Use of the American College of Radiology BI-RADS Guidelines by Community Radiologists: Concordance of Assessments and Recommendations Assigned to Screening Mammograms

Constance Lehman[1]
Sarah Holt[2]
Susan Peacock[3]
Emily White[3]
Nicole Urban[3]

OBJECTIVE. This study evaluated the use of the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) by community radiologists by determining the concordance of assessment categories and recommendations assigned to screening mammograms.

MATERIALS AND METHODS. The study comprised the interpretations of 82,620 consecutive screening mammograms by 18 radiologists between January 1, 1995, and December 31, 1998. For all mammograms, assessment categories and recommendations were compared to determine whether they were in accordance with BI-RADS guidelines. Overall patterns of discordance were analyzed, and comparisons of discordant patterns by assessment category, patient age, breast density, and year of examination were made.

RESULTS. The overall discordance between BI-RADS assessments and recommendations was low (3%). The assessment with the highest discordance was "probably benign finding" (category 3), at 53.5%. Mammograms obtained in 1998 were almost half as likely to have assessment–recommendation discordance compared with those obtained in 1995 (2.4% vs 4.5%, respectively; odds ratio = 0.52; $p < 0.001$). Mammograms of women with dense breast tissue were 30% more likely to have lesions assigned discordant assessments and recommendations compared with those of women with fatty tissue (3.4% vs 2.7%, respectively; odds ratio = 1.3; $p < 0.001$). No differences in the patterns of discordance were found between mammograms of women younger than 50 years and those of women 50 years old and older ($p = 0.10$).

CONCLUSION. There has been improvement in the accurate application of BI-RADS since its introduction. However, variation in the pairing of BI-RADS assessments and recommendations persists. Continued efforts to educate radiologists about the use of BI-RADS and to clarify BI-RADS terms would promote maximum consistency in this use of this reporting method.

[1]Department of Radiology, University of Washington, Seattle Cancer Care Alliance, 825 Eastlake Ave., G4-830, Seattle, WA 98109-1023. Address correspondence to C. Lehman.

[2]Department of Orthopedics, University of Washington, Harborview Medical Center, 325 Ninth Ave., Box 359798, Seattle, WA 98104-2499.

[3]Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109-1024.

In response to concerns raised both within and outside of the radiology community, the American College of Radiology developed a task force on breast cancer in the late 1980s and appointed a committee to develop guidelines for standardized reporting of mammographic findings. This work was published as the *Breast Imaging Reporting and Data System*, and this classification system is referred to as BI-RADS [1, 2]. This standardized system helps clinicians understand more clearly the disposition of their patients; aids in auditing mammography practices; and facilitates research efforts, particularly toward the development of large mammography databases.

The BI-RADS lexicon provides a dictionary of terms to use when describing lesions seen on mammography. In addition to these definitions, BI-RADS provides specific final assessment categories and the associated recommendations to be assigned to each mammogram. The standardized assessment categories used to describe findings on mammography include "need additional imaging evaluation" (category 0), "negative" (category 1), "benign finding" (category 2), "probably benign finding" (category 3), "suspicious abnormality" (category 4), and "highly suggestive of a malignancy" (category 5). Recommendations that are appropriate for each of the six assessment categories are provided by BI-RADS.

Early studies indicated the BI-RADS lexicon not only helped standardize the language used in interpreting mammograms, but also could improve the sensitivity and specificity of mammographic interpretations [3, 4]. In a study of five radiologists using BI-RADS lesion descriptors for masses and calcifications, moderate inter- and intraobserver agreements

were found in the interpretations of 60 mammograms [5]. Three subsequent independent reports indicated the BI-RADS lesion descriptors and assessment categories may improve the positive predictive value of mammographic interpretations [6–8]. Since its inception, BI-RADS has become widely used in both academic and private practice settings, and certain elements of BI-RADS are required for certification through the Mammography Quality Standards Act [9].

To date, a report about the use of the BI-RADS assessment categories and the associated recommendations in community practice has not been published. As databases from various community practices are pooled, final assessment categories 1 through 5 are used to calculate overall positive predictive values, sensitivity, and specificity. However, whether these assessment categories are used in a consistent manner in community practices is unclear. It is possible that the same assessment categories are associated with different recommendations for follow-up. The purpose of this study was to evaluate the use of the BI-RADS lexicon by community radiologists over time by determining the concordance of assessment categories with recommendations assigned to screening mammograms.

Materials and Methods

The Mammography Tumor Registry (MTR), established in 1994, is a linked cancer surveillance system that combines participating Washington state facilities' mammography records with breast cancer cases identified by the Cancer Surveillance System and the Washington State Cancer Registry. The MTR is used for medical outcome audits of facilities and radiologists, while serving as a foundation for collaborative interdisciplinary research on breast cancer detection and progression.

The data in this study were limited to MTR facilities that use the BI-RADS coding scheme and that reported both assessments and recommendations. Mammographic interpretations came from 18 radiologists in three facilities. In order that the data represent the range of practice patterns represented in the MTR, the data were randomly sampled so that no facility contributed more than 50% and no radiologist contributed more than 10% of the total examinations to the data set. Facilities included in the analyses contributed data between January 1, 1995, and December 31, 1998. Examinations were designated as diagnostic if the facility recorded the examination type as diagnostic or if a lump or bloody nipple discharge was reported. Examinations were considered screening if the examination type was reported as screening and if no specific symptoms or clinical findings were reported. Only screening mammograms were included in this study.

For each examination, the assessment category and recommendation were compared to determine whether the association was in accordance with the published guidelines of the American College of Radiology BI-RADS manual [1]. Not all the recommendations listed in the BI-RADS manual are specifically listed in the MTR database. For example, although BI-RADS specifically lists magnification views, spot compression views, and spot magnification views, and additional projections as possible recommendations associated with category 0, each of these specific variables is not included in the MTR database. Instead, the generic variable "additional views" is used in the MTR database. Table 1 lists concordant assessment–recommendation pairings for BI-RADS and the corresponding recommendations used in the MTR database. In summary, for a lesion described as BI-RADS category 0, the recommendation of needs additional imaging evaluation or comparison with prior mammograms was considered concordant. For findings characterized as BI-RADS category 1 or 2, either a recommendation of normal-interval follow-up or no specific recommendation was accepted as concordant. For BI-RADS category 3 lesions, short-term follow-up must have been recommended to be considered concordant. For BI-RADS categories 4 and 5, a recommendation for tissue sampling or surgical consultation must have been made.

Of the 82,620 mammograms used for this analysis, 588 (0.7%) were assigned two or more recommendations. Most of the examinations with multiple recommendations (79.3%) had an assessment of category 0 in addition to another assessment. For mammograms with two or more recommendations, if one of the recommendations was concordant with the assessment, then that recommendation was used in our analysis and the other recommendation was dropped. If none of the recommendations was concordant, we prioritized the recommendations on the basis of severity and selected the most severe recommendation for our analysis.

We divided the analysis set into groups to examine discordant patterns. Patterns of discordance were compared by assessment category, patient age, breast density, and year of examination. Examination results

TABLE I	Definitions for Assigning Concordance Between BI-RADS Assessments and Recommendations		
BI-RADS Assessment	Definition of BI-RADS Category	Concordant Recommendations	
		BI-RADS	Mammography Tumor Registry
Category 0	Need additional imaging evaluation	Additional projections Magnification views Spot compression views Spot magnification views Sonography Old films for comparison Ductography	Additional views Sonography Immediate prior film comparison Routine film comparison
Category 1	Negative	Normal-interval follow-up Decision for biopsy should be based on clinical assessment	Normal-interval follow-up No specific recommendation made
Category 2	Benign finding	Normal-interval follow-up Decision for biopsy should be based on clinical assessment	Normal-interval follow-up No specific recommendation made
Category 3	Probably benign finding; short-term (1–11 months) follow-up suggested	Follow-up at short interval (1–11 months)	Short-term follow-up
Category 4	Suspicious abnormality; biopsy should be considered	Biopsy should be considered Needle localization and biopsy	Fine-needle aspiration Cyst aspiration Biopsy
Category 5	Highly suggestive of malignancy; appropriate action should be taken	Histology using core biopsy Take appropriate action Cytologic analysis	Core needle biopsy Surgical biopsy Surgical consultation Immediate unspecified follow-up

Note.—BI-RADS = *Breast Imaging Reporting and Data System* [1].

for women younger than 50 years old and 50 years old and older, women with dense breast tissue versus those with fatty breast tissue, and those examined in 1995–1996 versus those examined in 1997–1998 were compared. Examinations that reported heterogeneously or extremely dense breast tissue were classified as dense, whereas examinations that reported breast tissue as entirely fat or scattered fibroglandular densities were classified as fatty.

Results of the recommendations by BI-RADS assessment category were put into tabular form for comparison. Discordance values for each assessment and recommendation were compared as a whole group and as several subgroups. To test whether discordance was significantly different among subgroups, we compared the proportions of overall discordance for patient age ($<$50 years vs $\geq$50 years), year of examination (1995–1996 vs 1997–1998), and breast density (fatty vs dense) using the Pearson's correlation coefficent and the chi-square test. An odds ratio was produced to further illustrate the magnitude of the difference. In our study, the crude odds ratio estimates the odds or likelihood that an assessment and recommendation for a given breast density, patient age, or year of examination would be discordant. Confidence intervals were based on the standard error of the coefficient and normal approximation. To analyze the trends of discordance over time, we analyzed the year of examination as a continuous variable and used logistic regression.

Results

Table 2 shows the assessments and recommendations given for the 82,620 mammograms in the study. The overall concordance of mammographic assessments with their appropriate recommendations was 97.1%. The categories negative (category 1), benign finding (category 2), and additional imaging required (category 0) were highly concordant (96–98%) with the appropriate recommendations. The assessment category with the lowest concordance was probably benign finding (category 3) at 53.5%. Thirty-three percent of the cases with an assessment of probably benign finding were assigned a recommendation of normal-interval follow-up, 17.0% were assigned a recommendation for additional imaging or sonography, and 2.8% were assigned a recommendation for biopsy.

Of the cases with an assessment of suspicious abnormality (category 4), 84.5% had a concordant recommendation of either a biopsy or an appropriate action. The remaining 15.5% of these cases had a discordant recommendation including 7.9% with a recommendation for additional imaging; 2.6%, for sonography; 2.6%, for normal-interval follow-up; and 1.3%, for short-interval follow-up. Of the cases assessed as highly suggestive of malignancy (category 5), 90.6% had a concordant recommendation. However, 7.6% of the mammograms assessed as highly suggestive of malignancy were associated with a recommendation for normal-interval follow-up. A total of 18 cases had an assessment of BI-RADS category 4 or 5, but these cases had been assigned the recommendation of normal-interval follow-up.

Tables 3 and 4 show the assessments and recommendations given for mammograms during the first half (1995–1996) and second half (1997–1998) of the study. The overall discordance of mammographic assessments and recommendations decreased over time—from 3.6% to 2.4% ($p < 0.001$). Mammograms obtained in 1998 were half as likely to have assessment–recommendation discordance than those obtained in 1995 (2.4% vs 4.5%, respectively; odds ratio = 0.52; $p < 0.001$). The decrease in

TABLE 2	Concordance and Discordance of BI-RADS Assessments and Recommendations Made for 82,620 Screening Examinations							
BI-RADS Assessment	Definition of BI-RADS Category	Recommendations						Total Discordance for Assessment (%)
		Normal-Interval Follow-Up or None	Additional Views	Sonography	Short-Interval Follow-Up	Invasive	Other	
Category 1	Negative finding							
%		**98.7**	0.6	0.0	0.7	0.0	0.0	1.3
No.		**69,602**	387	27	470	13	29	
Category 2	Benign finding							
%		**97.6**	0.6	0.1	1.1	0.1	0.5	2.4
No.		**5,080**	33	4	59	7	24	
Category 0	Need additional imaging evaluation							
%		3.4	**74.6**	**21.9**	0.0	0.1	0.1	3.6
No.		143	**3,187**	**936**	1	5	2	
Category 3	Probably benign finding							
%		33.0	11.8	5.2	**46.5**	2.8	0.7	53.5
No.		701	250	111	**987**	59	16	
Category 4	Suspicious abnormality							
%		2.6	7.9	2.6	1.3	**84.5**	1.1	15.5
No.		10	30	10	5	**322**	4	
Category 5	Highly suggestive of malignancy							
%		7.6	0.9	0.0	0.0	**90.6**	0.9	9.4
No.		8	1	0	0	**96**	1	
Total discordance for recommendation (%)		1.1	18.0	14.0	35.2	16.7	100.0	2.9

Note.—Concordant data are set in **boldface** type. Percentage values indicate the proportion of recommendations within each assessment level. BI-RADS = *Breast Imaging Reporting and Data System* [1].

discordance between 1995–1996 and 1997–1998 primarily resulted from the more accurate association of assessments with the appropriate recommendations for negative, benign finding, and probably benign finding (categories 1–3). Nevertheless, in 1997–1998, almost 50% of the lesions described as probably benign finding were associated with a discordant recommendation: normal-interval follow-up was recommended for 30.8%; additional imaging, for 9.8%; sonography, for 4.4%; and biopsy, for 3.3%. We found no improvement over time in concordance patterns for suspicious abnormality or highly suggestive of malignancy, categories 4 and 5, respectively.

Table 5 shows the association of patient age, breast density, and year of examination with patterns of discordance. The patterns of discordance for mammograms of women younger than 50 years versus those of women 50 years old and older did not differ ($p = 0.10$). However, women with dense breast tissue had a 30% increased likelihood of having discordant assessments and recommendations compared with women with fatty breast tissue (3.4% vs 2.7%, respectively; odds ratio = 1.3; $p < 0.001$). Ex-

aminations during the latter half of the study (1997–1998) were significantly less likely to have discordant assessment–recommendation pairings compared with examinations during the first half (1995–1996) of the study (2.4% vs 3.6%, respectively; odds ratio = 0.62; $p < 0.001$). In addition, logistic regression analysis showed a constant linear improvement over the 4 years of the study ($p < 0.001$).

Discussion

We evaluated the use of BI-RADS by community radiologists by assessing the concordance of assessment categories and recommendations assigned to screening mammograms. We found that the overall concordance of BI-RADS assessments and recommendations was high (97.1%). In addition, overall concordance improved over the 4 years of our study. However, we did find a disproportionately greater variation of recommendations associated with the assessment probably benign (BI-RADS category 3). This pattern of variation was persisted across all of the study years.

Taplin et al. (unpublished data) investigated the use in 1997 of BI-RADS by seven registries within the breast cancer surveillance consortium. Our data are consistent with their findings of high concordance with recommendations given to negative and benign assessments, but moderate concordance (41%) with recommendations given to probably benign assessments. From our results, improvement would be predicted to continue over time, particularly with continued educational efforts and possible refinements to the lexicon.

There are two issues of possible concern when a discordant assessment and recommendation are given. The primary concern is the disposition of the patient. For BI-RADS to achieve maximum benefit, the specific recommendation given for the appropriate management of the patient needs to be clear. For some of the discordant patterns, the final disposition of the patient may have been correct. For example, when additional views are recommended for an assessment of probably benign finding, the patient may have the additional views at another site or another time. Based on those additional views, the as-

TABLE 3	Concordance and Discordance of BI-RADS Assessments and Recommendations Made for 36,529 Screening Examinations Performed in 1995 and 1996							
BI-RADS Assessment	Definition of BI-RADS Category	Recommendations						Total Discordance for Assessment (%)
		Normal-Interval Follow-Up or None	Additional Views	Sonography	Short-Interval Follow-Up	Invasive	Other	
Category 1	Negative finding							
%		**98.4**	0.6	0.1	0.9	0.0	0.0	1.6
No.		**30,944**	199	18	276	7	5	
Category 2	Benign finding							
%		**96.6**	0.9	0.1	2.1	0.1	0.3	3.5
No.		**1,652**	15	2	35	2	5	
Category 0	Need additional imaging evaluation							
%		1.2	**76.3**	**22.1**	0.1	0.2	0.1	1.6
No.		22	**1,419**	**411**	1	4	2	
Category 3	Probably benign finding							
%		34.6	13.2	5.8	**43.6**	2.4	0.4	56.4
No.		428	163	72	**540**	30	5	
Category 4	Suspicious abnormality							
%		1.4	7.5	3.3	1.9	**84.5**	1.4	15.5
No.		3	16	7	4	**180**	3	
Category 5	Highly suggestive of malignancy							
%		5.1	1.7	0.0	0.0	**91.5**	1.7	8.5
No.		3	1	0	0	**54**	1	
Total discordance for recommendation (%)		1.4	21.7	19.4	36.9	15.5	100.0	3.6

Note.—Concordant data are set in **boldface** type. Percentage values indicate the proportion of recommendations within each assessment level. BI-RADS = *Breast Imaging Reporting and Data System* [1].

TABLE 4	Concordance and Discordance of BI-RADS Assessments and Recommendations Made for 46,091 Screening Examinations Performed in 1997 and 1998							
BI-RADS Assessment	Definition of BI-RADS Category	Recommendations						Total Discordance for Assessment (%)
		Normal-Interval Follow-Up or None	Additional Views	Sonography	Short-Interval Follow-Up	Invasive	Other	
Category 1	Negative finding							
%		**98.9**	0.5	0.0	0.5	0.0	0.1	1.1
No.		**38,658**	188	9	194	6	24	
Category 2	Benign finding							
%		**98.1**	0.5	0.1	0.7	0.1	0.5	2.0
No.		**3,428**	18	2	24	5	19	
Category 0	Need additional imaging evaluation							
%		5.0	**73.2**	**21.7**	0.0	0.0	0.0	5.1
No.		121	**1,768**	**525**	0	1	0	
Category 3	Probably benign finding							
%		30.8	9.8	4.4	**50.5**	3.3	1.2	49.6
No.		273	87	39	**447**	29	11	
Category 4	Suspicious abnormality							
%		4.2	8.3	1.8	0.6	**84.5**	0.6	15.5
No.		7	14	3	1	**142**	1	
Category 5	Highly suggestive of malignancy							
%		10.6	0.0	0.0	0.0	**89.4**	0.0	10.6
No.		5	0	0	0	**42**	0	
Total discordance for recommendation (%)		1.0	14.8	9.2	32.9	18.2	100.0	2.4

Note.—Concordant data are set in **boldface** type. Percentage values indicate the proportion of recommendations within each assessment level. BI-RADS = *Breast Imaging Reporting and Data System* [1].

sessment–recommendation pairing of probably benign finding and short-interval follow-up or the pairing of negative and normal-interval follow-up may have then been given. The end result is the appropriate management of the patient, although an original erroneous assessment of category 3 (probably benign finding) rather than category 0 (needs additional imaging evaluation) was given.

Similarly, when a probably benign finding is paired with normal-interval follow-up (33%), it is possible that the patient management was correct in that the patient was in the second or third year of follow-up for stability of a probably benign lesion. At that time, a 1-year rather than 6-month follow-up interval was deemed appropriate. We wondered if this scenario could have been the case in our sample. We investigated further the 701 mammograms with an assessment of probably benign finding and normal-interval follow-up or no recommendation. We found that 283 (40%) of the 701 mammograms were of patients who had a mammogram obtained within the previous 2 years. However, of these 283 patients, only 12 (4%) had a

TABLE 5	Association of Patient Age, Breast Density, and Year of Examination with Discordant BI-RADS Assessment and Recommendations			
Factor	Mammograms (n = 81,587)		Odds Ratio (Confidence Interval)	
	No.	%	Unadjusted [a]	Adjusted [b]
Age (yr)				
<50	28,524	35.0	1.00	1.00
≥50	53,063	65.0	0.93 (0.86–1.01)	1.00 (0.91–1.09)
p			0.1	0.1
Breast density				
Fatty	57,628	70.6	1.00	1.00
Dense	23,959	29.4	1.29 (1.18–1.41)	1.28 (1.17–1.40)
p			<0.001	<0.001
Year of examination				
1995–1996	36,389	44.6	1.00	1.00
1997–1998	45,198	55.4	0.62 (0.57–0.67)	0.62 (0.57–0.67)
p			<0.001	<0.001

Note.—Analysis includes all mammograms with valid responses for assessment, recommendation, age, density, and year of examination. BI-RADS = *Breast Imaging Reporting and Data System* [1].

[a] Odds of discordance between assessment and recommendation of mammogram in this group versus reference group.

[b] Odds ratio is adjusted for other factors listed, which calculates the influence of each factor in relation to the others.

mammogram with a recommendation of short-interval follow-up within the 2 years before the mammogram with the assessment–recommendation pairing of probably benign finding and normal-interval follow-up was obtained. Finally, mammograms of lesions being followed up every 6 months should not have been in our study sample because these images should have been reported as diagnostic mammograms rather than as screening mammograms. Only the first screening mammogram with a lesion described as a probably benign finding should have been in our study sample, and all of those initial examinations should have been given a final assessment of probably benign finding and the recommendation of short-interval follow-up.

Clarifying the criteria for the probably benign finding category would be a helpful addition to the BI-RADS lexicon. A specific case to illustrate potential confusion is the reporting of an initial screening mammogram showing a mass. The initial screening mammogram would be reported as showing a category 0 lesion, need additional imaging evaluation. Subsequent spot compression views and sonograms would then be obtained, and only after those evaluations could the mass be assessed as a probably benign finding with the recommendation of short-interval follow-up. The mass is found to have remained stable at the first 6-month follow-up, and the lesion is again given an assessment and recommendation of probably benign finding and short-interval follow-up. At the second evaluation (1 year from the initial screening mammogram), the mass is said to have remained stable. At this point, a 1-year follow-up rather than a 6-month follow-up is deemed appropriate. On the basis of the mammographic findings, should the mass be assigned the assessment and recommendation of benign finding and normal-interval follow-up before 2–3 years of stability has been established, or should the mass be assigned the assessment and recommendation of probably benign finding and normal-interval follow-up? This issue could be clarified in future publications of the BI-RADS lexicon.

For some discordant patterns that we observed, the disposition of the patient is more concerning. For example, why normal-interval follow-up was recommended for 18 patients who were said to have a category 4 or 5

lesion on screening mammography is not clear. This discordant recommendation of normal-interval follow-up represents 4% of the patients with a lesion designated as a suspicious abnormality or as highly suggestive of malignancy. In these 18 cases, we examined previously obtained mammograms and reviewed each patient's cancer history at the time of and after mammography; however, these efforts failed to yield an explanation for the discordance in any of these cases. For example, these cases did not represent patients for whom a cancer diagnosis was already known and the normal-interval follow-up recommendation was for the contralateral breast. We also did not find any cases for which these aberrant category 4 or 5 assessments were followed up within 6 months by a cancer diagnosis, suggesting that the assessment and not the normal-interval follow-up recommendation guided the events. For discordant recommendations like these to be avoided, mammography sites in the course of their site audit could flag discordant assessment–recommendation pairings to identify potential miscommunications to patients or to referring clinicians regarding follow-up care.

The second issue in regard to concordance patterns is whether the appropriate data are used for audit and research purposes. Most practices use assessment categories alone in their auditing practices. However, with discordant recommendations assigned to assessments, two different practices could yield different audit results—even though the same recommendations were made to the patient. Similarly, in multicenter trials, pooling data from multiple sites involves decisions to pool either the assessment or the recommendation variable or both. Our findings indicate that one cannot always assume a given assessment leads to a specific recommendation.

Whether discordant assessments and recommendations primarily affect the outcome of the patient or the accuracy of audit data is unclear. This study suggests that there is significant variation in the pairing of probably benign finding with different recommendations. Fewer than half of the mammograms assessed as probably benign finding were recommended for short-interval follow-up.

A particular strength of this study is that it was conducted in community-based mammography practices using a large database from radiologists who used the BI-RADS

system over a defined period of time. A limitation of our study is the absence of specific follow-up data. In a study of this size for which data are collected from multiple community practices, it is not possible to assess details of individual discordant examinations or details of specific patient histories or outcomes. However, providing information back to individual sites can facilitate further detailed evaluation by each specific practice.

Overall, we conclude that in community practice the concordance between BI-RADS assessments and recommendations is high and that the degree of concordance has improved over time. We did find significant variation in recommendations associated with the categories probably benign finding and suspicious abnormality, categories 3 and 4, respectively. Attention to these variations is important for both auditing and research purposes and for further possible refinements of the BI-RADS lexicon and in training programs about the use of the BI-RADS lexicon.

References

1. American College of Radiology. *Breast imaging reporting and data system (BI-RADS),* 3rd ed. Reston, VA: American College of Radiology, **1998**
2. D'Orsi CJ. The American College of Radiology mammography lexicon: an initial attempt to standardize terminology. *AJR* **1996**;166:779–780
3. Getty DJ, Pickett RM, D'Orsi CJ, Swets JA. Enhanced interpretation of diagnostic images. *Invest Radiol* **1988**;23:240–252
4. D'Orsi CJ, Getty DJ, Swets JA, Pickett RM, Seltzer SE, McNeil B. Reading and decision aids for improved accuracy and standardization of mammographic diagnosis. *Radiology* **1991**;184:619–622
5. Baker JA, Kornguth PJ, Floyd CE Jr. Breast Imaging Reporting and Data System standardized mammography lexicon: observer variability in lesion description. *AJR* **1996**;166:773–778
6. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR* **2000**;174:1769–1777
7. Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, Dershaw DD. The Breast Imaging Reporting and Data System: positive predictive value of mammographic features and final assessment categories. *AJR* **1998**;171:35–40
8. Lacquement MA, Mitchell D, Hollingsworth, AB. Positive predictive value of the Breast Imaging Reporting and Data System. *J Am Coll Surg* **1999**;189:34–40
9. Quality Mammography Standards: Final Rule of Department of Health and Human Services and Food and Drug Administration, 21 CFR §16 and 900, **1997**