Diagnostic Architectural and Dynamic Features at Breast MR Imaging: Multicenter Study[1]

Mitchell D. Schnall, MD, PhD
Jeffrey Blume, PhD
David A. Bluemke, MD, PhD
Gia A. DeAngelis, MD
Nanette DeBruhl, MD
Steven Harms, MD
Sylvia H. Heywang-Köbrunner, MD, PhD
Nola Hylton, PhD
Christiane K. Kuhl, MD
Etta D. Pisano, MD
Petrina Causer, MD
Stuart J. Schnitt, MD
David Thickman, MD
Carol B. Stelling, MD
Paul T. Weatherall, MD
Constance Lehman, MD, PhD
Constantine A. Gatsonis, PhD

Purpose: To prospectively determine the prevalence and predictive value of three-dimensional (3D) and dynamic breast magnetic resonance (MR) imaging and contrast material kinetic features alone and as part of predictive diagnostic models.

Materials and Methods: The study protocol was approved by the institutional review board or ethics committees of all participating institutions, and informed consent was obtained from all participants. Although study data collection was performed before HIPAA went into effect, standards that would be compliant with HIPAA were adhered to. Data from the International Breast MR Consortium trial 6883 were used in the analysis. Women underwent 3D (minimum spatial resolution, 0.7 × 1.4 × 3 mm; minimal temporal resolution, 4 minutes) and dynamic two-dimensional (temporal resolution, 15 seconds) MR imaging examinations. Readers rated enhancement shape, enhancement distribution, border architecture, enhancement intensity, presence of rim enhancement or internal septations, and the shape of the contrast material kinetic curve. Regression was performed for each feature individually and after adjustment for associated mammographic findings. Multivariate models were also constructed from multiple architectural and dynamic features. Areas under the receiver operating characteristic curve (A_z values) were estimated for all models.

Results: There were 995 lesions in 854 women (mean age, 53 years ± 12 [standard deviation]; range, 18–80 years) for whom pathology data were available. The absence of enhancement was associated with an 88% negative predictive value for cancer. Qualitative characterization of the dynamic enhancement pattern was associated with an A_z value of 0.66 across all lesion architectures. Focal mass margins (A_z = 0.76) and signal intensity (A_z = 0.70) were highly predictive imaging features. Multivariate models were constructed with an A_z value of 0.880.

Conclusion: Architectural and dynamic features are important in breast MR imaging interpretation. Multivariate models involving feature assessment have a diagnostic accuracy superior to that of qualitative characterization of the dynamic enhancement pattern.

[1] From the Department of Radiology, Hospital of the University of Pennsylvania, 3400 Spruce St, Philadelphia, Pa 19104 (M.D.S.). The complete list of affiliations is at the end of this article. From the 2004 RSNA Annual Meeting. Received December 14, 2004; revision requested February 2, 2005; revision received March 10; final version accepted March 23. Supported by National Cancer Institute grants U01-CA74696 and U01-CA74680.

Breast magnetic resonance (MR) imaging is emerging as an important tool for the detection and characterization of breast cancer. The value of breast MR imaging is derived primarily from the high sensitivity of contrast material enhancement in the detection of breast cancer (1–3). The characterization of lesions as benign or malignant on the basis of MR imaging characteristics remains a challenge. The specificity of breast MR imaging has been reported to be between 20% and 100% in a diagnostic population (4–13). The specificity in screening populations is likely to be lower. There remains debate in the literature regarding interpretation strategy and the relative importance of architectural features and kinetic imaging curves in the discrimination of benign from malignant contrast enhancement. Previous studies aimed at identifying the predictive value of imaging features have focused primarily on the architectural or contrast material kinetic imaging features (4–13). Fewer investigators have attempted to integrate kinetic and temporal features (14–16).

In addition, there is little generalizable multicenter data to guide the integration of imaging features into a comprehensive interpretation strategy. The International Breast MR Consortium has conducted a large single-arm multicenter cohort trial to investigate breast MR imaging in women with a suspicious clinical finding or finding at conventional imaging work-up (17). The purpose of our study was to prospectively determine the prevalence and predictive value of three-dimensional (3D) and dynamic breast MR imaging and contrast material kinetic features alone and as part of predictive diagnostic models.

Materials and Methods

The data used in this analysis were collected in accordance with International Breast MR Consortium protocol 6883 by a group of 18 institutions in the United States, Canada, and Germany. The study protocol was approved by the institutional review board or ethics committee of all participating institutions and was funded by the U.S. National Cancer Institute. Informed consent was obtained from all participants. All institutions operated with a valid assurance from the Office for Human Research Protections of the U.S. Department of Health and Human Services. Although our study data collection was performed before the Health Insurance Portability and Accountability Act, or HIPAA, went into effect, we nevertheless adhered to standards that would be compliant with HIPAA. The gadolinium-based contrast agents used in the study were provided by GE Healthcare (Waukesha, Wis), Berlex Laboratories (Wayne, NJ), and Bracco Diagnostics (Milan, Italy). The authors had full control of the data and information submitted for publication.

Entrance Criteria

Patients with a suspicious or highly suspicious imaging finding (Breast Imaging Reporting and Data System grade of 4 or 5) at conventional imaging or suspicious clinical findings that were believed to require biopsy at the time of presentation were eligible for participation. Conventional imaging included mammography, and, when clinically appropriate, ultrasonography (US) and galactography. All participants older than 40 years were required to have undergone mammography within 6 weeks before or after their MR imaging examination if mammography had not previously been performed. For the purposes of the study, mammograms were interpreted by one of the study mammographers. This was required to be a different radiologist from the one who interpreted the MR images. Patients younger than 40 years with clinical findings could undergo US alone—a practice that was consistent with standard clinical care. Patients who had recently undergone fine-needle aspiration biopsy were eligible; however, the study mammogram and MR image readers remained blinded to the fine-needle aspiration biopsy results. Patients who had undergone core or excisional biopsy in the affected breast within 6 months before study entry were excluded. Additional exclusion criteria included a history of breast cancer in the affected breast, pregnancy, and any contraindication to MR imaging.

MR Imaging

Participants initially underwent a high-spatial-resolution 3D contrast material–enhanced MR imaging examination. The protocol for the 3D examination consisted of a T2-weighted rapid (fast or turbo) spin-echo sequence with 4-mm or thinner sections and a repetition time msec/echo time msec of approximately 4000/90. This was followed by a 3D T1-weighted gradient-echo sequence performed before and after the intravenous administration of 0.1 mmol (0.2 mL) of gadolinium chelate per kilogram of body weight. The gadolinium chelates used at this examination were the same ones used at the two-dimensional examination. The tubing used to deliver the contrast material was primed before the injection was initiated. The injection was administered over 10 seconds through a 20- or 22-gauge peripheral intravenous catheter and was followed by a flush of 20 mL of saline. Imaging was initiated after gadolinium chelate injection but before the saline flush. The imaging protocol con-

Published online
10.1148/radiol.2381042117

Radiology 2006; 238:42–53

Abbreviations:
A_z = area under receiver operating characteristic curve
CI = confidence interval
DCIS = ductal carcinoma in situ
3D = three-dimensional

Author contributions:
Guarantors of integrity of entire study, M.D.S., G.A.D., C.A.G.; 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; literature research, M.D.S., S.H., S.H.H., E.D.P., C.A.G.; clinical studies, M.D.S., D.A.B., G.A.D., N.D., S.H., S.H.H., N.H., C.K.K., P.C., D.T., C.B.S., P.T.W., C.L.; statistical analysis, J.B., E.D.P., C.A.G.; and manuscript editing, M.D.S., J.B., D.A.B., G.A.D., N.D., S.H., N.H., P.C., C.B.S., P.T.W., C.L., C.A.G.

Address correspondence to M.D.S.
(e-mail: *Mitchell.Schnall@uphs.upenn.edu*).

See Materials and Methods for pertinent disclosures.

sisted of a 3D gradient-echo sequence with a repetition time of 20 msec or less, an echo time of 4.5 msec or less, and a flip angle of 30°–45°. Imaging was performed with a 16–18-cm field of view over a minimum matrix of 256 × 128 and with 32–128 sections of 3 mm or less. The number of sections acquired and the section thickness depended on the size of the breast. The total imaging time for this acquisition was required to be less than 4 minutes. Fat suppression or image subtraction was used in all cases.

Patients with enhancing abnormalities were asked to return for dynamic MR imaging 18 hours or more after the first MR imaging examination. This time-resolved study consisted of the acquisition of a progressive saturation data set so that T1 relaxation time could be estimated (repetition times: 100, 200, 400, and 1200 msec) followed by the acquisition of sequential two-dimensional images at a 15-second temporal resolution over a total of 5 minutes after the intravenous administration of gadolinium chelate. So that time-resolved data from different sites could be directly compared, all sites used the same timing parameters for the dynamic time course and T1 measurements. Both acquisitions were developed from a two-dimensional spoiled gradient-echo sequence that involved an echo time of 4–5 msec (maintaining fat and water in phase at 1.5 T) and a 90° flip angle. The progressive saturation acquisition included repetition times of 0.1, 0.2, 0.4, and 1.2 seconds. All of the dynamic images were acquired with a repetition time of 0.1 second. The acquisition matrix for the dynamic series was 256 × 128, and the section thickness was 4 mm for all images. The acquisition of the first dynamic images began concurrently with the initiation of the injection of 0.1 mmol of gadolinium chelate per kilogram. Sites used Omniscan (GE Healthcare), Prohance (Bracco), or Magnevist (Berlex) administered over 10 seconds. The contrast material injection was followed by a 20-mL saline flush.

MR Image Interpretation

Twenty-seven radiologists (including M.D.S., D.A.B., G.A.D., N.D., S.H., S.H.H., C.K.K., E.D.P., P.C., D.T., and P.T.W.) from participating institutions were trained in the imaging and qualitative kinetic features before the investigation was initiated. All readers had more than 2 years of experience in interpreting breast MR images. A self-evaluation test and an atlas of sample cases with ratings were distributed. MR images were interpreted at host institutions in a normal clinical context (with access to mammograms if mammograms were available at the time of interpretation). MR image interpretation included an assessment of lesion architectural features. The feature list was designed to adhere as closely as possible to the standard being developed at that time by the American College of Radiology working group on breast MR imaging standards (18). This included a rating of the lesion architecture and contrast material kinetics if available. The architectural rating included an initial description of the enhancement type (focal, ductal, area, patchy), followed by specific descriptors within each enhancement type. A list of key architectural features is included in Table 1. Features were rated as either present or absent by using a four-point confidence scale (definitely present, possibly present, possibly absent, definitely absent).

In patients who underwent high-temporal-resolution dynamic imaging, a region of interest was chosen from the dynamic images for assessment of the lesion enhancement kinetics. Readers were instructed to choose the area of the lesion that demonstrated the greatest degree of early enhancement so that time–signal intensity curves could be generated. Region of interest sizes typically ranged from 2 to 10 mm. In addition, signal intensities were measured by using progressive saturation data in the same region of interest on the images obtained with different repetition times. Gadolinium concentration curves were calculated from the time–signal intensity curves and the estimated T1 with the assumption of fast exchange and a relaxivity of 4.5 $L \cdot mmol^{-1} \cdot sec^{-1}$ (19).

The kinetic data were analyzed qualitatively as washout, persistent enhancement, or plateau-shaped curves. Various quantitative metrics were extracted from the enhancement curve, including initial enhancement rate, maximal enhancement rate and amplitude, and enhancement rate at various time points. In addition, gadolinium concentration was calculated from the enhancement curves and T1 data.

Pathologic Evaluation

Each site was responsible for comparing lesions identified in pathology specimens with the MR and other imaging findings. Participating sites sent representative slides and their report for each lesion to a pathology core facility for review when these slides and reports were available. The slides were reviewed by a pathologist (S.J.S.) with 20 years of experience in breast disease. Otherwise, the pathologic diagnosis was extracted from local clinical pathology reports. In cases that resulted in disagreement with local pathologists, the individual site pathologists were notified by letter.

All pathology interpretations were categorized as indicating benign atypical findings, atypical findings, in situ carcinoma, or invasive carcinoma. Within each of these major categories, specific diagnoses were also collected.

Table 1

Architectural Features

Finding	Feature
Focal mass	Margin
	Shape
	Homogeneity
	Magnitude
	Enhancing rim
	Nonenhancing internal septation
	Internal septation on T2-weighted images
	T2 signal intensity
	Associated area enhancement
Area enhancement	Distribution: segmental, regional, or diffuse
	Form: stippled, clumped, or confluent
Ductal enhancement	Branching
	Border

Statistical Analysis

Initially, frequency tables of the MR imaging features were generated. The eligible and analysis populations were investigated to check the generalizability of the inferences of the analysis and potential effects of missing data.

Univariate logistic regression models were fit for each feature in an attempt to characterize the predictive ability of each feature for identifying a cancerous lesion. Invasive cancer and ductal carcinoma in situ (DCIS) were considered to be positive diagnoses, while all other diagnoses were considered to be negative. In addition, the predictive ability of each model was evaluated by using the area under the receiver operating characteristic curve (A_z) for the model's linear predictor (also referred to as the summary C-index measure in SAS software [SAS, Cary, NC]). This procedure was then repeated in models that adjusted for various patient characteristics: presence of a palpable mass, presence of calcifications on mammogram, presence of mass on mammogram, and mammographic density of breast parenchyma. The goal of this analysis was to determine if a particular patient characteristic changed the overall predictive ability of a feature. To account for clustering due to institutions (readers), generalized estimating equation techniques (20,21) were used. To assess how a combination of features or even what combination of features would be most predictive, several candidate multivariate models were developed and compared by using cross-validation techniques to enable some generalization beyond the study population.

Specifically, candidate models were developed in the following manner: Prominent combinations or clusterings of features were investigated with the classification and regression trees, or CART, algorithm (22). Optimal tree size was determined with cross validation, and several base models were chosen on the basis of scientific and statistical criteria. Base models suggested by the CART algorithm were then put into standard backward and forward stepwise selection procedures that included all feature variables (including kinetics variables), patient characteristics (eg, age, presence of palpable mass, density of breast parenchyma, presence of calcifications, presence of mass), and all possible three-way interactions present in the model, if indicated by the data. In this sense, the CART analysis provided our base models, and scientific subject matter and statistical model selection procedures were used to refine them.

After a set of candidate models were created, the Akaike information criterion and Bayesian information criterion (estimates of the goodness of fit of each model) and the A_z value were calculated for each model. These summary statistics were compared to arrive at the "best" models. Missing covariate values (primarily patient characteristics) were imputed by using the multiple imputations technique (23). As indicated, several imputed data sets were generated, and parameter values and a summary index were averaged across these data sets. To validate the A_z value for each model and correct its bias, we used the bootstrap method (24).

The best models were fit by using generalized estimating equation techniques to obtain standard errors that accounted for site clustering. Final estimates were obtained by averaging the parameter estimates from the multiple imputed data sets and adjusting the covariance matrix properly (25,26). The predictive ability of these models is described by the A_z values corrected for bias that were calculated in the previous step.

Table 2

Final Pathologic Diagnosis in and Clinical Manisfestation of 995 Lesions

Parameter	No. of Lesions*
Final pathologic diagnosis	
Invasive cancer	422 (42.4)
DCIS	77 (7.7)
Benign atypical lesion	63 (6.3)
Benign lesion	433 (43.5)
Clinical presentation	
Palpable lesion	456 (45.8)
Any mammographic finding	764 (76.8)
Mammographic finding of cancer	320 (32.2)
Mammographic finding of mass	412 (41.4)
Mammographic finding of mass and cancer	67 (6.7)
Mammographic finding of architectural distortion	98 (9.8)
Mammographic finding of focal asymmetric density	146 (14.7)

* Data in parentheses are percentages.

Results

Patient Population

A total of 1004 women were enrolled in the protocol, and 976 of them completed the MR imaging examinations. Pathology data were available for 854 women, who formed the case set used in the current analysis (analysis set). The mean age of the patients was 53 years ± 12 (standard deviation) (range, 18–80 years). The analysis set was compared with the total eligible population in terms of baseline and demographic variables. There was no indication of bias owing to missing pathologic diagnoses. Missing covariate information in the analysis set was minimal and imputed as described previously.

Pathologic Results

There were a total of 995 lesions in the 854 patients for which a pathologic diagnosis was known. Three hundred eighty-one patients with 410 lesions underwent a dynamic study. A breakdown of the 995 lesions according to presentation and final pathologic diagnosis is given in Table 2. In 913 (91.8%) of the lesions, the pathologic diagnosis was established by the core pathologist. The

rate of disagreement between the core and the local site pathologist in terms of identifying cancer (DCIS or invasive carcinoma) was less than 0.5% for those cases reviewed by the core pathologist, supporting the use of individual site pathology reports in this analysis. Twenty-three of the 995 lesions were excluded from the architectural feature analysis owing to incomplete feature data.

Univariate Models

The frequency and probability of cancer for the major architectural types are included in Table 3. The two most prevalent subtypes were focal mass and no enhancement, together representing approximately 84% (832 of 995) of all cases. Twenty-five (12%; 95% CI: 8.5%, 16.8%) of the 208 cases classified as showing no enhancement were associated with a pathologic diagnosis of cancer. Of the 25 cancers that were not observed to enhance, 12 were DCIS and 13 were invasive cancer. The lack of observed enhancement had a 94% negative predictive value (195 of 208 cases; 95% CI: 89.6%, 96.6%) for invasive cancer and an 88% negative predictive value (183 of 208 cases; 95% CI: 82.8%, 92.1%) for any cancer.

The A_z values for the univariate models for major predictive features within each enhancement subtype are included in Table 4. Images and time–signal intensity curves illustrating some of the key features used in this analysis are given in the Figure. The most predictive features for malignancy were margin (for focal masses), signal intensity, homogeneity, and qualitative enhancement kinetics. An analysis that adjusted for various patient characteristics, including the presence of calcifications, presence of associated palpable finding, presence of mass at mammography, and radiographic breast density revealed only a marginal effect on the diagnostic accuracy (ie, A_z value) of each image feature.

The subset of participants with area and ductal enhancement was too small to permit a meaningful analysis adjusted for the patient characteristics listed here. However, univariate analyses of the subset of area enhancement could be performed for the following variables: area enhancement distribution, area enhancement form, and the amount of enhancement (Table 4). Area enhancement distribution was the most predictive variable for malignancy, with an A_z value of 0.78.

The best-performing kinetic features are listed in Table 4. Qualitative assessment of the enhancement curve had the highest A_z value (ie, 0.66). This increased to 0.69 in women with a mammogram density at presentation and decreased minimally to 0.65 in

Table 3

Probability of Cancer and Frequency of Major Architectural Features

Feature	Percentage Likelihood of Malignancy*	Prevalence†	Prevalence in Lesions with Pathologic Verification‡
Focal mass	63.1 (52.4, 72.7)	819	624
Ductal enhancement	58.5 (42.5, 72.9)	51	41
Area enhancement	52.0 (32.5, 71.0)	95	75
Patchy enhancement	50.0 (35.2, 64.8)	26	24
No enhancement	12.0 (8.5, 16.8)	254	208

* Data in parentheses are 95% confidence intervals (CIs).

† In terms of total number of lesions with a given feature.

‡ Data are numbers of lesions. These lesions ultimately comprised the analysis set.

Table 4

A_z Values for Univariate Models

Feature	A_z Uni*	A_z Dens†	A_z Calc‡	A_z Palp§
Focal mass feature				
Margin	0.76	0.79‖	0.79‖	0.81
Shape	0.66	0.69‖	0.69	0.71
Intensity	0.70	0.75	0.76‖	0.77
Homogeneity	0.62	0.65	0.62‖	0.67
Associated area enhancement	0.56	0.59‖	0.55	0.63
Rim	0.58	0.62	0.59‖	0.65
Internal nonenhancing septations	0.55	0.58	0.54‖	0.64
Area enhancement feature				
Distribution	0.78	0.84	0.80	0.80
Form	0.64	NC	NC	NC
Intensity	0.71	0.83‖	0.81‖	0.83
Ductal enhancement feature				
Border	0.72	NC	NC	NC
Probability of branching	0.78#	NC	NC	NC
Dynamic feature**				
Qualitative kinetics	0.66	0.69	0.65	0.68
Maximum enhancement rate (%/sec)	0.54	0.65	0.69	0.63
Percentage enhancement at 1 minute	0.60	0.63	0.64	0.66

Note.—NC = model did not converge.

* A_z for the univariate model.

† A_z for model adjusted for mammographic breast density.

‡ A_z for model adjusted for presence of calcification on the mammogram.

§ A_z for model adjusted for presence of a palpable lesion.

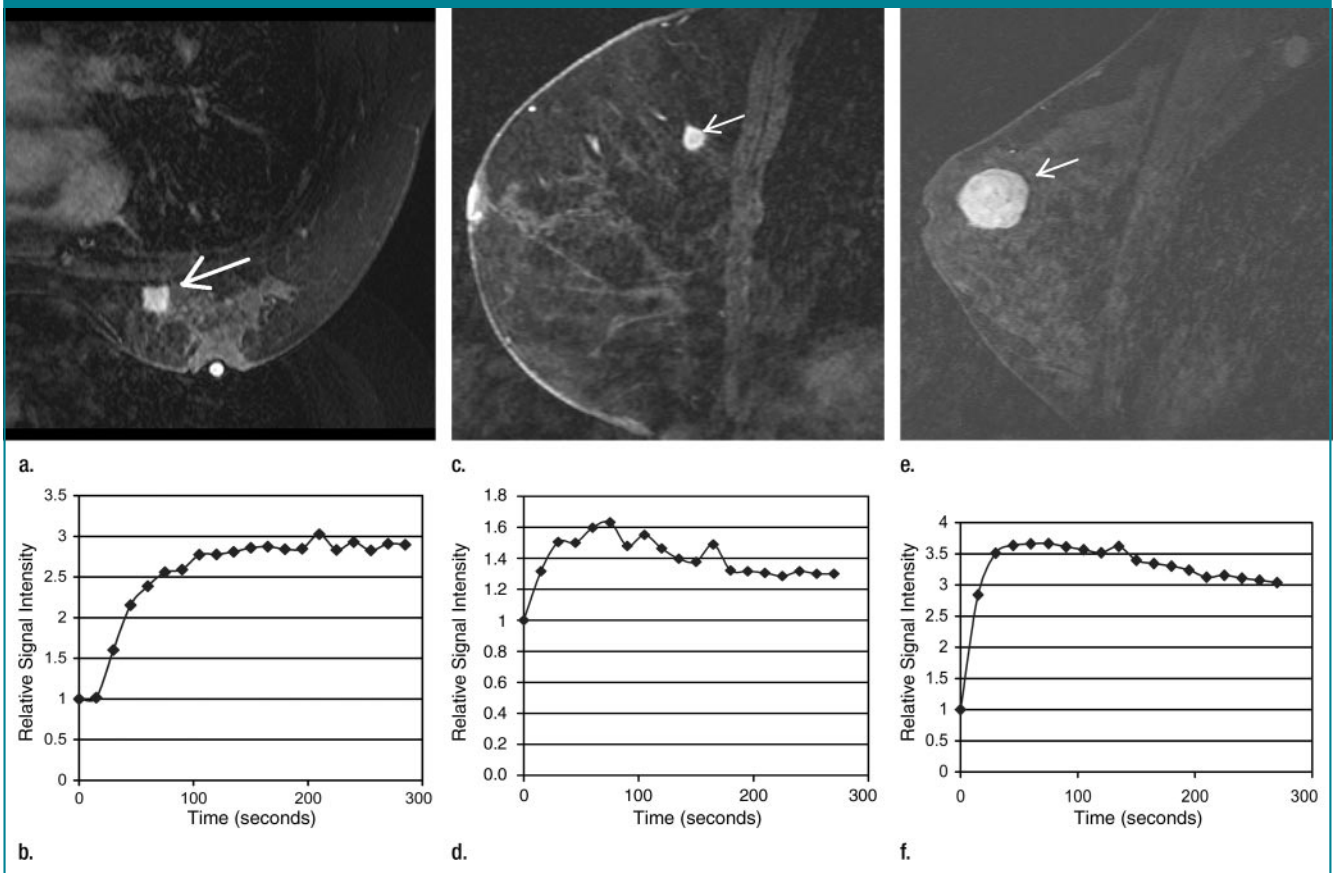
‖ Only main effect of stratification variable could be included because the interaction could not be supported.

Quasi-complete separation—model verified by other means.

** Based on all enhancing lesions.

women with mammogram calcification at presentation. Among the quantitative kinetic features, the feature that best predicted diagnosis was the maximum enhancement rate, which had an A_z value of 0.64. Of note, 55 (76%) of 72 kinetic curves that were described as washout were associated with a cancer diagnosis. In addition, 65 (45%) of 144 curves that were described as persistent were also associated with a cancer diagnosis. This accounted for 26% of the 248 cancers evaluated with the kinetic study. Thus, excluding cancer on the basis of a persistent enhancement curve alone would lead to false-negative results.

The A_z values of the univariate models provide an overall assessment of the potential influence of assessment of a specific feature on the overall classification of lesions as benign or malignant. In this sense, the A_z value is a summary measure of each feature's diagnostic accuracy. Features with marginal A_z values may still yield high positive and negative predictive values. Thus, although only a small subset of lesions will have these features, when present, these features are highly associated with a specific diagnosis. Specific low-prevalence feature ratings that appeared to be particularly predictive included associated regional enhancement among focal lesions (with a prevalence of 14.4% [90 of 624 cases] and a positive predictive value of 81% [73 of 90 cases; 95% CI: 71.5%, 88.6%]) and rim enhancement that was classified as definitely present (with a prevalence of 15.9% [99 of 624 cases] and a positive predictive value of 84% [83 of 99 cases; 95% CI: 75.1%, 90.5%]). The presence of nonenhancing internal septations was not a strong feature—only eight (47%) of 17 lesions



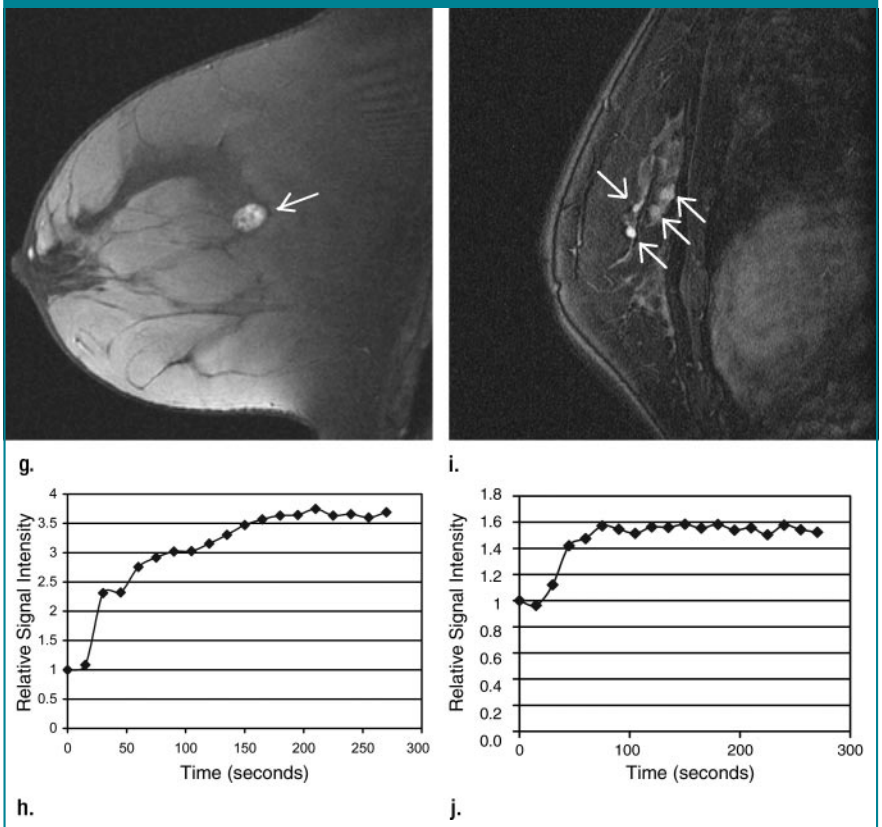
Enhancing lesions from five patients assessed according to the protocol guidelines described in the Materials and Methods section. **(a, b)** First patient. **(a)** Transverse 3D fat-suppressed spoiled gradient-echo postcontrast MR image shows an enhancing lesion (arrow) in the right breast that was described as a focal mass with irregular margins (no rim enhancement). **(b)** The time–signal intensity curve in this patient was characterized as persistent. The pathologic diagnosis was invasive cancer. **(c, d)** Second patient. **(c)** Sagittal 3D fat-suppressed spoiled gradient-echo postcontrast MR image shows an enhancing lesion (arrow) in the breast that was described as a focal mass with irregular margins. Rim enhancement was rated as probably present. **(d)** The time–signal intensity curve in this patient was characterized as a washout curve. Pathologic examination revealed invasive cancer. **(e, f)** Third patient. **(e)** Sagittal 3D fat-suppressed spoiled gradient-echo postcontrast MR image shows an enhancing lesion (arrow) that was described as a focal mass with a round shape and smooth borders. Internal septation was rated as definitely present. **(f)** The time–signal intensity curve in this patient was characterized as a washout curve. Pathologic examination revealed a benign lesion *(Figure continues).*

rated as having definitely present nonenhancing internal septations represented cancer.

Multivariate Models

Data from the best-performing multivariate models derived from forward and backward stepwise selection procedures (denoted as models 1 and 2, respectively) are detailed in Table 5. The mean A_z value for model 1 was 0.873 (range, 0.871–0.875), while the mean A_z value for model 2 was 0.880 (range, 0.877–0.884). For model 1, the average Akaike information criterion was 563.8 (range, 558.5–568.2) and the average Bayesian information criterion was 634.7 (range, 629.5–639.1), while for model 2, the average Akaike information criterion was 562.2 (range, 553.0–567.4) and the average Bayesian information criterion was 664.2 (range, 655.1–669.4). These were the two "best" candidate models, and they were indeed very similar. The Appendix details how these models were used to arrive at a predicted probability of cancer for an individual given her clinical profile and provides an example. CIs for this probability are not available in closed form; hence, they are not included here.

A multivariate model could not be developed for lesions with area or ductal enhancement because the data were too sparse.



(continued): **(g, h)** Fourth patient. **(g)** Individual image from two-dimensional dynamic MR acquisition (100/4.4; matrix, 128 × 256; section thickness, 4 mm) shows an enhancing lesion (arrow) that was described as a focal mass with an oval shape and smooth margin. **(h)** The time–signal intensity curve in this patient was characterized as persistent. Pathologic examination revealed a benign lesion. **(i, j)** Fifth patient. **(i)** Sagittal 3D fat-suppressed spoiled gradient-echo postcontrast MR image shows an enhancing focus that was described as showing clumped enhancement in a segmental distribution. Arrows point to the multiple foci of enhancement that comprise the finding. **(j)** The time–signal intensity curve in this patient was characterized as a plateau curve. Pathologic examination revealed invasive cancer.

Discussion

The results of this study clarify the important features contributing to the classification of lesions at breast MR imaging. As expected, a lack of enhancement was strongly predictive of benignity. There were 208 lesions that were reported as showing no enhancement, of which 25 were cancer. The 25 nonenhancing cancers included 12 cases of DCIS and 13 cases of invasive cancer. Thus, 16% of 77 DCIS lesions and 3% of 422 invasive cancers were coded as showing no enhancement. The bias of nonenhancing malignancies toward DCIS relative to the total percentage of malignancies represented in the population is expected. Previous investigators estimated the sensitivity of MR imaging in the detection of DCIS to be approximately 70% (27,28); thus, our finding of cases of DCIS that were associated with no reported enhancement is expected. However, the negative predictive value of nonenhancement for invasive cancer of 94% (95% CI: 89.6%, 96.6%) observed in this study is lower than what has been typically reported for single-institution studies (1,3,10). Although the image analysis paradigm used was very different from that used in the present study, in another multicenter study, Heywang-Kobrunner et al (29) found that 4% of invasive cancers (all smaller than 5 mm) could not be identified retrospectively at MR imaging. Thus, the absence of observed enhancement at MR imaging does not appear to exclude invasive cancer. It should be noted that many of the invasive cancers that were reported to not enhance in the current study were small or had small invasive components. A complete description of this cohort is beyond the scope of this article and will be reported separately.

Once detected, lesions need to be classified by type of enhancement. This classification is subjective and affects the performance of individual features within each major category. A study performed by the MR imaging lexicon committee (30) revealed that readers can differentiate masses from non–mass enhancement and were particu-

larly good at identifying mass enhancement. However, classification of types of non–mass enhancement (eg, ductal, regional) was not reliable.

In the present study, focal mass was the most prevalent architectural classification, accounting for 624 (62.7%) of 995 cases. The margin feature of the focal mass category was the single most predictive feature, with an A_z value for the univariate model of 0.76. This is consistent with results described in prior reports of the importance of margin analysis in breast MR image interpretation (10,15). The MR imaging lexicon committee reported a moderate agreement among readers ($\kappa = 0.55$) for this feature (30), indicating that margins can be rated relatively reliably. The results of the reader study published by the international working group on breast MR imaging indicated that although lesion shape was relatively predictive, readers had trouble differentiating shape from margin; hence, lesion shape was ultimately dropped from the breast MR imaging lexicon (30). In the present study, qualitative enhancement intensity was also highly predictive, with an A_z value in the univariate model of 0.70. The qualitative intensity was determined on the high-spatial-resolution 3D images, which were required to be acquired within 4 minutes of the contrast material injection. Because rectilinear k-space sampling in the y direction was used by most study sites, signal intensity on the high-spatial-resolution images most accurately reflects the signal intensity at 2 minutes or less.

Table 5

Odds Ratios Calculated with Best-performing Multivariate Models

Parameter	Model 1 Odds Ratio*	Model 2 Odds Ratio
Age	1.07 (1.03, 1.10)	1.07 (1.04, 1.11)
Palpable lesion		
Yes	2.16 (1.38, 3.37)	2.18 (1.44, 3.30)
No	1	1
Calcification		
Yes	NA	1.97 (1.07, 3.62)
No	NA	1
Qualitative kinetics		
Persistent	0.46 (0.29, 0.74)	0.46 (0.29, 0.71)
Plateau	1.69 (0.95, 3.00)	1.67 (0.93, 3.02)
Washout	2.14 (0.99, 4.65)	2.30 (1.05, 5.07)
Indeterminant	0.70 (0.31, 1.59)	0.64 (0.29, 1.40)
Not performed	1	1
Focal enhancement margin		
Smooth	1	1
Scalloped	3.19 (1.28, 7.97)	3.18 (1.28, 7.95)
Irregular	3.95 (1.87, 8.37)	4.30 (1.93, 9.61)
Spiculated	17.73 (7.85, 40.05)	19.94 (7.98, 49.80)
Enhancement intensity		
Minimal	1	1
Moderated	3.78 (1.15, 12.43)	6.15 (1.22, 31.11)
Marked	9.40 (2.72, 32.46)	26.08 (6.11, 111.26)
Enhancement rim		
None	1	1
Possible	0.96 (0.41, 2.28)	1.02 (0.42, 2.43)
Probable	0.82 (0.31, 2.14)	0.80 (0.28, 2.33)
Definite	2.64 (1.32, 5.26)	2.91 (1.47, 5.76)
Predominant signal intensity on T2-weighted images		
Low	NA	1
Intermediate	NA	1.74 (0.23, 13.29)
High	NA	12.61 (1.80, 88.18)
Associated area enhancement		
Yes	2.67 (1.38, 5.16)	2.53 (1.42, 4.49)
No	1	1
Enhancement signal intensity and predominant signal intensity on T2-weighted images		
Moderate and intermediate	NA	1.08 (0.13, 9.20)
Marked and intermediate	NA	0.45 (0.04, 5.53)
Moderate and high	NA	0.11 (0.03, 0.50)
Marked and high	NA	0.05 (0.01, 0.25)

Note.—Data in parentheses are 95% CIs. Model 1 was derived from a forward stepwise selection procedure; model 2, from a backward stepwise selection procedure.

* NA = not applicable. (These parameters were not used in model 1.)

The performance of the focal mass multivariate model in the present study was similar to the reported performance of the overall assessment of the radiologist (17). The readers involved in the present study were experienced in breast MR imaging and underwent additional training at the protocol initiation meetings. The fact that the performance of the multivariate models approaches reader performance would suggest that much of the key information used by readers to make their assessment is captured in these models. The focal mass multivariate models provide important insight into those features that should be considered in the diagnostic classification of enhancing focal masses. As suggested by results with the univariate models, margin and qualitative enhancement intensity (at 2 minutes or less after contrast agent injection) are the most important features in classifying focal masses. The next most important feature is the qualitative assessment of the kinetic curve. The relative risk of cancer for a lesion that has a washout curve as compared with the risk for a lesion that has a persistent curve is approximately five to one—a substantial difference.

Several low-prevalence features, in-

cluding rim enhancement or associated area enhancement, were highly correlated with a cancer diagnosis when present but were only observed in a small fraction of the cases, resulting in a modest A_z value. However, these features had an effect in the multivariate models. When definitely present, these features were associated with a greater risk of malignancy by a factor of between two and three. Although the value of rim enhancement is well known, to our knowledge, the importance of associated area enhancement has not been previously reported. In addition, it should be pointed out that the absence of these findings is not as valuable (ie, their absence is not a strong predictor of benignity). Presumably, this is related to the many factors—technical and physiologic—that contribute to making these features visible. In contradiction to results in prior single-institution studies (10), nonenhancing internal septation was not highly correlated with a benign outcome in our study. This may be related to the difficulty of generalizing individual features in a multicenter study. In model 2, the odds ratios in favor of cancer for high signal intensity on T2-weighted images were significant. However, model 1 had nearly identical performance but did not include this feature, which suggests that this feature may not be critical to consider and may be highly correlated with other features in the model.

The focal mass multivariate models are expected to serve as an important teaching and decision-support tool. Penn et al (31) have shown that a computer-aided diagnosis system involving the use of some similar variables was able to improve the performance of two of five readers. Further study is needed to assess and document the ability of these models to improve the performance of radiologists of various experience levels.

In the present study, the classification of enhancement as ductal was only modestly predictive of cancer (58.5% [24 of 41 cases; 95% CI: 42.5, 72.9]). Liberman et al (32) reported a positive predictive value of 26% for the ductal enhancement feature; however, this was in a heterogeneous patient population that had a lower prevalence of disease than did the diagnostic population targeted in the International Breast MR Consortium study. The tendency for ductal enhancement to represent cancer is consistent with the results of Nunes et al (10); however, the percentage of lesions with ductal enhancement that were cancerous in the current study was less than the 84% published by Nunes et al. The decrease in the predictive value of ductal enhancement in this multicenter study relative to that in the single-institution study of Nunes et al is probably the result of the ability of the large number of readers in the current study to reliably rate lesions as showing ductal enhancement. This inference is based on the MR imaging lexicon committee's report of a κ value of 0.35 (poor agreement) for the classification of enhancement distribution as linear branching (30). Refining feature descriptions and education will be a key to further improving reader performance.

For the lesions with area enhancement, the distribution was most predictive of diagnosis, with an A_z value in the univariate model of 0.78. The overall effect of the intensity and form of enhancement was more modest; however, a closer look at the data reveals specific feature ratings that were predictive. Stippled enhancement was associated with a low incidence of malignancy (25% [95% CI: 7.3%, 52.4%]), while other form ratings (heterogeneous, clumped, and homogeneous) were associated with a likelihood of cancer of 53% (95% CI: 34.3%, 71.7%), 60% (95% CI: 36.1%, 80.9%), and 67% (95% CI: 9.4%, 99.2%), respectively. In addition, a segmental distribution was associated with a 78% likelihood of cancer (95% CI: 56.3%, 92.5%); and a regional distribution, with a 21% likelihood of cancer (95% CI: 8.0%, 39.7%). These data support the practice of assuming that low-level stippled enhancement in a regional distribution indicates benignity, while segmental or clumped enhancement is cause for concern. The data also suggest that an interpreter should consider the intensity of area enhancement in his or her assessment.

Limitations of this study included patient population bias. The patient population was intended to represent as much as possible a true diagnostic population in which the results of conventional imaging did not conclusively indicate benign disease. In our cohort, 50% of lesions were malignant, a higher percentage than the reported biopsy yield expected in the target cohort (33). This would indicate an enrichment of this cohort with cancer lesions. Findings at clinical presentation in this population, including the fact that 45.8% of the lesions were palpable and 76.8% had mammographic findings (some women had both palpable lesions and lesions with mammographic findings) and the fact that 32.2% of the mammographic findings were findings of calcification, were similar to those expected in the target population. Another potential weakness was that this analysis included cases for which central pathology review was not available and the pathologic diagnosis was established in the local site pathology report. The high level of agreement between the core pathologist and the local site pathologist for the cases reviewed at the core facility supports the inclusion in this analysis of cases in which only pathology reports were available. This inclusion is further supported by the work of Collins et al (34), who showed 96% agreement between core and site pathologists in a multicenter trial of image-guided breast biopsy. This agreement was independent of biopsy type.

In addition, the analysis of the kinetic curves was limited to qualitative analysis and simple descriptive statistics. This analysis was aimed at evaluating predominant clinical practices. Advanced modeling will be considered for future work.

In conclusion, architectural and kinetic features are valuable in interpreting breast MR images. The most important feature was the presence of enhancement of any type; however, the absence of enhancement did not exclude invasive cancer. Other critical features included qualitative enhancement intensity and qualitative characterization of the dynamic curve. In addition, margin, rim enhancement, and associated area enhancement are important features of focal masses that should be considered in lesion characterization. Multivariate models developed from feature ratings have performances that

approach that of the overall assessment of highly experienced readers. The multivariate models can serve an important role as a training and decision support system for assisting less experienced radiologists practicing breast MR imaging.

Appendix

The parameter estimates for the multivariate models are as follows. With the forward stepwise selection procedure, the probability of disease (π) was modeled as follows:

$$\pi = \exp(\beta_0 + A \cdot 0.0634 + X\beta) \div (1 + \exp[\beta_0 + A \cdot 0.0634 + X\beta]),$$

where A is the age of the patient in years and $X\beta$ denotes the summation of the coefficients for the categorical covariates listed in Table A1.

For example, in a 50-year-old patient with a palpable lesion that had a spiculated focal enhancement margin, a washout enhancement curve, marked enhancement intensity, a definite enhancement rim, and associated area enhancement, the probability of disease would be calculated as follows:

$$\pi = \exp(-6.2835 + 50 \cdot 0.0634 + 2.8753 + 0.7619 + 2.2412 + 0.9690 + 0.9829) \div (1 + \exp[-6.2835 + 50 \cdot 0.0634 + 2.8753 + 0.7619 + 2.2412 + 0.9690 + 0.9829]) = 0.99.$$

With the backward stepwise selection procedure, π was calculated as follows:

$$\pi = \exp(\beta_0 + A \cdot 0.0683 + X\beta) \div (1 + \exp[\beta_0 + A \cdot 0.0683 + X\beta]),$$

Table A1

Coefficients for Categorical Covariates in Forward Stepwise Model

Covariate	Coefficient
Intercept β_0	−6.2835
Palpable lesion	0.7694
Impalpable lesion	Reference condition
Focal enhancement margin	
Scalloped	1.1614
Irregular	1.3749
Spiculated	2.8753
Smooth	Reference condition
Qualitative kinetics	
Persistent	−0.7715
Plateau	0.5255
Washout	0.7619
Indeterminant	−0.3556
Kinetic analysis not performed	Reference condition
Enhancement intensity	
Marked	2.2412
Moderate	1.3309
Minimal	Reference condition
Enhancement rim	
Definite	0.9690
Probable	−0.2019
Possible	−0.0378
None	Reference condition
Associated area enhancement	0.9829
No associated area enhancement	Reference condition

Table A2

Coefficients for Categorical Covariates in Backward Stepwise Model

Covariate	Coefficient
Intercept β_0	−7.5957
Palpable lesion	0.7775
Impalpable lesion	Reference condition
Calcifications present	0.6771
No calcifications present	Reference condition
Focal enhancement margin	
Scalloped	1.1580
Irregular	1.4592
Spiculated	2.9926
Smooth	Reference condition
Qualitative kinetics	
Persistent	−0.7850
Plateau	0.5153
Washout	0.8336
Indeterminant	−0.4526
Kinetic analysis not performed	Reference condition
Enhancement intensity	
Marked	3.2612
Moderate	1.8170
Minimal	Reference condition
Enhancement rim	
Definite	1.0669
Probable	−0.2174
Possible	0.0149
None	Reference condition
High predominant signal intensity on T2-weighted images	2.5342
Intermediate predominant signal intensity on T2-weighted images	0.5525
Low predominant signal intensity on T2-weighted images	Reference condition
Associated area enhancement	0.9275
No associated area enhancement	Reference condition
Marked enhancement signal intensity and high predominant signal intensity on T2-weighted images	−2.9402
Moderate enhancement signal intensity and high predominant signal intensity on T2-weighted images	−2.1727
Marked enhancement signal intensity and intermediate predominant signal intensity on T2-weighted images	−0.7954
Moderate enhancement signal intensity and intermediate predominant signal intensity on T2-weighted images	0.0759

where Xβ denotes the summation of the coefficients for the categorical covariates listed in Table A2.

Author affiliations: Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, Pa (M.D.S.); Center for Statistical Sciences, Brown University, Providence, RI (J.B., C.A.G.); Russell H. Morgan Department of Radiology and Radiological Sciences, Johns Hopkins University School of Medicine, Baltimore, Md (D.A.B.); Department of Radiology, University of Virginia Health System, Charlottesville, Va (G.A.D.); Department of Radiological Sciences, UCLA School of Medicine, Los Angeles, Calif (N.D.); Department of Radiology, University of Arkansas for Medical Sciences, Little Rock, Ark (S.H.); Technical University Munich, Munich, Germany (S.H.H.); Magnetic Resonance Science Center, Department of Radiology, University of California, San Francisco, Calif (N.H.); Department of Radiology, University of Bonn, Bonn, Germany (C.K.K.); Department of Radiology, University of North Carolina Chapel Hill, Chapel Hill, NC (E.D.P.); University of Toronto-Sunnybrook Cancer Care Center, Toronto, Ontario, Canada (P.C.); Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Mass (S.J.S.); Radiology Imaging Associates, Porter Adventist Hospital, Denver, Colo (D.T.); Division of Diagnostic Imaging, The University of Texas M. D. Anderson Cancer Center, Houston, Tex (C.B.S.); Department of Radiology, University of Texas-Southwestern Medical Center, Dallas, Tex (P.T.W.); and University of Washington Medical Center, Seattle, Wash (C.L.).

Acknowledgments: Gadolinium contrast agents were provided by GE Healthcare, Berlex Laboratories, and Bracco Diagnostics. The authors acknowledge the efforts of Cynthia Olson from the American College of Radiology (ACR) in project management, Sophia Sabina and Robert Sole from the ACR in data management, Brad Snyder from Brown University (Providence, RI) in programming the analyses, and Paul Stomper, MD, for participation in the mammogram quality control program. We acknowledge the support of Mark Elliot, PhD, from the Metabolic Magnetic Resonance Research and Computing Center of the University of Pennsylvania (grant RR-P41-02305) in the analysis of the kinetic data.

References

1. Harms SE, Flamig DP, Hesley KL, et al. MR imaging of the breast with rotating delivery of excitation off resonance: clinical experience with pathologic correlation. Radiology 1993;187:493–501.

2. Bone B, Pentek Z, Perbeck L, Veress B. Diagnostic accuracy of mammography and contrast-enhanced MR imaging in 238 histologically verified breast lesions. Acta Radiol 1997;38:489–496.

3. Kuhl CK, Schmutzler R, Leutner CC, et al. Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. Radiology 2000;215:267–279.

4. Gilles R, Guinebretiere JM, Lucidarme O, et al. Nonpalpable breast tumors: diagnosis with contrast-enhanced subtraction dynamic MR imaging. Radiology 1994;191:625–631.

5. Heywang SH, Wolf A, Pruss E, et al. MR imaging of the breast with Gd-DTPA: use and limitations. Radiology 1989;171:95–103.

6. Perman WH, Heiberg EM, Grunz J, et al. A fast 3D-imaging technique for performing dynamic Gd-enhanced MRI of breast lesions. Magn Reson Imaging 1994;12:545–551.

7. Kelcz F, Santyr GE, Cron GO, Mongin SJ. Application of a quantitative model to differentiate benign from malignant breast lesions detected by dynamic, gadolinium-enhanced MRI. J Magn Reson Imaging 1996;6:743–752.

8. Fischer U, Kopka L, Brinck U, Korabiowska M, Schauer A, Grabbe E. Prognostic value of contrast-enhanced MR mammography in patients with breast cancer. Eur Radiol 1997;7:1002–1005.

9. Kaiser WA, Zeitler E. MR imaging of the breast: fast imaging sequences with and without Gd-DTPA—preliminary observations. Radiology 1989;170(3 pt 1):681–686.

10. Nunes LW, Schnall MD, Orel SG, et al. Breast MR imaging: interpretation model. Radiology 1997;202:833–841.

11. Boetes C, Barentsz JO, Mus RD, et al. MR characterization of suspicious breast lesions with a gadolinium-enhanced TurboFLASH subtraction technique. Radiology 1994;193:777–781.

12. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology 1999;211:101–110.

13. Hulka CA, Smith BL, Sgroi DC, et al. Benign and malignant breast lesions: differentiation with echo-planar MR imaging. Radiology 1995;197:33–38.

14. Schnall MD, Rosten S, Englander S, Orel SG, Nunes LW. A combined architectural and kinetic interpretation model for breast MR images. Acad Radiol 2001;8:591–597.

15. Szabo BK, Aspelin P, Wiberg MK, Bone B. Dynamic MR imaging of the breast: analysis of kinetic and morphologic diagnostic criteria. Acta Radiol 2003;44:379–386.

16. Wedegartner U, Bick U, Wortler K, Rummeny E, Bongartz G. Differentiation between benign and malignant findings on MR-mammography: usefulness of morphological criteria. Eur Radiol 2001;11:1645–1650.

17. Bluemke DA, Gatsonis CA, Chen MH, et al. Magnetic resonance imaging of the breast prior to biopsy. JAMA 2004;292:2735–2742.

18. Ikeda DM. Progress report from the American College of Radiology Breast MR Imaging Lexicon Committee. Magn Reson Imaging Clin N Am 2001;9:295–302.

19. Tofts PS, Shuter B, Pope JM. Ni-DTPA doped agarose gel: a phantom material for Gd-DTPA enhancement measurements. Magn Reson Imaging 1993;11:125–133.

20. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13–22.

21. Karim R, Carey V. Splus GEE code from Statlib, 1992. http://lib.stat.cmv.edu/. Accessed November 1, 2005.

22. Breiman L, Friedman J, Olshen R, Stone C. Classification and regression trees. Belmont, Calif: Wadsworth, 1984.

23. van Buuren S, Oudshoorn CGM. Multivariate imputation by chained equations: MICE V1.0 user's manual. Report PG/VGZ/00.038. Leiden, the Netherlands: TNO Prevention and Health, 2000.

24. Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 1983;78:316–331.

25. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996;91:473–518.

26. Rubin DB, Schafer JL. Efficiently creating multiple imputations for incomplete multivariate data. ASA Proceedings of the Statistical Computing Section, 83–88. Alexandria, Va: ASA, 1990.

27. Boetes C, Strijk SP, Holland R, Barentsz JO, Van Der Sluis RF, Ruijs JH. False-negative MR imaging of malignant breast tumors. Eur Radiol 1997;7:1231–1234.

28. Orel SG, Mendonca MH, Reynolds C, Schnall MD, Solin LJ, Sullivan DC. MR imaging of ductal carcinoma in situ. Radiology 1997;202:413–420.

29. Heywang-Kobrunner SH, Bick U, Bradley WG Jr, et al. International investigation of breast MRI: results of a multicentre study (11 sites) concerning diagnostic parameters for contrast-enhanced MRI based on 519 histopathologically correlated lesions. Eur Radiol 2001;11:531–546.

30. Ikeda DM, Hylton NM, Kinkel K, et al. Development, standardization, and testing of a

lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. J Magn Reson Imaging 2001;13:889–895.

31. Penn AI, Kumar NA, Thompson SF, Schnall MD, Wand F, Gatsonis C. Preliminary performance analysis of breast MRI CAD system. In: Sonka M, Hanson KM, eds. Proceedings of SPIE: medical imaging 2001—image processing. Vol 4322. Bellingham, Wash: International Society for Optical Engineering, 2001; 1944–1953.

32. Liberman L, Morris EA, Dershaw DD, Abramson AF, Tan LK. Ductal enhancement on MR imaging of the breast. AJR Am J Roentgenol 2003;181:519–525.

33. Hall FM, Storella JM, Silverstone DZ, et al. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. Radiology 1988; 167:353–358.

34. Collins LC, Connolly JL, Page DL, et al. Diagnostic agreement in the pathologic interpretation of image-guided breast core needle biopsies. Am J Surg Pathol 2004;28:126–131.