance was summarized using the area under the receiver operating characteristic curve (AUC), calculated using internal validation techniques (bootstrap and cross-validation) to account for model training.

Results: At the examination level, 181 additional cancers were identified among 1396 total preoperative MRI examinations (median patient age, 56 years; range, 25–94 years), resulting in a positive predictive value for biopsy of 43% (181 true-positive findings of 419 core-needle biopsies). In univariate analysis, no patient or tumor feature was associated with a true-positive outcome ($P = .022$) and younger age ($< 50$ years, $P = .025$) were associated with false-positive examinations. Machine learning approaches provided weak performance for predicting true-positive, false-positive, and negative examinations (AUC range, 0.50–0.57).

Conclusion: Commonly used patient and tumor factors driving expert opinion for the use of preoperative MRI provide limited predictive value for determining preoperative MRI outcomes in women.

Supplemental material is available for this article.

© RSNA, 2020

The use of preoperative breast MRI has increased substantially over the past 2 decades due to its ability to detect additional clinically and mammographically occult disease prior to definitive treatment in multiple clinical trials (1–8). In meta-analyses, the prevalence of additional cancer detected with preoperative MRI is 16% to 20% in the ipsilateral breast and 4.1% to 5.5% in the contralateral breast (8–10). However, preoperative MRI also prompts unnecessary biopsies, with a 2:1 ratio of true-positive to false-positive findings in the affected breast and a 0.6:1 ratio in the contralateral breast (9). The false-positive rate of preoperative MRI can lead to additional patient and health care system burdens, including avoiding costs, patient anxiety, and treatment delay. As a result, there is wide heterogeneity in utilization of preoperative MRI both between and within institutions based on surgeon, radiologist, and oncologist preferences.

In the absence of clear data to identify which patients are more likely to benefit from or be adversely affected by preoperative MRI, several medical societies have developed recommendations based on consensus expert opinion on when to perform preoperative MRI. The American Society of Breast Surgeons (11), National Comprehensive Cancer Network (12), European Society of Breast Cancer Specialists (13), and European Society of Breast Imaging (14) guidelines all recommend the use of preoperative MRI in cases in which clinical or conventional imaging findings are suspicious for additional malignancy, with particular consideration of patients with invasive lobular carcinoma (ILC) or dense breast tissue. However, these guidelines are not widely adopted, with institution-level variation in preoperative MRI use dependent on surgeon preference and based loosely on patient factors such as age, breast density, and tumor histopathologic characteristics.

Given the lack of data-driven recommendations, better risk stratification tools are needed to determine which women would be most likely to benefit from preoperative MRI. When such medical decision making cannot be
Patient and Tumor Factors in Predicting Preoperative Breast MRI Outcomes

Abbreviations

AUC = area under the receiver operating characteristic curve, CNB = core-needle biopsy, DCIS = ductal carcinoma in situ, HER2 = human epidermal growth factor receptor type 2, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, LASSO = least absolute shrinkage and selection operator, OR = odds ratio

Summary

In the patient population described in this study, the major factors driving expert opinion for the use of preoperative breast MRI—age, breast density, and tumor phenotype—were not predictive for identifying additional disease.

Key Points

- Preoperative breast MRI allows for identification of additional breast cancers in women newly diagnosed with breast cancer, with an additional cancer detection rate of 13% (181 cancers of 1396 examinations) and a positive predictive value for biopsy of 43% (181 cancers of 419 core-needle biopsies).
- Patient and tumor features, including age, mammographic density, and lobular phenotype, were not found to be valuable for predicting true-positive (area under the receiver operating characteristic curve [AUC] range, 0.50–0.53), false-positive (AUC range, 0.54–0.56), or negative outcomes of preoperative breast MRI (AUC range, 0.55–0.56) even when examined using machine learning models.

Materials and Methods

Study Design

We performed a retrospective analysis of a large clinical breast MRI database linked to mammographic density and pathologic outcomes. All features recorded in the MRI database were abstracted by the interpreting radiologist (including H.R. and C.I.L.) at the time of MRI interpretation. This Health Insurance Portability and Accountability Act–compliant study was approved by our institution’s institutional review board, which waived the need for informed consent. We identified all patients who underwent preoperative MRI to further evaluate a newly diagnosed breast cancer at core-needle biopsy (CNB) from January 2005 to February 2015. Any breast MRI performed after primary breast cancer surgery or after neoadjuvant chemotherapy were excluded. Patients with bilateral breast cancers diagnosed prior to the preoperative MRI were also excluded from the study (n = 32), since inclusion of such patients would create challenges in determining rates of ipsilateral and contralateral true-positive, false-positive, and negative results. The initial database query yielded 1596 unique women who underwent preoperative MRI following the diagnosis of unilateral breast cancer during the study period. Patient and index tumor features at CNB were extracted from the MRI database, including age, mammographic breast density, tumor type (ductal carcinoma in situ [DCIS], invasive ductal carcinoma [IDC], or ILC), tumor grade (Nottingham or nuclear grade), estrogen and progesterone receptor status, human epidermal growth factor receptor type 2 (HER2)/neu receptor status, and Ki-67 proliferation index.

All preoperative MRI examinations were performed less than 6 months after a CNB diagnosis of breast cancer, defined as DCIS and/or invasive breast cancer, but before the performance of any surgery to treat malignancy. Of the initial 1596 patients, 200 women were excluded due to one or more missing variables that included age (n = 1), breast density (n = 96), Nottingham grade (n = 40), hormone receptor status (n = 41), HER2/neu receptor status (n = 67), and DCIS nuclear grade (n = 22), leaving 1396 women with complete data for analysis (Fig 1). Of these 1396 patients, 941 (67%) were 50 years old or older (median age, 56 years; range, 25–94 years), 887 (64%) had extremely or heterogeneously dense breasts, and 1086 (78%) had an invasive breast cancer diagnosis. A more detailed overview of patient characteristics is found in Table 1. It should be noted that a portion of the patients ultimately included in this study (n = 310) were also included in a published article on the use of preoperative MRI to improve surgical outcomes of DCIS (16).

At least one additional CNB was performed after the preoperative MRI (a nonnegative preoperative MRI outcome) on the side ipsilateral to the index tumor in 314 of the 1396 women (22%), on the contralateral side in 152 of the 1396 women (11%), and on either side in 419 of the 1396 women (30%; 47 women underwent CNB on both sides). The median time between preoperative MRI and subsequent CNB was 7 days (interquartile range, 4–12 days).

Pathologic Features

For the purposes of analyses, cancers that were described as mixed ductal and lobular at CNB were categorized as ILC, while cancers that were not specified (eg, invasive mammary carcinoma—not otherwise specified) were categorized as IDC since most invasive breast cancers are ductal phenotype. It should be noted that although some breast cancer phenotypes are reclassified at surgical excision, we utilized CNB pathologic assessment prior to MRI for determining preoperative
MRI Acquisition
Breast MRI examinations were performed with a GE LX 1.5-T scanner (GE Healthcare, Waukesha, Wis) prior to 2010 and with a Philips Achieva Tx 3-T scanner (Philips Healthcare, Best, the Netherlands) between January 2010 and February 2015 using a dedicated bilateral breast coil in accordance with the American College of Radiology (ACR) Breast MRI Accreditation program. Each bilateral MRI examination included a precontrast fat-suppressed T2-weighted fast spin-echo sequence followed by a T1-weighted dynamic contrast material–enhanced sequence with one precontrast and at least three postcontrast fat-suppressed three-dimensional fast gradient-echo acquisitions. Initial postcontrast acquisitions were centered between 90 seconds and 120 seconds after contrast material administration, and final delayed acquisitions were centered between 4.5 minutes and 7.5 minutes after contrast material administration, depending upon protocol. Gadolinium-based contrast agent (before November 2010, Omniscan, GE Healthcare; starting November 2010, ProHance, Bracco Diagnostics, Princeton, NJ) was power injected (0.1 mmol per kilogram of body weight at 2 mL/sec) followed by a 20-mL saline flush.

MR Image Interpretation
The MR image clinical interpretations were all performed within 2 business days of the date of the MRI by fellowship-trained breast radiologists (13 total radiologists [including H.R. and C.I.L.], with 1–20 years of breast MRI experience at the time of interpretation). ACR Breast Imaging Reporting and Data System assessments and recommendations (19) were entered at the time of the MRI interpretation into our MRI database linked to histopathologic outcomes. It should be noted that because this study utilized data recorded at the MRI value since excision data are not available at the time of preoperative MRI decision making. Outcomes of any CNB performed after preoperative MRI but prior to initial surgery were also obtained from the MRI database and were classified by worst pathology with the following hierarchy (beginning with the worst): any type of invasive carcinoma, DCIS, high-risk lesions, benign. High-risk lesions were defined as atypical ductal hyperplasia, lobular neoplasia (lobular carcinoma in situ and atypical lobular hyperplasia), radial scar and/or complex sclerosing lesion, flat epithelial atypia, or any other “atypical” lesion. Because papillomas without atypia that are concordant with imaging features do not routinely undergo surgical excision at our institution, they were not considered high-risk lesions for this study.

Figure 1: CONSORT diagram for the study. CNB = core-needle biopsy, DCIS = ductal carcinoma in situ, HER2 = human epidermal growth factor receptor type 2.
time of the original interpretation and prior to any subsequent biopsies, no blinding of outcomes was required.

**Outcome Definitions**

Outcomes were based on CNBs performed within 180 days of the preoperative MRI (to ensure CNBs were associated with the MRI) on lesions within the breast only, specifically excluding axillary lymph node biopsies. Multiple CNBs performed after preoperative MRI but before definitive surgery were grouped by worst outcome per breast using the aforementioned hierarchy. We report multiple preoperative MRI outcomes at the examination level (both breasts), ipsilateral breast level, and contralateral breast level. A true-positive outcome (ie, an additional cancer detected) was defined as a preoperative MRI that resulted in any CNB diagnosis of additional DCIS and/or invasive carcinoma prior to definitive surgery and regardless of the number of CNBs performed. Similarly, a false-positive outcome was defined as a preoperative MRI examination resulting in any number of CNBs but none of which resulted in malignancy (ie, benign or high-risk pathologies). Finally, a negative preoperative MRI outcome was defined as no CNBs performed between preoperative MRI and surgery. Positive predictive value for biopsy of preoperative MRI examinations was calculated at the examination level and was defined as the number of examinations with at least one CNB yielding a malignant outcome performed after the preoperative MRI divided by the total number of examinations with at least one CNB performed after preoperative MRI.

It should be noted that negative examinations were not further divided into true-negative or false-negative examinations because it was challenging to define a clinically relevant reference standard for this patient population. Unlike in a purely screening setting in which a false-negative examination is defined as an interval cancer, in the diagnostic setting a false-negative examination would also relate to the total amount of disease identified at surgery versus the amount described at MRI. Because women underwent varying types of surgeries, including mastectomy and partial mastectomy, and results such as total span of disease, additional sites of disease, and margin status are dependent on surgical technique, defining what constituted a false-negative examination was not practical.

**Machine Learning**

Multivariate models for predicting true-positive, false-positive, and negative outcomes were developed using two machine learning approaches, independent of the univariate analyses described below. In the first approach, least absolute shrinkage and selection operator (LASSO)—based logistic regression was used for variable selection and fitting multivariate models for each outcome (20,21). In this case, each selected predictor was associated with an odds ratio (OR) to represent its effect on the outcome. In the second approach, we fitted models using AutoPrognosis, an algorithmic framework for automating the design of machine learning–based clinical prognostic models. AutoPrognosis uses an advanced Bayesian optimization technique to (automatically) design a prognostic model made out of a weighted ensemble of machine learning pipelines. Each machine learning pipeline comprises design choices for data imputation, feature processing, classification, and calibration algorithms (and their hyperparameters). The design space of AutoPrognosis contains 5460 possible pipelines (seven possible imputation algorithms, nine feature processing algorithms, 20 classification algorithms, and three calibration methods), and the algorithms that constitute the design space of AutoPrognosis have been reported previously (22).

**Table 1: Patient and Index Tumor Characteristics**

| Variable                          | Value          |
|----------------------------------|----------------|
| Age (y) (continuous)             | 56 (25–94)     |
| Breast density                   |                |
| Almost entirely fatty             | 48 (3.4)       |
| Scattered fibroglandular densities | 461 (33.0)   |
| Heterogeneously dense            | 743 (53.2)     |
| Extremely dense                  | 144 (10.3)     |
| Tumor type                       |                |
| Pure DCIS                        | 310 (22.2)     |
| Invasive                         | 1086 (77.8)    |
| Invasive, type (n = 1086)        |                |
| IDC                              | 932 (85.8)     |
| ILC                              | 154 (14.2)     |
| Invasive, Nottingham grade (n = 1086) |          |
| 1                                | 297 (27.3)     |
| 2                                | 407 (37.5)     |
| 3                                | 382 (35.2)     |
| Invasive, HR status (n = 1086)   |                |
| Negative                         | 186 (17.1)     |
| Positive                         | 900 (82.9)     |
| Invasive, HER2/neu (n = 1086)*    |                |
| Nonpositive                      | 931 (85.7)     |
| Positive                         | 155 (14.3)     |
| Invasive, Ki-67 (n = 1086)       |                |
| Low                              | 189 (17.4)     |
| Intermediate                     | 304 (28.0)     |
| High                             | 358 (33.0)     |
| Unknown                          | 235 (21.6)     |
| DCIS, nuclear grade (n = 775)†    |                |
| Low                              | 103 (13.3)     |
| Intermediate                     | 300 (38.7)     |
| High                             | 372 (48.0)     |

Note.—Data are numbers with percentages in parentheses or medians with ranges in parentheses. Some percentages may not add up to 100% due to rounding. DCIS = ductal carcinoma in situ, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, HER2 = human epidermal growth factor type 2, HR = hormone receptor.

* HER2/neu by fluorescence in situ hybridization test if available, otherwise by immunohistochemistry; 52 equivocal or indeterminate cases were combined with the negative group to form the nonpositive group.

† Includes 310 with DCIS only and 465 with both DCIS and invasive components.
Two machine learning approaches were used because each had particular strengths and weaknesses. The LASSO-based approach was selected because the underlying logistic regression structure and regularization provides an interpretable model (eg, each predictor is associated with an OR representing the strength of its relationship with the outcome) that is relatively stable with less variability than nonlinear approaches. The AutoPrognosis approach was selected because its more flexible underlying structure allows it to learn more complex relationships, if present, although with the trade-off that the relationship between each predictor and outcome is less apparent.

Statistical Analysis
Univariate analysis of patient features (ie, age, breast density) and index tumor characteristics (ie, tumor type, grade, hormone receptor, HER2, Ki-67) was performed to estimate how well any individual feature could predict negative, true-positive, and false-positive outcomes. Fisher exact test was used for this exploratory assessment without adjustment for the number of comparisons.

Overall performance of each multivariate deep learning model was summarized using the area under the receiver operating characteristic curve (AUC) (ie, C statistic), internally validated using a bootstrap optimism-adjustment (LASSO models), or fivefold cross-validation (AutoPrognosis models) (23). The bootstrap and cross-validation approaches both attempt to produce unbiased estimates of performance by repeatedly leaving some data out of modeling training for testing. The bootstrap optimism-adjustment involved drawing 1000 bootstrap samples, refitting the models on the new sample, estimating the difference in AUC estimated using the bootstrap sample and the original sample (optimism bias), and subtracting that bias from the original AUC estimate. The cross-validation approach involved randomly partitioning the original sample into five equal-sized subsamples and then iteratively refitting the models using four of five subsamples and estimating the AUC using the left-out subsample not used in training (all five such estimates were averaged to produce the final estimate). The incremental impact of each predictor on AutoPrognosis model performance was summarized as the change in the AUC (ΔAUC) between the full model with all predictors available and a reduced model with that predictor excluded. All computations were performed using JMP version 12/1/0 (SAS, Cary, NC), R version 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria), and Python version 3.6 (Python Software Foundation, https://www.python.org/).

Results Summary of CNB True- and False-Positive Results after Preoperative MRI
To determine the degree of true-positive and false-positive outcomes after preoperative MRI, the total number of patients who underwent CNB with a malignant (true-positive) diagnosis and nonmalignant (false-positive) diagnosis were determined. Of the 1396 women, a total of 419 patients underwent CNB after preoperative MRI. There were 154 of 1396 (11%) true-positive findings on the ipsilateral side, 30 of 1396 (2%) true-positive findings on the contralateral side, and 181 of 1396 (13%) true-positive findings at the examination level. Furthermore, there were 160 of 1396 (11%) false-positive findings on the ipsilateral side, 122 of 1396 (9%) false-positive findings on the contralateral side, and 238 of 1396 (17%) false-positive findings at the examination level. At the examination level, 181 of 419 women underwent a CNB due to preoperative MRI recommendation that yielded a malignant outcome, resulting in a positive predictive value for biopsy of 43% (95% confidence interval [CI]: 38%, 48%). A flowchart of these outcomes is shown in Figure 1, and examples of a preoperative MRI false-positive, true-positive, and negative outcomes are provided in Figures 2, 3, and 4, respectively.

Breast Characteristics Associated with Predictive Outcomes at Preoperative MRI
Next, we aimed to determine if specific breast characteristics were associated with negative, true-positive, and false-positive results at preoperative MRI. Overall, patients with dense breasts were less likely to have a negative examination (68% vs...
74%; $P = .025$) and more likely to have a false-positive examination (19% vs 14%; $P = .022$) (Table 2). Similarly, women younger than 50 years were less likely than women aged 50 years or older to have negative examinations (65% vs 72%; $P = .009$) and were more likely to have false-positive examinations (21% vs 15%; $P = .010$). Women with nonpositive (negative or equivocal) HER2/neu disease were less likely than women with positive HER2/neu disease to have a negative examination (69% vs 77%; $P = .029$) and were more likely to have a false-positive examination (18% vs 11%; $P = .047$). There were no other statistically significant associations between patient and index tumor characteristics and examination level outcomes during the univariate analysis (Table 2). Univariate analysis of ipsilateral and contralateral outcomes demonstrated similar associations between age and outcomes, with detailed results shown in Table E1 (supplement).

**Multivariate Modeling of Predictors of Outcomes**

Multivariate modeling of the three examination-level outcomes based on the LASSO machine learning algorithm are
shown in Table 3. The LASSO algorithm did not select any predictors for true-positive outcomes. The predictors selected by LASSO for both false-positive and negative predictive MRI outcomes were age (OR = 0.94 and 1.06, respectively, per 10-year increase), breast density (OR = 1.19 and 0.84, respectively, per one-level increase), and HER2/neu status (OR = 0.74 and 1.34, respectively). Although the LASSO identified factors that were predictive of false-positive and negative examinations, the bootstrap-adjusted AUC estimates were modest at 0.56 (95% CI: 0.52, 0.61; \( P = 0.024 \)) and 0.55 (95% CI: 0.52, 0.60; \( P = 0.035 \)), respectively. Results were similar for models predicting ipsilateral and contralateral outcomes, with bootstrap-adjusted AUC estimates ranging from 0.50 to 0.57 (Table E2 [supplement]).

The AutoPrognosis-based multivariate models also had modest performance for predicting true-positive, false-positive, and negative examinations, with cross-validated AUCs of 0.53, 0.54, and 0.56, respectively (Table 4). The \( \Delta \)AUCs for individual predictors were most often plus or minus 0.01. Similar to the LASSO models for false-positive and negative preoperative MRI outcomes, breast density (\( \Delta \)AUC = +0.01 for both) and HER2/neu status (\( \Delta \)AUC = +0.01 and +0.02, respectively) were identified as incrementally predictive although the final AUCs were all less than or equal to 0.55. Cross-validated AUCs ranged from

### Table 2: Univariate Analysis of Patient and Index Tumor Characteristics for Predicting True-Positive, False-Positive, and Negative MRI Examinations

| Variable                        | No. of Patients | Either Side |  |  |  |
|---------------------------------|-----------------|-------------|---|---|---|
|                                |                 | TP          | FP | Negative | P Value* |
| Age (y)                         |                 | .50         | .010 | .009 |
| <50                             | 455             | 63 (13.8)   | 95 (20.9) | 297 (65.3) |
| ≥ 50                            | 941             | 118 (12.5)  | 143 (15.2) | 680 (72.3) |
| Breast density                  |                 | .68         | .022 | .025 |
| Not dense                       | 509             | 63 (12.4)   | 71 (13.9) | 375 (73.7) |
| Dense                           | 887             | 118 (13.3)  | 167 (18.8) | 602 (67.9) |
| Tumor type                      |                 | .34         | .49 | .94 |
| Pure DCIS                       | 310             | 35 (11.3)   | 57 (18.4) | 218 (70.3) |
| Invasive                        | 1086            | 146 (13.4)  | 181 (16.7) | 759 (69.9) |
| Invasive, type (n = 1086)       |                 | .44         | .48 | > .99 |
| IDC                             | 932             | 122 (13.1)  | 159 (17.1) | 651 (69.8) |
| ILC                             | 154             | 24 (15.6)   | 22 (14.3) | 108 (70.1) |
| Invasive, Nottingham grade (n = 1086) |          | .17         | .91 | .43 |
| 1                               |                 | 297 (11.4)  | 52 (17.5) | 211 (71.0) |
| 2                               |                 | 407 (16.0)  | 67 (16.5) | 275 (67.6) |
| 3                               |                 | 382 (12.3)  | 62 (16.2) | 273 (71.5) |
| Invasive, HR status (n = 1086)   |                 | .19         | .52 | .66 |
| Negative                        | 186             | 19 (10.2)   | 34 (18.3) | 133 (71.5) |
| Positive\(^1\)                  | 900             | 127 (14.1)  | 147 (16.3) | 626 (69.6) |
| Invasive, HER2/neu (n = 1086)   |                 | .53         | .047 | .029 |
| Nonpositive                     | 931             | 128 (13.7)  | 164 (17.6) | 639 (68.6) |
| Positive                        | 155             | 18 (11.6)   | 17 (11.0) | 120 (77.4) |
| Invasive, Ki-67 (n = 1086)      |                 | .21         | .79 | .16 |
| Low                             | 189             | 21 (11.1)   | 28 (14.8) | 140 (74.1) |
| Intermediate                    | 304             | 51 (16.8)   | 55 (18.1) | 198 (65.1) |
| High                            | 358             | 42 (11.7)   | 61 (17.0) | 255 (71.2) |
| Unknown                         | 235             | 32 (13.6)   | 37 (15.7) | 166 (70.6) |
| DCIS, nuclear grade (n = 775)   |                 | .63         | .76 | .63 |
| Low                             | 103             | 12 (11.7)   | 21 (20.4) | 70 (68.0) |
| Intermediate                    | 300             | 44 (14.7)   | 56 (18.7) | 200 (66.7) |
| High                            | 372             | 46 (12.4)   | 65 (17.5) | 261 (70.2) |

Note.—Values are numbers of patients with percentages in parentheses with outcome in each subgroup. Percentages may not add up to 100% due to rounding. DCIS = ductal carcinoma in situ, FP = false positive, HER2 = human epidermal growth factor receptor type 2, HR = hormone receptor, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, TP = true positive.

* Fisher exact test comparing groups as indicated by the Variable column.

\(^1\) Positive indicates estrogen and/or progesterone receptor positive, and negative indicates neither estrogen nor progesterone positive.
0.53 to 0.55 for ipsilateral outcome models and from 0.56 to 0.62 for contralateral outcome models (Table E3 [supplement]). Taken together, results from both multivariate models indicated that age, breast density, and HER2/neu status did not significantly contribute specific examination-level outcomes.

**Discussion**

In our retrospective study using a large clinical database of patient and tumor features from a decade’s worth of preoperative MRI examinations performed in women with newly diagnosed breast cancer, we confirm the benefit of preoperative MRI to detect additional cancers but failed to identify which women will benefit most from its routine use. Specifically, we found that age, breast density, and tumor prognostic markers were not associated with a true-positive preoperative MRI outcome. Even more sophisticated machine learning techniques did not yield any clinically significant associations (AUCs ≤ 0.57 for all models) between these characteristics and examination-level true-positive preoperative MRI outcomes. Our findings are important as they do not support application of commonly used patient- and tumor-level factors (eg, mammographic density, age, ILC diagnosis) that currently drive expert recommendations for preoperative MRI.

While we found no patient or tumor features to predict which patients will benefit from preoperative MRI through identification of additional disease, we did find that patients with higher mammographic breast density, younger age, and HER2/neu positivity were more likely to experience a false-positive finding prompted by preoperative MRI. Our observed association of higher mammographic breast density with increased false-positive outcomes without increased true positives corroborates previous studies, including those by Deurloo et al (24) and Siegmann et al (25), who found that changes in clinical management due to preoperative MRI were independent of breast density. Similarly, our findings of younger women experiencing more false-positive findings from preoperative MRI is also similar to published results from Tillman et al (26). As younger women usually have denser breast parenchyma, the general argument that these women are more likely than older women with nondense breast tissue to benefit from preoperative MRI is not supported by our analysis.

A particular strength of our study was that our analysis benefited from a decade’s worth of preoperative MRI examinations that were routinely performed at our institution in women younger than 70 years with no contraindication to MRI. Because our site does not follow defined criteria for preoperative breast MRI eligibility, we were able to assess whether guidelines, such as those by the National Comprehensive Cancer Network that recommend preferential use of preoperative MRI among women with specific tumor subtypes (eg, ILC), could be supported by evidence. Previous literature supporting this recommendation is mixed (27–29), and we did not find that tumor

| Parameter | TP (n = 181) | FP (n = 238) | Negative (n = 977) |
|-----------|-------------|-------------|-------------------|
| Variable  |             |             |                   |
| Age, per 10-year increase | ... | 0.94 | 1.06 |
| Breast density (F, SD, HD, ED)*, per 1-level increase | ... | 1.19 | 0.84 |
| Invasive tumor type (vs pure DCIS) | ... | ... | ... |
| If invasive | ... | ... | ... |
| ILC (vs IDC only) | ... | ... | ... |
| Nottingham grade (1–3)*, per 1-level increase | ... | ... | ... |
| HR positive (vs negative)* | ... | ... | ... |
| HER2/neu positive (vs nonpositive) | ... | 0.74 | 1.34 |
| Ki-67 (low, intermediate, high)*, per 1-level increase | ... | ... | ... |
| Ki-67 unknown or unavailable (vs known and low) | ... | ... | ... |
| If DCIS | DCIS nuclear grade (low, intermediate, high)*, per 1-level increase | ... | ... | ... |
| Model performance | Bootstrap-adjusted AUC | 0.50 | 0.56 | 0.55 |
| (95% CI)** | NA | (0.52, 0.61) | (0.52, 0.60) |
| P value† | > .99 | .024 | .035 |

Note.—Values are odds ratios from the corresponding LASSO-based logistic regression model; an ellipsis indicates that the LASSO did not select the corresponding variable (ie, the effective odds ratio = 1). AUC = area under the curve, CI = confidence interval, DCIS = ductal carcinoma in situ, ED = extremely dense, F = almost entirely fatty, FP = false positive, HD = heterogeneously dense, HER2 = human epidermal growth factor receptor type 2, HR = hormone receptor, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, LASSO = least absolute shrinkage and selection operator, SD = scattered fibroglandular densities, TP = true positive.
* Coded as 0, 1, 2, etc, and treated as a linear/continuous variable in the model.
† Positive indicates estrogen and/or progesterone receptor positive, and negative indicates neither estrogen nor progesterone positive.
‡ 95% confidence intervals and corresponding P value (test vs AUC = 0.5) for bootstrap-adjusted AUC.
ble that our estimates of false-positive findings are slightly higher than our estimates of false-negative findings. The general screening population is approximately 50% breast density (typical frequency in the general population). It is also possible that there was likely some selection bias present in our study, as 63% of our cohort had mammographically dense breasts (typical frequency in the general population). However, for most of the study period, ours was the only National Cancer Institute–Designated Comprehensive Cancer Center in the Pacific Northwest, receiving referrals from multiple surrounding states and providing substantial geographic diversity among our study population who routinely underwent preoperative MRI for newly diagnosed cancer. Despite this, it should be acknowledged that there was likely some selection bias present in our study, as 63% of our cohort had mammographically dense breasts (typical frequency in the general screening population is approximately 50%). It is also possible that our estimates of false-negative findings are slightly higher due to our inability to reconcile final pathology of high-risk lesions at surgical excision, as up to 15% to 20% of such lesions upgrade to cancer at final excision. Finally, we could not assess additional clinical outcomes, including the effect of false-positive MRI examinations on surgeries, concordance of MRI size to pathology size, surgical re-excision rate, breast cancer recurrence rate, and breast cancer–specific survival.

In summary, our study adds to the literature on the utility of preoperative MRI based on patient and tumor characteristics by taking a machine learning approach to analyzing a decade of data at a major cancer referral center. We found that the major factors driving expert opinion for the use of preoperative MRI—age, breast density, and tumor phenotype—are not predictive for identifying additional disease at preoperative MRI. If these factors are included in conventional imaging tumor size in our analysis, which is a factor suggested by some experts for determining the utility of preoperative MRI. Nevertheless, we focused on all other major factors currently used in preoperative MRI decision making, including age, breast density, and tumor molecular characteristics. Second, this is a single-institution retrospective study, and its results may not reflect preoperative MRI outcomes at other institutions. However, for most of the study period, ours was the only National Cancer Institute–Designated Comprehensive Cancer Center in the Pacific Northwest, receiving referrals from multiple surrounding states and providing substantial geographic diversity among our study population who routinely underwent preoperative MRI for newly diagnosed cancer. Despite this, it should be acknowledged that there was likely some selection bias present in our study, as 63% of our cohort had mammographically dense breasts (typical frequency in the general screening population is approximately 50%). It is also possible that our estimates of false-negative findings are slightly higher.

Table 4: AutoPrognosis-based Prediction Models for Examination-level Outcomes after Preoperative MRI

| Parameter                          | Outcome |   |   |
|-----------------------------------|---------|---|---|
|                                   | TP (n = 181) | FP (n = 238) | Negative (n = 977) |
| Age                               |         |   |   |
| Breast density                    |         |   |   |
| Invasive tumor type (vs pure DCIS)|         |   |   |
| If invasive                       |         |   |   |
| ILC (vs IDC only)                 |         |   |   |
| Nottingham grade (1–3)            | +0.01   |   |   |
| HR positive (vs negative)         |         |   |   |
| HER2/neu positive (vs nonpositive)| +0.01   |   |   |
| Ki-67 (unknown/unavailable, low,  | -0.01   |   |   |
| intermediate, high)               |         |   |   |
| DCIS                              |         |   |   |
| DCIS nuclear grade (low, intermediate, high) | +0.01 | -0.01 |   |
| Model performance                 | 0.53    | 0.54 | 0.56 |

Note.—Each value is the estimated change in the AUC between the full model with all predictors available and a reduced model with that predictor excluded; a positive change indicates improvement in performance attributable to that predictor; an ellipsis indicates an AUC change of 0.00. AUC = area under the curve, DCIS = ductal carcinoma in situ, FP = false positive, HER2 = human epidermal growth factor receptor type 2, HR = hormone receptor, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, TP = true positive.

The use of preoperative MRI—age, breast density, and tumor phenotype—are not predictive for identifying additional disease at preoperative MRI. If these factors are included in conventional imaging tumor size in our analysis, which is a factor suggested by some experts for determining the utility of preoperative MRI. Nevertheless, we focused on all other major factors currently used in preoperative MRI decision making, including age, breast density, and tumor molecular characteristics. Second, this is a single-institution retrospective study, and its results may not reflect preoperative MRI outcomes at other institutions. However, for most of the study period, ours was the only National Cancer Institute–Designated Comprehensive Cancer Center in the Pacific Northwest, receiving referrals from multiple surrounding states and providing substantial geographic diversity among our study population who routinely underwent preoperative MRI for newly diagnosed cancer. Despite this, it should be acknowledged that there was likely some selection bias present in our study, as 63% of our cohort had mammographically dense breasts (typical frequency in the general screening population is approximately 50%). It is also possible that our estimates of false-negative findings are slightly higher.
References

1. Houssami N, Hayes DF. Review of preoperative magnetic resonance imaging (MRI) in breast cancer: should MRI be performed on all women with newly diagnosed, early stage breast cancer? CA Cancer J Clin 2009;59(5):290–302.

2. Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. Radiology 2004;233(3):830–849.

3. Esserman L, Hylton N, Yassa L, Barclay J, Frankel S, Sickles E. Utility of magnetic resonance imaging in the management of breast cancer: evidence for improved preoperative staging. J Clin Oncol 1999;17(1):110–119.

4. Liberman L, Morris EA, Dershaw DD, Abramson AF, Tan IK. MR imaging of the ipsilateral breast in women with percutaneously proven breast cancer. AJR Am J Roentgenol 2003;180(4):901–910.

5. Mumtaz H, Hall-Craggs MA, Davidson T, et al. Staging of symptomatic primary breast cancer with MR imaging. AJR Am J Roentgenol 1997;169(2):417–424.

6. Schnall MD, Blume J, Bluemke DA, et al. MRI detection of distinct incidental cancer in women with primary breast cancer studied in IBMC 6883. J Surg Oncol 2005;92(1):32–38.

7. Sardanelli F, Giuseppetti GM, Panizza P, et al. Sensitivity of MRI versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard. AJR Am J Roentgenol 2004;183(4):1149–1157.

8. Houssami N, Ciatto S, Macaskill P, et al. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: systematic review and meta-analysis in detection of multifocal and multicentric cancer. J Clin Oncol 2008;26(19):3248–3258.

9. Plana MN, Carreira C, Murial A, et al. Magnetic resonance imaging in the preoperative assessment of patients with primary breast cancer: systematic review of diagnostic accuracy and meta-analysis. Eur Radiol 2012;22(1):26–38.

10. Brennan ME, Houssami N, Lord S, et al. Magnetic resonance imaging of the contralateral breast in women with newly diagnosed breast cancer: systematic review and meta-analysis of incremental cancer detection and impact on surgical management. J Clin Oncol 2009;27(33):5640–5649.

11. American Society of Breast Surgeons. Official Statement: Consensus Guideline on Diagnostic and Screening Magnetic Resonance Imaging of the Breast. https://www.breastsurgeons.org/docs/statements/Consensus-Guideline-on-Diagnostic-and-Screening-Magnetic-Resonance-Imaging-of-the-Breast.pdf. Accessed February 2, 2020.

12. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines). Breast Cancer Version 4.2020—May 8, 2020. https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf. Accessed June 17, 2020.

13. Sardanelli F, Boetes C, Botisch B, et al. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. Eur J Cancer 2010;46(S8):1296–1316.

14. Mann RM, Kuhl CK, Kinkel K, Boetes C. Breast MR: guidelines from the European Society of Breast Imaging. Eur Radiol 2008;18(7):1307–1318.

15. Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920–1930.

16. Lam DL, Smith J, Partridge SC, et al. The impact of preoperative breast MRI on surgical management of women with newly diagnosed ductal carcinoma in situ. Acad Radiol 2020;27(4):478–486.

17. Buist DS, Porter PL, Lehmam C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40–49 years. J Natl Cancer Inst 2004;96(19):1432–1440.

18. Yeatman TJ, Cantor AB, Smith TJ, et al. Tumor biology of infiltrating lobular carcinoma: implications for management. Ann Surg 1995;224(4):549–559; discussion 559–561.

19. D’Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS Atlas: Breast Imaging Reporting and Data System. Reston, Va: American College of Radiology, 2013.

20. Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J R Stat Soc B 2011;73(3):273–282.

21. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 2000;19(8):1059–1079.

22. Alaa AM, van der Schaar M. AutoPrognosis: automated clinical prognostic modeling via Bayesian optimization with structured kernel learning. ArXiv 2018/02.07207 [preprint] https://arxiv.org/abs/1802.07207. Posted February 20, 2018. Accessed October 10, 2019.

23. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouw Y, Habbema JD. Internal validation of predictive model: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001;54(8):774–781.

24. Deurloo EE, Klein Zegeling WK, Teertstra HJ, et al. Contrast-enhanced MRI in breast cancer patients eligible for breast-conserving therapy: complementary value for subgroups of patients. Eur Radiol 2006;16(3):692–701.

25. Sigmund KC, Baar A, Vogel U, Kraemer B, Hahn M, Claussen CD. Risk-benefit analysis of preoperative breast MRI in patients with primary breast cancer. Clin Radion 2009;64(4):403–413.

26. Tillman GF, Orel SG, Schnall MD, Schultz DJ, Tan JE, Solin LJ. Effect of breast magnetic resonance imaging on the clinical management of women with early-stage breast carcinoma. J Clin Oncol 2002;20(16):3413–3423.

27. Houssami N, Turner RM, Morrow M. Meta-analysis of pre-operative magnetic resonance imaging (MRI) and surgical treatment for breast cancer. Breast Cancer Res Treat 2017;165(2):273–283.

28. Bansal GJ, Santosh D, Davies EL. Selective magnetic resonance imaging (MRI) in invasive lobular breast cancer based on mammographic density: does it lead to an appropriate change in surgical treatment? Br J Radiol 2016;89(1060):20150679.

29. Sardanelli F. Overview of the role of pre-operative breast MRI in the absence of evidence on patient outcomes. Breast 2010;19(1):3–6.